Springer Protocols

Elena García-Fruitós *Editor*

Insoluble
Proteins

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

For further volumes:
http://www.springer.com/series/7651

Insoluble Proteins

Methods and Protocols

Edited by

Elena García-Fruitós

CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra (Barcelona), Spain;
Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina,
Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

☼ Humana Press

Editor
Elena García-Fruitós
CIBER de Bioingeniería
Biomateriales y Nanomedicina (CIBER-BBN)
Bellaterra (Barcelona), Spain

Departament de Genètica i de Microbiologia
Institut de Biotecnologia i de Biomedicina
Universitat Autònoma de Barcelona
Bellaterra (Barcelona), Spain

Preface

Since recombinant proteins are necessary for a wide range of applications for both biotechnological and pharmaceutical industries, the interest in the recombinant protein production field has been growing exponentially in the last several years. In this context, although some of these proteins are easily produced and purified, many of them show important bottlenecks in the production and purification process with insolubility being one of the most important ones. Thus, this volume of the *Methods in Molecular Biology* series aims to provide the scientific community with detailed and reliable state-of-the-art protocols that are used in order to successfully produce and purify recombinant proteins prone to aggregate. The main objective of this book is to help those working in the recombinant protein production field by describing a wide number of protocols and examples. The book is organized into 24 chapters that describe not only the recombinant protein production in different expression systems but also different purification and characterization methods to finally obtain these difficult-to-obtain proteins. Chapters 1–13 are focused on the description of protein production methods using both prokaryotic and eukaryotic expression systems. Chapters 14–17 describe purification protocols using insoluble proteins, while Chapters 18–23 are useful to find information regarding the characterization of insoluble proteins. Finally, Chapter 24 aims to give a general overview of interesting applications of insoluble proteins.

I would like to stress that this book has been written by a multidisciplinary team, which adds value to its content since it has been analyzed from different points of view.

Finally, I would like to thank all the authors for their great job. The publication of this book would not have been possible without the effort of all of them. I would also like to thank Prof. John Walker for giving me the opportunity to edit this book and for his full support through the whole process.

Bellaterra (Barcelona), Spain   *Elena García-Fruitós*

Contents

Part IX   Insoluble Protein Characterization

Part X   Insoluble Protein Applications

Contributors

WASSIM ABDULRAHMAN • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

LUCIA ACHBERGEROVÁ • *Department of Carbohydrate Enzymology, Laboratory of Synthetic Biology, Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Bratislava, Slovak Republic; Institute of Chemistry, Center of Excellence for White-Green Biotechnology, Slovak Academy of Sciences, Bratislava, Slovak Republic*

GENNARO APUZZO • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Institute of Protein BiochemistryNational Research Council, Naples, Italy*

RANJANA ARYA • *School of Biotechnology, Jawaharlal Nehru University, New Delhi, India*

MIGUEL J. BAGAJEWICZ • *School of Chemical, Biological and Materials Engineering, University of Oklahoma, Norman, OK, USA*

STEFAN BAGBY • *Department of Biology and Biochemistry, University of Bath, Bath, UK*

FRANK BERNHARD • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*

CATHERINE BIRCK • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

JASON T. BOOCK • *School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA*

ROOP S. BORA • *Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia*

SYLVAIN BOUTIGNY • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

SÍLVIA BRONSOMS • *Servei de Proteòmica i Biologia Estructural, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*

OLIVIA CANO-GARRIDO • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

JONATHAN M.G. CHAMBERLAIN • *Institute of Biomedical and Biomolecular Science, University of Portsmouth, Portsmouth, UK*

JOSÉ LUIS CORCHERO • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

AGUSTÍN CORREA • *Recombinant Protein Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay*

DIMAS TADEU COVAS • *Regional Blood Center, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil; Department of Clinical Medicine, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil*

MATTHEW P. DELISA • *School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA*

CESAR DÍEZ-GIL • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

SILVIA M. DOGLIA • *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy; Department of Physics, University of Milano-Bicocca, Milan, Italy; Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia (CNISM), Milan, Italy*

VOLKER DÖTSCH • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*

ELISA ELIZONDO • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

LOUISE FAIRALL • *Department of Biochemistry, University of Leicester, Leicester, UK*

NEUS FERRER-MIRALLES • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

MARCELA CRISTINA CORRÊA DE FREITAS • *Regional Blood Center, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil*

ANNIE FRELET-BARRAND • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France; CEA Saclay, iBiTec-S, Service de Bioénergétique Biologie Structurale & Mécanismes, UMR8221 CNRS, Gif-Sur-Yvette, France*

ASSAF FRIEDLER • *Department of Organic Chemistry, The Hebrew University of Jerusalem, Jerusalem, Israel*

LIANWU FU • *Department of Cell, Developmental and Integrative Biology, Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, AL, USA*

ELENA GARCÍA-FRUITÓS • *CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBERBBN), Bellaterra (Barcelona), Spain; Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain*

MARIA GIULIANI • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Novartis Vaccines and Diagnostics, Siena, Italy*

SHAVETA GOYAL • *Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore*

ROGER G. HARRISON • *School of Chemical, Biological and Materials Engineering, University of Oklahoma, Norman, OK, USA*

ERIK HENRICH • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*

EVA HRABÁROVÁ • *Department of Carbohydrate Enzymology, Laboratory of Synthetic Biology, Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Bratislava, Slovak Republic; Institute of Chemistry, Center of Excellence for White-Green Biotechnology, Slovak Academy of Sciences, Bratislava, Slovak Republic*

KEISUKE ITO • *Biomolecular Engineering Laboratory, Graduate School of Integrated Pharmaceutical and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan*

LEI KAI • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany; Institute of Botany, Applied Plant Science, Darmstadt University of Technology, Darmstadt, Germany*

YASUAKI KAWARASAKI • *Biomolecular Engineering Laboratory, Graduate School of Integrated Pharmaceutical and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan*

ISABELLE KOLB-CHEYNEL • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

SIMON E. KOLSTOE • *Institute of Biomedical and Biomolecular Science, University of Portsmouth, Portsmouth, UK*

TAKESHI KUROSE • *Biomolecular Engineering Laboratory, Graduate School of Integrated Pharmaceutical and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan*

MARIO LEBENDIKER • *Protein Purification Facility, Wolfson Centre for Applied Structural Biology, Jerusalem, Israel*

LIANGZHONG LIM • *Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore*

ZHANGLIN LIN • *Department of Chemical Engineering, Tsinghua University, Beijing, China*

MICHAL MAES • *Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem, Israel*

GENNARO MARINO • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Institute of Protein Biochemistry, National Research Council, Naples, Italy*

DARIO MIZRACHI • *School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA*

EVELYN MORENO • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

LUCAS MOYET • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

JOZEF NAHÁLKA • *Department of Carbohydrate Enzymology, Laboratory of Synthetic Biology, Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Bratislava, Slovak Republic; Institute of Chemistry, Center of Excellence for White-Green Biotechnology, Slovak Academy of Sciences, Bratislava, Slovak Republic; Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic*

ANTONINO NATALELLO • *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy; Department of PhysicsUniversity of Milano-Bicocca, Milan, Italy; Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia (CNISM), Milan, Italy*

SUSANNA NAVARRO • *Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*

JOANNE E. NETTLESHIP • *OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell Oxford, Didcot, UK; Division of Structural BiologyUniversity of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, UK*

PABLO OPPEZZO • *Recombinant Protein Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay*

ERIKA ORBÁN • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*

JUDIT OSZ-PAPAI • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

RAYMOND J. OWENS • *OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell Oxford, Didcot, UK; Division of Structural Biology, University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, UK*

AMULYA K. PANDA • *National Institute of Immunology, New Delhi, India*

ERMENEGILDA PARRILLI • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Institute of Protein Biochemistry, National Research Council, Naples, Italy*

MIREIA PESARRODONA • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

SPELA PETERNEL • *EN-FIST Centre of Excellence, Ljubljana, Slovenija*

VIRGÍNIA PICANÇO-CASTRO • *Regional Blood Center, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil*

MAREIKE G. POSNER • *Department of Biology and Biochemistry, University of Bath, Bath, UK*

ARNAUD POTERSZMAN • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

DAVIDE PROVERBIO • *Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe-University of Frankfurt/Main, Frankfurt/Main, Germany*

HAINA QIN • *Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore*

LAURA RADU • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

NAHID RAHMAN-HUQ • *OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell Oxford, Didcot, UK; Division of Structural Biology, University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, UK*

IMMA RATERA • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

YAMINI REDDIVARI • *OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell Oxford, Didcot, UK; Division of Structural Biology, University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, UK*

NORBERT ROLLAND • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

JAMAL S.M. SABIR • *Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia*

PAOLO SACCARDO • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

KULVINDER S. SAINI • *Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia*

DANIEL SALVI • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

FILOMENA SANNINO • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Institute of Protein Biochemistry, National Research Council, Naples, Italy*

EMELINE SAUTRON • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

JOHN W.R. SCHWABE • *Department of Biochemistry, University of Leicester, Leicester, UK*

DAPHNÉ SEIGNEURIN-BERNY • *CNRS, UMR 5168, Laboratoire de Physiologie Cellulaire & Végétale, Grenoble, France; Univ. Grenoble Alpes, LPCV, Grenoble, France; CEA, DSV, iRTSV, LPCV, Grenoble, France; INRA, LPCV, USC1359, Grenoble, France*

JOAQUIN SERAS-FRANZOSO • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

ANUPAM SINGH • *National Institute of Immunology, New Delhi, India*

JIANXING SONG • *Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore*

KAMILLA SWIECH • *Department of Pharmaceutical Sciences, Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil; Regional Blood Center, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil*

ELIZABETH SZTUL • *Department of Cell, Developmental and Integrative Biology, Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, AL, USA*

WITOLD TATKIEWICZ • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain*

SEBASTIÁN A. TREJO • *Servei de Proteòmica i Biologia Estructural, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*

NATHALIE TROFFER-CHARLIER • *Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, Illkirch, France*

MARIA LUISA TUTINO • *Department of Chemical Sciences, University of Naples "Federico II", Naples, Italy; Institute of Protein Biochemistry, National Research Council, Naples, Italy*

UGUTZ UNZUETA • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

ABHISHEK UPADHYAY • *Department of Biology and Biochemistry, University of Bath, Bath, UK*

VAIBHAV UPADHYAY • *National Institute of Immunology, New Delhi, India*

ESTHER VÁZQUEZ • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

JAUME VECIANA • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

NORA VENTOSA • *Department of Molecular Nanoscience and Organic Materials, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

SALVADOR VENTURA • *Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*

ANNA VILLAR-PIQUE • *Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*

ANTONIO VILLAVERDE • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

DUJDUAN WARAHO-ZHMAYEV • *School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA; Biological Engineering Program, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangmod, Toongkru, Bangkok, Thailand*

PETER J. WATSON • *Department of Biochemistry, University of Leicester, Leicester, UK*

LEI XING • *Department of Chemical Engineering, Tsinghua University, Beijing, China*

ZHIKUN XU • *Departament de Genètica i de Microbiologia, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*

WANGHUI XU • *Department of Chemical Engineering, Tsinghua University, Beijing, China*

QING ZHAO • *Department of Chemical Engineering, Tsinghua University, Beijing, China*

BIHONG ZHOU • *Department of Chemical Engineering, Tsinghua University, Beijing, China*

Chapter 1

General Introduction: Recombinant Protein Production and Purification of Insoluble Proteins

Neus Ferrer-Miralles, Paolo Saccardo, José Luis Corchero, Zhikun Xu, and Elena García-Fruitós

Abstract

Proteins are synthesized in heterologous systems because of the impossibility to obtain satisfactory yields from natural sources. The production of soluble and functional recombinant proteins is among the main goals in the biotechnological field. In this context, it is important to point out that under stress conditions, protein folding machinery is saturated and this promotes protein misfolding and, consequently, protein aggregation. Thus, the selection of the optimal expression organism and the most appropriate growth conditions to minimize the formation of insoluble proteins should be done according to the protein characteristics and downstream requirements.

Escherichia coli is the most popular recombinant protein expression system despite the great development achieved so far by eukaryotic expression systems. Besides, other prokaryotic expression systems, such as lactic acid bacteria and psychrophilic bacteria, are gaining interest in this field. However, it is worth mentioning that prokaryotic expression system poses, in many cases, severe restrictions for a successful heterologous protein production. Thus, eukaryotic systems such as mammalian cells, insect cells, yeast, filamentous fungus, and microalgae are an interesting alternative for the production of these difficult-to-express proteins.

Key words Recombinant proteins, Protein expression, Protein purification, Aggregation, Solubility, Heterologous system, Insoluble proteins

1 Protein Folding

1.1 Protein Synthesis and Folding

Protein expression in cells is a highly regulated process that permits to build the whole essential protein apparatus for the cells. Nucleic acid codons, through the ribosomal machinery, lead to the formation of linear amino acid sequences that will result in a 3D polypeptide structure. The formation process of this defined spatial structure is called protein folding. Since 1961, when Anfinsen showed that the DNA sequence owns the information for the final tridimensional structure, a lot has been learnt and discussed about the protein folding phenomena [1]. Nowadays, the folding process

can be described as the way by which the proteins reach the most favored status at the bottom of an energetic funnel, rolling down into different energetics status.

The number of theoretical conformations that a relatively small protein can reach is really high. As an example, a 100 amino acid peptide can fold in 1,030 possible conformations. Folding for accidental scanning among all conformation permitted, but not functionally exact, could take up to 1,011 year. Despite of this statistics, inside the cells, the protein emerging from ribosome folds spontaneously and rapidly, under hydrophobic driving forces [2].

One of the major issues in protein folding is that, until the whole protein is synthesized, the N-terminal overhanging polypeptide chains lacks complete information for a correct folding. It is now clear that in vivo, newly growing synthesized proteins must be protected in order to avoid misfolding or aggregation until the whole translation is complete. Moreover, once the synthesis is complete, proteins should be immediately folded in order to avoid nonspecific interaction with other components of the crowded cytoplasmic environment. Otherwise, if proteins are required to be moved to another cellular compartment, they must maintain the unfolded state in order to permit the membrane translocation to the appropriate subcellular target site.

During the evolution, cells have developed a protein quality control system, which control protein synthesis, folding, unfolding, and turnover. This system is constituted by a class of highly conserved proteins called chaperones and also by a clearance mechanism, which act together [3]. Chaperones synthesis can be induced by heat shock, among other factors, and, because of that, they are called heat shock proteins (Hsps) [2]. Most of chaperones interact with other regulatory and cooperating proteins which support their functions and are extremely important for the cells, especially under stress situations. There are two major classes of chaperones, Hsp70s and Hsp60s (or chaperonins). Both are characterized by being ATP hydrolysis-dependent to assist the specific protein folding in eukaryotic and prokaryotic cells [4, 5]. Despite their analogy on ATP hydrolysis and substrate binding, they show a completely different mechanism of action.

1.2 Soluble and Insoluble Proteins

In biotechnology, proteins are synthesized in heterologous systems because of the impossibility to obtain satisfactory yields from natural sources. Expressing and purifying the maximum amount of recombinant active protein as possible are among the main goals in this field. In this context, it is important to point out that the selection of the optimal expression organism, as well as the most appropriate growth conditions, should be done according to the protein characteristics and downstream requirements [2].

Under stress situations, such as thermal or oxidative stress, or under protein overexpression conditions, protein folding machinery

is saturated, and this promotes protein misfolding and, consequently, protein aggregation. Other causes of aggregation could be a mutation in the protein primary structure sequence due to a RNA/DNA mutation or to a translational misincorporation, or the high concentration of the newly synthesized protein [6–8]. Thus, aggregation process is a common phenomenon observed during recombinant protein production. These protein-based aggregates are generally present in low copy number in the cell cytoplasm or periplasm, and they are formed by a wide range of different conformational populations, including those polypeptides that are partially folded as well as by proteins that have reached their native form [9]. Protein aggregates are found in both eukaryotic and prokaryotic cells under homologous or heterologous protein overexpression being its formations favored at high growth temperatures. Specifically, protein aggregates formed in bacteria are known as inclusion bodies (IBs) (*see* Subheading 3.2, Chapters 4, 6, and 16), while in mammalian cells they are named aggresomes (*see* Subheading 3.3 and Chapter 17).

In contrast to what occurs during IB formation, aggresomes are not aggregates of only single protein species; chaperones, chaperonins residues, and proteasome subunits are also found in aggresome immunohistochemical analysis [10, 11]. It is being thought that concentrating aggregates in a defined area have the function to remove aggregates from cytosol and also promote their disposal by autophagy [12].

Besides aggregation, it is important to emphasize that in many cases, misfolded proteins can be degraded through the proteasome complex.

2 Expression Systems for Recombinant Protein Expression (Fig. 1)

Since the production of soluble and functional proteins through a cost-effective and easily scalable process is one of the main challenges nowadays, most of the efforts in this context are aimed at developing and optimizing gene expression systems to minimize the formation of insoluble proteins.

2.1 Prokaryotic Systems

2.1.1 The Preferred Expression System: *Escherichia coli* (See Chapter 2)

E. coli is the most popular recombinant protein expression system despite the great development achieved so far by eukaryotic expression systems. The key of success is related to the easy of handle, reduced cost, high yield, and the possibility to optimize downstream processes by affordable scaling-up processes. In addition, a large amount of protein expression tools are available. In fact, the use of *E. coli* as the preferred expression system is patent in the amount of released PDB entries from proteins obtained in this host organism, representing more than 88 % of the stored structures while only in 12 % of them an *E. coli* gene is expressed,
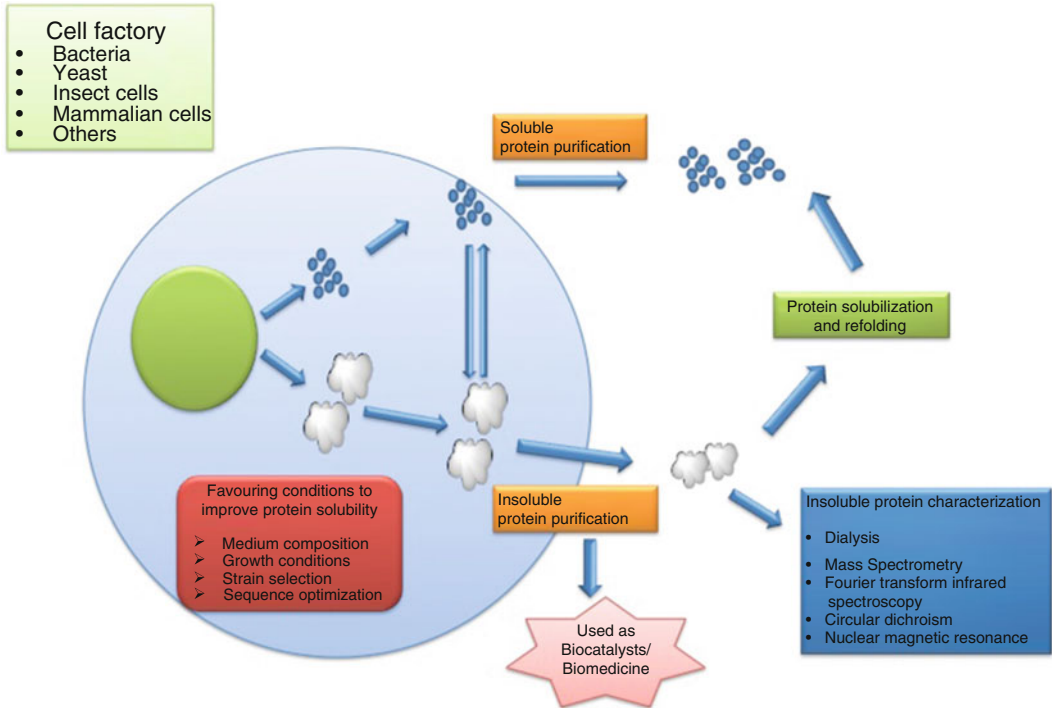
Fig. 1 General scheme of recombinant protein production and purification of insoluble proteins

demonstrating the great power of this expression system for heterologous protein production.

However, as has been already broadly discussed [13, 14], this expression system poses severe restrictions for heterologous proteins such as full-length mammalian proteins and the difficult-to-express membrane proteins [15–17]. In that sense, the *E. coli* expression system of biopharmaceutical proteins on the market drops to 30 % [18] when compared to eukaryotic expression systems. The limitation of this prokaryotic expression system relies in the reduced capacity to fulfill certain specific posttranslational modifications of the eukaryotic world which can be related to protein solubility and/or biological activity. In other instances, the protein is not even transcribed or translated, and, in most of the cases, aggregation takes place, making the purification process from the soluble cellular fraction a laborious or impossible task.

In summary two main problems are encountered when producing recombinant proteins in *E. coli*: reduced or lack of heterologous gene expression and aggregation.

The strategies to improve protein yield mainly relies on the gene design aimed to optimize the rate of transcription, the stability of the mRNA, and the rate of translation [19–24]. On the other hand, improving solubility involves changes in cellular metabolism or/and the protein quality control system.

In the *E. coli* cell, the protein quality control system is composed by two key elements: chaperones and proteases that control the correct folding of proteins and eliminate reluctant protein species that cannot be properly processed, respectively. There are two types of chaperones depending on their effect on protein folding. On the one hand, holding chaperones detect and bind to unfolded or partially folded protein species to let the quality control system to try to fold the polypeptide. Trigger factor binds to nascent polypeptides, and the small heat shock proteins IbpA and IbpB bind to hydrophobic patches in partially folded proteins. On the other hand, two sets of folding chaperones interact with partially folded polypeptides to assist them in their proper folding (GroEL with the accessory protein GroES and DnaK and co-chaperone DnaJ and GrpE). The GroELS complex has a broad specificity and is essential for cell viability, while DnaKJE complex shows substrate preference for nascent polypeptides and is not essential. Finally, the quality control system removes unfolded or folding reluctant proteins by cellular proteases as Lon, ClpA, and ClpB, releasing small peptides in the cytosol that can be recycled in protein synthesis.

This finely tuned system seems to be overcome when overexpression of a recombinant gene takes place in an *E. coli* cell as in many other expression systems, and the limiting step in protein production and solubility might be related to the limitation of one or more protein factors involved in protein folding. For that reason, many chaperone cocktails have been co-expressed with the gene of interest as a strategy to compensate for the stress produced to the cell. However, the outcome of the supplementation of chaperones is variable, and not a single, universal cocktail has been described being a matter of trial and error process for each and every protein that has to be attempted to be produced.

In the case of cellular metabolism, one of the most explored variables has been media formulation that has a great impact in protein yield [25] as well as in protein solubility [26]. During gene expression induction, expressing cells suffer metabolic stress derived from the reduced access to oxygen, substrates, and also pH changes among others. In addition, limiting cofactors may have a great impact in the proper protein folding and stabilization even in the presence of optimized media formulations [27]. In that scenario, the establishment of optimal growth conditions in fermentation systems guarantees the reproducibility of the process, although controlled batch experiments give not negligible results [25, 28].

Additionally, reduction of growth temperature has a positive effect over solubility since hydrophobic interactions are promoted at high temperatures and expression of chaperones is induced. In summary, less newly synthesized polypeptides are produced, having less hydrophobic interactions and more access to the folding machinery. Obviously, at low growth temperatures, protein yield is compromised, yet protein solubility has been demonstrated to be favored [29].

As it would be discussed in the following sections, aggregation of proteins in *E. coli* does not seem to be a dead-end for some recombinant proteins as it was assumed in the near past. On the one hand, the recovery of functional recombinant protein coming from IBs after denaturing-refolding processes [30–32] has been widely documented, while on the other hand, biologically active proteins are detected in the formed IB during recombinant gene overexpression [33–40]. In this latter case, solubilization protocols allow the partial recovery of the entrapped protein [41], and more interestingly, innovative biotechnological applications of intact IB are underway to use them as biocatalysts, nanopills, or cell proliferation factors in regenerative medicine among others [42–46].

In any case, several approximations to improve the solubility/aggregation rate of recombinant proteins in *E. coli* have been developed as solubility has been linked to conformational quality and biological activity. However, it is important to note that this link might be a simplistic view since the soluble cellular fraction has been demonstrated to contain a wide spectrum of soluble protein species, reaching threshold protein conformations in which proteins have a high tendency to aggregate and therefore accumulating and forming part of IBs [47]. In agreement with this observation, it has been described lost in the specific activity of the produced recombinant protein while gaining solubility in some cases [48, 49]. This phenomenon has been ascribed to the way in which the cellular quality control system copes with overexpression of recombinant proteins imposing solubility over folding efficiency.

2.1.2 Lactic Acid Bacteria

It is important to emphasize that recombinant production of difficult-to-express proteins, including those mostly insoluble and prone to aggregate, is one of the main important challenges in the biotechnology field [50]. In this context, the use of lactic acid bacteria (LAB) as a recombinant cell factory is gaining importance, especially for the recombinant production of membrane proteins, which are known to have a huge tendency to precipitate, being mostly insoluble (*see* Chapter 8). LAB, being a prokaryotic expression system, not only present the same advantages as *E. coli* (cheap and easily scalable system) but also an important added value, since they do not contain endotoxins in their membrane, which are pyrogenic in humans and other mammals [50–53]. Thus, since the presence of bacterial endotoxins in proteins is becoming one major concern by regulatory agencies [54], many efforts are being addressed to the development of alternative expression systems, being LAB, classified as Generally Recognized As Safe (GRAS), an excellent candidate.

Additionally, it is worth stressing that the endotoxin removal process has not only important associated costs but also presents a risk to destroy protein folding and protein function, being a step that should be avoided, especially for those proteins that are difficult to isolate such as insoluble proteins [55].

Besides, LAB have in general an efficient protein secretion system, being another important characteristic to be considered in the protein production system when overexpressing proteins difficult to produce and purify.

Thus, as described in Chapter 8, it is important to point out that nowadays it is already possible to find commercial proteins produced in recombinant LAB, being *Bacillus subtilis* and *Lactococcus lactis* widely used as host microorganisms.

2.1.3 Pseudoalteromonas haloplanktis

Protein aggregation is mainly driven by stereospecific interactions between solvent-exposed hydrophobic patches [9, 56]. Such interactions are weakened when temperature decreases. Thus, the production of recombinant proteins in psychrophilic bacteria (cultured at 4 °C or below) represents an exciting model to improve the quality/solubility of recombinant proteins. In this context, a few cold-adapted bacterial species are under early but intense exploration as cell factories, with *Pseudoalteromonas haloplanktis* TAC125 as a representative example. *P. haloplanktis* TAC125 is a Gram-negative bacterium isolated from an Antarctic coastal seawater sample [57], being able to duplicate in the range of 0–30 °C [58], and even at lower temperatures, making it one of the faster growing psychrophiles so far characterized, and an attractive host as cell factory (*see* Chapter 13).

P. haloplanktis TAC125 versatility has been improved by the development of genetically engineered strains with improved features as cell factories [59, 60]. *P. haloplanktis* TAC125 was also the first Antarctic bacterium in which an efficient gene expression technology was set up, by the proper assembly of psychrophilic molecular signals [58, 61] into a modified *E. coli* cloning vector [62]. Several generations of cold-adapted expression vectors allow the production of recombinant proteins either by constitutive [61] or inducible profiles [63] and address the product toward any cell compartment or to the extracellular medium [64].

Beneficial effects in using this cold-adapted platform with respect of the conventional mesophilic *E. coli* have been reported during the production of antibody fragments [65, 66] or in the production of some "difficult-to-express" proteins such as the human nerve growth factor, h-NGF [67], or the alpha-glucosidase from *Saccharomyces cerevisiae* [63]. While when produced in *E. coli* the h-NGF fails to fold and accumulates into IBs [68], its production in *P. haloplanktis* TAC125 results in fully soluble and periplasmically translocated protein, accumulating in almost fully dimeric form [67]. In the same line, alpha-glucosidase from *S. cerevisiae* is largely insoluble when expressed in *E. coli*, but its recombinant production in *P. haloplanktis* TAC125 renders a recombinant enzyme totally soluble and highly active [63].

Observation that insoluble aggregates of recombinant proteins have never been observed in P. *haloplanktis* TAC125 (even at high

expression levels [63]) suggests that its cellular physicochemical conditions and/or folding processes are quite different from those observed in mesophilic bacteria [69].

Recently a synthetic medium for *P. haloplanktis* TAC125 growth was obtained, and the new optimized medium was used for *P. haloplanktis* TAC125 chemostat cultivation [66]. Moreover, a *P. haloplanktis* TAC125 fed-batch fermentation strategy could be established, which is feasible to be used in lab-scale or for industrial purposes [70]. The next challenges for the industrial application of *P. haloplanktis* TAC125 as nonconventional system for protein production include the development of efficient fermentation scheme to upscale the production in automated bioreactors.

2.2 Eukaryotic Systems

2.2.1 Mammalian Cells as Expression System

Prokaryotic protein expression systems, such as *E. coli*, often fail to produce correctly folded, functional eukaryotic proteins. The expression of these proteins greatly benefits of using a eukaryotic expression system, such as mammalian cells, due to their ability to perform proper posttranslational modifications, usually essential for the functionality of therapeutic proteins.

Recent advances have significantly improved the expression levels in mammalian cell lines, reaching up to a few grams of recombinant antibodies per liter in stably transfected Chinese hamster ovary (CHO) cells [71, 72]. The development of a process for recombinant protein production in mammalian cells usually follows a well-established scheme. Unfortunately, this process can take several months, being the major drawback of the stable CHO cell lines. Thus, faster and cheaper approaches for recombinant protein production are needed when many proteins (or several variants of a single protein) must be rapidly evaluated as potential biotechnological or biopharmaceuticals products. For that purpose, a different strategy (called "transient gene expression" or TGE) is preferred. In TGE, recombinant gene is not incorporated into the host cell genome, and selection and isolation of stable transfectants is bypassed so that protein expression is obtained rapidly but only for a limited period of time. By TGE approach, it is possible to produce milligram quantities of recombinant proteins within days or weeks [73].

CHO cells have become the standard mammalian host cells used for the production of recombinant proteins, since it grows rapidly, offers process versatility, can be cultured as either an adherent or a suspension-adapted culture, and is capable of growing in protein-free medium [74, 75]. Apart from CHO, other cell lines as mouse myeloma (NS0), baby hamster kidney (BHK), human embryonic kidney (HEK-293), or human retina-derived (PERC6) cells have proved to be good alternatives. Volumetric yields of secreted recombinant proteins are usually higher when using HEK-293 cells [76]. Thus, HEK-293 cells have also been adapted to grow in serum-free medium, and it has been

demonstrated the feasibility of transfecting these cells in suspension and in large-scale volumes (*see* Chapters 11 and 12) [77, 78].

As in any other expression system, high expression levels are pursued when using mammalian cells as a "cell factory." However, if the synthesis rate of the recombinant protein exceeds its combined folding and degradation rates, some of the protein will be unable to reach its native, soluble form and will accumulate into insoluble aggregates, in subcellular structures called "aggresomes," as described by Johnston in the late 1990s [79]. Cells have special machinery responsible for the transport of such protein aggregates, in a microtubule-dependent manner, to the centrosome, forming there the aggresome [6, 79, 80].

Aggresomes formation is usually related to overexpression of recombinant proteins. For example, overexpression of the cystic fibrosis transmembrane conductance regulator and presenilin-1 [79], mutant forms of superoxide dismutase [79], synphilin 1 [81], or a chimera between green fluorescent protein (GFP) and a fragment of p115, a membrane protein [10] led to their accumulation into aggregates. Therefore, protein misfolding and aggregation into aggresomes are issues that must be considered in the design of biotechnological procedures.

Finally, recent data revealed that aggresomes formed by different mutants of GFP were fluorescent [10, 82, 83], indicating that protein embedded into such aggresomes is not completely inactivated by aggregation. Such observations could have interesting theoretical and practical implications: for example, aggresomes formed by proteins with biomedical or biotechnological interest could be used as nanopills or as immobilized biocatalysts. However, these new, putative applications of aggresomes have not been further investigated.

2.2.2 Insect Cells

Insect cells expression system is an appealing alternative for many biotechnological applications since insect cells perform similar posttranslational modifications present in mammalian proteins. However, in the case of glycosylation, the metabolic pathways diverge, and in biomedical applications, these differences need to be analyzed. For instance, when recombinant proteins obtained from insect cells are intended to be included in vaccine formulations, this difference might represent a positive adjuvant effect [84–86]. In some other applications, mainly when the recombinant protein is intended to be repeatedly administered as a therapeutic component or vehicle, undesired immunostimulation might be triggered [87]. Some efforts have been made to obtain transgenic insect cell lines capable of performing humanized glycosylation patterns (MimicTM Sf9 insect cells from Life Technologies), although resulting proteins show insufficient terminal glycosylation [88].

Two different approaches can be followed with this expression system. On the one hand, stable insect cell lines provide

continuous production of the protein of interest in the same way as mammalian cell stable cell lines do, while on the other hand, the insect cell-baculovirus tandem offers an important issue that has not been solved yet by any of the other expression systems which corresponds to the expression of up to four genes at a time in the same infected cell (pABAC) using viral strong promoters (*see* Chapter 10).

Recombinant protein aggregation in that expression system has been also documented although less studied, and it has been associated with the accumulation of the recombinant protein in cell aggresomes [89, 90]. Several strategies have been tried to improve protein solubility including lowering growth temperature, using softer cell lysis methods and adding high salt concentrations or detergents to the lysis buffer [91]. In addition, co-expression of chaperones offers an alternative as in the *E. coli* expression system [92]. However, since this expression system is mostly used for the secretion of recombinant proteins, the study of the insoluble cell fraction remains mostly unexplored, and the real impact of protein aggregation needs to be further investigated.

2.2.3 Yeast

Yeast cells combine the eukaryotic ability to perform posttranslational modifications with the bacterial capacity to grow to high cell densities, usually rendering higher yields of recombinant proteins and better scalability than mammalian cells. Moreover, yeasts are able to secrete recombinant proteins to the extracellular medium, which is a major advantage during downstream processes (*see* Chapter 9).

S. cerevisiae, known for ages as a beer and bread producer, is one of the most common yeast used to produce therapeutic proteins [18]. *S. cerevisiae* genome was the first from a eukaryotic organism to be sequenced, their genetics and physiology are widely known, and tools for molecular biology are very well established.

Several therapeutic proteins have been produced and commercialized using *S. cerevisiae* as expression system. As an example, Ardiani and coauthors [93] reviewed the use of recombinant *S. cerevisiae* cells engineered to express viral or tumoral antigens as therapeutic vaccines. Fusion of carrier proteins to therapeutic proteins or peptides is receiving increasing interest because of its potential advantages over the first generation of therapeutic products. In general, such fusions goal is to increase the circulation half-life of the protein of interest [94, 95].

All the commercialized therapeutic proteins produced in *S. cerevisiae* to date are non-glycosylated, although yeast glycoprotein expression is potentially envisaged as a main source of human glycoproteins in the future [96, 97]. In that respect, a huge effort is being done to generate a collection of new strains with humanized sugar contents, starting with a *S. cerevisiae* mutant strain with a deletion in the alpha-1,6-mannosyltransferase OCH1 gene [98].

Despite the successful commercialization of several therapeutic proteins obtained in this system, *S. cerevisiae* has been reported to show limitations in the soluble production of particular protein species [99–101]. This can be exemplified by the failing expression of virus surface glycoproteins (namely, mumps or measles hemagglutinin [102]) that renders inactive aggregates.

Apart from *S. cerevisiae*, a number of alternative yeast expression systems have been developed (reviewed in [103]). Among them, the methylotrophic yeast *Pichia pastoris* as a cellular host for the expression of recombinant proteins has become increasing popular in recent times. *P. pastoris* was originally developed as a single-cell protein production system by Philips Petroleum (Bartlesville, OK, USA) but was subsequently adapted for heterologous protein expression. More than 120 recombinant proteins have been expressed in this host, many of them being of human or mammalian origin [104].

Recombinant protein production in *P. pastoris* has several advantages over other eukaryotic and prokaryotic expression systems: rapid growth rate; ease high cell-density fermentation; high levels of productivity; elimination of endotoxin and bacteriophage contamination; ease genetic manipulation; absence of known human pathogenicity in the spectrum of *P. pastoris* lytic viruses; diverse posttranslational modifications including polypeptide folding, glycosylation, methylation, acylation, or proteolytic adjustment; and the ability to engineer secreted proteins that can be purified from growth medium without harvesting the yeast cells themselves [105].

Several products from *P. pastoris* like human serum albumin, insulin, interferon-alpha, and hepatitis B vaccine are marketed in India and/or Japan [106]. Several reviews [107–109] have described different recombinant proteins with application in diverse areas expressed in *P. pastoris*.

As mentioned before, *S. cerevisiae* is by far the best studied yeast with respect to molecular and cell biology, including protein folding and secretion. However, some evidence shows that its secretion pathway differs more from higher eukaryotes than that of *P. pastoris*. The regulation pattern of unfolded protein response, the major regulon controlling folding limitations, shows significant differences between these two yeasts [110]. With the advent of humanized yeast strains and their ability to control glycosylation, development of a significant number of biopharmaceuticals produced in yeast-based expression systems can be easily envisaged.

2.2.4 Trichoderma reesei

Trichoderma reesei is an efficient secretory filamentous fungus with reported production yields in excess of 100 g/L [111], of industrially applicable native enzymes. This fungus is a soil-based microorganism able to utilize cellulose as its source of nutrition, allowing for both low-cost fermentation media and also strong induction when using the cellobiohydrolase I (cbh1) promoter [112].

Therapeutic protein production in *T. reesei* is an emerging but promising field, particularly considering that the major N-glycan form synthesized by *T. reesei* GlcMac2 MAN5 [113, 114] is a suitable precursor for mammalian glycosylation. Thus, the possibilities for humanization of the *T. reesei* glycosylation pathway are better than, for example, in yeast systems. In that respect, human N-acetylglucosaminyltransferase I has already been expressed in *T. reesei* to transfer a GLcNAc residue to the GlcNac2Man5 fungal glycans [115]. In terms of potential for pharmaceutical protein production, *T. reesei* is not only well established in large-scale fermentation but is also already approved as a GRAS organism for food applications, thereby presenting a platform for progression toward regulatory approval for therapeutic uses.

Since a wide number of proteins have an important tendency to aggregate, an interesting approach under current study is promoting or favoring the accumulation of the recombinant protein into intracellular insoluble aggregates by the fusion of specific signals, such as the ZERA peptide [116] or the endogenous hydrophobin [117] fusion partners. Such systems allow for accumulation of the fusion protein within a protein body structure (similar to IBs or aggresomes). Then, purification can be achieved by utilizing the highly hydrophobic properties of the fusion, by mechanical gravity separation, or by two-phase extraction for ZERA peptide or hydrophobin fusions, respectively, with the need for additional downstream purification.

Development of improved *T. reesei* strains for production of therapeutic proteins must concentrate on both overcoming the bottlenecks of not only expression and purification but also refining the molecular mechanisms involved in determining tertiary structural characteristics, in order to yield molecules of high efficacy and immunogenic compatibility to humans.

2.2.5 Microalgae

The term "microalgae" includes a diverse photosynthetic group of both prokaryotic (cyanobacteria) and eukaryotic organisms. Historically, microalgae have been used in applications ranging from enhancing the nutritional value of animal feed to as producers of highly valuable molecules, like polyunsaturated fatty acid oils, pigments, or human nutritional supplements [118, 119]. Apart from these traditional uses, during the last years, microalgae have received the attention of researchers as an alternative to current recombinant protein expression systems [120–122], due to the feasibility of microalgae to be genetically modified and express heterologous genes. In this context, microalgae show the benefits of plants (they share the same basic photosynthetic mechanism), together with the high productivities of microbial systems. Being most microalgae photoautotrophs, they require only light, water, and basic nutrients for their culture. Some microalgae can also be grown as heterotrophs in fermenters without light as energy source,

thus requiring a supply of sugars for energy and as a carbon source. At the same time, microalgae can be grown in large-scale liquid cultures (either in controlled, closed bioreactors or in open ponds). The potential for large-scale culture (on scales ranging from a few milliliters to 500,000 liters in a cost-effective manner) makes microalgae a desirable target as cell factories for the synthesis of high-value therapeutic proteins. More advantages making microalgae ideal candidates for recombinant protein production include the fact that (1) transgenic algae can be generated quickly, requiring only a few weeks between the generation of transformants and their scale up to production volumes, (2) both chloroplast and nuclear genomes of microalgae can be genetically transformed, and (3) green algae fall into the GRAS category. Since there is no gene flow by pollen or other vehicles of gene escaping, transgenic microalgae are harmless to the environment [123].

Regarding economic issues, and according to a recombinant antibody production study, the cost of production per gram of functional antibody was $150, $0.05, and $0.002 (USD) in mammalian, plant, and microalgae expression systems, respectively, data that makes microalgae-based expression systems very appealing for biotechnological industries [124].

Despite the increasing examples of successful transformation of different microalgae species, *Chlamydomonas*, *Chlorella*, *Volvox*, *Haematococcus*, and *Dunaliella* remain the most widely used [125, 126]. However, current work is mainly performed with *Chlamydomonas reinhardtii*, as it is the best characterized microalgae specie, and for which stable genetic transformation at both chloroplast [127] and nuclear [128, 129] level was first reported. In order to achieve high expression levels of protein in *C. reinhardtii* chloroplast, codon-optimized reporter genes has been developed [130, 131] and used to examine a variety of promoter and translational elements [132]. Using this strategy, GFP accumulation up to 0.5 % of total soluble protein (TSP) was achieved in transgenic chloroplasts [131, 132].

Considerable progress has been made in metabolic engineering toward increasing the expression of naturally produced compounds, with varying levels of success [126, 133]. The expanding genetic engineering toolbox for microalgae has allowed the expression of fully functional antibodies [134, 135], therapeutics [136, 137], and bactericides [138]. However, many obstacles still remain to be solved before microalgae can be seen as standard expression systems. So far, success essentially remains anecdotal, and no wide-ranging system or protocol leading to high-level expression has been fully established.

2.2.6 Transgenic Animals and Plants

The expanding recombinant protein market seems to be limited by the achieved yield of the conventional expression systems described above. Therefore, production systems derived from transgenic

animals and plants have been developed with the aim to increase the production potential. In the case of animals, the most promising systems are proteins secreted in the milk (first approved biopharmaceutical of that type is the anticoagulant human antithrombin, from goat) and semen or accumulated in white yolk of hen eggs. Unfortunately, aggregation of the recombinant protein in animals is not determined. Due to the fact that the protein of interest is recovered in a secreted form, protein accumulation in producing cells is not analized. Therefore, information related to the amount of protein retained in the insoluble fraction of the producing cell is not considered in the production studies.

In the case of plants, the recombinant protein is usually produced at low levels, and the purification process tends to be relatively expensive and complicated [139]. Interestingly, plants have specialized tissues (seeds) that are able to store recombinant proteins at a high purity [140]. In addition, unlike for most of the expression systems, delivery of recombinant protein to aggregated structures in plants offers many advantages. Plants make use of protein aggregation to accumulate proteins for storage purposes in specific cell compartments. For instance, proteins of interest can be sent to protein bodies derived from endoplasmic reticulum by including a proline-rich domain in the recombinant gene [141]. The recovery from this specialized membranous structure simplifies downstream processing by increasing capture of the recombinant protein. In addition, this technology can be transferred to non-seed tissues in plants and also to other eukaryotic expression systems [142, 143].

2.2.7 Cell-Free System

Synthesis of proteins without the entire machinery of a living cell, better known as cell-free protein system (CFPS), is an emerging technology for simple and effective protein productions (*see* Chapter 7) [144, 145].

This platform takes advantage of catalytic components and the necessary elements for transcription, translation, and protein folding that are extracted from crude lysates of *E. coli* (ECE), rabbit reticulocytes (RRL), wheat germ (WGE), or insect cells (ICE) principally. Production in CFPS is quick and simple and not restricted by the eventually toxic effects of the final product.

The appropriate cell-free system should be chosen depending on protein complexity, posttranslational modifications, downstream process, and yield required [146, 147].

Proteins' tendency to aggregate when overexpressed in prokaryote's expression system can be often overcome changing to eukaryote CFPS organism. This technology also permits to perform denaturing and refolding process as well as permits to add components that assist in protein folding, avoiding aggregation [148, 149].

The biggest problem noticed by using those extracts is that cell lysate contains cellular proteins as proteases or nucleases, and nucleic acids not required for protein of interest expression. These components can act in an unpredictable and often unknown way, causing problems to the reaction.

The "PURE" in vitro system, which consists in purified translation factor components of *E. coli*, has been shown as an efficient alternative to the crude lysates [150, 151].

Recently, the CFPS evolution has permitted protein yield at milligram levels thanks also to the development of various reaction schemes as the continuous exchange, continuous flow, hollow fiber systems, or batch-type improvements [152–157].

Another interesting aspect is that the final product purification is simplified [158].

Despite these improvements, costs, lack of wide experience of use, and the problem in reproducing the folding environment, resulting in a non-correct protein folding, are still limiting factors of this technique.

3   Purification and Characterization

3.1   Protein Purification

As already mentioned, in Subheadings 1 and 2, many insoluble proteins aggregate during the expression process, being necessary to optimize different parameters. However, in other cases, protein aggregation occurs during downstream purification processes. In these situations, it is crucial to develop a suitable and optimized purification protocol for each protein (*see* Chapter 14). In fact, it is important not only to carefully evaluate the best purification protocol but also the appropriate characterization (*see* Chapter 21) and dialysis and storage conditions (*see* Chapter 18).

3.2   Inclusion Bodies Purification

IBs are protein aggregates formed in both bacterial cell cytoplasm and periplasm when overexpressing insoluble proteins, which have a huge tendency to aggregate. IBs show, in general, a sphere-like shape varying between 0.1 to 0.8 μm in diameter, depending on the cell host dimension, growth conditions, and protein sequence (*see* Chapter 22). Interestingly, they show higher stability than that of their soluble counterparts and are essentially formed by the protein of interest. In this regard, it is important to stress that it was found that IBs show a spongelike organization, which combine both active and inactive protein forms. Inactive forms correspond to proteins which adopt an amyloid-like organization forming a protease K-resistant fibrillar scaffold that is fully embedded by functional proteins [159–161] (*see* Chapters 19 and 20).

Thus, since these insoluble aggregates are mainly composed by the target protein, they are an important source of the protein of interest. In this regard, several protocols that aim obtaining

protein from IBs have been developed in the last decades. These protocols include denaturation and refolding processes, which allow the purification of variable amounts of the recombinant protein of interest (*see* Chapter 15).

Given that in recent years several groups have shown the huge potential of IBs for diverse applications such as catalysis and tissue engineering, some researchers have focused their efforts on the development of protocols for obtaining highly pure IBs to be used in the applications mentioned above (*see* Chapters 16 and 24). In this regard, this book reviews the newest and most widely used for both isolation of soluble proteins from IBs and protocols for IB purification.

3.3 Aggresomes Purification

Many studies have attempted to identify proteins associated with aggresomes (*see* Chapter 17). For that, the simplest approach would be to isolate the aggresome and determine its composition, since factors involved in the formation of such aggregates may be physically associated with them. However, a detailed analysis of the components associated with aggresomes is hampered by the difficulties found for their isolation. Aggresome heterogeneity in size and charge precludes application of conventional biochemical methods (e.g., gel filtration, or ion exchange or affinity chromatography) for their isolation.

Protein aggregate isolation has been attempted by using the ionic detergent insolubility of amyloids [162] or density gradient fractionation [163]. Such methods may be useful to address certain questions, but they are inadequate to identify aggregate-associated proteins. SDS treatment dissociates most of the associated proteins, and the use of density gradients (apart of being a tedious procedure) renders a high number of nonspecifically associated polypeptides. Due to the abovementioned difficulties, most of the published studies regarding aggresome composition rely on immunocytochemistry of cells overexpressing certain recombinant proteins and that, consequently, produce cytoplasmic aggregates [164–166]. In such studies, aggresomes isolation is bypassed, since detection of their components is performed directly on cells producing such structures.

As an alternative, other approaches have been proposed to obtain cellular insoluble fractions containing aggresomes. For example, a protocol for the isolation of aggresomes formed by a GFP fusion protein has been proposed by García-Mata and colleagues [10]. In this protocol, pellets of cells producing aggresomes were washed with phosphate buffered saline (PBS) and lysed for 30 min on ice with different detergent-containing buffers. Lysates were then passed through a 27 gauge needle, and finally the insoluble material was recovered by centrifugation. Such isolated insoluble material allowed to gain insight into formation and composition of aggresomes, but the presence of some ionic detergents like SDS in some of the buffers used for cell lysis could render misleading results.

Finally, Wang and colleagues have developed a protocol for the isolation of aggregates based on affinity purification without involvement of a solid phase [167, 168]. This highly reproducible procedure yielded a fraction of polyQ aggregates of diverse size and charge, which could be separated by 2D gel and analyzed by mass spectrometry to identify aggregate-associated proteins. Also, the method allows for semiquantitative comparison of the identified proteins. Another advantage of this protocol is that it avoids exposure of the aggregates to extreme pH or ionic strength as well as to ionic detergents, thus preserving putative weak protein interactions.

The apparition of future new applications for protein aggresomes will surely result in the development of faster and better purification protocols for such aggregates, for example, those based on magnetic micro- and nanoparticles.

4   General Overview

The development of recombinant DNA technology represented a breakthrough in the treatment of some human diseases, increasing life expectancy and life quality of patients. The first therapeutic recombinant protein product, human insulin, was approved in 1982, opening up a new pharmaceutical market with an unceasing demand and steady global sale increase [169, 170].

Industrial market also benefits from the recombinant DNA technology in the enzyme and agricultural industry [171].

In this arena, different expression systems have been established to fulfil the production needs [172]. It is widely accepted that prokaryotic expression systems represent a cost-effective alternative when comparing with eukaryotic expression systems [173]. However, at least in the use of therapeutic proteins, regulatory and functional constraints impose the use of the eukaryotic expression systems [174, 175].

In any case, the recombinant protein production process copes with similar limitations in any of the available expression systems [171]. In the extremely crowded cell cytosol, the appearance of a great amount of newly synthesized polypeptide chains challenges the folding machinery, and, consequently, protein aggregation is detected [172].

In some instances, modifications in the growth parameters can modulate the ratio of the amount of protein in the soluble/insoluble cell fraction, but in some cases, the valuable recombinant protein is reluctant to solubilize [176] (*see* Chapter 23).

In this book, the reviewed strategies to improve protein solubility are disclosed in addition to established approaches to obtain soluble protein from protein aggregates. In addition, novel applications for the use of protein aggregates in nanomedicine are also shown.

Acknowledgments

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223–230

2. Mogk A, Mayer MP, Deuerling E (2002) Mechanisms of protein folding: molecular chaperones and their application in biotechnology. Chembiochem 3:807–814

3. Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol 14:630–642

4. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332

5. Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. Cell 125:443–451

6. Kopito RR (2000) Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol 10:524–530

7. Wetzel R (1994) Mutations and off-pathway aggregation of proteins. Trends Biotechnol 12:193–198

8. Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3:R9–23

9. Speed MA, Wang DI, King J (1996) Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. Nat Biotechnol 14: 1283–1287

10. Garcia-Mata R, Bebok Z, Sorscher EJ et al (1999) Characterization and dynamics of aggresome formation by a cytosolic GFP-chimera. J Cell Biol 146:1239–1254

11. Wigley WC, Fabunmi RP, Lee MG et al (1999) Dynamic association of proteasomal machinery with the centrosome. J Cell Biol 145:481–490

12. Glick D, Barth S, Macleod KF (2010) Autophagy: cellular and molecular mechanisms. J Pathol 221:3–12

13. Francis DM, Page R (2010) Strategies to optimize protein expression in E. coli. Curr Protoc Protein Sci. Chapter 5:Unit-29

14. Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb Cell Fact 4:1

15. Wagner S, Klepsch MM, Schlegel S et al (2008) Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A 105:14371–14376

16. Nannenga BL, Baneyx F (2011) Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. Protein Sci 20:1411–1420

17. Freigassner M, Pichler H, Glieder A (2009) Tuning microbial hosts for membrane protein production. Microb Cell Fact 8:69

18. Ferrer-Miralles N, Domingo-Espin J, Corchero JL et al (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8:17

19. Gopal GJ, Kumar A (2013) Strategies for the production of recombinant protein in Escherichia coli. Protein J 32:419–425

20. Schumann W, Ferreira LCS (2004) Production of recombinant proteins in Escherichia coli. Genet Mol Biol 27:442–453

21. Care S, Bignon C, Pelissier M et al (2008) The translation of recombinant proteins in E-coli can be improved by in silico generating and screening random libraries of a 70/96 mRNA region with respect to the translation initiation codon. Nucleic Acids Res 36:e6

22. Graslund S, Nordlund P, Weigelt J et al (2008) Protein production and purification. Nat Methods 5:135–146

23. Hatfield G, Roth DA (2007) Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering (TM). Biotechnol Annu Rev 13:27–42

24. Menzella HG (2011) Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. Microb Cell Fact 10:15

25. Hortsch R, Weuster-Botz D (2011) Growth and recombinant protein expression with Escherichia coli in different batch cultivation media. Appl Microbiol Biotechnol 90: 69–76

26. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307:249–264

27. Apiyo D, Wittung-Stafshede P (2002) Presence of the cofactor speeds up folding of desulfovibrio desulfuricans flavodoxin. Protein Sci 11:1129–1135

28. Jenzsch M, Gnoth S, Kleinschmidt M et al (2007) Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. J Biotechnol 128:858–867

29. Song JM, An YJ, Kang MH et al (2012) Cultivation at 6–10 °C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli. Protein Expr Purif 82:297–301

30. Vaks L, Benhar I (2014) Production of stabilized scFv antibody fragments in the E. coli bacterial cytoplasm. Methods Mol Biol 34: 171–184

31. Mollania N, Khajeh K, Ranjbar B et al (2013) An efficient in vitro refolding of recombinant bacterial laccase in Escherichia coli. Enzyme Microb Technol 52:325–330

32. Ramakrishnan B, Qasba PK (2013) Invitrofolding of beta-1,4galactosyltransferase and polypeptide-alpha-N-acetylgalactosaminyltransferase from the inclusion bodies. Methods Mol Biol 1022:321–333 (Clifton, NJ)

33. Sans C, Garcia-Fruitos E, Ferraz RM et al (2012) Inclusion bodies of fuculose-1-phosphate aldolase as stable and reusable biocatalysts. Biotechnol Prog 28:421–427

34. Garcia-Fruitos E, Aris A, Villaverde A (2007) Localization of functional polypeptides in bacterial inclusion bodies. Appl Environ Microbiol 73:289–294

35. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27

36. Nahalka J, Mislovicova D, Kavcova H (2009) Targeting lectin activity into inclusion bodies for the characterisation of glycoproteins. Mol Biosyst 5:819–821

37. Nahalka J, Patoprsty V (2009) Enzymatic synthesis of sialylation substrates powered by a novel polyphosphate kinase (PPK3). Org Biomol Chem 7:1778–1780

38. Nahalka J (2008) Physiological aggregation of maltodextrin phosphorylase from Pyrococcus furiosus and its application in a process of batch starch degradation to alpha-D-glucose-1-phosphate. J Ind Microbiol Biotechnol 35:219–223

39. Peternel S, Komel R (2010) Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. Microb Cell Fact 9:66

40. Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A (2007) Recombinant protein solubility – does more mean better? Nat Biotechnol 25:718–720

41. Peternel S, Grdadolnik J, Gaberc-Porekar V et al (2008) Engineering inclusion bodies for non denaturing extraction of functional proteins. Microb Cell Fact 7:34

42. Diez-Gil C, Krabbenborg S, Garcia-Fruitos E et al (2010) The nanoscale properties of bacterial inclusion bodies and their effect on mammalian cell proliferation. Biomaterials 31:5805–5812

43. Liovic M, Ozir M, Zavec AB et al (2012) Inclusion bodies as potential vehicles for recombinant protein delivery into epithelial cells. Microb Cell Fact 11:67

44. Cano-Garrido O, Rodriguez-Carmona E, Diez-Gil C et al (2013) Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. Acta Biomater 9:6134–6142

45. Villaverde A, Garcia-Fruitos E, Rinas U et al (2012) Packaging protein drugs as bacterial inclusion bodies for therapeutic applications. Microb Cell Fact 11:76

46. Vazquez E, Corchero JL, Burgueno JF et al (2012) Functional inclusion bodies produced in bacteria as naturally occurring nanopills for advanced cell therapies. Adv Materials 24: 1742–1747

47. Martinez-Alonso M, Garcia-Fruitos E, Villaverde A (2008) Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli. Biotechnol Bioeng 101: 1353–1358

48. Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N et al (2007) Divergent genetic control of protein solubility and conformational quality in Escherichia coli. J Mol Biol 374:195–205

49. Martinez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E et al (2008) The functional quality of soluble recombinant polypeptides produced in Escherichia coli is defined by a

wide conformational spectrum. Appl Environ Microbiol 74:7431–7433

50. Garcia-Fruitos E (2012) Lactic acid bacteria: a promising alternative for recombinant protein production. Microb Cell Fact 11:157

51. Zweers JC, Barak I, Becher D et al (2008) Towards the development of Bacillus subtilis as a cell factory for membrane proteins and protein complexes. Microb Cell Fact 7:10

52. Westers L, Westers H, Quax WJ (2004) Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. Biochim Biophys Acta 1694:299–310

53. Petsch D, Anspach FB (2000) Endotoxin removal from protein solutions. J Biotechnol 76:97–119

54. Williams SF, Martin DP, Horowitz DM et al (1999) PHA applications: addressing the price performance issue: I. Tissue engineering. Int J Biol Macromol 25:111–121

55. Parlane NA, Grage K, Lee JW et al (2011) Production of a particulate hepatitis C vaccine candidate by an engineered Lactococcus lactis strain. Appl Environ Microbiol 77:8516–8522

56. Carrio M, Gonzalez-Montalban N, Vera A et al (2005) Amyloid-like properties of bacterial inclusion bodies. J Mol Biol 347: 1025–1037

57. Birolo L, Tutino ML, Fontanella B et al (2000) Aspartate aminotransferase from the Antarctic bacterium Pseudoalteromonas haloplanktis TAC 125. Cloning, expression, properties, and molecular modelling. Eur J Biochem 267:2790–2802

58. Tutino ML, Duilio A, Parrilli R et al (2001) A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. Extremophiles 5:257–264

59. Cusano AM, Parrilli E, Marino G et al (2006) A novel genetic system for recombinant protein secretion in the Antarctic Pseudoalteromonas haloplanktis TAC125. Microb Cell Fact 5:40

60. Parrilli E, Giuliani M, Marino G et al (2010) Influence of production process design on inclusion bodies protein: the case of an Antarctic flavohemoglobin. Microb Cell Fact 9:19

61. Duilio A, Madonna S, Tutino ML et al (2004) Promoters from a cold-adapted bacterium: definition of a consensus motif and molecular characterization of UP regulative elements. Extremophiles 8:125–132

62. Tutino ML, Parrilli E, Giaquinto L et al (2002) Secretion of alpha-amylase from Pseudoalteromonas haloplanktis TAB23: two different pathways in different hosts. J Bacteriol 184:5814–5817

63. Papa R, Rippa V, Sannia G et al (2007) An effective cold inducible expression system developed in Pseudoalteromonas haloplanktis TAC125. J Biotechnol 127:199–210

64. Parrilli E, De VD, Cirulli C et al (2008) Development of an improved Pseudoalteromonas haloplanktis TAC125 strain for recombinant protein secretion at low temperature. Microb Cell Fact 7:2

65. Dragosits M, Frascotti G, Bernard-Granger L et al (2011) Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. Biotechnol Prog 27:38–46

66. Giuliani M, Parrilli E, Ferrer P et al (2011) Process optimization for recombinant protein production in the psychrophilic bacterium Pseudoalteromonas haloplanktis. Process Biochem 46:953–959

67. Vigentini I, Merico A, Tutino ML et al (2006) Optimization of recombinant human nerve growth factor production in the psychrophilic Pseudoalteromonas haloplanktis. J Biotechnol 127:141–150

68. Rattenholl A, Lilie H, Grossmann A et al (2001) The pro-sequence facilitates folding of human nerve growth factor from Escherichia coli inclusion bodies. Eur J Biochem 268:3296–3303

69. Piette F, D'Amico S, Struvay C et al (2010) Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium Pseudoalteromonas haloplanktis TAC125. Mol Microbiol 76: 120–132

70. Wilmes B, Hartung A, Lalk M et al (2010) Fed-batch process for the psychrotolerant marine bacterium Pseudoalteromonas haloplanktis. Microb Cell Fact 9:72

71. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393–1398

72. Figueroa B Jr, Ailor E, Osborne D et al (2007) Enhanced cell culture performance using inducible anti-apoptotic genes E1B-19K and Aven in the production of a monoclonal antibody with Chinese hamster ovary cells. Biotechnol Bioeng 97:877–892

73. Geisse S (2009) Reflections on more than 10 years of TGE approaches. Protein Expr Purif 64:99–107

74. Derouazi M, Girard P, Van TF et al (2004) Serum-free large-scale transient transfection of CHO cells. Biotechnol Bioeng 87: 537–545

75. Rosser MP, Xia W, Hartsell S et al (2005) Transient transfection of CHO-K1-S using serum-free medium in suspension: a rapid mammalian protein expression system. Protein Expr Purif 40:237–243

76. Baldi L, Hacker DL, Adam M et al (2007) Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. Biotechnol Lett 29:677–684

77. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res 30:E9

78. Girard P, Derouazi M, Baumgartner G et al (2002) 100-liter transient transfection. Cytotechnology 38:15–21

79. Johnston JA, Ward CL, Kopito RR (1998) Aggresomes: a cellular response to misfolded proteins. J Cell Biol 143:1883–1898

80. Kopito RR, Sitia R (2000) Aggresomes and Russell bodies. Symptoms of cellular indigestion? EMBO Rep 1:225–231

81. Zaarur N, Meriin AB, Gabai VL et al (2008) Triggering aggresome formation. Dissecting aggresome-targeting and aggregation signals in synphilin 1. J Biol Chem 283:27575–27584

82. Garcia-Mata R, Gao YS, Sztul E (2002) Hassles with taking out the garbage: aggravating aggresomes. Traffic 3:388–396

83. Shimohata T, Sato A, Burke JR et al (2002) Expanded polyglutamine stretches form an 'aggresome'. Neurosci Lett 323:215–218

84. Kim MC, Song J, Min O et al (2013) Virus-like particles containing multiple M2 extracellular domains confer improved cross-protection against various subtypes of influenza virus. Mol Ther 21:485–492

85. Loureiro S, Ren J, Phapugrangkul P et al (2011) Adjuvant-free immunization with hemagglutinin-Fc fusion proteins as an approach to influenza vaccines. J Virol 85:3010–3014

86. Huetter J, Roedig JV, Hoeper D et al (2013) Toward animal cell culture-based influenza vaccine design: viral hemagglutinin N-glycosylation markedly impacts immunogenicity. J Immunol 190:220–230

87. De Groot AS, Scott DW (2007) Immunogenicity of protein therapeutics. Trends Immunol 28:482–490

88. Legardinier S, Duoner-Cerutti M, Devauchelle G et al (2005) Biological activities of recombinant equine luteinizing hormone/chorionic gonadotropin (eLH/CG) expressed in Sf9 and Mimic insect cell lines. J Mol Endocrinol 34:47–60

89. Radner S, Celie PH, Fuchs K et al (2012) Transient transfection coupled to baculovirus infection for rapid protein expression screening in insect cells. J Struct Biol 179:46–55

90. Heath CM, Windsor M, Wileman T (2001) Aggresomes resemble sites specialized for virus assembly. J Cell Biol 153:449–455

91. Bernard A, Payton M, Radford KR (2001) Protein expression in the baculovirus system. Curr Protoc Neurosci. Chapter 4:Unit 4.19

92. Martinez-Alonso M, Toledo-Rubio V, Noad R et al (2009) Rehosting of bacterial chaperones for high-quality protein production. Appl Environ Microbiol 75:7850–7854

93. Ardiani A, Higgins JP, Hodge JW (2010) Vaccines based on whole recombinant Saccharomyces cerevisiae cells. FEMS Yeast Res 10:1060–1069

94. Evans L, Hughes M, Waters J, et al. (2010) The production, characterisation and enhanced pharmacokinetics of scFv-albumin fusions expressed in Saccharomyces cerevisiae. Protein Expr Purif. 73(2): 113–124

95. Kim BJ, Zhou J, Martin B, et al. (2010) Transferrin fusion technology: a novel approach to prolonging biological half-life of insulinotropic peptides. J Pharmacol Exp Ther. 334(3): 682–692

96. Chiba Y, Akeboshi H (2009) Glycan engineering and production of 'humanized' glycoprotein in yeast cells. Biol Pharm Bull 32:786–795

97. Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 22:1409–1414

98. Nakayama K, Nagasu T, Shimma Y et al (1992) OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. EMBO J 11:2511–2519

99. Gasser B, Saloheimo M, Rinas U et al (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11

100. Porro D, Sauer M, Branduardi P et al (2005) Recombinant protein production in yeasts. Mol Biotechnol 31:245–259

101. Porro D, Branduardi P (2009) Yeast cell factory: fishing for the best one or engineering it? Microb Cell Fact 8:51

102. Ciplys E, Samuel D, Juozapaitis M et al (2011) Overexpression of human virus surface glycoprotein precursors induces cytosolic unfolded protein response in Saccharomyces cerevisiae. Microb Cell Fact 10:37

103. Boer E, Steinborn G, Kunze G et al (2007) Yeast expression platforms. Appl Microbiol Biotechnol 77:513–523

104. Cregg JM, Cereghino JL, Shi J et al (2000) Recombinant protein expression in Pichia pastoris. Mol Biotechnol 16:23–52

105. Li P, Anumanthan A, Gao XG et al (2007) Expression of recombinant proteins in Pichia pastoris. Appl Biochem Biotechnol 142: 105–124

106. Shekhar C (2008) Pichia power: India's bio-tech industry puts unconventional yeast to work. Chem Biol 15:201–202

107. Rabert C, Weinacker D, Pessoa A et al (2013) Recombinants proteins for industrial uses: utilization of Pichia pastoris expression system. Braz J Microbiol 44:351–356

108. Gasser B, Prielhofer R, Marx H et al (2009) Pichia pastoris: protein production host and model organism for biomedical research. Future Microbiol 8:191–208

109. Bollok M, Resina D, Valero F et al (2009) Recent patents on the Pichia pastoris expression system: expanding the toolbox for recombinant protein production. Recent Pat Biotechnol 3:192–201

110. Graf A, Gasser B, Dragosits M et al (2008) Novel insights into the unfolded protein response using Pichia pastoris specific DNA microarrays. BMC Genomics 9:390

111. Cherry JR, Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. Curr Opin Biotechnol 14:438–443

112. Harkki A, Uusitalo J, Bailey M et al (1989) A novel fungal expression system – secretion of active calf chymosin from the filamentous fungus Trichoderma reesei. Nat Biotechnol 7:596–603

113. Stals I, Sandra K, Devreese B et al (2004) Factors influencing glycosylation of Tricho-derma reesei cellulases. II: N-glycosylation of Cel7A core protein isolated from different strains. Glycobiology 14:725–737

114. Stals I, Sandra K, Geysens S et al (2004) Factors influencing glycosylation of Trichoderma reesei cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. Glycobiology 14:713–724

115. Maras M, De Bruyn A, Vervecken W et al (1999) In vivo synthesis of complex N-glycans by expression of human N-acetylgluco-saminyltransferase I in the filamentous fungus Trichoderma reesei. FEBS Lett 452: 365–370

116. Saloheimo M, Lund M, Penttila ME (1999) The protein disulphide isomerase gene of the fungus Trichoderma reesei is induced by endoplasmic reticulum stress and regulated by the carbon source. Mol Gen Genet 262: 35–45

117. Collen A, Saloheimo M, Bailey M et al (2005) Protein production and induction of the unfolded protein response in Trichoderma reesei strain Rut-C30 and its transfor-mant expressing endoglucanase I with a hydrophobic tag. Biotechnol Bioeng 89: 335–344

118. Spolaore P, Joannis-Cassan C, Duran E et al (2006) Commercial applications of microal-gae. J Biosci Bioeng 101:87–96

119. Apt KE, Behrens PW (1999) Commercial developments in microalgal biotechnology. J Phycol 35:215–226

120. Walker TL, Purton S, Becker DK et al (2005) Microalgae as bioreactors. Plant Cell Rep 24:629–641

121. Specht E, Miyake-Stoner S, Mayfield S (2010) Micro-algae come of age as a platform for recombinant protein production. Biotechnol Lett 32:1373–1383

122. Gong Y, Hu H, Gao Y et al (2011) Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. J Ind Microbiol Biotechnol 38:1879–1890

123. Janssen M, Tramper J, Mur LR et al (2003) Enclosed outdoor photobioreactors: light regime, photosynthetic efficiency, scale-up, and future prospects. Biotechnol Bioeng 81:193–210

124. Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active anti-body in algae. Proc Natl Acad Sci U S A 100: 438–442

125. Griesbeck C, Kobl I, Heitzer M (2006) Chlamydomonas reinhardtii: a protein expression system for pharmaceutical and biotechno-logical proteins. Mol Biotechnol 34:213–223

126. Rosenberg JN, Oyler GA, Wilkinson L et al (2008) A green light for engineered algae: redi-recting metabolism to fuel a biotechnology revolution. Curr Opin Biotechnol 19:430–436

127. Boynton JE, Gillham NW, Harris EH et al (1988) Chloroplast transformation in Chlamydomonas with high velocity micro-projectiles. Science 240:1534–1538

128. Debuchy R, Purton S, Rochaix JD (1989) The argininosuccinate lyase gene of Chlamydomonas reinhardtii: an important tool for nuclear transformation and for cor-relating the genetic and molecular maps of the ARG7 locus. EMBO J 8:2803–2809

129. Fernandez E, Schnell R, Ranum LP et al (1989) Isolation and characterization of the nitrate reductase structural gene of Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A 86:6449–6453

130. Mayfield SP, Schultz J (2004) Development of a luciferase reporter gene, luxCt, for Chlamydomonas reinhardtii chloroplast. Plant J 37:449–458

131. Franklin S, Ngo B, Efuet E et al (2002) Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. Plant J 30:733–744

132. Barnes D, Franklin S, Schultz J et al (2005) Contribution of 5′- and 3′-untranslated regions of plastid mRNAs to the expression of Chlamydomonas reinhardtii chloroplast genes. Mol Genet Genomics 274:625–636

133. Potvin G, Zhang Z (2010) Strategies for high-level recombinant protein expression in transgenic microalgae: a review. Biotechnol Adv 28:910–918

134. Franklin SE, Mayfield SP (2005) Recent developments in the production of human therapeutic proteins in eukaryotic algae. Expert Opin Biol Ther 5:225–235

135. Grayson WL, Martens TP, Eng GM et al (2009) Biomimetic approach to tissue engineering. Semin Cell Dev Biol 20:665–673

136. Boehm R (2007) Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. Ann N Y Acad Sci 1102:121–134

137. Weathers PJ, Towler MJ, Xu J (2010) Bench to batch: advances in plant cell culture for producing useful products. Appl Microbiol Biotechnol 85:1339–1351

138. Li SS, Tsai HJ (2009) Transgenic microalgae as a non-antibiotic bactericide producer to defend against bacterial pathogen infection in the fish digestive tract. Fish Shellfish Immunol 26:316–325

139. Joensuu JJ, Conley AJ, Linder MB et al (2012) Bioseparation of recombinant proteins from plant extract with hydrophobin fusion technology. Methods Mol Biol 824:527–534

140. Tian L, Sun SS (2011) A cost-effective ELP-intein coupling system for recombinant protein purification from plant production platform. PLoS One 6:e24183

141. Torrent M, Llop-Tous I, Ludevid M (2009) Protein body induction: a new tool to produce and recover recombinant proteins in plants. Methods Mol Biol 483:193–208

142. Wilken LR, Nikolov ZL (2012) Recovery and purification of plant-made recombinant proteins. Biotechnol Adv 30:419–433

143. Torrent M, Llompart B, Lasserre-Ramassamy S et al (2009) Eukaryotic protein production in designed storage organelles. BMC Biol 7:5

144. Katzen F, Chang G, Kudlicki W (2005) The past, present and future of cell-free protein synthesis. Trends Biotechnol 23:150–156

145. Swartz J (2006) Developing cell-free biology for industrial applications. J Ind Microbiol Biotechnol 33:476–485

146. Carlson ED, Gan R, Hodgman CE et al (2012) Cell-free protein synthesis: applications come of age. Biotechnol Adv 30:1185–1194

147. Endo Y, Sawasaki T (2006) Cell-free expression systems for eukaryotic protein production. Curr Opin Biotechnol 17:373–380

148. Dadashipour M, Fukuta Y, Asano Y (2011) Comparative expression of wild-type and highly soluble mutant His103Leu of hydroxynitrile lyase from Manihot esculenta in prokaryotic and eukaryotic expression systems. Protein Expr Purif 77:92–97

149. Kudou M, Ejima D, Sato H et al (2011) Refolding single-chain antibody (scFv) using lauroyl-L-glutamate as a solubilization detergent and arginine as a refolding additive. Protein Expr Purif 77:68–74

150. Shimizu Y, Inoue A, Tomari Y et al (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755

151. Shimizu Y, Kanamori T, Ueda T (2005) Protein synthesis by pure translation systems. Methods 36:299–304

152. Spirin AS, Baranov VI, Ryabova LA et al (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. Science 242:1162–1164

153. Kim DM, Choi CY (1996) A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. Biotechnol Prog 12:645–649

154. Kigawa T, Yabuki T, Yoshida Y et al (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett 442:15–19

155. Yamamoto YI, Nagahori H, Yao SL et al (1996) Hollow fiber reactor for continuous flow cell-free protein production. J Chem Eng Jpn 29:1047–1050

156. Kawarasaki Y, Nakano H, Yamane T (1994) Prolonged cell-free protein synthesis in a batch system using wheat germ extract. Biosci Biotechnol Biochem 58:1911–1913

157. Kawarasaki Y, Kawai T, Nakano H et al (1995) A long-lived batch reaction system of cell-free protein synthesis. Anal Biochem 226:320–324

158. Jewett MC, Swartz JR (2004) Rapid expression and purification of 100 nmol quantities of active protein using cell-free protein synthesis. Biotechnol Prog 20:102–109

159. Garcia-Fruitos E, Sabate R, de Groot NS et al (2011) Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J 278:2419–2427

160. Wang L (2009) Towards revealing the structure of bacterial inclusion bodies. Prion 3:139–145

161. Mitraki A (2010) Protein aggregation from inclusion bodies to amyloid and biomaterials. Adv Protein Chem Struct Biol 79:89–125

162. Doi H, Mitsui K, Kurosawa M et al (2004) Identification of ubiquitin-interacting proteins in purified polyglutamine aggregates. FEBS Lett 571:171–176

163. Suhr ST, Senut MC, Whitelegge JP et al (2001) Identities of sequestered proteins in aggregates from cells with induced polyglutamine expression. J Cell Biol 153:283–294

164. Zhang X, Qian SB (2011) Chaperone-mediated hierarchical control in targeting misfolded proteins to aggresomes. Mol Biol Cell 22:3277–3288

165. Beaudoin S, Goggin K, Bissonnette C et al (2008) Aggresomes do not represent a general cellular response to protein misfolding in mammalian cells. BMC Cell Biol 9:59

166. Tanaka M, Kim YM, Lee G et al (2004) Aggresomes formed by alpha-synuclein and synphilin-1 are cytoprotective. J Biol Chem 279:4625–4631

167. Wang Y, Meriin AB, Costello CE et al (2007) Characterization of proteins associated with polyglutamine aggregates: a novel approach towards isolation of aggregates from protein conformation disorders. Prion 1:128–135

168. Meriin AB, Wang Y, Sherman MY (2010) Isolation of aggresomes and other large aggregates. Curr Protoc Cell Biol. Chapter 3:Unit-9

169. Calo-Fernandez B, Martinez-Hurtado JL (2012) Biosimilars: company strategies to capture value from the biologics market. Pharmaceuticals 5:1393–1408

170. Weinacker D, Rabert C, Zepeda AB et al (2013) Applications of recombinant in the healthcare industry. Braz J Microbiol 44:1043–1048

171. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27:297–306

172. Brondyk WH (2009) Selecting an appropriate method for expressing a recombinant protein. Methods Enzymol 463:131–147

173. Chen R (2012) Bacterial expression systems for recombinant protein production: E. coli and beyond. Biotechnol Adv 30:1102–1107

174. Grillberger L, Kreil TR, Nasr S et al (2009) Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells. Biotechnol J 4:186–201

175. Berkowitz SA, Engen JR, Mazzeo JR et al (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. Nat Rev Drug Discov 11:527–540

176. Lebendiker M, Danieli T (2014) Production of prone-to-aggregate proteins. FEBS Lett 588:236–246

Part I

Recombinant Protein Production in Escherichia coli

Chapter 2

Overcoming the Solubility Problem in *E. coli*: Available Approaches for Recombinant Protein Production

Agustín Correa and Pablo Oppezzo

Abstract

Despite the importance of recombinant protein production in academy and industrial fields, many issues concerning the expression of soluble and homogeneous product are still unsolved. Although several strategies were developed to overcome these obstacles, at present there is no magic bullet that can be applied for all cases. Indeed, several key expression parameters need to be evaluated for each protein. Among the different hosts for protein expression, *Escherichia coli* is by far the most widely used. In this chapter, we review many of the different tools employed to circumvent protein insolubility problems.

Key words Recombinant proteins, Protein expression, *E. coli*, High-throughput screening, Inclusion bodies, Directed evolution

1 Introduction

With the advances in genome sequencing nowadays, over 1,900 genomes are publicly available (http://www.microbesonline.org) generating massive information in this area. A typical microbial genome codes for between 1,500 and 8,000 proteins while in eukaryotic genomes is around 10,000–60,000 proteins. Despite all this information, and in contrast with nucleic acids, obtaining the target protein from the natural host in a soluble, homogeneous state and enough quantities for biochemical and structural studies is very uncommon. This makes the production of the target protein in a recombinant form the method of choice. Different expression hosts are available for recombinant expression, including bacterial, fungal, or eukaryotic host cells. Among these, the use of the enterobacterium *Escherichia coli* is the most commonly employed with approximately 60 % of all recombinant proteins in the literature and nearly 30 % of the currently approved recombinant therapeutic proteins produced on it [1, 2]. This is mainly due to the low cost, fast growth, easy handling, high yield of target protein and the extensive knowledge of the genetics of *E. coli*.

However, when working with eukaryotic proteins, it has being estimated that approximately only 30 % of the cloned genes can be expressed in a soluble form in *E. coli* where the rest of the targets are degraded, expressed as insoluble aggregates known as inclusion bodies (IBs) or undetectable in cell extracts [3]. This is especially the case for membrane proteins or those requiring posttranslational modifications for folding or function. In order to overcome these limitations, several *E. coli* strains were developed as well as vectors carrying promoters with different strengths, fusion of the gene of interest with molecular tags that can aid in the purification and/or soluble production of the target protein, or the co-expression of chaperones or biological partners that can improve protein folding and stability [4, 5]. Furthermore, with the advent of the high-throughput screening (HTS) technology, all these variables can be evaluated in a simultaneous, fast, automated, and reliable manner in order to find the combination of the parameters that enable a soluble protein production [6]. Despite all this, soluble and homogeneous expressions of the target protein are not always the case. In this regard, many efforts were done with some success in the refolding of insoluble proteins from IBs [3, 7].

As an alternative strategy, the introduction of rational or, moreover, random mutations into the gene of interest in order to obtain a variant with stabilized properties or increased soluble expression has shown to be an attractive and effective approach in the soluble expression of target proteins, thus being a suitable last resource when everything else fails [8, 9].

2   Common Problems When Expressing Recombinant Proteins in *E. coli*

One of the main reasons why eukaryotic proteins often fail to be produced as soluble proteins in *E. coli* is the requirement of posttranslational modifications for correct folding. So a first step could be a sequence-based prediction of these modifications. In this regard, the ExPASy server (http://www.expasy.org) contains numerous bioinformatic tools that can estimate with a good accuracy the presence or not of posttranslational modifications like N- or O-glycosylation sites, phosphorylation sites, and protein localization, among others [10]. All this information can give us an idea of the possible success and help us in the strategy to follow for protein expression. Other factors that can have an impact in the soluble expression of the target are the codon usage, the sequence at the translation initiation region (TIR), as well as the correct formation of disulfide bridges. A brief description of the strategies designed to obtain soluble recombinant proteins is given in the next sections, and finally, in-depth information of them will be given in the following chapters of this book.

*2.1 Effects of DNA/
RNA Sequence
in Protein Expression*

The presence of uncommon codons for *E. coli* can have a strong influence in the gene expression. Because of the heterologous nature of the target protein, the target gene may have codons that are in low abundance in this host. This can lead into growth arrest, premature translation termination, and low yield of protein production, among others [11].

In order to overcome this problem, two different approaches have been proposed: (1) the substitution of the rare codons present in the gene sequence by de novo gene synthesis or (2) the expression of the gene of interest in an *E. coli* strain that is supplemented by tRNAs that are present in low abundance. In the former case, several algorithms were developed in order to optimize the gene sequence to the host codon usage [12, 13]. More recently, a software was developed in order to not only evaluate the codon frequency but also the codon pair usage or codon context. This approach suggests that codon pair usage and codon context can be as important as the individual codon usage [14]. For the second strategy, several commercially available strains have been developed that co-express tRNAs for rare codons, like BL21 CodonPlus (Novagen) and Rosetta™ (Invitrogen). The use of such strains demonstrated to be effective for the soluble expression of several targets [15, 16].

Finally, changing the rare codons can increase the translation rate, but in some cases this can lead to protein aggregation and misfolding as it was demonstrated for several proteins expressed in *E. coli* [17]. This suggests that translation pauses can be necessary in some cases for proper folding of individual domains [18]; thus the procedure of gene optimization or the use of a codon optimized for an *E. coli* strain cannot be used as a general rule.

Also at the DNA sequence level, it has been shown that the sequence at the 5′ of the gene can have an important impact in the levels of protein expression due to the generation of secondary structures in the messenger RNA that can hamper the translation by the ribosome complex. In this regard, it was shown that sequences immediately after the start codon up to position +25 can have a profound effect in protein expression. For these reasons, there are some programs that enable the optimization of the TIRs in order to improve protein expression by defining silent mutations in the first seven codons [19]. More recently a predictive method for designing synthetic ribosome binding sites was developed where different translation initiation rates can be targeted, thus enabling the rational control and fine tuning of recombinant protein expression [20, 21].

In the same way, in bacteria, the half-life of mRNA is much shorter than in eukaryotic cells. It was shown that mutation in the gene coding for RNaseE confers increased mRNA stability [22]. A BL21 derivative strain containing such mutation is commercialized by Invitrogen under the name of BL21 Star.

2.2 Disulfide Bonds: A Common Posttranslational Modification Implies a Common Problem for Recombinant Protein Expression

Disulfide bonds correspond to a covalent linkage between two sulfur atoms from two cysteine residues. They are frequently essential for proper folding, stability, and/or function of the target protein, thus a very important feature to take into account when expressing a target gene [23]. The presence of disulfide bonds can be predicted by web-based servers in order to estimate if the target protein can require such posttranslational modification [24, 25].

Disulfide bonds are formed in oxidizing environments like the eukaryotic endoplasmic reticulum or the bacterial periplasm. The formation of disulfide bridges in the periplasmic of E. coli, requires the action of DsbC system were DsbA catalyze disulfide formation whiles DsbC catalyze the isomerization of incorrectly formed disulfide bridges. The cycle can be restarted by the actions of the membrane proteins DsbB and DsbD that recycle DsbA and DsbC, respectively [26]. The expression of recombinant proteins in the periplasm of *E. coli* has allowed the correct formation of disulfide bridges of several targets [26, 27]. Purification of proteins from the periplasm is usually easier than purification of proteins from total cell lysates, since the periplasm contains a less complex protein mixture than the cytoplasm [28]. Targeting proteins to the periplasm of *E. coli* can be achieved by the addition of an N-terminal leader peptide that, depending on its nature, can use the Sec (relatively slow, posttranslational translocation) or the SRP (fast, cotranslational translocation) pathways that transport proteins through the inner plasma membrane as unfolded precursors [29, 30]. There is another translocation system: the twin-arginine translocation pathway, named Tat pathway, that, in contrast with the aforementioned pathways, catalyzes the translocation of proteins in their folded state [31].

However, one common drawback of periplasmic expression is that the translocation machinery can be saturated, which can be toxic for the host cell and decrease the final yield of the target protein. By using a strain where the expression intensity can be precisely controlled like Lemo21(DE3) (New England Biolabs), the saturation of the translocation machinery can be avoided, and thus these negative effects are minimized [32, 33].

As an alternative to periplasmic expression, engineered *E. coli* strains that contain a more oxidizing cytoplasm were developed in order to improve disulfide bond formation in this compartment. These strains contain mutations in the genes for glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) involved in the maintenance of the reduced environment in the cytoplasm and a mutation in the peroxiredoxin gene *ahpC* essential for restoring growth in these mutants [23, 34]. One strain containing such mutations and used for the expression of disulfide bridges containing proteins is Origami, commercialized by Novagen [35]. However, a common problem for using such strains is the lack of disulfide bond isomerization. In this regard, a strain containing the trxB/gor/ahpC

mutations that express the DsbC isomerase in the cytoplasm of *E. coli* was developed and commercialized by New England Biolabs known as Shuffle, allowing the soluble expression of some disulfide-containing proteins in its cytoplasm [36]. Recently, by the co-expression of the sulfhydryl oxidase from *S. cerevisiae* Erv1p and the *E. coli* disulfide isomerase DsbC, disulfide bonds were generated in proteins expressed in the *E. coli* cytoplasm with the reducing pathways intact. Moreover, for some cases it was shown that the addition of a catalyst for the formation of disulfide bonds could be more effective than the removal of the reducing pathways [37, 38]. In the same sense, after making N-terminal fusions with DsbC with 28 different small disulfide-rich proteins, it was found that the strain BL21(DE3)pLysS was much more efficient in producing soluble and oxidized folded proteins in comparison to Origami B(DE3)pLysS or Shuffle T7 Express lysY cells [39]. Interestingly, when one of the fusions was used to evaluate if the disulfide bound formation occurred in the cytoplasm of BL21(DE3)pLysS cells or during the extraction and purification steps, it was found that this process occurred ex vivo [39].

3   Boosting Protein Purification and/or Expression

3.1   The Use of Fusion Tags/ Proteins

With the advent of genetic engineering, the target gene can be easily cloned in frame with different affinity and/or solubility-enhancing tags that can be exploited to increase protein solubility and yield and facilitate protein purification or downstream processing. In this regard, we can separate the fusion tags in three main categories. In the first category, referred as affinity tags, we found short tags that can be placed as N-terminus or C-terminus of the partner and can be recognized by special matrices or molecules serving for affinity purification of the fusion protein. In the second category, extremely soluble proteins with chaperone activities or thermostable characteristics in some cases are fused in order to transfer some of these properties to the fusion partner and improve the folding and/or the final yield of the target protein. Usually these tags are expressed as N-terminal fusions and are termed solubility-enhancing tags. Finally, we also have proteins that can offer a double purpose, in one hand, can be recognized by other molecules, thus serving for purification purposes, and, in the other hand, can improve the soluble production of the target protein, thus improving the target protein purity and yield [4, 40, 41].

3.2   Affinity Tags

Among the affinity tags, the His-tag is one of the most commonly used for purification of recombinant proteins in *E. coli*. This small tag (0.84 kDa) consists of 6–10 histidines in tandem and can reversibly interact with metal ions most commonly Ni or Co immobilized in a metal chelate matrix (Ni-NTA, Qiagen; Sepharose 6,

GE; or Talon resins, Clontech) [42], thus allowing mild elution conditions like the use of a competitor such as imidazole. The His-tag has several advantages like its small size and relatively inert nature, making it compatible with most downstream applications. Because the ternary structure of the His-tag is not necessary for metal coordination, it is possible to purify the protein in denaturing conditions or even perform the refolding procedure on column [43, 44]. Also the purification scheme has been automated in small and large-scale formats and has been used widely in HTS protocols, demonstrating the versatility of this tag [45–47].

As a disadvantage, when working with low-yield expressed proteins, it was shown that increasing the culture volume does not correlate with an increase in recovery. Moreover, there is a decrease in recovery because of the presence of small chelators mainly associated with the periplasm of *E. coli* that can decrease the binding capacity of the purification resin [48]. This can be improved by removing the periplasmic material before cell lysis [48]. Also it was shown that several histidine-rich *E. coli* proteins can bind to the column (like ArnA, SlyD, and GlmS), especially when working with low-expressing protein targets [49]. This reduces the purity of the target protein, consequently requiring the addition of more purification steps, thus reducing the final yield.

In this regard, some *E. coli* strains that are mutants in some of these proteins have been developed in order to overcome this issue [50, 51], and one is commercially available as NiCo21 (New England Biolabs).

Another strategy is the use of an alternative affinity tag such as Strep-tag II. This is also a small tag consisting of eight residues (WSHPQFEK) and can be specifically recognized by an engineering version of streptavidin (Strep-Tactin) [52]. Elution can be done as for the case of His-tag using mild conditions by competition with D-desthiobiotin for the Strep-tag II. Despite the binding capacity of the Strep-Tactin containing media can be lower when comparing to Sepharose 6 resins for His-tagged proteins, for example, its greater specificity makes it a good option when working with proteins that are expressed in very low quantities [52, 53]. Purification schemes for the Strep-tag II include prepacked columns as well as 96× well plates (www.iba-lifesciences.com). A variation of the Strep-tag II named Twin-Strep-tag® was recently developed and exhibited higher stability and affinity for the interaction with the Strep-Tactin. This tag consists of two *Strep*-tag®II-binding sequences connected by a short linker and showed to be more suitable for purification of diluted samples [54].

3.3 Solubility-Enhancing Tags

A common strategy to overcome the solubility problem is to fuse the target protein with a very stable and soluble one that can drive the resulting expression. It was shown for many proteins that were not soluble when expressed alone, that when expressed

as a fusion with other protein can be produced in a soluble and homogeneous state. Moreover, after cleavage and removal of the fusion partner, the target remained soluble demonstrating the utility of this approach [6, 39, 55, 56].

Among the commonly used solubility-enhancing fusion proteins, we can find the maltose-binding protein (MBP), glutathione S-transferase (GST), thioredoxin A (TrxA), disulfide isomerase C (DsbC), small ubiquitin-like modifier protein (SUMO), and N-utilization substance A (NusA).

An attractive feature of MBP and GST is that they can be used also as affinity tags. MBP is a 42 kDa protein expressed in the *E. coli* periplasm and can bind strongly to amylose resins, and elution can be done with free maltose [57]. For the case of GST, it is a 26 kDa protein from *Schistosoma japonicum* that can bind to glutathione resins, and elution is achieved by the application of reduced glutathione allowing a single-step purification process [58]. Despite GST protein is widely used, it has been shown to be a poor solubility enhancer, since in many cases after cleavage of the fusion, the target protein precipitates [6, 55, 59]. However, expression can be improved for some proteins or peptides, and the purification by glutathione resins makes it still an attractive option. Vectors for the expression of GST fusions can be found in the pGEX series from GE Healthcare or pET41a-c/pET42a-c from Novagen.

MBP was fused to either N- or C-terminus, where the expression and folding of eukaryotic fusion proteins was increased in many cases [59–61]. Vectors for MBP fusion can be found in the pMAL series from New England Biolabs or pIVEX from Roche. Also if the natural signal peptide of MBP is present (MalE$_{ss}$), expression can be directed to the periplasm of *E. coli*. This was used recently for the successful expression of disulfide-rich venom peptides [27].

TrxA is an 11.6 kDa *E. coli* thermostable (Tm: 85 °C) oxido-reductase that is expressed in very high yields. When used as a fusion tag, some of these properties can be transferred to the target protein improving its folding, solubility, and stability [62–64]. Moreover in a comparative study, all positive hits with Trx-fusions, remained still soluble after tag cleavage [6]. Expression vectors for fusion with Trx are pET32a-c from Novagen.

SUMO is a yeast protein (11.2 kDa) that when used as N-terminal fusion protein during prokaryotic expression can promote folding and soluble expression of the target protein [65–67]. Another advantage of this fusion is that it can be cleaved by a specific and efficient protease (yeast Ulp1) which recognize tertiary structure elements and a Gly-Gly-containing motif in the C-terminus of SUMO and can leave a native N-terminus on the target (except for proline) [66].

The *E. coli* disulfide isomerase DsbC (25 kDa) has isomerase and chaperonin activities [34] and has been successfully used as a fusion partner for the soluble expression of disulfide-containing

targets as mentioned earlier [39, 68]. The pET40 (Novagen) expression vector allows fusion with DsbC.

Finally, the transcription elongation and anti-termination factor of *E. coli* NusA (55 kDa) have also demonstrated to be useful for enhancing soluble protein expression [69]. In a comparative study after using several aggregation-prone target proteins, it was shown that the solubility-enhancing properties of NusA were comparable and similar to the well-studied MBP validating its utility [70]. Fusion with NusA can be achieved with the pET43.1a-c and pET44a-c vector series from Novagen.

Because TrxA, SUMO, DsbC, and NusA do not facilitate purification on their own, they are used in conjunction with small affinity tags like the aforementioned His-tag or Strep-tag II to enable protein purification. It is important to underline that despite some trends in fusion proteins were found in several studies, there is no rule for which is the best suited for the protein of interest, so it is better to test several different fusions in order to find the best option.

3.4 Tag Removal   Once the protein is expressed, in most of the cases, it is necessary to remove the fusion tag. This can be achieved by incorporating an aminoacidic sequence between the fusion tag and the protein of interest that can be recognized by a specific protease. Several proteases appear as possible options for tag removal like enterokinase (DDDDK′X), factor Xa (IE/DGR′X, where X can be any residue except for R or P), thrombin (LVPR′GS), PreScission™ protease (GE Healthcare, LEVLFQ′GP), and tobacco etch virus (TEV) protease (ENLYFQ′G), among others [41]. Between these, TEV protease is a very specific protease with several advantages like that it can be produced in the lab with high yields in *E. coli* [71], and cleavage can be done at 4 °C. Moreover, despite reducing conditions are optimal for cleavage (usually 1 mM DTT), if avoided, cleavage can still occur [27] which is preferable for disulfide-containing proteins. Also, it was demonstrated that the last glycine residue from the cleavage recognition site can be substituted by all residues except for proline, but at the expense of cleavage efficiency, allowing the release of a target protein with a native N-terminus [72].

Finally, fusion proteins were not only used for expression/purification purposes, but they were also used to obtain the crystallographic structures of several targets. This last brings the additional advantage that if the structure of the fusion is known, it can also help in the structure determination process of the target protein. Such is the case for some fusions with MBP, GST, Trx, and GFP, among others [73–76].

3.5 Cloning Methods   In order to succeed in the soluble expression of a "difficult" target protein, a recommended strategy is to test different fusion proteins, which requires the cloning of the gene of interest in several vectors. Doing this by restriction-based methods can be a complicated task,

principally when different restriction sites are present and even further when working with several targets at the same time. Nowadays some methodologies were developed as alternatives to the restriction-based cloning to facilitate the easy transfer of a DNA fragment into several vectors. Commercial kits like Gateway (Invitrogen) [77] and In-FusionTM (Clontech) [78] are efficient recombination-based cloning methods. For the case of Gateway, a suite of vectors for the easy transfer of the same DNA fragment between vectors is available. More recently, a cloning method based only in PCR reactions was developed and initially termed as RF cloning (RF, restriction free) [79]. In this method, the DNA is amplified with primers that contain complementary sequences with the site of insertion in the destination plasmid. So after the first PCR, the generated megaprimer is used in a second PCR to amplify the whole plasmid, inserting in this reaction the gene of interest in the desired position. The advantage is that insertion can be done at any position in the destination vector avoiding extra sequences to be added to the gene of interest, and if several vectors contain the same insertion sequence, the same megaprimer can be used in all of them, facilitating the cloning stage and allowing an automated HTS cloning approach [4, 79]. So by using a vector containing a fusion protein, just by inserting in the same position of the fusion other genes (like MBP, GST, SUMO, etc.), one can easily make its own suite of expression vectors where the site of insertion for the target gene is conserved along all vectors [80]. Recently, an improved protocol for RF cloning termed Transfer-PCR was developed where the generation of the megaprimer and subsequent integration of the PCR product into the destination vector occur in a single PCR reaction [81]. A web-based tool was developed for the correct design of the primers for RF cloning and is freely available (http://www.rf-cloning.org) [82]. The use of this kind of tools for molecular cloning is very useful for the generation of the genetic constructs necessary for finding a condition for soluble expression.

3.6 Expression Conditions

At the culture level, several parameters like induction temperature and medium composition can have an important effect in soluble protein yields. It was shown that lower temperature during induction (16–25 °C) can increase the final yield of soluble protein. It was assumed that a slower translation rate could favor the correct folding of the protein [83]. However, the lower temperature can also decrease the final biomass, so if the protein is well expressed, this can hamper the final yield [6]. In general, it is necessary to evaluate different temperatures to find the optimal condition. At the medium level, several media have been used for protein expression: Luria Broth (LB), 2xYT, Terrific Broth (TB), Super Broth (SB), autoinduction medium, and others. Among these media, the autoinduction medium, developed by Studier [84], has been used with success for protein expression screening in a wide range of

scales because it produces a high level of biomass. Thus, there is no need to monitor the growth; induction of cultures in well plates occurs at a comparable growth phase, which is preferable in HTS experiments; and there is a tighter control of protein induction improving expression of toxic proteins [6, 84]. A disadvantage of this medium is that it is adversely affected by aeration level. This can be reduced by decreasing the level of lacI repressor provided by the expression vector [85]. Recently, it was demonstrated that the oxygen sensitivity of expression in autoinduction medium can be practically obviated. This was achieved by using a glucose fed-batch-based autoinduction medium where the glycerol carbon source was substituted with the EnBase system [86]. This system is based on a soluble polysaccharide component within the medium and slow release of the glucose units from the polymer chain by an added specific biocatalyst [85, 87]. This kind of rich media allows an increase in the biomass production, so expression conditions can be evaluated in a reduced format like a 24× deep wells, enhancing the sensibility of automated HTS screenings for soluble protein production [4, 85, 88]. Also and as it was mentioned along the text, several strains should be used in order to find the proper condition; thus a combination of temperature and strain should be included in the screening. These in conjunction with the use of different constructs (i.e., fusion tags) make a considerable number of conditions to evaluate. In this regard, the HTS methods have had a pivotal role in making this kind of approaches possible [4, 6, 88, 89].

4   Inclusion Bodies' Renaturation

Frequently proteins accumulate, as insoluble aggregates in the cytoplasm or periplasm of *E. coli* known as inclusion bodies (IBs). As dramatically as it seems, this is not always a negative issue. Some advantages of expressing the protein as IBs are the high yield of its expression and the homogeneity in composition where in some cases the recombinant target can account for more than 90 % of the proteins in that fraction, facilitating the purification of the target after renaturation [90]. Renaturation conditions involve the evaluation of several parameters like pH, ionic strength, temperature, and addition of low molecular weight compounds, among others. In this regard, several approaches in a 96× well format have been developed to facilitate the optimization of the refolding conditions, and automated HTS protocols for protein refolding were proposed [7, 91]. Apart from the mentioned parameters, these can be combined with several methods to perform the refolding process like dilution, dialysis, or in-column refolding methods [7, 43, 92].

An attractive and counter-intuitive strategy is to introduce a tag that reduces the solubility of the fusion protein and can direct the expressed protein into insoluble IBs. This is particularly useful if the target protein is toxic to the host when soluble and correctly folded. In this regard, a mutant variant of the N-terminal autoprotease N^{pro}, of classical swine fever virus termed EDDIE, when fused to the N-terminus of the target protein can reduce its solubility in such a way that the fusion accumulates as IBs. When changing from chaotropic to kosmotropic conditions, the protease becomes active and can perform the autocleavage of the fusion, leaving a native N-terminus in the target protein [93, 94]. The comprehension of IBs nature has dramatically changed in the last years. Often, it was assumed that IBs were made of inert aggregates composed of denatured or partially folded polypeptides rather from mature native molecules. Nevertheless, in the last decades, it was shown in several cases that IBs can be made with native and active proteins [95–98]. This opens the possibility of using them in downstream applications without the need of performing protein renaturation in applications where protein aggregation is not an impediment, thus facilitating production/purification and reducing costs [99–101].

5   Protein Characterization

Obtaining the protein in a soluble state does not assure proper folding of the target protein. A common scenario is to find that the protein is soluble but forms aggregates. This is indicative of unfolded regions. A last purification step by size-exclusion chromatography (SEC) is recommendable to not only remove some remaining impurities but also to assess the oligomeric state of the sample. Protein quality assessment, can be implemented at the analytical level, with microgram quantities of protein by coupling for example, Ni Sepharose 6 beads or His MultiTrap FF 96-well plates (GE Healthcare) with the minicolumns for analytical SEC (ASEC), Superdex™ 5/150 GL (GE Healthcare), when still evaluating different expression conditions [88, 102, 103]. By using an autosampler for ASEC, the characterization step can be completely automated requiring only 14 min for each sample [102]. Also, sometimes it is necessary to evaluate different combination of additives, like for the case of membrane proteins, a combination of different detergents and/or lipids and genetic constructs in order to find a condition that gives a soluble and homogeneous sample. This kind of screening requires the purification of microgram to milligram of protein. A very useful alternative is to make GFP covalent fusions with the target protein and performing fluorescence-detection size-exclusion chromatography (FSEC). By using this approach, it is possible to determine the soluble expression

level, oligomeric state, thermostability, and approximate molecular mass using only nanogram quantities of unpurified protein, allowing working directly with the soluble extracts [104, 105]. Recently, a similar approach was developed where instead of fusing the target protein with GFP, a special fluorescent probe that can specifically recognize the His-tag was used, thus overcoming the limitations that can be associated in some cases with GFP fusions like the presence of false positives or protein aggregation issues following fusion cleavage [106].

6   Directed Evolution for Soluble Protein Expression

Despite the evaluation of many expression and growth conditions, it is often impossible to obtain the target protein in a soluble and stable manner. Under these circumstances, instead of exploring more expression parameters, one can change the physical properties of the target by making mutations or deletions in the target sequence in order to improve the solubility/stability of the recombinant protein. When structural and functional information are available, these sequence modifications can be achieved by rationally designed site-directed mutagenesis [107, 108]. Unfortunately, for most of the interesting targets, structural information is not available so rational design is not possible. In these cases, an interesting alternative is the use of directed evolution. This approach is based on an iterative process consisting of a first step of sequence diversification followed by a second step of selection of the improved mutants. The diversification process is usually achieved by random mutagenesis (error-prone PCR, chemical mutagenesis, or a mutator *E. coli* strain) [109] and/or in vitro recombination (DNA shuffling) [110]. In the directed evolution approaches, a library of mutants generated by a random process is screened for the solubility/stability of the target protein. So, after mutation occurs, one must select those few mutants with the improvements in the desired phenotype among the millions of futile mutants generated. In this regard, one can perform the selection by analyzing the activity of a reporter protein (reporter tag) or in special cases the activity of the target protein [111].

One folding reporter tag that was used successfully for the evolution of active and soluble mutant variants is the GFP-folding reporter [112, 113]. In this system, the test protein is expressed as an N-terminal fusion with GFP. So the fluorescence of *E. coli* cells is directly related to the productive folding of the fused protein [112]. In this way, the isolation of the brightest cells in the search for the mutations that improve solubility can be done using simple colony-plating techniques or fluorescence-activated cell sorting (FACS) in a flow cytometer. Later this system was improved even further by the design of a self-complemented split GFP [114]

derived from an exceptionally well-folded variant of GFP, "super-folder GFP" [115]. In this case, the target protein is fused as an N-terminal fusion to a small GFP fragment (residues 215–230, GFP11), while the GFP detector fragment (residues 1–214, GFP1-10) is expressed separately in another vector. So if the target protein is expressed in a soluble form, the GFP11 fragment can interact with GFP1-10, leading to the development of fluorescence [114].

In a different approach, the target protein can be expressed as an N-terminal fusion with a selectable marker such as the chloramphenicol acetyltransferase (CAT; 25 kDa), thus conferring resistance to chloramphenicol. It was observed that if the fusion protein is expressed in a soluble form, the cell is resistant to higher concentrations of chloramphenicol than when it is expressed in an insoluble form [116]. By using this method, it was possible to obtain soluble variants of the membrane-associated human cytochrome P450 (1A2), confirming the usefulness of this method [117].

More recently another antibiotic was used as a selectable marker but in a split manner linking in vivo protein stability to antibiotic resistance. In this case the target protein is inserted into the TEM1-β-lactamase (resistance to β-lactam antibiotics) as part of a tripartite fusion [8]. The antibiotic-resistance gene is separated between residues 196 and 197, for the insertion of the target protein gene. Thus, when protein is expressed in a soluble and stable form, the two fragments of β-lactamase can interact and thereby confer resistance to β-lactam antibiotics [8]. This method showed a low false-positive rate and, as for the CAT, is based on a selection rather than a screening for obtaining improved mutants.

Another elegant approach is the colony filtration (CoFi) blot. This is based in the fact that IBs can be separated from soluble proteins by filtration at the colony level. So after transforming bacteria with the mutant library, colonies are transferred to a filter membrane where protein expression is induced and cells are then lysed. Soluble proteins can diffuse through the filter and bind to the nitrocellulose membrane for detection [118, 119]. An anti-His antibody can be used for the detection of His-tagged soluble variants making it an easy to adopt method. Cornvik and colleagues randomized the N-terminal region of 32 mammalian proteins, and mutants were selected for soluble expression using this methodology. By this approach, the success rate for soluble expression was increased from 34 to 68 %, showing the high potential of this methodology [118].

Just as in the HTS, usually many different expression conditions for the same protein are evaluated; in the directed evolution approach, a library of mutants generated by a random process is screened for the solubilization/stabilization of the target protein. The key issues in this strategy are the diversity of the library and the selection/isolation method employed for finding the mutant with the improved characteristics.

7 Conclusions and Future Perspectives

Although a lot of progress has been made in recombinant protein expression, this field is still far for the generation of a universal protocol, so many different parameters are necessary to be evaluated for each target. The development of robotic technologies has facilitated the evaluation of an important number of different conditions reducing cost and effort through miniaturization of experiments. At present there are novel technological approaches (strain engineering, fusion technologies, and protein purification, among others), which are key factors that should be used in the lab to increase the success for the production of a soluble and homogeneous target protein.

Acknowledgments

References

1. Sorensen HP (2010) Towards universal systems for recombinant gene expression. Microb Cell Fact 9:27

2. Huang CJ, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. J Ind Microbiol Biotechnol 39:383–399

3. Yang Z, Zhang L, Zhang Y et al (2011) Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. PLoS One 6:e22981

4. Correa A, Oppezzo P (2011) Tuning different expression parameters to achieve soluble recombinant proteins in E. coli: advantages of high-throughput screening. Biotechnol J 6:715–730

5. Samuelson JC (2011) Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations. Methods Mol Biol 705:195–209

6. Vincentelli R, Cimino A, Geerlof A et al (2011) High-throughput protein expression screening and purification in Escherichia coli. Methods 55:65–72

7. Vincentelli R, Canaan S, Campanacci V et al (2004) High-throughput automated refolding screening of inclusion bodies. Protein Sci 13:2782–2792

8. Foit L, Morgan GJ, Kern MJ et al (2009) Optimizing protein stability in vivo. Mol Cell 36:861–871

9. Hart DJ, Waldo GS (2013) Library methods for structural biology of challenging proteins and their complexes. Curr Opin Struct Biol 23:403–408

10. Artimo P, Jonnalagedda M, Arnold K et al (2012) ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 40:W597–W603

11. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22: 346–353

12. Puigbo P, Guzman E, Romeu A et al (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35:W126–W131

13. Villalobos A, Ness JE, Gustafsson C et al (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7:285

14. Chung BK, Lee DY (2012) Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol 6:134

15. Burgess-Brown NA, Sharma S, Sobott F et al (2008) Codon optimization can improve expression of human genes in Escherichia coli: a multi-gene study. Protein Expr Purif 59:94–102

16. Tegel H, Tourle S, Ottosson J et al (2010) Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). Protein Expr Purif 69: 159–167

17. Rosano GL, Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. Microb Cell Fact 8:41

18. Marin M (2008) Folding at the rhythm of the rare codon beat. Biotechnol J 3:1047–1057

19. Voges D, Watzele M, Nemetz C et al (2004) Analyzing and enhancing mRNA translational efficiency in an Escherichia coli in vitro expression system. Biochem Biophys Res Commun 318:601–614

20. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27:946–950

21. Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498:19–42

22. Makino T, Skretas G, Georgiou G (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microb Cell Fact 10:32

23. Salinas G, Pellizza L, Margenat M et al (2011) Tuned Escherichia coli as a host for the expression of disulfide-rich proteins. Biotechnol J 6:686–699

24. Ferre F, Clote P (2005) DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 33:W230–W232

25. Lin HH, Tseng LY (2010) DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines. Nucleic Acids Res 38:W503–W507

26. Berkmen M (2012) Production of disulfide-bonded proteins in Escherichia coli. Protein Expr Purif 82:240–251

27. Klint JK, Senff S, Saez NJ et al (2013) Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the periplasm of E. coli. PLoS One 8:e63865

28. Mergulhao FJ, Summers DK, Monteiro GA (2005) Recombinant protein secretion in Escherichia coli. Biotechnol Adv 23:177–202

29. den Blaauwen T, Driessen AJ (1996) Sec-dependent preprotein translocation in bacteria. Arch Microbiol 165:1–8

30. Luirink J, Sinning I (2004) SRP-mediated protein targeting: structure and function revisited. Biochim Biophys Acta 1694:17–35

31. Natale P, Bruser T, Driessen AJ (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. Biochim Biophys Acta 1778:1735–1756

32. Wagner S, Klepsch MM, Schlegel S et al (2008) Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A 105:14371–14376

33. Schlegel S, Rujas E, Ytterberg AJ et al (2013) Optimizing heterologous protein production in the periplasm of E coli by regulating gene expression levels. Microb Cell Fact 12:24

34. de Marco A (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb Cell Fact 8:26

35. Bessette PH, Aslund F, Beckwith J et al (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc Natl Acad Sci U S A 96: 13703–13708

36. Lobstein J, Emrich CA, Jeans C et al (2012) SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact 11:56

37. Hatahet F, Nguyen VD, Salo KE et al (2010) Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. Microb Cell Fact 9:67

38. Nguyen VD, Hatahet F, Salo KE et al (2010) Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E. coli. Microb Cell Fact 10:1

39. Nozach H, Fruchart-Gaillard C, Fenaille F et al (2013) High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in E. coli. Microb Cell Fact 12:37

40. Walls D, Loughran ST (2011) Tagging recombinant proteins to enhance solubility and aid purification. Methods Mol Biol 681:151–175

41. Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J 7:620–634

42. Murphy MB, Doyle SA (2005) High-throughput purification of hexahistidine-tagged proteins expressed in E. coli. Methods Mol Biol 310:123–130

43. Zhu XQ, Li SX, He HJ et al (2005) On-column refolding of an insoluble His6-tagged recombinant EC-SOD overexpressed in Escherichia coli. Acta Biochim Biophys Sin (Shanghai) 37:265–269

44. Li M, Su ZG, Janson JC (2004) In vitro protein refolding by chromatographic procedures. Protein Expr Purif 33:1–10

45. Schafer F, Romer U, Emmerlich M et al (2002) Automated high-throughput purification of 6xHis-tagged proteins. J Biomol Tech 13:131–142

46. Vincentelli R, Canaan S, Offant J et al (2005) Automated expression and solubility screening of His-tagged proteins in 96-well format. Anal Biochem 346:77–84

47. Steen J, Uhlen M, Hober S et al (2006) High-throughput protein purification using an automated set-up for high-yield affinity chromatography. Protein Expr Purif 46:173–178

48. Magnusdottir A, Johansson I, Dahlgren LG et al (2009) Enabling IMAC purification of low abundance recombinant proteins from E. coli lysates. Nat Methods 6:477–478

49. Bolanos-Garcia VM, Davies OR (2006) Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta 1760:1304–1313

50. Robichon C, Luo J, Causey TB et al (2011) Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. Appl Environ Microbiol 77:4634–4646

51. Andersen KR, Leksa NC, Schwartz TU (2013) Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins 81:1857–1861

52. Schmidt TG, Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protoc 2:1528–1535

53. Lichty JJ, Malecki JL, Agnew HD et al (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41:98–105

54. Schmidt TG, Batz L, Bonet L et al (2013) Development of the Twin-Strep-tag(R) and its application for purification of recombinant proteins from cell culture supernatants. Protein Expr Purif 92:54–61

55. Hammarstrom M, Hellgren N, van Den Berg S et al (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci 11:313–321

56. Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17:353–358

57. Pattenden LK, Thomas WG (2008) Amylose affinity chromatography of maltose-binding protein: purification by both native and novel matrix-assisted dialysis refolding methods. Methods Mol Biol 421:169–189

58. Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67:31–40

59. Dyson MR, Shadbolt SP, Vincent KJ et al (2004) Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnol 4:32

60. Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8:1668–1674

61. Cho HJ, Lee Y, Chang RS et al (2008) Maltose binding protein facilitates high-level expression and functional purification of the chemokines RANTES and SDF-1alpha from Escherichia coli. Protein Expr Purif 60:37–45

62. LaVallie ER, Lu Z, Diblasio-Smith EA et al (2000) Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. Methods Enzymol 326:322–340

63. Kim S, Lee SB (2008) Soluble expression of archaeal proteins in Escherichia coli by using fusion-partners. Protein Expr Purif 62:116–119

64. LaVallie ER, DiBlasio EA, Kovacic S et al (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology (N Y) 11:187–193

65. Marblestone JG, Edavettal SC, Lim Y et al (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci 15:182–189

66. Malakhov MP, Mattern MR, Malakhova OA et al (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics 5:75–86

67. Butt TR, Edavettal SC, Hall JP et al (2005) SUMO fusion technology for difficult-to-express proteins. Protein Expr Purif 43:1–9

68. Zhang Z, Li ZH, Wang F et al (2002) Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in Escherichia coli. Protein Expr Purif 26:218–228

69. De Marco V, Stier G, Blandin S et al (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochem Biophys Res Commun 322:766–771

70. Nallamsetty S, Waugh DS (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expr Purif 45:175–182

71. van den Berg S, Lofdahl PA, Hard T et al (2006) Improved solubility of TEV protease by directed evolution. J Biotechnol 121:291–298

72. Kapust RB, Tozser J, Copeland TD et al (2002) The P1′ specificity of tobacco etch virus protease. Biochem Biophys Res Commun 294:949–955

73. Moon AF, Mueller GA, Zhong X et al (2010) A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci 19:901–913

74. Suzuki N, Hiraki M, Yamada Y et al (2010) Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr D Biol Crystallogr 66:1059–1066

75. Smyth DR, Mrozkiewicz MK, McGrath WJ et al (2003) Crystal structures of fusion proteins with large-affinity tags. Protein Sci 12:1313–1322

76. Corsini L, Hothorn M, Scheffzek K et al (2008) Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci 17:2070–2079

77. Esposito D, Garvey LA, Chakiath CS (2009) Gateway cloning for protein expression. Methods Mol Biol 498:31–54

78. Berrow NS, Alderton D, Sainsbury S et al (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res 35:e45

79. Unger T, Jacobovitch Y, Dantes A et al (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol 172:34–44

80. Correa A, Ortega C, Obal G, Alzari P, Vincentelli R, Oppezzo P (2014) Generation of a vector suite for protein solubility screening. Front Microbiol. 5: 67

81. Erijman A, Dantes A, Bernheim R et al (2011) Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. J Struct Biol 175:171–177

82. Bond SR, Naus CC (2012) RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res 40:W209–W213

83. Vera A, Gonzalez-Montalban N, Aris A et al (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng 96:1101–1106

84. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234

85. Blommel PG, Becker KJ, Duvnjak P et al (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. Biotechnol Prog 23:585–598

86. Ukkonen K, Mayer S, Vasala A et al (2013) Use of slow glucose feeding as supporting carbon source in lactose autoinduction medium improves the robustness of protein expression at different aeration conditions. Protein Expr Purif 91:147–154

87. Krause M, Ukkonen K, Haataja T et al (2010) A novel fed-batch based cultivation method provides high cell-density and improves yield of soluble recombinant proteins in shaken cultures. Microb Cell Fact 9:11

88. Vincentelli R, Romier C (2013) Expression in Escherichia coli: becoming faster and more complex. Curr Opin Struct Biol 23:326–334

89. Koehn J, Hunt I (2009) High-throughput protein production (HTPP): a review of enabling technologies to expedite protein production. Methods Mol Biol 498:1–18

90. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24:179–185

91. Dechavanne V, Barrillat N, Borlat F et al (2010) A high-throughput protein refolding screen in 96-well format combined with design of experiments to optimize the refolding conditions. Protein Expr Purif 75:192–203

92. Clark EDB (1998) Refolding of recombinant proteins. Curr Opin Biotechnol 9:157–163

93. Achmuller C, Kaar W, Ahrer K et al (2007) N(pro) fusion technology to produce proteins with authentic N termini in E. coli. Nat Methods 4:1037–1043

94. Ke T, Liang S, Huang J et al (2012) A novel PCR-based method for high throughput prokaryotic expression of antimicrobial peptide genes. BMC Biotechnol 12:10

95. Tokatlidis K, Dhurjati P, Millet J et al (1991) High activity of inclusion bodies formed in Escherichia coli overproducing Clostridium thermocellum endoglucanase D. FEBS Lett 282:205–208

96. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactiva-

tion of enzymes and fluorescent proteins. Microb Cell Fact 4:27

97. de Groot NS, Ventura S (2006) Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. J Biotechnol 125:110–113

98. Peternel S, Grdadolnik J, Gaberc-Porekar V et al (2008) Engineering inclusion bodies for non denaturing extraction of functional proteins. Microb Cell Fact 7:34

99. Garcia-Fruitos E (2010) Inclusion bodies: a new concept. Microb Cell Fact 9:80

100. Garcia-Fruitos E, Vazquez E, Diez-Gil C et al (2012) Bacterial inclusion bodies: making gold from waste. Trends Biotechnol 30:65–70

101. Villaverde A, Garcia-Fruitos E, Rinas U et al (2012) Packaging protein drugs as bacterial inclusion bodies for therapeutic applications. Microb Cell Fact 11:76

102. Low C, Moberg P, Quistgaard EM et al (2013) High-throughput analytical gel filtration screening of integral membrane proteins for structural studies. Biochim Biophys Acta 1830:3497–3508

103. Sala E, de Marco A (2010) Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography. Protein Expr Purif 74:231–235

104. Hattori M, Hibbs RE, Gouaux E (2012) A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. Structure 20:1293–1299

105. Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14:673–681

106. Backmark AE, Olivier N, Snijder A et al (2013) Fluorescent probe for high-throughput screening of membrane protein expression. Protein Sci 22:1124–1132

107. Dale GE, Broger C, Langen H et al (1994) Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. Protein Eng 7:933–939

108. Eijsink VG, Bjork A, Gaseidnes S et al (2004) Rational engineering of enzyme stability. J Biotechnol 113:105–120

109. Rasila TS, Pajunen MI, Savilahti H (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. Anal Biochem 388:71–80

110. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391

111. Roodveldt C, Aharoni A, Tawfik DS (2005) Directed evolution of proteins for heterologous expression and stability. Curr Opin Struct Biol 15:50–56

112. Waldo GS, Standish BM, Berendzen J et al (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17:691–695

113. Pedelacq JD, Piltch E, Liong EC et al (2002) Engineering soluble proteins for structural genomics. Nat Biotechnol 20:927–932

114. Cabantous S, Terwilliger TC, Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol 23:102–107

115. Pedelacq JD, Cabantous S, Tran T et al (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 24:79–88

116. Maxwell KL, Mittermaier AK, Forman-Kay JD et al (1999) A simple in vivo assay for increased protein solubility. Protein Sci 8:1908–1911

117. Sieber V, Martinez CA, Arnold FH (2001) Libraries of hybrid proteins from distantly related sequences. Nat Biotechnol 19:456–460

118. Dahlroth SL, Nordlund P, Cornvik T (2006) Colony filtration blotting for screening soluble expression in Escherichia coli. Nat Protoc 1:253–258

119. Cornvik T, Dahlroth SL, Magnusdottir A et al (2005) Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. Nat Methods 2:507–509

<div style="text-align: right">

Chapter 3

</div>

Optimization of Culture Parameters and Novel Strategies to Improve Protein Solubility

Ranjana Arya, Jamal S.M. Sabir, Roop S. Bora, and Kulvinder S. Saini

Abstract

The production of recombinant proteins, in soluble form in a prokaryotic expression system, still remains a challenge for the biotechnologist. Innovative strategies have been developed to improve protein solubility in various protein overexpressing hosts. In this chapter, we would focus on methods currently available and amenable to "desired modifications," such as (a) the use of molecular chaperones; (b) the optimization of culture conditions; (c) the reengineering of a variety of host strains and vectors with affinity tags; and (d) optimal promoter strengths. All these parameters are evaluated with the primary objective of increasing the solubilization of recombinant protein(s) during overexpression in *Escherichia coli*.

Key words Protein solubility, Inclusion bodies, Chaperones, Host strain, Fusion tags, Culture parameters, Glucose, Temperature

1 Introduction

Escherichia coli remain one of the favorite, extensively used, robust, and versatile systems for the expression of recombinant proteins. The well-characterized genetics, rapid growth rate, inexpensive cell culture, and simplicity to handle and manipulate offer major advantages of overexpressing a protein of interest in *E. coli* [1]. However, one major limitation in the bulk production of recombinant protein is accumulation of heterologous proteins, as insoluble aggregates, in the form of inclusion bodies (IBs) [2]. Isolation of recombinant protein from IBs usually requires denaturing conditions, followed by several renaturing steps [3]. While protocols for denaturation and renaturation are described elsewhere in this book, the focus of this chapter is to discuss strategies and provide guidelines to improve protein solubility in *E. coli*. The major emphasis is on protocols for chaperone co-expression and optimization of culture conditions. In addition, we briefly review a select variety of

host strains and vectors with affinity tags, differential promoter strength, and codon usage that can be synergistically applied to produce the desired recombinant protein(s) in soluble form.

2   Materials

Different chaperone-containing plasmids are mentioned in Table 1, different *E. coli* host strains are mentioned in Table 2, different vectors with fusion protein are outlined in Table 3, and different vectors with promoter strengths are described in Table 4.

2.1   Buffers
(See Notes 1 and 2)

1. Lysis buffer: 1× PBS, pH 7.4, 100 mM NaCl, 5 mM DTT, 1 mM PMSF, 10 % glycerol, 0.5 % Triton X-100.

2. Equilibration buffer/wash buffer: 1× PBS, pH 7.4, 100 mM NaCl, 5 mM DTT, 10 % glycerol.

3. Elution buffer: 50 mM Tris–Cl, pH 8.0, 100 mM NaCl, 5 mM DTT, 10 % glycerol, reduced glutathione (10–100 mM)—*make fresh every time* (6.8 mg reduced glutathione per mL of elution buffer, i.e., ~10 mM) (*see* **Note 3**).

4. LB medium: 10 g tryptone, 5 g yeast extract, 10 g NaCl. Final volume to 1 L with MQ water. Autoclave.

5. LB agar: 10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar. Final 1 L volume with MQ water. Autoclave.

6. SOC medium: 2 g tryptone, 0.5 g yeast extract, 1 mL 1 M NaCl, 0.25 mL 1 M KCl, 1 mL 1 M $MgCl_2$, 1 mL 1 M $MgSO_4$, 1 mL 2 M glucose. Final volume to 100 mL MQ water.

7. LB + 20 % glycerol: 80 mL LB, 20 mL glycerol.

8. LB-ampicillin agar (per L): 1 L of LB agar, autoclaved, and wait till it cools down to 55 °C. Add 10 mL of 10 mg/mL filter-sterilized ampicillin and pour into petri dishes (~25 mL/100 mm plate) (*see* **Notes 4** and **5**).

9. TE buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA.

10. 4× SDS-PAGE loading buffer (10 mL): 50 mM Tris–HCl, pH 6.8, 2 % SDS, 10 % glycerol, 1 % β-mercaptoethanol, 12.5 mM EDTA, 0.02 % bromophenol blue, 2.6 mL H_2O.

11. Lysis buffer: 10 mM Tris–Cl, 300 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 10 % glycerol, 1 mg/mL lysozyme, protease inhibitor cocktail.

12. SDS-PAGE-10× running buffer: 144 g glycine, 31 g Tris base, 10 g SDS. MQ H_2O to make up the final volume to 1 L.

13. Staining solution (1 L): 2.5 g CBB (R 250), 450 mL methanol, 100 mL acetic acid, 450 mL MQ H_2O.

14. Destaining solution (1 L): 450 mL methanol, 100 mL acetic acid, 450 mL MQ H_2O.

Table 1
Use of molecular chaperones to increase protein solubility

Chaperones	Plasmid	Marker	Inducer	Protein expressed	Reference
dnaK/dnaJ/grp, groEL/groES	pG-KJE8 (TaKaRa)	Chloramphenicol	L-Arabinose tetracycline	Human crystalline ALDH3A1; human papillomavirus 16 E7 oncoprotein; cold-active lipase, Lip-948; Japanese cedar pollen, Cryj2	[6, 8, 23, 24]
groEL/groES	pGro7 (TaKaRa)	Chloramphenicol	L-Arabinose	Human crystalline ALDH3A1; human papillomavirus 16 E7 oncoprotein; Japanese cedar pollen, Cryj2; secreted xylanase precursor, xynB; human interferon gamma	[6, 8, 23, 25, 26]
groEL/groES	pGro6	Kanamycin	Tryptophan	Japanese cedar pollen, Cryj2	[6]
groEL/groES	pGro11	Chloramphenicol	Tetracycline	Japanese cedar pollen, Cryj2	[6]
groEL/groES	pGro11	Kanamycin	L-Arabinose	Japanese cedar pollen, Cryj2	[6]
dnaK/dnaJ/grpE	pKJE7 (TaKaRa)	Chloramphenicol	L-Arabinose	Human crystalline ALDH3A1; human papillomavirus 16 E7 oncoprotein; cold-active lipase, Lip948; Japanese cedar pollen, Cryj2; secreted xylanase precursor, xynB	[6, 8, 23–25]
dnaK/dnaJ/grpE	pKJE5	Kanamycin	Tryptophan	Japanese cedar pollen, Cryj2	[6]
groEL/groES/tig	pG-Tf2 (TaKaRa)	Chloramphenicol	Tetracycline	Human crystalline ALDH3A1; human papillomavirus 16 E7 oncoprotein; cold-active lipase, Lip948	[8, 23, 24]
Tig	pTf16 (TaKaRa)	Chloramphenicol	L-Arabinose	Human crystalline ALDH3A1; human papillomavirus 16 E7 oncoprotein	[8, 24]
grpE, clpB	pBB540 (Addgene)	Chloramphenicol	IPTG	Human IL-6	[27]

(continued)

Table 1
(continued)

Chaperones	Plasmid	Marker	Inducer	Protein expressed	Reference
dnaK/dnaJ/ groEL/groES	pBB542 (Addgene)	Spectinomycin	IPTG	Human IL-6	[27]
dnaK/dnaJ	pBB535 (Addgene)	Spectinomycin	IPTG	VP1 capsid protein foot-and-mouth disease virus	[28]
Skp	pAR3 (ATCC)	Chloramphenicol	L-Arabinose	Fab antibody	[29]
GroEL/ES	pAG	Chloramphenicol	L-Arabinose	Fab antibody	[29]
Trigger factor	pBAD33-Tig	Chloramphenicol	L-Arabinose	Fab antibody	[29]
sHSP-Lo18	pACYC184 (NEB)	Chloramphenicol & tetracycline	IPTG	β-glucosidase	[30]
DnaK	pXCK-K	Ampicillin	IPTG	Prion protein, PrP; varicella zoster virus protein ORF21p	[31]
GroEL	pXCK-EL	Ampicillin	IPTG	Prion protein, PrP; varicella zoster virus protein ORF21p	[31]
IbpB	pET-XynBIbpB	Ampicillin	IPTG	XynB of *Streptomyces olivaceovirdis*	[32]
Skp and FkpA	pACYCDuet™-1	Chloramphenicol	Arabinose	scFvD1.3	[11]
DsbA/DsbC, SurA, FkpA	pTUM4.1	Chloramphenicol	Arabinose		

Table 2
Use of different host strains to enhance protein solubility

Bacterial strain	Company	Features	Growth requirement
BL21(DE3)	Novagen	• Contains DE3 lysogen that expresses T7 polymerase upon IPTG induction • Deficient of lon and ompT proteases • Suitable for expression of nontoxic genes • No tight control over the expression • High IPTG conc. for induction	1 % glucose to the medium to reduce growth rates
BL21(DE3)-pLysS	Novagen	• Contains DE3 lysogen that expresses T7 polymerase upon IPTG induction • Produces T7 lysozyme to reduce basal level expression of the gene of interest • Suitable for expression of toxic genes • pLysS contains the p15A origin thus compatible with plasmids containing the ColE1 or pMB1 origin (i.e., pUC- or pBR322-derived plasmids)	Chloramphenicol 34 μg/mL
BL21 Star-pLysS	Invitrogen	• Mutated strain of RNase E thus decreases RNase degradation and increases protein expression • Tight regulation of expression by pLysS for both soluble and insoluble protein production	Chloramphenicol 34 μg/mL
BL21-SI	Invitrogen	• T7 polymerase in the DE3 is under control of the salt-inducible proU promoter • Induction of protein production at 0.1–0.5 M of NaCl • Protein solubility	LBON medium (LB w/o NaCl), 30 °C. Induction by NaCl, if controlled by T7 promoter. Otherwise—IPTG
BL21-AI	Invitrogen	• Contains a chromosomal copy of the T7 RNA polymerase gene under the tight control of the arabinose-inducible araBAD promoter	Tetracycline 12.5 μg/mL Induction by arabinose. (If using pET vectors, IPTG is also required). Glucose will repress this induction
Tuner	Novagen	• Contains a mutation in the lac permease (lacZY) gene • Allows uniform entry of IPTG into all cells in the population that produces a concentration-dependent, homogeneous level of induction • By adjusting the concentration of IPTG, expression can be regulated from very low levels up to the robust, fully induced levels commonly associated with pET vectors	None

(continued)

Table 2
(continued)

Bacterial strain	Company	Features	Growth requirement
Tuner pLysS	Novagen	• Contains the pLysS plasmid (tighter control over expression) in addition to the lac permease mutation	Chloramphenicol 34 µg/mL
Origami	Novagen	• K12 derivatives that have mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes • Greatly enhances disulfide bond formation in the cytoplasm	Kanamycin 15 µg/mL Tetracycline 12.5 µg/mL
Origami B	Novagen	• Similar to Origami strains, except that derived from a lacZY mutant of BL21 • Thus combine the desirable characteristics of BL21, Tuner, and Origami hosts in one strain	Kanamycin 15 µg/mL Tetracycline 12.5 µg/mL
Rosetta	Novagen	• BL21 lacZY (Tuner) derivatives designed to enhance the expression of eukaryotic proteins • These strains supply tRNAs for rare codons AUA, AGG, AGA, CUA, CCC, GGA on a compatible chloramphenicol-resistant plasmid • The tRNA genes are driven by their native promoters	Chloramphenicol 34 µg/mL
Rosetta pLysS	Novagen	• Rare tRNA genes are present on the same plasmids that carry the T7 lysozyme	Chloramphenicol 34 µg/mL
Rosetta-gami-pLysS	Novagen	• Origami derivatives • Enhanced disulfide bond formation resulting from trxB/gor mutations • Enhanced expression of eukaryotic proteins that contain rare codons of *E. coli* • Supply tRNAs for AGG, AGA, AUA, CUA, CCC, GGA on a compatible chloramphenicol-resistant plasmid • tRNA genes are driven by their native promoters • Rare tRNA genes are present on the same plasmids that carry the T7 lysozyme	Kanamycin 15 µg/mL Tetracycline 12.5 µg/mL Chloramphenicol 34 µg/mL
AD494	Novagen	• Thioredoxin reductase (trxB) mutants of the K12 strain • Enable disulfide bond formation in the cytoplasm • Enhances the proper protein folding	Kanamycin 15 µg/mL
BL21trxB	Novagen	• Contains thioredoxin reductase mutation (trxB) in the protease-deficient BL21 background	Kanamycin 15 µg/mL

HMS174	Novagen	• Contains recA mutation in a K12 background. Like BLR • Stabilize certain target genes whose products may cause the loss of the DE3 prophage	
NovaBlue(DE3)	Novagen	• K12 strain suited for initial cloning and expression host due to its high transformation efficiency and recA, endA mutations • High yields of excellent quality plasmid DNA • Contains DE3 for T7 polymerase expression. Tight regulation of expression due to the presence of lacI^q repressor encoded by the F episome	Tetracycline 12.5 µg/mL
BLR	Novagen	• recA⁻ derivative of BL21 • Improves plasmid monomer yields • Stabilize target plasmids containing repetitive sequences or whose products may cause the loss of the DE3 prophage	Tetracycline 12.5 µg/mL
C41(DE3)	Lucigen	• Effective in expressing toxic and membrane proteins from all classes of organisms including viruses, eubacteria, archaea, yeasts, plants, insects, and mammals • Derived from BL21(DE3), contains at least one uncharacterized mutation that prevents cell death associated with expression of toxic recombinant proteins	
C43(DE3)	Lucigen	• Effective in expressing toxic and membrane proteins from all classes of organisms including viruses, eubacteria, archaea, yeasts, plants, insects, and mammals	
M15	Qiagen	• Contain pREP4 plasmid-encoding lac repressor • Tightly regulated expression prior to IPTG induction	Kanamycin
Lemo21(DE3)	NEB	• Contains features of BL21(DE3) • Tunable expression of difficult clones by varying the level of lysozyme (lysY), the natural inhibitor of T7 RNA polymerase	L-Rhamnose 0–2,000 µM
SHuffle® T7 Express	NEB	• Constitutively expresses a chromosomal copy of the disulfide bond isomerase DsbC • Promotes the correct folding of mis-oxidized proteins • Expresses a chromosomal copy of T7 RNA Polymerase • Nonspecific endonuclease I (endA1) activity is eliminated for highest quality plasmid preparations • Deficient in proteases Lon and OmpT resistance to phage T1 (fhuA2)	

(continued)

Table 2
(continued)

Bacterial strain	Company	Features	Growth requirement
BL21-AI™ One Shot®	Invitrogen	• Contains a chromosomal insertion of the gene-encoding T7 RNA polymerase into the araB locus of the araBAD operon • Regulation under the control of the arabinose-inducible araBAD promoter • Useful for the expression of toxic genes	
SoluBL21	Genlantis	• Optimized for expressing insoluble proteins in soluble form • Used for toxic proteins • Compatible with all T7 promoter-based expression vectors	
BL21 CodonPlus	Stratagene	• Cells carry extra copies of the argU, ileY, and leuW tRNA genes which recognize the AGA/AGG (arginine), AUA (isoleucine), and CUA (leucine) codons, respectively	Chloramphenicol 34 µg/mL
BL21-CodonPlus (DE3)-RIL	Stratagene	• Encodes T7 RNA polymerase under control of the lacUV5 promoter for easy protein expression • Use with pET or pCAL vectors	Chloramphenicol 34 µg/mL
BL21-CodonPlus -RP	Stratagene	• For extremely toxic proteins • Used with lambda CE6 under the control of T7 RNA polymerase	Chloramphenicol 34 µg/mL
BL21-CodonPlus (DE3)-RP-X	Stratagene	• Encodes T7 RNA polymerase under control of the lacUV5 promoter for easy expression • Use with pET or pCAL vectors • Used for methionine labeling of cells	Chloramphenicol 34 µg/mL
BL21-CodonPlus (DE3)-RILP	Stratagene	• Contain extra copies of rare E. coli argU, ileY, leuW, proL tRNA genes which corrects for codon bias • Improves expression of sequences from other organisms	Chloramphenicol 34 µg/mL

Table 3
Use of fusion tags/proteins to enhance protein solubility

	Source species	Size	Effect on solubility	Matrix	Elution condition
Affinity tags to improve yield					
Glutathione S-transferase	*Schistosoma japonicum*	26 kDa	Increased	Glutathione	5–10 mM reduced glutathione
Maltose-binding protein (MBP)	*E. coli* K12	42 kDa	Increased	Amylose	10 mM maltose
Thioredoxin A	*E. coli*	11.6 kDa	Increased	Phenylarsine oxide	
Small ubiquitin-related modifier (SUMO)	*Saccharomyces cerevisiae*	11 kDa	Increased	IMAC when used with hexa-His tag	Imidazole 20–250 mM or low pH
N-utilization substance A (NusA)	*E. coli*	55 kDa	Increased	IMAC when used with hexa-His tag	Imidazole 20–250 mM or low pH
Protein disulfide isomerase I (DsbA)	*E. coli*	21.1 kDa	Increased	IMAC when used with hexa-His tag	Imidazole 20–250 mM or low pH
Mistic	*B. subtilis*	13 kDa	Increased		
Ketosteroid isomerase (KSI)		13 kDa	Decreased		
TrpΔLE	*E. coli*	27 kDa	Decreased		
Affinity tags to facilitate purification					
c-Myc	Human c-Myc proto-oncogene	1.20 kDa	No effect	Monoclonal antibody	Low pH
HA	Human influenza virus hemagglutinin	9 AA	No effect	Anti-HA agarose	Low pH
FLAG		1.01 kDa	No effect	Anti-FLAG M2 MAb agarose FLAG peptide	pH 3.0 or 2–5 mM, FLAG peptide
1D4	C-terminus of bovine rhodopsin	9 AA	No effect	Antibody-based matrix	Low pH
polyArg		80 kDa	No effect	Cation exchange resins	NaCl linear gradient from 0 to 400 mM at alkaline pH>8.0
polyHis		84 kDa	No effect	Ni^{2+}-NTA	Imidazole 20–250 mM or low pH
Strep tag	Random genetic library	1.06 kDa	No effect	Modified streptavidin	2.5 mM desthiobiotin
Calmodulin-binding peptide		2.96 kDa	No effect	Immobilized calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain		3–20 kDa	No effect	Cellulose matrices	Guanidine HCl or urea >4 M

Table 4
Use of different promoters to improve protein solubility

Promoters	Regulation	Induction	Level of expression
Lac	lacI, lacIq	IPTG	Low
lacUV5	lacI, lacIq	IPTG	Low
tac (hybrid)	lacI, lacIq	IPTG	Allows accumulation of protein to about 15–30 % of total cell protein
trc (hybrid)	lacI, lacIq	IPTG	Allows accumulation of protein to about 15–30 % of total cell protein
Trp	Addition of fructose to the growth medium increases down regulation under non-induced conditions	Tryptophan starvation or addition of B-indole acrylic acid	
araBAD	araC	L-Arabinose	Slightly weaker than the tac promoter
phoA	phoB (positive) phoR (negative)	Phosphate starvation	
recA	lexA	Nalidixic acid	
proU		Osmolarity	
Cst-1		Glucose starvation	
tetA		Tetracycline	
cadA	cadR	pH	
Nar	Fnr	Anaerobic conditions	
pL	l cIts857	Thermal (shift to 42 °C)	

http://wolfson.huji.ac.il/expression/bac-strains-prot-exp.htmL

3   Methods

3.1   Chaperone Co-expression

IBs are usually comprised of misfolded protein aggregates. Under normal cellular conditions, these molecular chaperones interact reversibly with nascent polypeptide chains to prevent aggregation during the folding process. Some chaperones prevent polypeptides from aggregation, while other chaperones assist in refolding and solubilization of misfolded proteins [4]. The most abundant and physiologically important cytoplasmic chaperones in *E. coli* include trigger factor, DnaK, DnaJ, GrpE, GroEL, and GroES. These different chaperones have been used singly, or in combination, to increase the protein solubility in recombinant systems [5]. Two major chaperone combinations, i.e., DnaK–DnaJ–GrpE and GroEL–GroES play distinct but cooperative roles in protein

folding [6]. While DnaK–DnaJ–GrpE promotes release of unfolded proteins, GroEL–GroES prevent degradation of peptides [7]. Trigger factor associates with GroEL and strengthens GroEL-substrate binding to facilitate protein folding or degradation [8]. Other valuable chaperones include ClpB (Hsp100) that along with DnaK solubilizes and disaggregates protein [9]. On the other hand, small heat shock proteins, IpbA and IpbB, protect aggregation of heat-denatured proteins in an ATP-independent manner [10]. For each recombinant protein, different combinations of heat shock proteins and chaperones should be tested to find the most appropriate one. Co-expression of Skp and FkpA chaperones improves solubility of antibody fragments [11]. Various studies have used these molecular chaperones for improving protein solubility and avoiding protein aggregation as outlined in Table 1. Key points to remember are:

1. Choose the right combination of plasmid, strain, and chaperones required to co-express with the target protein.

2. Transform the chosen *E. coli* strain with appropriate plasmid for selective expression of different chaperone combinations (*see* **Note 6**).

3. Remove competent cells from –80 °C and place directly on ice. Thaw cells for 5–10 min.

4. Gently mix cells by tapping the tube. Aliquot 50 μL cells into the transformation tubes.

5. Add 1–50 ng of DNA encoding the chaperones (or 1 μL control DNA) into 50 μL competent cells. Gently tap tube to mix.

6. Place the tubes on ice for 30 min.

7. Heat shock the cells for 90 s in a 42 °C water bath. Do not shake.

8. Add 900 μL of room temperature (RT) medium to each transformation reaction.

9. Incubate at 37 °C for 1 h with shaking (225–250 rpm).

10. Spread on LB agar plates containing appropriate antibiotic.

11. Incubate the plates at 37 °C overnight (O/N) (12–16 h).

12. Pick up the single colony of transformed strain and make them competent for next transformation. Use the correct selection marker for selection of positive clones.

13. Transform the positive clone with the plasmid carrying target gene having different ori and selectable marker than the chaperone plasmid.

14. Isolate the colonies from double-/triple-resistance marker plates.

15. Inoculate single colony in 3 mL LB and incubate at 37 °C for 12–16 h.

16. Inoculate with 1 % inoculum in 100 mL LB medium carrying the selectable marker at 37 °C with shaking.

17. When the OD_{600nm} of the culture reaches 0.6–0.8, induce with the appropriate inducer.

18. Incubate further at 37 °C or lower temperature for 4–6 h.

19. Lyse the cells with lysis buffer (*see* **Note 7**).

20. Sonicate the cells three times with 10 s pulse and 15 s interval on ice and centrifuge at $10,000 \times g$ and 4 °C for 15 min.

21. Collect the pellet and supernatant and analyze for protein expression in SDS-PAGE.

3.2 Use of Different Host Strains

A wide variety of host strains are available with distinct features and advantages to obtain protein in soluble fraction as listed in Table 2 [12]. Some host strains such as BL21(DE3) carry mutations that inhibit production of cellular proteases and therefore enhance expression of nontoxic genes [13, 14]. Tuner strains from Novagen allow uniform entry of IPTG into the cells leading to homogeneous expression of recombinant protein (www.novagen.com). Origami strains have mutations in thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which greatly enhance disulfide bond formation in the cytoplasm resulting in greater protein solubility [15]. Since expression of eukaryotic proteins in *E. coli* is limited due to unfavorable codon usage, a number of strains have been generated, such as Rosetta-gami and BL21-CodonPlus-RILP, that supply tRNA for rare codons and allow correct folding and expression of heterologous proteins [16, 17]. Some strains have constitutive expression of disulfide isomerase (DsbC) to allow proper disulfide bond formation, thereby inhibiting accumulation of misfolded proteins [18]:

1. Transform the strain of interest with desired plasmid and select positive clones on appropriate resistance marker (transformation protocol is the same as described above in Subheading 3.1).

2. Express the protein as described (*see* **steps 15–21** in Subheading 3.1).

3.3 Use of Fusion Tags/Proteins to Improve Solubility

A popular approach to solubilize an aggregation-prone protein is to fuse it with a highly soluble partner. In this regard, several fusion partners have been explored in *E. coli* with various pros and cons. Different vectors available commercially possess these fusion tags at N- or C-terminus (reviewed in Table 3). Among these fusion partners, maltose-binding protein (MBP) and glutathione-S-transferase (GST) fusion proteins have been extensively used to improve the solubility of recombinant protein [19]. Hexa-histidine tag along with MBP has also been used to solubilize the proteins, but their folding is either spontaneous or chaperone-mediated [20].

1. Clone the gene of interest in a vector containing a GST fusion tag at N- or C-terminus by using right combination of restriction enzymes.

2. Transform BL21 or appropriate strain compatible with the plasmid-carrying target gene as described above (*see* Subheading 3.1).

3. Inoculate a starter culture of 10 mL in a 50 mL flask containing appropriate antibiotics with a single transformed colony.

4. Shake O/N at 37 °C at 200 rpm.

5. Add 1 % of starter culture in 1 L of LB broth containing appropriate antibiotics.

6. Shake for 2–3 h at 37 °C, till the OD_{600nm} reaches 0.6–0.8.

7. Induce culture with 0.1 mM IPTG.

8. Shake for 2.5 h at 37 °C.

9. Spin at $7,000 \times g$ at 4 °C for 10 min.

10. Remove the supernatant and either begin lysis or store pellet at −80 °C.

11. Resuspend the bacterial pellet from 1 L culture in 40 mL chilled lysis buffer (25:1). Transfer to a 50 mL conical tube and leave on ice for 10–20 min.

12. Break cells by sonication: incubate on ice with 800 μg/mL lysozyme for 0.5–1 h; sonicate for 20 s with 1 min interval (6–7 pulses). Save 100–200 μL sample as total lysate (Sample 1).

13. Spin the sample at $15,000 \times g$ at 4 °C for 30 min and/or filter through 0.45 μm filter before applying to glutathione sepharose medium. Save 100–200 μL supernatant as Sample 2.

14. Prepare the glutathione beads: Cut a P1000 tip to take 0.5 mL of the slurry (50 %) and transfer to a 15 mL Falcon tube, bring the volume to 5 mL with cold equilibration buffer, and spin the beads at $2,500 \times g$ at 4 °C for 3 min (*see* **Note 8**).

15. Remove the supernatant and wash the beads at least two more times with equilibration buffer. After the final wash, bring the volume to 0.8 mL with equilibration buffer and gently tap the beads to resuspend them.

16. Add the supernatant obtained from **step 3** to the equilibrated beads and rotate at 4 °C for 1.5–2 h. Spin at $2,500 \times g$ at 4 °C for 3 min (*see* **Note 9**).

17. Remove the supernatant leaving the bound beads in the tube (0.5–1 mL) and save supernatant as Sample 3.

18. Wash beads thrice with 10 mL of wash buffer (rotate 10 min each wash).

19. Repeat the spin and remove the supernatant leaving bound beads (0.5–1 mL).

20. Elution: Using a cut 1 mL tip, resuspend the beads and transfer to a 1.5 mL microfuge tube.

21. Add 500 μL elution buffer containing 10 mM reduced glutathione.

22. Gently tap the beads and spin at 2,500×*g* at 4 °C for 3 min.

23. Let beads settle on ice and remove the supernatant. Save supernatant and label as eluent 1.

24. Add 500 μL elution buffer containing 20 mM reduced glutathione and repeat the **steps 12** and **13**. Similarly, repeat for the gradient of reduced glutathione (10–100 mM).

25. Run all the collected samples and eluents on SDS-PAGE for protein analysis.

3.4 Use of Different Culture Conditions

Overexpression of protein at low temperature usually improves both protein solubility and activity. A temperature range between 16 and 23 °C is often recommended to check for protein expression [21], and we use this to optimize culture conditions in our laboratory. We have expressed recombinant UDP *N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase (GNE) at low temperature in *E. coli* with increased solubility (unpublished data).

3.4.1 Growth at Lower Temperatures

Day 1: Primary Culture

1. Pick up a single colony of *E. coli* from a previously streaked agar plate with a sterile toothpick.

2. Inoculate single colony into 10 mL LB broth containing appropriate selection marker at standard concentration.

3. Incubate at 37 °C O/N with shaking at 200 rpm.

Day 2: Secondary Culture

4. Add 1 % primary culture to autoclaved LB broth, for instance, 5–500 mL of LB broth. Add requisite amount of drug. Prepare five flasks, if required to test five different temperatures.

5. Induce the culture with 1 mM IPTG when the OD_{600nm} of the culture reaches 0.6–0.8. Keep 1 mL aliquot as uninduced control sample. Pellet the uninduced cells by spinning at 13,000×*g* during 10 min at 4 °C and freeze at –20 °C.

6. Incubate the remaining cultures at different temperatures; 16, 25, 30, and 37 °C to find the optimum conditions where maximum expression of recombinant protein is observed in soluble fraction.

7. To further optimize the culture conditions, incubate the cultures for different time points, e.g., 3, 6, 9, and 12 h (*see* **Note 10**). Pellet the cells at 13,000×*g* at 4 °C for 10 min.
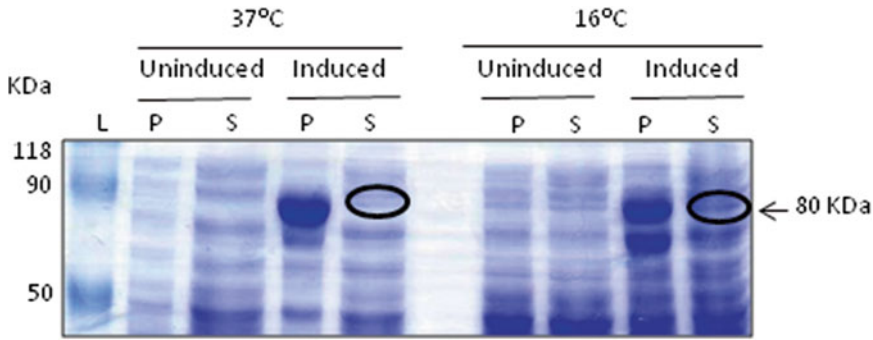
Fig. 1 Expression of recombinant GNE at low temperature. Recombinant GNE cloned in pET30a was induced for expression using 1 mM IPTG for 9 h at 37 and 16 °C. Cell lysates were prepared as described in Subheading 3, and equal amount of protein was loaded on SDS-PAGE followed by staining with Coomassie blue. *L* molecular weight ladder, *P* pellet, *S* supernatant

      8. Wash the cells with 50 mM Tris–Cl, pH 7.5, and centrifuge at $13,000 \times g$ at 4 °C for 5 min.

      9. Redissolve the pellet in 400 μL lysis buffer and sonicate three times with 10 s pulse and 15 s intervals on ice.

   10. Centrifuge at $10,000 \times g$ for 10 min at 4 °C and collect the pellet and supernatant.

   11. Add 4× SDS-PAGE buffer to both pellet and supernatant and analyze the samples on 10 % SDS-PAGE.

As shown in Fig. 1, the expression of GNE in soluble fraction increased up to 60 % at low temperature compared to 37 °C.

3.5 Use of Different Promoters

Various vectors with weak or strong promoters are commercially available to express recombinant proteins in *E. coli* as listed in Table 4. Using a weak promoter, e.g., trc instead of T7, increases protein solubility [12].

      1. Inoculate 2 mL of LB containing ampicillin (50 μg/mL) with a single recombinant *E. coli* colony.

      2. Grow O/N at 37 °C with shaking.

      3. Next day, inoculate 400 mL of LB containing ampicillin (50 μg/mL) with 1 % O/N inoculum.

      4. Grow the culture at 37 °C with shaking to an $OD_{600nm} = 0.6$ (the cells should be in mid-log phase).

      5. Remove a 1 mL aliquot of cells prior to IPTG induction (controls), centrifuge the sample in a microcentrifuge, and aspirate the supernatant. Freeze at –20 °C. This will be the time zero sample.

6. Add IPTG to a final concentration of 1 mM and grow bacteria at 37 °C with shaking. Take samples at 1 h intervals for 5 h (or more). Centrifuge each sample and store both the supernatant and the pellet at 4 °C. For long-term storage (longer than 5 h), store the samples at –20 °C.

7. When all time points have been collected, resuspend each pellet in 100 μL of 20 mM phosphate buffer at neutral pH and freeze in liquid nitrogen or methanol/dry ice (exercise caution when handling liquid nitrogen, it can cause severe burns if it comes in contact with the skin; wear appropriate protective gloves). Thaw the frozen lysate at 42 °C. Repeat this freeze-thaw 2–3 additional times and pellet the insoluble protein in a refrigerated microcentrifuge for 10 min at maximum speed.

8. Remove the supernatant to a fresh-labeled tube. Resuspend the pellet in 100 μL of Laemmli buffer/50 μL 4× SDS-PAGE buffer. To 100 μL of supernatant sample, add 4× SDS-PAGE buffer.

9. Analyze 10–20 μL of both the supernatant and pellet samples on a 10 % SDS polyacrylamide gel.

10. Stain the gel with Coomassie blue for protein analysis.

3.6 Addition of Glucose in the Growth Media

Alterations in media composition can be used to enhance protein solubility. Minimal medium is supplemented with 1 % glucose to improve solubility of protein. Addition of cofactors, buffers, polyols, ethanol, and low molecular weight thiols to the growth media has been shown to increase protein solubility [22].

Day 1: Primary Culture

1. Pick up a single colony of *E. coli* from a previously streaked agar plate with a sterile toothpick.

2. Inoculate single colony into 10 mL LB broth containing appropriate selection marker at standard concentration.

3. Incubate at 37 °C O/N with shaking at 200 rpm.

Day 2: Secondary Culture

1. Add 1 % of primary culture to autoclaved LB broth, for instance, 5–500 mL of LB broth (add requisite amount of drug accordingly).

2. When the OD_{600nm} reaches 0.6–0.8, pellet the cells by centrifuging at $6,000 \times g$ for 5 min at RT.

3. Resuspend the cells in equal volume of minimal medium. Divide the cells in two halves. Spin at $6,000 \times g$ for 5 min at RT.

4. Resuspend one half in minimal medium (control) and the other half in minimal medium containing 0.5–1 % glucose.

5. Induce the cultures with 1 mM IPTG (*see* **Note 11**).

6. Incubate the remaining cultures at 37 °C for 3–12 h (*see* **Note 12**). Pellet the cells at $13,000 \times g$ at 4 °C for 10 min.

7. Wash the cells with 50 mM Tris–Cl, pH 7.5 and centrifuge at $13,000 \times g$ at 4 °C for 5 min.

8. Redissolve the pellet in 400 μL lysis buffer.

9. Sonicate three times with 10 s pulse and 15 s interval on ice.

10. Centrifuge at $10,000 \times g$ for 10 min at 4 °C and collect the pellet and supernatant.

11. Add 4× SDS-PAGE buffer to both pellet and supernatant and analyze the samples on 10 % SDS-PAGE

4   Notes

1. For buffer preparation use high-purity water and chemicals and filter all buffers through a 0.45 μm filter before use.

2. Spin at $15,600 \times g$ for 10 min before pipetting protein for further experiments.

3. 1–20 mM DTT may be included in the binding and elution buffers to reduce the risk of oxidation of free-SH groups on GST, which may cause aggregation of the tagged target protein, resulting in lower yield of GST-tagged protein.

4. All antibiotics should be filter sterilized and checked by growing sensitive strains.

5. Addition of antibiotics to LB agar should not be done at high temperature.

6. High-efficiency competent cells should be used for all transformation experiments.

7. If the lysate is too viscous, dilute it with lysis buffer, increase lysis treatment (sonication, homogenization), or add DNase/RNase to reduce the size of nucleic acid fragments.

8. All the steps in protein purification should be done at 4 °C or in a cold room.

9. The binding of the target protein to the beads can be done O/N.

10. Better yields are observed at low temperatures, if cultures are incubated for longer time: 9–12 h.

11. Keep 1 mL aliquot as uninduced control sample. Pellet the uninduced cells and freeze at –20 °C.

12. Incubate for different times to find maximum expression.

Acknowledgments

References

1. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307:249–264

2. Arya R, Bhattacharya A, Saini KS (2008) Dictyostelium discoideum—a promising expression system for the production of eukaryotic proteins. FASEB J 22:4055–4066

3. Khattar SK, Gulati P, Kundu PK et al (2007) Enhanced soluble production of biologically active recombinant human p38 mitogen-activated-protein kinase (MAPK) in Escherichia coli. Protein Pept Lett 14:756–760

4. Carrio MM, Villaverde A (2003) Role of molecular chaperones in inclusion body formation. FEBS Lett 537:215–221

5. de Marco A (2007) Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. Nat Protoc 2:2632–2639

6. Nishihara K, Kanemori M, Kitagawa M et al (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. Appl Environ Microbiol 64:1694–1699

7. Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb Cell Fact 4:1

8. Folwarczna J, Moravec T, Plchova H et al (2012) Efficient expression of Human papillomavirus 16 E7 oncoprotein fused to C-terminus of Tobacco mosaic virus (TMV) coat protein using molecular chaperones in Escherichia coli. Protein Expr Purif 85:152–157

9. Betiku E (2006) Molecular chaperones involved in heterologous protein folding in Escherichia coli. Biotechnol Mol Biol 1:66–75

10. Guzzo J (2012) Biotechnical applications of small heat shock proteins from bacteria. Int J Biochem Cell B 44:1698–1705

11. Ow DSW, Lim DYX, Nissom PM et al (2010) Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. Microb Cell Fact 9:22

12. Samuelson JC (2011) Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations. In: Heterologous gene expression in E. coli. Methods Mol Biol 705:195–209

13. Phillips TAVRA, Neidhardt FC (1984) lon gene product of Escherichia coli is a heat-shock protein. J Bacteriol 159:283–287

14. Grodberg J, Dunn JJ (1988) ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol 170:1245–1253

15. Prinz WA, Aslund F, Holmgren A et al (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. J Biol Chem 272:15661–15667

16. Seidel HM, Pompliano DL, Knowles JR (1992) Phosphonate biosynthesis: molecular cloning of the gene for phosphoenolpyruvate mutase from Tetrahymena pyriformis and overexpression of the gene product in Escherichia coli. Biochemistry 31:2598–2608

17. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol 6:494–500

18. Bessette PH, Aslund F, Beckwith J et al (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc Natl Acad Sci U S A 96:13703–13708

19. Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J 7:620–634

20. Raran-Kurussi S, Waugh DS (2012) The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding

protein but their folding is either spontaneous or chaperone-mediated. PLoS One 7:e49589

21. San-Miguel T, Perez-Bermudez P, Gavidia I (2013) Production of soluble eukaryotic recombinant proteins in E coli is favoured in early log-phase cultures induced at low temperature. SpringerPlus 2:89

22. Pan SH, Malcolm BA (2000) Reduced background expression and improved plasmid stability with pET vectors in BL21(DE3). Biotechniques 29:1234–1238

23. Cui SS, Lin XZ, Shen JH (2011) Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948. Protein Expr Purif 77:166–172

24. Voulgaridou GP, Mantso T, Chlichlia K et al (2013) Efficient E. coli expression strategies for production of soluble human crystallin ALDH3A1. PLoS One 8:e56582

25. Jhamb KSDK (2012) Production of soluble recombinant proteins in Escherichia coli: effects of process conditions and chaperone co-expression on cell growth and production of xylanase. Bioresour Technol 123:135–143

26. Yan X, Hu S, Guan YX, Yao SJ (2012) Coexpression of chaperonin GroEL/GroES markedly enhanced soluble and functional expression of recombinant human interferon-gamma in Escherichia coli. Appl Microbiol Biotechnol 93:1065–1074

27. Nausch H, Huckauf J, Koslowski R et al (2013) Recombinant production of human interleukin 6 in Escherichia coli. PLoS One 8:e54933

28. Martinez-Alonso M, Vera A, Villaverde A (2007) Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells. FEMS Microbiol Lett 273:187–195

29. Levy R, Weiss R, Chen G et al (2001) Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones. Protein Expr Purif 23: 338–347

30. Ronez FDN, Arbault P, Guzzo J (2012) Co-expression of the small heat shock protein, Lo18, with b-glucosidase in Escherichia coli improves solubilization and reveals various associations with overproduced heterologous protein, GroEL/ES. Biotechnol Lett 34:935–939

31. Kyratsous CA, Silverstein SJ, DeLong CR et al (2009) Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in Escherichia coli. Gene 440:9–15

32. Su XY, Zhang S, Wang L et al (2009) Overexpression of IbpB enhances production of soluble active Streptomyces olivaceovirdis XynB in Escherichia coli. Biochem Biophys Res Commun 390:673–677

Chapter 4

Cleavable Self-Aggregating Tags (cSAT) for Protein Expression and Purification

Zhanglin Lin, Qing Zhao, Bihong Zhou, Lei Xing, and Wanghui Xu

Abstract

Rapid protein expression and purification remains a critical technological need, in particular as the number of proteins being identified is exploding. In this chapter, we describe a simple and rapid scheme for expression and purification of recombinant proteins using *Escherichia coli*, by taking advantage of two self-aggregating peptide fusion tags 18A (EWLKAFYEKVLEKLKELF) and ELK16 (LELELKLKLELELKLK) that can drive target proteins into active protein aggregates in vivo. In practice, a target protein is fused at the N-terminus of the self-cleavable *Mxe* GyrA intein, which is followed by the 18A or ELK16 tag. The fusion protein is first expressed in the form of active aggregate and then separated by centrifugation upon cell lysis. Subsequently, the DTT-mediated intein self-cleavage reaction releases the target protein into solution. These cleavable self-aggregating tags (cSAT, intein-18A/ELK16) provide a quick and efficient route for the production of proteins with modest purity (around 90 % in the case of intein-ELK16). Two application examples are included in the chapter.

Key words Cleavable self-aggregating tags, Protein purification, Self-assembling amphipathic peptides, Self-aggregating peptides, Active protein aggregates, Inclusion bodies, Intein

1 Introduction

With the advance of genomics and proteomics, there is a continuous need to develop high-throughput expression and purification techniques for recombinant proteins. Overexpression of heterologous proteins in bacteria often leads to formation of inactive protein aggregates in vivo, known as inclusion bodies (IBs) [1]. IBs have several outstanding characteristics such as high expression level, quick separation, and reduced degradation by endogenous proteases [2, 3]. However, tedious refolding procedures are often required to recover biologically active soluble proteins from purified IBs, and thus the application has been generally limited to expression of proteins or peptides that are

toxic to the cell [3–7]. Along this line, a few fusion carriers such as N^{pro} [3] and ketosteroid isomerase (KSI) [7] have also been developed to deliberately induce the formation of inactive IBs for target polypeptides that are otherwise solubly expressed.

In recent years, it has been gradually revealed that in some cases IBs can be highly active [8–11]. For example, the foot-and-mouth disease virus capsid protein VP1, the human β-amyloid peptide Aβ42(F19D), a maltose-binding protein mutant (MalE31), and the cellulose-binding domain of *Clostridium cellulovorans* (CBD$_{clos}$) can be used as fusion partners to drive proteins into active IBs [12]. We then further found that a number of short self-assembling amphipathic peptides, i.e., an α-helical peptide 18A (EWLKAFYEKVLEKLKELF), a β-sheet peptide ELK16 (LEL ELKLKLELELKLK), and small surfactant-like peptides (L$_6$KD, L$_6$K$_2$, DKL$_6$), can induce the formation of highly active protein aggregates in *E. coli* when fused to various target proteins [12–14]. These observations provide an exciting avenue for quick protein expression and purification without the need of refolding steps [13, 15]. To this end, we have combined the self-assembling peptides with the self-cleavable inteins to devise cleavable self-aggregating tags (cSAT) for simple, reliable, and cost-effective protein purification [13, 16]. These cSAT tags are much shorter and thus more economical than other tags that use protein fusion partners to similarly induce the formation of active IBs. In the following section, we describe in detail such a cSAT scheme in which a target protein is fused to the N-terminus of *Mxe* GyrA intein, which is followed by the aggregation inducer 18A or ELK16 via a PT linker PTPPTTPTPPTTPTPTP. The *Mxe* GyrA intein used here carries a mutation Asn198Ala, which silences its C-terminal cleavage activity, whereas its N-terminal cleavage activity can be induced by adding dithiothreitol (DTT) or other thio-reagents [17]. Three amino acid residues MRM (Met-Arg-Met) are added to the N-terminus of intein to facilitate its self-cleavage [18]. The resulting fusion protein is first expressed as active aggregate in *E. coli* and then separated by centrifugation from soluble impurities upon cell lysis. Subsequently, the target protein is released from the aggregate into solution via DTT-mediated intein self-cleavage at its N-terminus. This scheme typically yields target proteins at a modest purity (around 90 % for the intein-ELK16 tag) without any chromatography step, and it can be applied in a high-throughput manner. It has been successfully used for the production of several proteins and peptides, such as *Bacillus subtilis* lipase A (LipA), *Aspergillus fumigates* amadoriase II (AMA), *Bacillus pumilus* xylosidase (XynB), antimicrobial peptide histatin 1, and human glucagon-like peptide 1 (GLP1) [13]. Two examples (LipA and AMA) are presented here.

2   Materials

2.1   Reagents,
Buffers, and Solutions

1. Buffer B1 (lysis buffer, wash buffer): 20 mM Tris–HCl, pH 8.5, 500 mM NaCl, 1 mM disodium edentate (EDTA).

2. Buffer B3 (cleavage buffer): 20 mM Tris–HCl, pH 8.5, 500 mM NaCl, 1 mM EDTA, and 40 mM DTT.

3. Stacking gel 5 % (1.5 mL) for sodium dodecyl sulfate poly-acrylamide gel electrophoresis (SDS-PAGE): 190 μL 40 % (w/v) acrylamide, 190 μL Tris–HCl, pH 6.8, 15 μL 10 % (w/v) SDS, 15 μL 10 % (w/v) ammonium persulfate (APS) (prepared fresh), 2 μL N,N,N,N'-tetramethylethylenediamine (TEMED), and 1.2 mL ddH$_2$O.

4. Separation gel 12 % (5 mL) for SDS-PAGE: 1.5 mL 40 % acrylamide, 1.25 mL Tris–HCl, pH 8.8, 50 μL 10 % SDS, 50 μL 10 % APS (freshly prepared), 2 μL TEMED, and 2.2 mL ddH$_2$O.

5. SDS-PAGE buffer (10×): 250 mM Tris–HCl, 129 mM glycine, 1 % SDS.

6. LipA reaction buffer: 50 mM sodium phosphate buffer, pH 8.0, 1 mg/mL Arabic gum, and 2.07 mg/mL sodium deoxycholate.

7. AMA reaction mixture: 100 mM potassium phosphate buffer, pH 8.0, 2.7 purpurogallin units of peroxidase, 0.45 mM 4-ami-noantipyrine, 0.5 mM N-ethyl-N-(2-hydroxy-3-sulfopropyl)-m-toluidine (TOOS), and 5.0 mM d-fructosyl-glycine.

8. PCR reaction mixture using pfu polymerase (100 μL in total): 10 μL 10× pfu buffer, 8 μL dNTP mixture (2.5 mM), 100 ng DNA template, 4 μL forward primer (20 μM), 4 μL reverse primer (20 μM), and 2.5 μL pfu (2.5 U/μL), and then add ddH$_2$O to bring the volume to 100 μL.

9. Colony PCR reaction mixture (8 μL): 1 μL 10× PCR buffer, 0.8 μL dNTP mixture (25 mM), 0.1 μL forward primer (20 μM), 0.1 μL reverse primer (20 μM), 0.05 μL rTaq poly-merase, and 6 μL ddH$_2$O. 2 μL of the supernatant of lysed cells from an individual colony is added in each colony PCR reaction (see below).

10. Luria-Bertani (LB): 10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, and 950 mL of ddH$_2$O. Stir to dissolve all solids and bring the final volume to 1 L with ddH$_2$O before autoclaving.

11. LB-agar plates: add 15 g of agar per L of LB broth prepared as above before autoclaving.

2.2 Strains and Plasmid	1. *E. coli* BL21(DE3) strain (Novagen, Madison, WI).
	2. pET30a(+) expression plasmid (Novagen).
	3. pTWIN1 vector carrying the gene of *Mxe* GyrA intein (New England Biolabs, Beverly, MA).
2.3 Enzymes	1. Restriction enzymes and T4 DNA ligase.
	2. Pfu DNA polymerase.
	3. rTaq DNA polymerase (TaKaRa, Dalian, China).
	4. Shrimp alkaline phosphatase (SAP).
2.4 Kits and Apparatus	1. TIANquick Mini Purification Kit, TIANgel Mini Purification Kit, TIANpure Mini Plasmid Kit (Tiangen).
	2. Quantity One software (Bio-Rad Laboratories, Hercules, CA).
	3. Pierce® Bicinchoninic Acid (BCA) Protein Assay Kit (Thermo Scientific, Rockford, IL).
	4. SPECTRAMAX M2 microtiter reader (Molecular Device, CA).

3 Methods

The methods described below outline (1) plasmid construction, (2) protein expression, (3) protein purification, and (4) protein activity assays.

3.1 Plasmid Construction

The construction of the expression plasmids pET30a-target protein-Mxe GyrA intein-18A/ELK16 (pET30a-target protein-I-18A/ELK16) (Fig. 1) is described below in detail (*see* Subheading 3.1.1), which is derived from pET30a-LipA-I-18A/ELK16.

3.1.1 pET30a-LipA-I-18A/ELK16 Construction

Plasmid pET30a-LipA-PT linker-18A/ELK16 is first constructed, which is then used for the construction of pET30a-LipA-I-18A/ELK16 (*see* **Note 1**).

1. Assemble the full length of PT linker-18A using pfu DNA polymerase with the following set of oligonucleotides (*see* **Note 2**):

5′-AATGAAA**AAGCTT***CCGACCC*-3′

5′-*GGTGGCGTTGGCGTGGTCGGTGGGGTCGGAAGCTT*-3′

5′-*GCCAACGCCACCAACCACCCCAACCCCGACGCCG<u>G</u>*-3′

5′-<u>TTTCGTAGAACGCTTTCAGCCACTC</u>*CGGCGTCGGG*-3′

5′-<u>TGAAAGCGTTCTACGAAAAGGTCCTGGAGAAACTG</u>-3′

5′-TCGTT**CTCGAG**<u>TCAGAACAGTTCTTTCAGTTTCTCCAGGACC</u>-3′
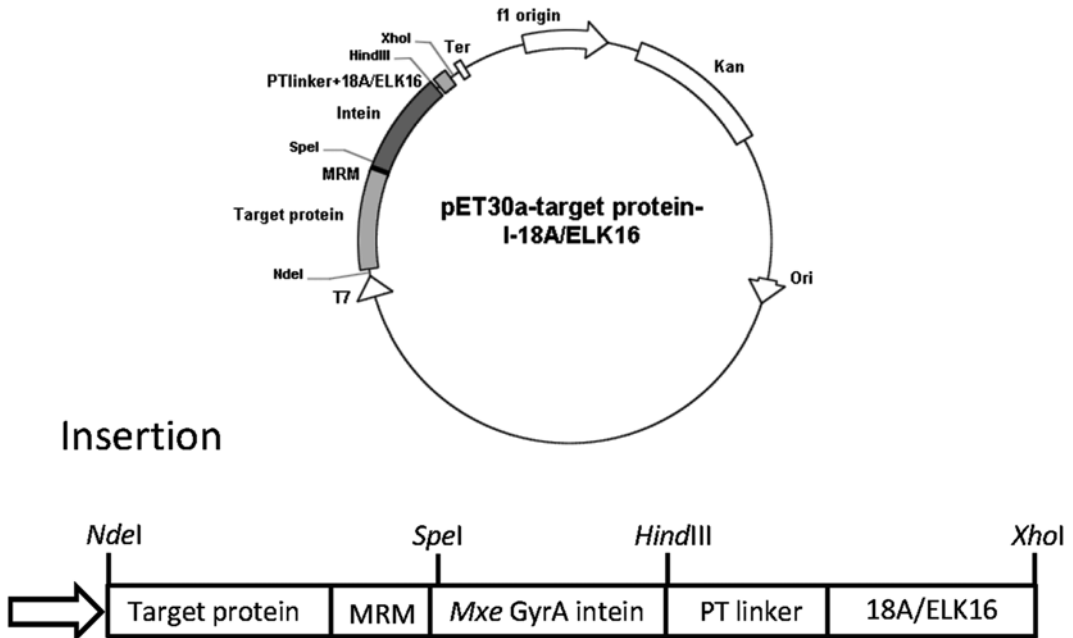
Fig. 1 Schematic illustration of the constructs pET30a-target protein-I-18A/ELK16, with the insertion detailed separately. Reproduced from ref. 13 (*see* acknowledgement)

The restriction endonuclease sites in these oligonucleotides are *Hind*III and *Xho*I (shown in bold), respectively. The sequences for the amphipathic α-helical octadecapeptide 18A and the PT-type linker are underlined and italicized, respectively.

2. Doubly digest the resulting PT linker-18A sequence with *Hind*III and *Xho*I at 37 °C for 12 h, purify the resulting product using a TIANquick Mini Purification Kit, and then ligate the fragment into the pET30a(+) plasmid which has been similarly digested with *Hind*III and *Xho*I and purified with a TIANgel Mini Purification Kit.

3. Thaw the chemically competent *E. coli* BL21(DE3) cells on ice. Transfer 10 μL of the ligation product into 100 μL of the competent cells in a microcentrifuge tube, mix the content by flicking, and incubate on ice for 30 min. Heat shock the cells at 42 °C for exactly 90 s and then place the cells on ice for 2 min. Subsequently, add 800 μL super optimal broth with catabolite repression (SOC) medium, and then transfer the mixture to a 15 mL tube. The cells are allowed to grow at 37 °C with shaking (220 rpm) for 45 min.

4. Plate the transformed *E. coli* BL21(DE3) cells on a Luria-Bertani (LB) agar plate supplemented with 50 mg/L kanamycin, and incubate it at 37 °C overnight (ON).

5. Select 2–3 positive clones by colony PCR (94 °C for 2 min, 19 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 20 s,

with a final extension at 72 °C for 10 min) using rTaq DNA polymerase and the following primers (forward and reverse, respectively):

5′-TCTCAGAAGCTTCCGACCCCACCGACCAC-3′

5′-TTCGATCTCGAGTCAGAACAGTTCTTTCAGT-3′

The DNA templates for the colony PCR are prepared as follows: cells are scratched from individual colonies, resuspended in 10 μL ddH$_2$O, lysed at 95 °C for 5 min, and then centrifuged at 15,000×g for 5 min. 2 μL of the supernatant is then used in one colony PCR reaction (in a total volume of 10 μL). Analyze the PCR products by gel electrophoresis. The positive clones should produce a DNA fragment of 132 bp.

6. Prepare the plasmid DNA using a TIANpure Mini Plasmid Kit, and send for sequencing to verify the construction. This yields pET30a-PT linker-18A.

7. Amplify the LipA gene from a previously constructed plasmid [12] by PCR (94 °C for 2 min, 30 cycles of 94 °C for 1 min, 59 °C for 1 min, and 72 °C for 1 min, with a final extension at 72 °C for 10 min) using pfu DNA polymerase and the following primers (forward and reverse, respectively):

5′-ACGACGA**CATATG**CACCATCACCATCACCAC CCCACCCCTATGGCTGAACACAATCCAGT-3′

5′-AAATTT**AAGCTT**ATTCGTATTCTGGCCCCCGC-3′

The restriction endonuclease sites in these primers are *Nde*I and *Hind*III (shown in bold), respectively.

8. Doubly digest the LipA gene with *Nde*I and *Hind*III at 37 °C for 12 h and purify. Ligate the product into the pET30a-PT linker-18A plasmid which has been similarly doubly digested with *Nde*I and *Hind*III to yield pET30a-LipA-PT linker-18A.

9. Amplify the sequence for LipA-PT linker from pET30a-LipA-PT linker-18A with a reverse primer carrying an extra sequence for ELK16 and using pfu DNA polymerase and the following primers (forward and reverse, respectively):

5′-ACGACGA**CATATG**GCTGAACACAATCCAGT-3′

5′-TCGTT**CTCGAG**TCA<u>TTTCAGCTTTAATTCTAAT TCCAGTTTTAACTTCAGTTCAAGTTCCAG</u>*CGGCGTCG GGGTTGGGGTGGTTGG*-3′

The restriction endonuclease sites in these primers are *Nde*I and *Xho*I (shown in bold), respectively. The sequences for ELK16 and part of the PT-type linker are underlined and italicized, respectively. The PCR conditions are as follows: 94 °C for 2 min, 30 cycles of 94 °C for 1 min, 59 °C for 1 min, and 72 °C for 1 min 30 s, with a final extension at 72 °C for 10 min.

10. Doubly digest the product with *Nde*I and *Xho*I at 37 °C for 12 h and purify. Ligate the product into the pET30a(+) plasmid, which has been doubly digested with *Nde*I and *Xho*I to yield pET30a-LipA-PT linker-ELK16.

11. Amplify the LipA gene from pET30a-LipA-PT linker-ELK16 using pfu DNA polymerase and the following primers (forward and reverse, respectively):

    5′-GCGATA**CATATG**CACCATCACCATCA-3′

    5′-GCATCTCCCGTGATGCA*CATTCGCAT*ATTCGTA TTCTGGCCCC-3′

12. The restriction endonuclease site in the forward primer is *Nde*I (shown in bold). The sequence for the three extra residues MRM at the C-terminus of LipA is italicized (*see* **Note 3**). The PCR conditions are as follows: 94 °C for 2 min, 30 cycles of 94 °C for 1 min, 63 °C for 1 min, and 72 °C for 1 min 15 s, with a final extension at 72 °C for 10 min. Amplify the gene encoding *Mxe* GyrA intein from the pTWIN1 plasmid using pfu DNA polymerase and the following primers (forward and reverse, respectively):

    5′-GGGGCCAGAATACGAAT*ATGCGAATG*TGCATCA CGGGAGAT-3′

    5′-ATTTTA**AAGCTT**AGCGTGGCTGACGAACCCG TTC-3′

13. The restriction endonuclease site in the reverse primer is *Hind*III (shown in bold). The sequence for the extra three residues MRM at the N-terminus of *Mxe* GyrA intein is italicized. The forward primer has a 41 bp overlap with the reverse primer used in the previous step to amplify the LipA gene. Please also note that there is a *Spe*I site in the sequence of *Mxe* GyrA intein (Fig. 1), which is located 18 bp downstream from its 5′ end. The PCR conditions are as follows: 94 °C for 2 min, 30 cycles of 94 °C for 1 min, 70 °C for 1 min, and 72 °C for 1 min 15 s, with a final extension at 72 °C for 10 min. Overlap the gene of LipA with that of Mxe GyrA intein using pfu DNA polymerase to yield the fusion gene LipA-Mxe GyrA. 100 ng each of the overlapping DNA fragments containing the two genes as generated in the above two steps is used in the PCR assembly reaction in a total volume of 100 μL and with the following conditions: 94 °C for 2 min, 30 cycles of 94 °C for 1 min, 70 °C for 1 min, and 72 °C for 2 min 30 s, with a final extension at 72 °C for 10 min.

14. Double digest LipA-*Mxe* GyrA with *Nde*I and *Hind*III and purify. Ligate the product into similarly digested vector pET30a-LipA-PT linker-18A or pET30a-LipA-PT linker-ELK16 to yield pET30a-LipA-I-18A or pET30a-LipA-I-ELK16.

3.1.2 pET30a-Target Protein-I-18A/ELK16 Construction

As an example, the construction of pET30a-AMA-I-18A/ELK16 (for the target protein AMA) is described here (*see* **Notes 4** and **5**).

1. Amplify the gene of AMA by PCR from a plasmid constructed previously [12] using pfu DNA polymerase, with forward and reverse primers that introduce an *Nde*I site and a *Spe* I site, respectively:

   5′-TTCTGGA**CATATG**GCGGTAACCAAGTCATC-3′

   5′-GGTGGT**ACTAGT**GCATCTCCCGTGATGCA*CATT CGCAT*TAACTTGGAAATATCTCTATA-3′

   The restriction endonuclease sites *Nde*I and *Spe*I are shown in bold, respectively. The sequence for the extra MRM residues is italicized. Please note that the reverse primer carries the first 23 nucleotides of the *Mxe* GyrA gene that contains the *Spe*I site. The PCR conditions are as follows: 94 °C for 2 min, 30 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min 30 s, with a final extension at 72 °C for 10 min.

2. Doubly digest the gene with *Nde*I and *Spe* I at 37 °C for 12 h, and purify. Ligate the product into the similarly digested pET30a-LipA-I-18A/ELK16 to yield the expression vectors pET30a-AMA-I-18A/ELK16.

3.2 Protein Expression

1. Inoculate a single *E. coli* BL21(DE3) colony harboring pET30a-target protein-I-18A/ELK16 into LB medium supplemented with 50 mg/L kanamycin, and incubate ON with shaking (250 rpm) at 37 °C.

2. Subsequently, dilute the saturated ON culture 50-fold into fresh LB medium supplemented with 50 mg/L kanamycin, and incubate it at 37 °C for about 1.5 h with shaking (250 rpm) until the culture reaches an optical density at 600 nm (OD_{600nm}) of 0.4–0.6.

3. Add isopropyl β-d-1-thiogalactopyranoside (IPTG) to the culture at a final concentration of 0.2 mM to initiate protein expression. The culture is then continued for an additional 6 h at 23, 30, or 37 °C with shaking (250 rpm) (*see* **Note 6**).

4. Harvest the cells by centrifugation for 20 min at $6,000 \times g$ and proceed to the purification step, or store the pellets at –70 °C until use.

3.3 Protein Purification and SDS-PAGE Analysis

3.3.1 Protein Purification

The general purification scheme is shown in Fig. 2.

1. Resuspend the cell pellets with buffer B1 to a final concentration of 10 OD_{600nm}/mL. Sonicate 1 mL of the resuspended cells for 99 pulses of 2 s each with a 2 s interval in an ice-water bath.

2. Centrifuge at $15,000 \times g$ for 15 min at 4 °C and collect both the soluble and insoluble fractions.
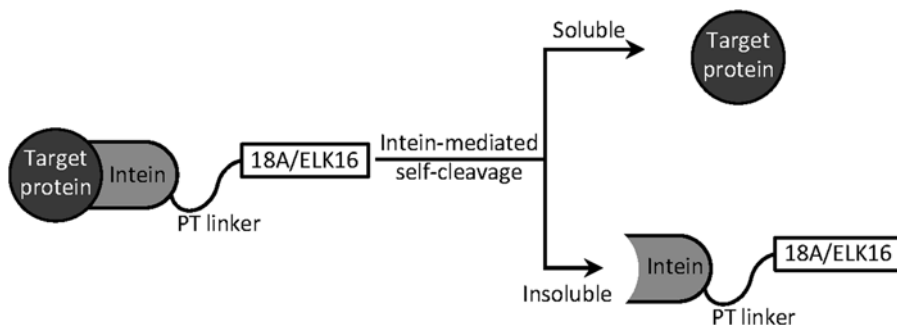
Fig. 2 Strategy for protein expression and purification using cleavable self-aggregating tags I-18A/ELK16 contained in pET30a-target protein-I-18A/ELK16. Reproduced from ref. 13 (*see* acknowledgement)

3. Wash the insoluble fraction 2–3 times in 1 mL of buffer B1, resuspend, and centrifuge at $15,000 \times g$ for 15 min at 4 °C.

4. Initiate the cleavage reaction by resuspending the precipitates with the same volume of buffer B3, and incubate the sample at 4 °C for 24 h (*see* **Notes 7** and **8**).

5. Centrifuge the sample at $15,000 \times g$ for 15 min at 4 °C to separate the soluble target protein from the insoluble contaminants. Collect both the soluble and insoluble factions.

3.3.2 SDS-PAGE Analysis

1. Prepare the protein samples by mixing 4 μL 6× protein loading buffer with 20 μL of the protein samples. Boil the mixtures for 10 min at 95 °C.

2. Load 12 μL of the boiled protein samples in each lane of a 12 % acrylamide gel, run SDS-PAGE gel, and stain it with Coomassie Brilliant Blue R250.

3. Estimate the protein amount of each band colorimetrically using BSA as the standard by a BCA Protein Assay Kit and using Quantity One software [19].

4. Figure 3 shows the SDS-PAGE results for LipA and AMA. The fusion proteins LipA-I-18A (46.6 kDa, Fig. 3a), LipA-I-ELK16 (46.1 kDa, Fig. 3b), AMA-I-18A (74.8 kDa, Fig. 3c), and AMA-I-ELK16 (74.3 kDa, Fig. 3d) were expressed as insoluble aggregates with yields of 34.1, 31.0, 19.1, and 23.2 μg/mg wet cell pellet, respectively (lanes lp in Fig. 3a, d). After DTT-mediated intein cleavage, LipA (21.0 kDa) and AMA (48.8 kDa) were released into the solution with yields of 10.4, 8.3, 7.9, and 4.0 μg/mg wet cell pellet, respectively (lanes cs in Fig. 3a, d). These yields are comparable to those of other quick purification schemes such as the classical his-tag purification [13] (*see* **Note 9**).

5. For the cSAT tag containing ELK16, upon intein self-cleavage, the I-ELK16 fusion remained almost totally insoluble (lanes
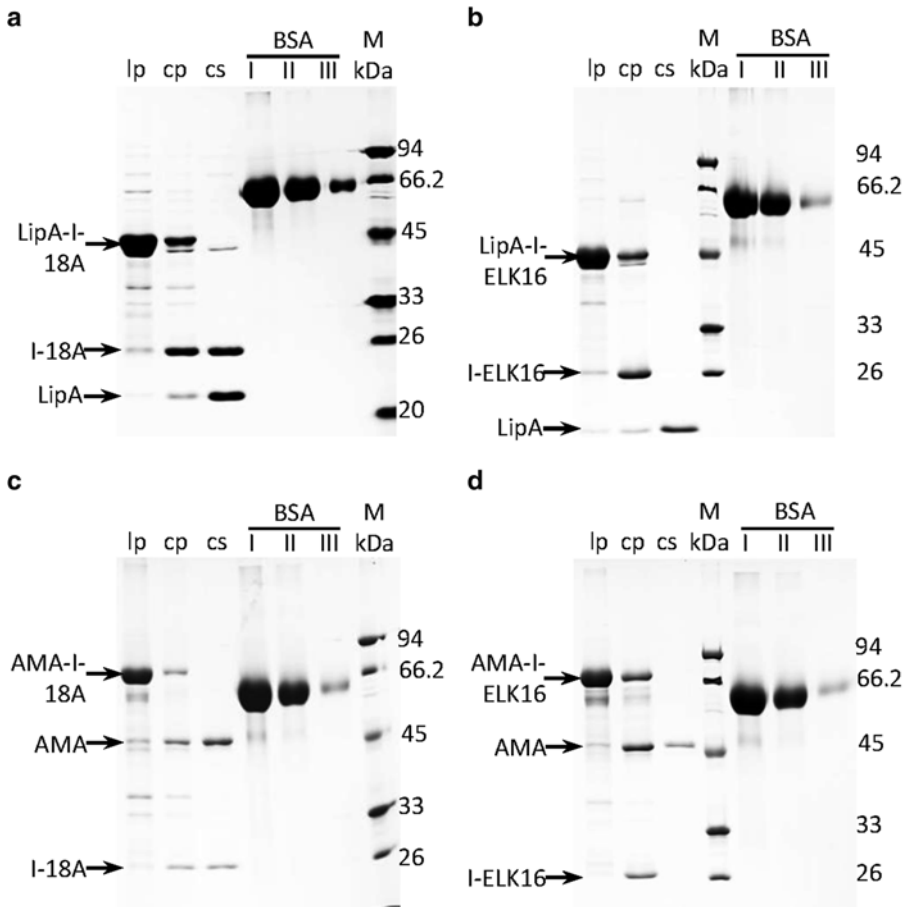
Fig. 3 Expression and purification of protein using cleavable self-aggregating tags I-18A/ELK16. (**a**) LipA-I-18A; (**b**) LipA-I-ELK16; (**c**) AMA-I-18A; (**d**) AMA-I-ELK16. Lanes: lp, insoluble fraction of cell lysate; cp, insoluble fraction of fusion protein upon DTT-mediated intein self-cleavage; cs, soluble fraction of fusion protein upon DTT-mediated intein self-cleavage; M, the molecular weight marker (14.4–94.0 kDa). BSA: bovine serum albumin (BSA) standards, at 6 (I) μg/lane, 3 (II) μg/lane, and 0.75 (III) μg/lane, respectively. Reproduced from ref. 13 (*see* acknowledgement)

cp and cs in Fig. 3b, d). The purity of the target protein was estimated to be about 90 % [13]. However, for the cSAT tag containing 18A, it can be seen that I-18A became partially soluble after DTT-induced intein cleavage, and thus contaminated the target protein in the supernatant (lanes cs in Fig. 3a, c). If the application of the target proteins is interfered by the presence of the I-18A fusion, an additional step of reverse phase-high performance liquid chromatography (RP-HPLC) is then needed.

3.4 Activity Assays of LipA and AMA

Measure the activity of LipA and AMA in a 96-well microplate based on the standard protocols with a SPECTRAMAX M2 microtiter reader [20, 21].

3.4.1 Activity Assay of LipA

1. Dissolve 6 mg of the substrate *p*-nitrophenyl palmitate (*p*NPP) into 2 mL 2-propanol, and then mix it with 38 mL of the LipA reaction buffer by vortexing. The final concentration of *p*NPP is 1.5 mg/mL.

2. Preheat the reaction mixture in a water bath at 37 °C and set the chamber temperature of the SPECTRAMAX M2 microtiter reader at 37 °C.

3. Add 5 μL of LipA (diluted fivefold) to 175 μL of the reaction mixture in a 96-well microplate; mix using a multichannel electronic pipettor rapidly but carefully to avoid generating bubbles.

4. Place the 96-well microplate into the SPECTRAMAX M2 microtiter reader; measure the activity by monitoring the formation of *p*-nitrophenol (*p*NP) using Abs_{405nm} (ε, 18.7 cm²/μmol).

5. One unit of enzyme activity of LipA is defined as the amount of enzyme that produces 1 μmol *p*NP per min.

6. We found that, although the LipA-I-18A/ELK16 aggregates showed little activity, the specific activity of LipA released from the aggregates was comparable to that reported for the native LipA [13].

3.4.2 Activity Assay of AMA

1. Prepare the AMA reaction mixture.

2. Preheat the mixture in a water bath at 37 °C, and set the chamber temperature of the SPECTRAMAX M2 microtiter reader at 37 °C.

3. Add 5 μL of AMA to 175 μL of the reaction mixture in a 96-well microplate, and mix using a multichannel electronic pipettor rapidly and carefully to avoid generating bubbles.

4. Place the 96-well microplate into the SPECTRAMAX M2 microtiter reader, and measure the activity by monitoring the formation of the quinone dye using Abs_{555nm} (ε, 39.2 cm²/μmol).

5. One unit of enzyme activity of AMA is defined as the amount of enzyme that produces 1 μmol H_2O_2 per min.

6. We found that the AMA-I-18A/ELK16 aggregates were active and the specific activity of AMA released from the aggregates was comparable to that reported for the native AMA [13].

4 Notes

1. The expression vectors pET30a-LipA-I-18A/ELK16 described here are constructed in a stepwise and somehow complicated manner due to historical reasons associated with the line of the

work. One might wish to simplify the construction by assembling all the DNA elements (plasmid backbone, target gene, the sequence of intein, linker, the sequence of 18A or ELK) in one step by using now standard assembling techniques such as the Gibson assembly, which joins overlapping DNA sequences in one step [22].

2. While, in the current expression vectors, 18A or ELK16 is fused to the C-terminus of intein via a PT-type linker PTPPTTPTPPTTPTPTP, a GS-type linker (GGGGS)₃ was also tested and was found to perform similarly as the PT linker.

3. In the current expression vectors, three amino acid residues Met-Arg-Met (MRM) are inserted between the C-terminus of the target protein and the N-terminus of the *Mxe* GyrA intein to facilitate the self-cleavage of intein, as suggested by the literature [18]. However, our preliminary experiments suggest that this tripeptide may not be necessary.

4. To facilitate cloning, when applicable, one may need to eliminate *Nde*I, *Hind*III, *Eco*RI, and *Xho*I sites from the target gene sequences using synonymous site-directed mutagenesis.

5. Dephosphorylation of the linearized vector (pET30a-LipA-I-18A/ELK16 doubly digested with *Nde*I and *Spe*I) by heat-liable SAP is sometimes necessary to increase the cloning efficiency.

6. The amounts of active protein aggregates and released target proteins are generally affected by the expression temperature. Thus, for a given target protein (or peptide), it should be optimized by a trial-and-error approach. For AMA and LipA, the optimal temperature is 30 °C. The expression time can also be changed for the same reason.

7. Since DTT is not stable in solution, freshly prepared buffer B3 is recommended. Alternatively, the buffer can be stored at –20 °C until use.

8. In our work, we have tested four different cleavage conditions (4 °C or 25 °C for 3 or 24 h, all at pH 8.5) with LipA as a model protein. In general, the self-cleavage efficiency of *Mxe* GyrA intein increases when increasing temperature or time. At 4 °C, it was found that the cleavage efficiency at 3 h was lower than that of 24 h. At 25 °C, the cleavage efficiency at 3 h was almost the same with that of 24 h. Furthermore, the cleavage efficiency of this intein in buffer B3 under different pH (5.6, 7.0, 8.5) was rather similar. Since proteins are more stable at low temperatures, the cleavage reaction is thus set at 4 °C for 24 h in buffer B3, pH 8.5 [13].

9. The possible contamination of nucleic acids can be checked by determining the OD_{260nm}/OD_{280nm} ratio of the released target protein.

Acknowledgments

References

1. Prouty W, Karnovsky M, Goldberg A (1975) Degradation of abnormal proteins in *Escherichia coli*. Formation of protein inclusions in cells exposed to amino acid analogs. J Biol Chem 250:1112–1122

2. Lilie H, Schwarz E, Rudolph R (1998) Advances in refolding of proteins produced in *E. coli*. Curr Opin Biotechnol 9:497–501

3. Achmueller C, Kaar W, Ahrer K et al (2007) N[pro] fusion technology to produce proteins with authentic N termini in *E. coli*. Nat Methods 4:1037–1043

4. Lee JH, Kim JH, Hwang SW et al (2000) High-level expression of antimicrobial peptide mediated by a fusion partner reinforcing formation of inclusion bodies. Biochem Biophys Res Commun 277:575–580

5. Wei QD, Kim YS, Seo JH et al (2005) Facilitation of expression and purification of an antimicrobial peptide by fusion with baculoviral polyhedrin in *Escherichia coli*. Appl Environ Microbiol 71:5038–5043

6. Yamaguchi S, Yamamoto E, Mannen T et al (2013) Protein refolding using chemical refolding additives. Biotechnol J 8:17–31

7. Kyle S, Aggeli A, Ingham E et al (2010) Recombinant self-assembling peptides as biomaterials for tissue engineering. Biomaterials 31:9395–9405

8. Tokatlidis K, Dhurjati P, Millet J et al (1991) High activity of inclusion bodies formed in *Escherichia coli* overproducing Clostridium thermocellum endoglucanase D. FEBS Lett 282:205–208

9. Worrall DM, Goss NH (1989) The formation of biologically active beta-galactosidase inclusion bodies in *Escherichia coli*. Aust J Biotechnol 3:28–32

10. García-Fruitós E, González-Montalbán N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27–32

11. Vera A, González-Montalbán N, Arís A et al (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng 96:1101–1106

12. Wu W, Xing L, Zhou B et al (2011) Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in *Escherichia coli*. Microb Cell Fact 10:9–16

13. Xing L, Wu W, Zhou B et al (2011) Streamlined protein expression and purification using cleavable self-aggregating tags. Microb Cell Fact 10:42–48

14. Zhou B, Xing L, Wu W et al (2012) Small surfactant-like peptides can drive soluble proteins into active aggregates. Microb Cell Fact 11:10–17

15. Mitraki A (2010) Protein aggregation: from inclusion bodies to amyloid and biomaterials. In: McPherson A (ed) Advances in protein chemistry and structural biology, 1st edn. Elsevier-Academic, San Diego, pp 89–125

16. Volkmann G, Mootz HD (2012) Recent progress in intein research: from mechanism to directed evolution and applications. Cell Mol Life Sci 70:1185–1206

17. Telenti A, Southworth M, Alcaide F et al (1997) The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal. J Bacteriol 179:6378–6382

18. Ge X, Yang DSC, Trabbic-Carlson K et al (2005) Self-cleavable stimulus responsive tags for protein purification without chromatography. J Am Chem Soc 127:11228–11229

19. Walker JM (2009) The bicinchoninic acid (BCA) assay for protein quantitation. In: Walker JM (ed) The protein protocols handbook, 2nd edn. Humana, New York, pp 11–15

20. Winkler UK, Stuckmann M (1979) Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by Serratia marcescens. J Bacteriol 138:663–670

21. Sakaue R, Kajiyama N (2003) Thermostabilization of bacterial fructosyl-amino acid

oxidase by directed evolution. Appl Environ Microbiol 69:139–145

22. Gibson DG, Young L, Chuang RY et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345

23. Xing L, Xu WH, Zhou BH et al (2013) Facile expression and purification of the antimicrobial peptide histatin 1 with a cleavable self-aggregating tag (cSAT) in *Escherichia coli*. Protein Expr Purif 88:248–253

Chapter 5

Beyond the Cytoplasm of *Escherichia coli*: Localizing Recombinant Proteins Where You Want Them

Jason T. Boock, Dujduan Waraho-Zhmayev, Dario Mizrachi, and Matthew P. DeLisa

Abstract

Recombinant protein expression in *Escherichia coli* represents a cornerstone of the biotechnology enterprise. While cytoplasmic expression in this host has received the most attention, achieving substantial yields of correctly folded proteins in this compartment can sometimes be met with difficulties. These issues can often be overcome by targeting protein expression to extracytoplasmic compartments (e.g., membrane, periplasm) or to the culture medium. This chapter discusses various strategies for exporting proteins out of the cytoplasm as well as tools for monitoring and optimizing these different export mechanisms.

Key words Cell factories, *Escherichia coli*, Extracellular protein production, Membrane protein expression, General secretory pathway, Twin-arginine translocation, YebF

1 Introduction

1.1 Protein Expression Using Escherichia coli Cell Factories

Over the past three decades, various host organisms have emerged as viable options for producing recombinant proteins with desired quality and quantity. Of these, the Gram-negative bacterium *Escherichia coli* remains one of the most popular given its extraordinary versatility [1]. *E. coli* is well known for (1) its rapid growth and ability to reach high cell densities using inexpensive substrates; (2) its well-characterized genetics; (3) the availability of large numbers of cloning/expression vectors and mutant strain collections (e.g., Keio [2]); and (4) the ease with which new strains can be engineered [3]. As a result, recombinant *E. coli* strains can express recombinant products in the cytoplasm at levels that reach up to 50 % of the total cellular protein. Nonetheless, many heterologous proteins are refractory to production in the cytoplasm, due in large part to either poor expression of the cloned gene or aberrant folding of the naive polypeptide. A variety of techniques have been developed to solve these problems, for example: (1) using plasmids

with different promoters and/or copy number, (2) using special-ized host strains and optimal growth temperatures, (3) changing the codon bias or the 5′ untranslated region of the cloned gene [4, 5], and (4) remodeling the folding environment by co-expression of molecular chaperones or alteration of the redox potential [6–8] (*see* Chapter 2). Even when these challenges can be overcome, product recovery is nontrivial given the large number of host proteins that accumulate in the cytoplasm alongside the protein of interest as well as host proteases that degrade the protein product. Further complications can arise from the tendency of some overex-pressed proteins to form cytoplasmic inclusion bodies, which must be subjected to expensive, labor-intensive denaturation/refolding processes to obtain biologically active proteins.

In *E. coli*, even though protein synthesis only takes place in the cytoplasm, about 40 % of all polypeptides are inserted into the inner and outer membranes, targeted to the periplasm or excreted into the growth medium [9]. As our understanding of these natu-ral mechanisms grows, so too does the number of applications that exploit these mechanisms for recombinant protein expression. This chapter focuses on such strategies, which effectively circumvent the problems associated with cytoplasmic production by targeting the protein of interest to an extracytoplasmic compartment (e.g., membrane, periplasm) or to the extracellular medium (for recent reviews on these topics, *see* [10–12]).

1.2 Secretory Expression

There are several features of the *E. coli* periplasm that make it attrac-tive for secretory protein expression and serve to refute the miscon-ception that lower yields are obligatory for secretory expression systems [10]. First, cleavage of the N-terminal export signal by a specific signal peptidase leads to formation of an N-terminal amino acid that identically matches that of the natural gene product. Second, recombinant proteins expressed in the periplasm are less prone to proteolysis as there is less protease activity present in this compartment compared to the cytoplasm [11]. Third, molecular chaperones and other folding modulators are naturally available in the periplasm—or can be co-expressed [13]—to assist the folding of newly synthesized proteins [1]. This includes the chaperones FkpA, SurA, and Skp as well as the network of redox enzymes (e.g., DsbA, DsbC) that catalyze disulfide bond formation and naturally reside in the *E. coli* periplasm [14]. Fourth, the periplasm contains only about 4 % of the total cellular proteins [15] which may reduce the cost and simplify the process of product purification.

Protein export from the cytoplasm to the periplasm involves complex secretion machineries known as translocases. Approximately 90 % of secreted proteins are exported in an unfolded state through the SecYEG translocase either via the posttranslational Sec pathway or the co-translational signal recog-nition particle (SRP) pathway (Fig. 1) [10, 11]. On the other hand, a smaller but still significant subset of proteins is exported in
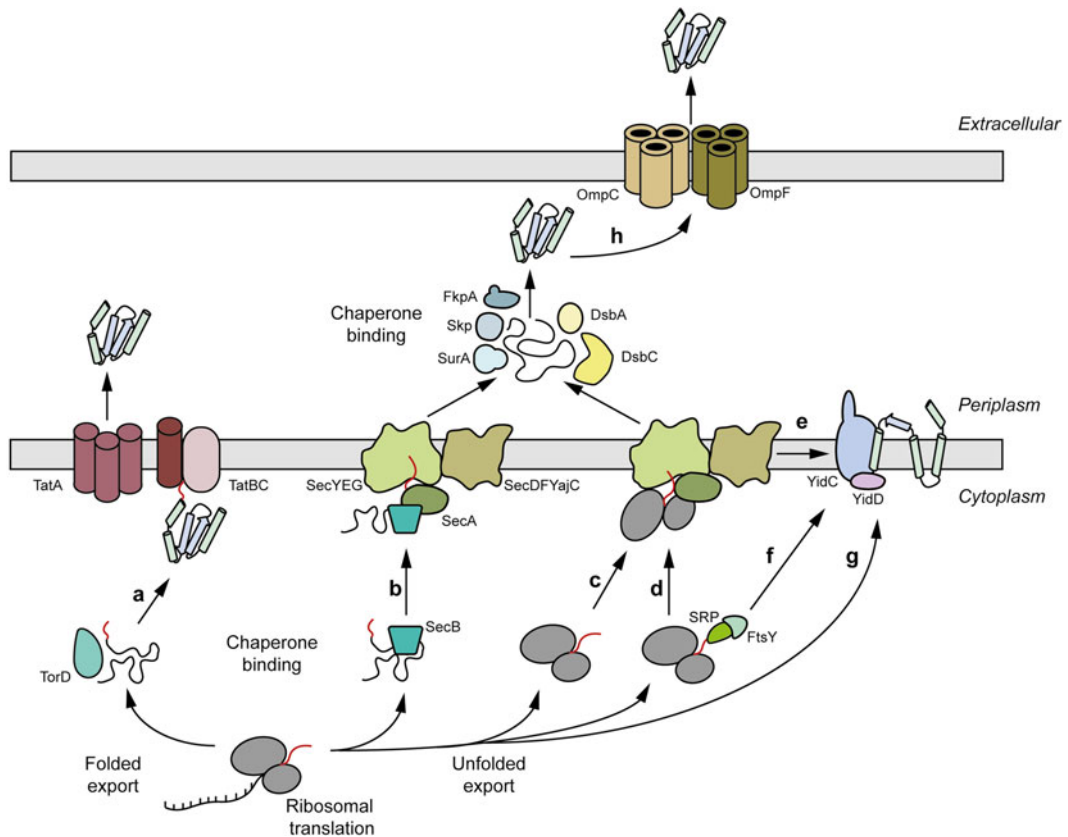
Fig. 1 The biogenesis of periplasmic, extracellular, and inner membrane proteins in *E. coli*. Proteins destined for the periplasm are translated with N-terminal signal peptides (*red line*) that direct Tat-, Sec-, or SRP-dependent export. These signal peptides are later removed by leader peptidase (Lep, not shown). Tat export is a posttranslational mechanism that involves completely folded substrates (**a**). Export is accomplished by the Tat translocase but the precise export mechanism is yet to be determined. Quality control along the Tat pathway includes proofreading chaperones (e.g., TorD) and direct sensing of substrate foldedness by the Tat translocase. Sec export is also a posttranslational mechanism, but instead involves unfolded substrates. Some Sec substrates remain unfolded with assistance from the SecB chaperone (**b**) while others are exported in a SecB-independent fashion (**c**). Sec export is accomplished by the Sec translocase, which together with the SecA ATPase ratchets unfolded Sec substrates into the periplasm through a narrow diameter pore formed by SecYEG. Once in the periplasm, molecular chaperones (e.g., FkpA, Skp, SurA) and enzymes of the disulfide bond formation pathway (e.g., DsbA, DsbC) promote the correct folding of newly translocated Sec substrates. SRP-dependent export is a co-translocational mechanism whereby ribosome nascent chain complexes (RNCs) are targeted to the membrane via the signal recognition particle (SRP) and its receptor FtsY (**d**). At the inner membrane, the RNC docks at the Sec translocase and the newly translated substrate is directly injected into the periplasm. Insertion of membrane proteins into the cytoplasmic membrane can also follow the SRP pathway (**e**). Following docking of the RNC, YidC mediates the transfer of transmembrane segments (TMs) from the Sec translocase into the lipid bilayer and can also assist membrane protein folding. Some membrane proteins bypass the Sec translocase and are targeted to YidC either via the SRP pathway (**f**) or directly (**g**). The SecDFYajC complex can play a role in the biogenesis of membrane proteins as well as the translocation and folding of secreted proteins. Likewise, YidD functions in the biogenesis of both YidC- and Sec-YidC-dependent membrane proteins. The FtsH complex is involved in quality control and degradation of membrane proteins (not shown). At least one protein, YebF, is first translocated into the periplasm by the Sec pathway and then translocated across the outer membrane in a process that appears to involve OmpF and OmpC (**h**)

a fully folded conformation through the TatABC translocase of the twin-arginine translocation (Tat) pathway [16]. Sorting of Sec, SRP, and Tat substrates is accomplished via distinct N-terminal signal peptides, which are required for targeting to the correct translocase. For each of these pathways, there are native quality control (QC) mechanisms that ensure proper structural integrity of substrate proteins, so they remain compatible with their respective translocases [17]. Importantly, these QC mechanisms can be leveraged to increase the probability of producing high-quality (e.g., correctly folded) protein products [18, 19]. It should also be mentioned that the development of genetic screens and selections specific for the different export pathways has made it possible to conveniently monitor and optimize periplasmic targeting of protein substrates as well as their folding efficiency [20–23]. Finally, although outside the scope of this chapter, it should be pointed out that asparagine-linked (N-linked) protein glycosylation has been engineered in *E. coli* [24–27], which now makes it possible to attach complex glycans of defined structure to target proteins that are expressed in the periplasm.

1.3 Extracellular Expression

Extracellular expression of heterologous proteins offers advantages over production inside of cells including ease of purification due to lack of contaminant proteins, elimination of proteases or cell crowding issues associated with poor protein expression, and exploitation of chemistries only possible outside of the cell (i.e., degradation of non-membrane-permeable substrates). Unfortunately, whereas *E. coli* has been the "workhorse" for making recombinant proteins in the cytoplasmic and periplasmic compartments as discussed above, it has historically been overlooked for applications requiring extracellular expression. This is because the chromosomal genes that, in other Gram-negative bacteria, are involved in extracellular expression (e.g., type II secretion system encoded by the *gsp* genes) are not expressed when *E. coli* is grown under standard laboratory conditions [28, 29]. Recently, however, this situation has been reversed with the discovery of endogenous proteins, such as YebF and the osmotically inducible protein Y (OsmY), that are naturally excreted by laboratory *E. coli* strains without compromising integrity of the outer membrane [30, 31]. Both of these proteins are first localized into the periplasmic compartment via the Sec pathway and, in the case of YebF, translocation across the outer membrane appears to involve OmpF/C (Fig. 1) [32]. Importantly, both OsmY and YebF have been used as carriers to deliver biotechnologically relevant fusion partners into the culture medium [30, 33] including enzymes that break down recalcitrant plant biomass [34, 35]. More recently, our laboratory developed a universal genetic assay that can be applied to diverse secretory pathways and allows rapid,

high-throughput screening of bacterial secretion phenotypes [36]. This platform provides a convenient tool for dissecting poorly understood aspects of extracellular secretion as well as assisting in the construction of engineered *E. coli* strains for efficient extracellular protein production. Finally, while alternative strategies for extracellular expression in *E. coli* have been reported including importing known secretion pathways (e.g., type II and type III secretion systems) from other organisms [37, 38] and selective leakage into the culture medium [39], this chapter focuses on utilizing YebF.

1.4 Membrane Protein Expression

Membrane proteins are a special case because, unlike the secreted and extracellular substrates discussed above, they are not soluble in aqueous solution. The extensive number of hydrophobic amino acids in their primary sequence and ultimate exposed hydrophobicity upon folding impose the need for interaction with the nonpolar environment of the bilayer interior. In *E. coli*, the biogenesis of most inner membrane proteins involves co-translational targeting to the membrane by the SRP-dependent pathway (Fig. 1) [40], which is also responsible for the export of secretory proteins containing highly hydrophobic signal peptides [10, 11]. Insertion into the membrane involves the Sec translocase, after which the membrane protein moves laterally from the translocase into the lipid bilayer, folds into the native conformation, and often assembles into oligomeric complexes [40]. Malfolded membrane proteins are identified and degraded by QC mechanisms such as FtsH. An additional factor, YidC, functions specifically in the biogenesis of inner membrane proteins, not only in association with the Sec translocase but also separately. Alternatively, membrane proteins can bypass the SecYEG translocon entirely and be targeted directly to YidC. Membrane proteins are notoriously difficult to produce at the high levels required for structural and biochemical characterization. While many different expression systems have been used to date, *E. coli* remains one of the best characterized and most versatile hosts for expressing membrane proteins recombinantly [12, 41]. Indeed, numerous prokaryotic and eukaryotic membrane proteins have been produced using *E. coli*, including those with complex topologies such as mammalian G-protein-coupled receptors (GPCRs) [42, 43]. While the yields for some of these membrane proteins remain low, chaperone co-expression strategies have been used to successfully improve membrane protein expression [44, 45]. Moreover, a number of high-throughput genetic tools are now available for monitoring and optimizing the localization, quantity, and quality of overexpressed membrane proteins in *E. coli* [42, 43, 46–51]. These assays have been used to reveal mechanistic information, as well as to construct improved membrane protein variants or genetically engineered *E. coli* strains for efficient heterologous membrane protein production.

2   Materials

2.1   Expression Plasmids

2.1.1   Plasmids

Plasmids for periplasmic, extracellular, and membrane protein expression (summarized in Figs. 2, 3, and 4, respectively) are available upon request from Prof. Matthew DeLisa, School of Chemical Engineering, Cornell University, Ithaca, NY 14853 (e-mail: md255@cornell.edu). Useful plasmids include the following:

1. pTrc99A-Peri.
2. pTrc99A-YepF.
3. pET22a-GlpF/MstX.
4. pRARE (Novagen; Cm^R) for use with target proteins containing codons rarely used in *E. coli*.
5. pTUM4 (Cm^R) for overexpression of four established periplasmic chaperones and folding catalysts: the thiol-disulfide oxidoreductases DsbA and DsbC that catalyze the formation and isomerization of disulfide bridges and the peptidyl-prolyl *cis/trans*-isomerases with chaperone activity, FkpA and SurA [13].
6. pBR-TatABC (Tet^R) for overexpressing the TatABC machinery and enhancing export efficiency by this pathway [52].
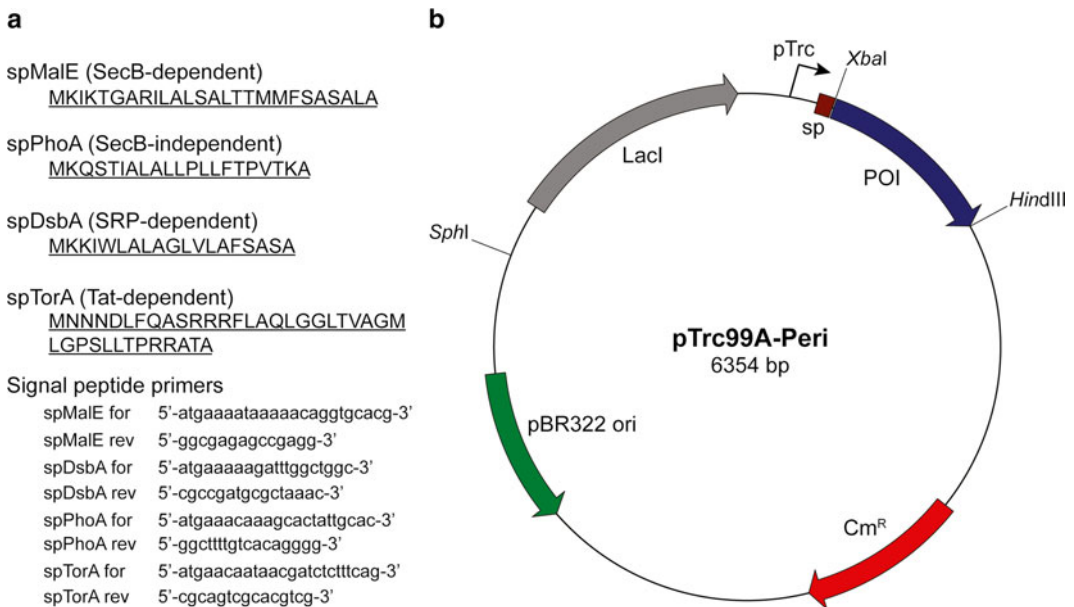
Fig. 2 Targeting expression to the periplasm. (**a**) Amino acid sequences of different signal peptide (sp) options that are commonly used for export into the periplasm. Also shown are the corresponding oligonucleotide primers that PCR amplify each sp. (**b**) A pTrc99A-based plasmid expressing a fusion between a chosen sp and a protein of interest (POI), resulting in accumulation of the POI in the periplasm [65]. The gene for the POI is inserted after the chosen sp using the *Xba*I and *Hin*dIII restriction sites. All restriction sites shown are unique

a

YebF Fusion Tag:

MKKRGAFLGLLLVSACASVFAANNETSKSVTFPKCEDLDA
AGIAASVKRDYQQNRVARWADDQKIVGQADPVAWVSLQDI
QGKDDKWSVPLTVRGKSADIHYQVSVDCKAGMAEYQRR

YebF Primers:

YebF for	5'-atgggaaaaagaggggcg-3'
YebF rev	5'-acgccgctgatattccg-3'
ΔspYebF for	5'-gccaataatgaaaccagcaag-3'

b
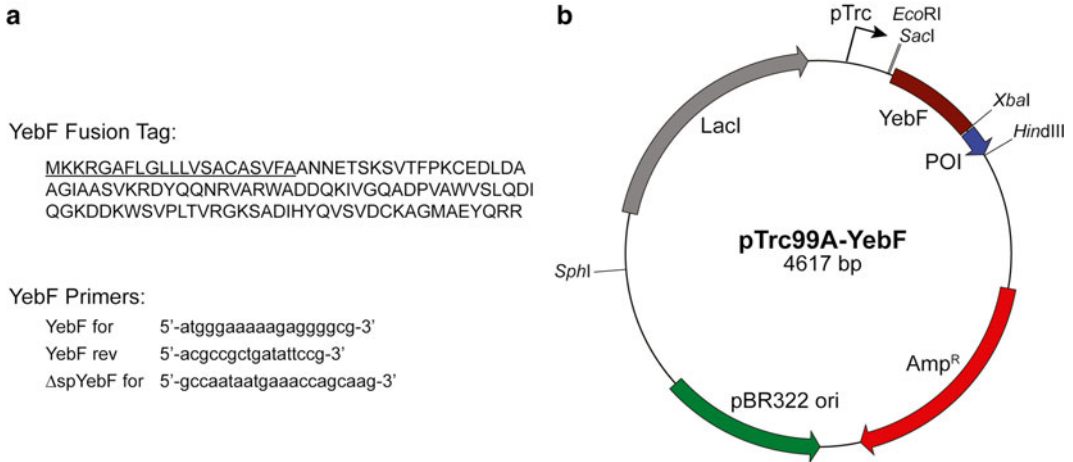
Fig. 3 Targeting expression to the extracellular medium. (**a**) Amino acid sequence of full-length YebF with its Sec-dependent signal peptide (sp) *underlined*. Also shown are oligonucleotide primers for PCR amplification of YebF as well as YebF without its native export signal (Δsp-YebF). (**b**) A pTrc99A-based plasmid for expressing a fusion between YebF and a protein of interest (POI), resulting in accumulation of the POI in the extracellular secretion [36]. The gene for the POI is inserted after full-length *yebF* using the *Xba*I and *Hin*dIII restriction sites. All restriction sites shown are unique

a

GlpF

MMSQTSTLKGQCIAEFLGTGLLIFFGVGCVAALKVAGASF
GQWEISVIWGLGVAMAIYLTAGVSGAHLNPAVTIALWLFA
CFDKRKVIPFIVSQVAGAFCAAALVYGLYYNLFFDFEQTH
HIVRGSVESVDLAGTFSTYPNPHINFVQAFAVEMVITAIL
MGLILALTDDGNGVPRGPLAPLLIGLLIAVIGASMGPLTG
FAMNPARDFGPKVFAWLAGWGNVAFTGGRDIPYFLVPLFG
PIVGAIVGAFAYRKLIGRHLPCDICVVEEKETTTPSEQKASL

MstX

MFCTFFEKHHRKWDILLEKSTGVMEAMKVTSEEKEQLSTA
IDRMNEGLDAFIQLYNESEIDEPLIQLDDDTAELMKQARD
MYGQEKLNEKLNTIIKQILSISVSEEGEKE

GlpF/MstX primers

GlpF for	5'-atgatgagtcaaacatcaaccttg-3'
GlpF rev	5'-cagcgaagctttttgttctg-3'
MstX for	5'-atgtttttgtacatttttgaaaaacatc-3'
MstX rev	5'-ttctttttctccttcttcagatactg-3'

b

Fig. 4 Targeting expression to the inner membrane. (**a**) Amino acid sequences of two commonly used membrane-protein expression enhancers, GlpF and MstX, and oligonucleotide primers used to amplify them. (**b**) A pET22a-based plasmid for expressing a fusion between GlpF (or MstX) and a protein of interest (POI), resulting in enhanced inner membrane expression of the POI. The gene for the POI is inserted after full-length *glpF* (or *mstX*) using the *Nhe*I and *Xho*I/*Eag*I restriction sites. All restriction sites shown are unique

7. pOFXbad-KJ2 (Spec^R) for overexpression of chaperones DnaKJ that improve Tat export efficiency [53].

8. pPspA (Amp^R) for enhancing export via the Tat pathway [54]; and pBAD-FtsH (Cm^R) for enhancing membrane protein expression [43].

2.1.2 Molecular Cloning Materials

1. Oligonucleotide primers for genes of interest.

2. DNA template for genes of interest.

3. PCR supplies: dNTPs, DNA polymerase, PCR buffer, and thermocycler.

4. Restriction enzymes: XbaI, HindIII, NheI, XhoI, EagI.

5. T4 DNA ligase and buffer.

6. Electrocompetent *E. coli* strain (e.g., DH5α or other *recA*-deficient host).

2.2 Cell Growth and Protein Expression

2.2.1 Strains

1. *E. coli* strain BL21(DE3) is commonly employed for the high-level expression of recombinant proteins using the T7 promoter of pET-based plasmids.

2. *E. coli* strain BW25113 and single-gene knockout mutants derived thereof (i.e., the Keio collection) [2] are employed for genetic analysis or expression optimization studies.

3. *E. coli* strain MC4100 and *tat*-deficient mutant strains B1LK0 (MC4100 Δ*tatC*) and DADE (MC4100 Δ*tatABCD*Δ*tatE*) are commonly used for expression analysis of Tat-dependent substrates.

4. *E. coli* strains with single-gene knockouts of *entE*, *nlpD*, or *tnaA* in BW25113 and BL21(DE3) backgrounds can be used for enhanced extracellular expression by the YebF pathway [36].

5. *E. coli* strain MC4100 *dnaJ*::Tn5 can be used for high-level expression of membrane proteins [55].

6. *E. coli* strains C41(DE3) and C43(DE3) [56], also referred to as the Walker strains, are commonly employed for membrane protein expression using pET-based plasmids.

7. *E. coli* strains Tuner™ (Novagen), BL21-AI (Invitrogen), KRX (Promega), and Lemo21(DE3) (New England Biolabs) allow for tunable expression of gene products which can help to avoid toxic overproduction (*see* **Note 1**).

2.2.2 Growth Media and Culture Materials

1. Luria-Bertani (LB): 10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, and 950 mL of ddH$_2$O. Alternatively 5 and 0.5 g of NaCl are used. The low salt formulations are ideal for cultures requiring salt-sensitive antibiotics (e.g. Blasticidin, Clonat, Hygromycin B, Puromycin and Zeocin). Stir to dissolve all solids and bring the final volume to 1 L with ddH$_2$O before autoclaving.

2. Terrific Broth (TB): 12 g of tryptone, 24 g of yeast extract, 4 mL glycerol, and 800 mL of ddH$_2$O. Stir to dissolve all

solids and bring the final volume to 900 mL with ddH$_2$O before autoclaving. A 10× buffered salt solution is prepared by dissolving 23.1 g of KH$_2$PO$_4$ (0.17 M final) and 125.3 g of K$_2$HPO$_4$ (0.72 M final) in a final volume of 1 L ddH$_2$O and filtered through a 0.22-μm filter. After autoclaving, add 100 mL of the 10× buffered salt solution to the 900 mL of sterile medium.

3. LB-agar plates: Add 15 g of agar per L of LB broth prepared as above before autoclaving.

4. Sterilized 200-mL baffled glass flasks.

5. Temperature-controlled incubator.

6. Spectrophotometer.

2.2.3 Antibiotics and Inducer

1. 100 mg/mL ampicillin (Amp) stock solution: Dissolve 1 g of Amp disodium salt in 10 mL of ddH$_2$O. Filter sterilize through a 0.22-μm membrane and store at –20 °C in 1-mL aliquots. For a 100 μg/mL working solution, dilute 1:1,000 (*see* **Note 2**).

2. 50 mg/mL carbenicillin (Carb) stock solution: Dissolve 0.5 g of Carb disodium in 10 mL of ddH$_2$O. Filter sterilize through a 0.22-μm membrane and store at –20 °C in 1-mL aliquots. For a 50 μg/mL working solution, dilute 1:1,000.

3. 50 mg/mL kanamycin (Kan) stock solution: Dissolve 0.5 g of Kan sulfate in 10 mL of ddH$_2$O. Filter sterilize through a 0.22-μm filter and store at –20 °C in 1-mL aliquots. For a 50 μg/mL working solution, dilute 1:1,000.

4. 20 mg/mL chloramphenicol (Cm) stock solution: Dissolve 0.2 g of Cm in 10 mL of ethanol. Filter sterilize through a 0.22-μm membrane and store at –20 °C in 1-mL aliquots. For a 20 μg/mL working solution, dilute 1:1,000.

5. 20 % (w/v) arabinose stock solution: Dissolve 2 g of L-arabinose in 8 mL of ddH$_2$O. Mix until all solid has dissolved, adjust the volume to 10 mL, and filter sterilize. Store at 4 °C. To induce at a concentration of 0.2 % (w/v), use a 1:100 dilution of the stock.

6. 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) stock solution: Dissolve 2.38 g of IPTG in 8 mL of ddH$_2$O. Mix until all solid has dissolved, adjust the volume to 10 mL, and filter sterilize. Store at –20 °C. To induce at a concentration of 1 mM, use a 1:1,000 dilution of the stock.

2.3   Cellular Fractionation

2.3.1   Preparation of Periplasmic and Cytoplasmic Fractions

1. Sucrose buffer: 50 mM Tris–HCl, pH 7.4, 1 mM EDTA, 20 % sucrose w/v, 5 mM MgCl$_2$. Dissolve 157.6 g of Tris–HCl in 1 L ddH$_2$O, adjust to a pH of 7.4 using NaOH pellets, and sterile filter to make a 1 M stock solution of Tris–HCl, pH 7.4. Add 50 mL of the 1 M Tris–HCl, pH 7.4, 2 mL of 0.5 M EDTA, 20 g sucrose, and 476 mg of MgCl$_2$ to 900 mL ddH$_2$O and dissolve. Adjust to a final volume of 1 L with ddH$_2$O.

2. 5 mM magnesium chloride: 476 mg of $MgCl_2$ in 1 L ddH_2O. Filter sterilize.

3. Phosphate-buffered saline (1× PBS): 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 in 1 L ddH_2O. Filter sterilize. Final pH should be ~7.4.

4. Sonicator, cell lysis solution, or other method of cell rupture.

5. Centrifuge.

6. 0.22-μm sterile filter.

2.3.2 Preparation of Extracellular Fraction

1. 100 % trichloroacetic acid (TCA) solution: Dissolve 500 g TCA in ~227 mL of ddH_2O.

2. Ice-cold acetone.

3. 1 M Tris–HCl buffer, pH 8: Dissolve 157.6 g of Tris–HCl in 1 L ddH_2O, sterile filter, and adjust to a pH of 8 using NaOH pellets.

4. Centrifuge.

5. 0.22-μm sterile filter.

6. Molecular-weight-cutoff columns: A molecular-weight cutoff that is at least twofold lower than the expected molecular mass of the protein target is desired.

2.3.3 Preparation of Membrane Fraction

1. Lysis buffer: 50 mM Tris–HCl, pH 7.2, 1 mM EDTA, 300 mM NaCl. Dissolve 157.6 g of Tris–HCl in 1 L ddH_2O, adjust to a pH of 7.2 using NaOH pellets, and sterile filter to make a 1 M stock solution of Tris–HCl, pH 7.2. Add 50 mL of 1 M Tris–HCl, pH 7.2, 2 mL of 0.5 M EDTA, and 17.5 g NaCl to 900 mL ddH_2O and dissolve. Adjust to a final volume of 1 L with ddH_2O and filter sterilize.

2. 10 % (w/v) 7-cyclohexyl-1-heptyl-β-D-maltopyranoside (Cymal-7) detergent: Dissolve 100 mg Cymal-7 in a final volume of 1 mL ddH_2O (*see* **Note 3**).

3. Centrifuge.

4. Sonicator.

5. Ultracentrifuge.

6. Rotating platform.

7. 0.22-μm sterile filter.

2.3.4 Preparation of Whole Cell, Soluble and Insoluble Fractions

1. PBS: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 in 1 L ddH_2O. Filter sterilize. Final pH should be ~7.4.

2. Tris–HCl, EDTA wash solution: 50 mM Tris–HCl, pH 8, 1 mM EDTA. Dissolve 157.6 g of Tris–HCl in 1 L ddH_2O, adjust to a pH of 8 using NaOH pellets, and sterile filter to make a 1 M stock solution of Tris–HCl, pH 8. Add 50 mL of the 1 M Tris–HCl, pH 8 and 2 mL of 0.5 M EDTA to 900 mL ddH_2O and dissolve. Adjust to a final volume of 1 L with ddH_2O.

3. PBS with 2 % (w/v) SDS: Add 10 g of SDS to 500 mL of 1×
   PBS. Sterile filter.
4. Sonicator, cell lysis solution, or other method of cell rupture.
5. Centrifuge.
6. Water bath at 100 °C.
7. 0.22-μm sterile filter.

2.3.5 Preparation of Sample Fractions for SDS-PAGE

1. 2× SDS loading dye: 100 mM Tris–HCl, pH 6.8, 4 % (w/v)
   SDS, 0.2 % w/v bromophenol blue, 20 % (v/v) glycerol.
   Dissolve 157.6 g of Tris–HCl in 1 L ddH$_2$O, adjust to a pH of
   6.8 using NaOH pellets, and sterile filter to make a 1 M stock
   solution of Tris–HCl, pH 6.8. Add 10 mL of the 1 M Tris–
   HCl, pH 6.8, 4 g of SDS, 200 mg bromophenol blue, and
   20 mL of glycerol to 50 mL ddH$_2$O and dissolve. Adjust to a
   final volume of 100 mL with ddH$_2$O.
2. β-Mercaptoethanol.
3. Water bath at 95 °C.
4. 4–20 % polyacrylamide protein gels.

3 Methods

3.1 Construction of Expression Plasmids

Here we provide instructions for adding a protein of interest (POI)
into the plasmids described in Figs. 2 and 4 via restriction enzyme-
based cloning. In addition to the suite of plasmids and cloning
strategies listed here, additional coding sequences are shown for
adding different fusion partners or signal peptides to the POI. We
also recommend the inclusion of an epitope tag such as FLAG
(DDYKDDDK) or c-Myc (KLISEEDL) and/or a polyhistidine
purification tag (6x-His: HHHHHH) at the C-terminus of fusion
proteins to visualize expression via Western blot or to separate
from native host proteins.

3.1.1 Cloning POIs for Periplasmic Expression

In this section, we provide instructions for creating a plasmid that
expresses a direct genetic fusion between a periplasmic-targeting
signal and POI. Adding fusion partners is a common way of target-
ing and increasing periplasmic production (*see* **Note 4**).

1. Oligonucleotide primers are designed to PCR amplify the gene
   encoding the POI. The forward primer should contain an *Xba*I
   site at the 5′ end and the reverse primer should contain a stop
   codon (e.g., TAA) followed by a *Hin*dIII site for cloning into
   the pTrc99A-Peri plasmid (Fig. 2). The pTrc99A backbone
   contains a hybrid *trp-lac* promoter and includes a signal pep-
   tide (sp) that targets the protein to the periplasm (*see* **Note 5**).
   PCR products are verified by agarose gel electrophoresis.

2. Use *Xba*I and *Hin*dIII as well as standard restriction enzyme-based cloning techniques to prepare the backbone and gene insert prior to ligation with T4 DNA ligase.

3. Transform the desalted ligation into electrocompetent cells (such as DH5α or other *recA*-deficient strain) and plate on LB agar containing appropriate antibiotics (Cm for pTrc99A-Peri).

3.1.2 Cloning POIs for Extracellular Expression

1. Oligonucleotide primers are designed to PCR amplify the gene encoding the POI. The forward primer should contain an *Xba*I site at the 5′ end and the reverse primer should contain a stop codon (e.g., TAA) followed by a *Hin*dIII site for cloning into the pTrc99A-YebF plasmid (Fig. 3). PCR products are verified by agarose gel electrophoresis. Alternative extracellular carrier proteins have been identified [33] and can be used in place of YebF (*see* **Note 6**).

2. Use *Xba*I and *Hin*dIII as well as standard restriction enzyme-based cloning techniques to prepare the backbone and gene insert prior to ligation with T4 DNA ligase.

3. Transform the desalted ligation into electrocompetent cells (such as DH5α or other *recA*-deficient strain) and plate on LB agar containing appropriate antibiotics (Amp or Carb for pTrc99A-YebF).

3.1.3 Cloning POIs for Inner Membrane Expression

In this section we provide instructions for creating a plasmid that expresses a genetic fusion between an N-terminal membrane protein solubility enhancer, *E. coli* glycerol-conducting channel protein (GlpF) [57] or *Bacillus subtilis* membrane-integrating protein mistic (MstX) [58], and a membrane POI. To create a plasmid for expressing the membrane POI without a fusion partner, several plasmids have been used successfully for this purpose (*see* **Note 7**).

1. Oligonucleotide primers are designed to PCR amplify the gene encoding the POI. The forward primer should contain an *Nhe*I site at the 5′ end and the reverse primer should contain a stop codon (e.g., TAA) followed by an *Xho*I/*Eag*I site for cloning into the pET22a-GlpF/MstX plasmids (Fig. 4). PCR products are verified by agarose gel electrophoresis.

2. Use *Nhe*I and *Xho*I/*Eag*I as well as standard restriction enzyme-based cloning techniques to prepare the backbone and gene insert prior to ligation with T4 DNA ligase.

3. Transform the desalted ligation into electrocompetent cells (such as DH5α or other *recA*-deficient strain) and plate on LB agar containing appropriate antibiotics (Amp or Carb for pET22a-GlpF/MstX).

3.2 Cell Growth and Protein Expression

Optimal conditions for periplasmic, extracellular, and membrane protein expression such as temperature, induction time, and co-expression of chaperones should be tested with small culture volumes.

1. Grow overnight (ON) cultures of BL21(DE3) (or other desired *E. coli* strain) harboring the pTrc99A- or pET-based (or other backbone) plasmids encoding the periplasmic, extra-cellular, or membrane POI in 5 mL of LB (or TB) supplemented with the appropriate antibiotics at 37 °C (*see* **Note 8**). If desired, a second co-expression plasmid encoding an accessory factor that enhances expression of the target can be co-transformed into the host strain.

2. Using 200-mL baffled shake flasks, inoculate 50 mL of antibiotic-supplemented LB (or TB) with ON cultures so that the optical density at 600 nm (OD_{600nm}) is ~0.05 (*see* **Note 9**).

3. Grow the cells to an OD_{600nm} of ~0.5 at 37 °C (*see* **Note 8**) and induce protein synthesis by adding inducer L-arabinose or IPTG to a final concentration of 0.2 % (w/v) or 1.0 mM, respectively (*see* **Note 10**). Protein expression is commonly carried out at 30 °C for 3–6 h (*see* **Note 11**). Collect a 1-mL sample for the preparation of whole-cell fractions, a 10-mL sample for preparation of soluble and insoluble fractions, and a 10-mL sample for the preparation of soluble periplasmic, supernatant, or membrane fractions (s*ee* **Note 12**).

3.3 Cellular Fractions

3.3.1 Preparation of Periplasmic and Cytoplasmic Fractions

1. Centrifuge 10 mL of an *E. coli* cell suspension for 10 min at 4 °C and $5,000 \times g$ to collect the cells. Discard the supernatant.

2. Resuspend the cells in 1 mL sucrose buffer and transfer to a microcentrifuge tube. Incubate for 10 min at room temperature. Centrifuge for 10 min at $10,000 \times g$. Discard the supernatant.

3. Resuspend pellet gently in 250 μL of ice-cold 5 mM $MgCl_2$. Place cells on ice for 10 min. Centrifuge for 10 min at 4 °C and $10,000 \times g$. Retain the supernatant as the periplasmic fraction.

4. Resuspend the pellet in 1 mL of PBS to wash. Centrifuge for 10 min at $10,000 \times g$. Discard the supernatant.

5. Resuspend the pellet in 250 μL of PBS. Sonicate three times for 30 s each, keeping the tube on ice between each sonication. Centrifuge for 10 min at $16,000 \times g$. Retain the supernatant as the cytoplasmic fraction.

3.3.2 Preparation of Extracellular Fraction

1. Centrifuge 10 mL of an *E. coli* cell suspension for 10 min at 4 °C and $5,000 \times g$ to collect the cells. Retain the supernatant and filter through a 0.22-μm filter to remove any cells that were not pelleted by centrifugation.

2. Add 1.1 mL of 100 % TCA to the clarified supernatant. Precipitation is carried out on ice or at 4 °C ON; however, times as short as 2 h are sufficient to precipitate protein. Centrifuge at $16,000 \times g$ for 10 min. Discard supernatant.

3. Wash with 1 mL ice-cold acetone. Gently remove acetone and let air-dry. The acetone wash has been omitted with similar results and less loss of precipitated protein. Centrifuge for 1 min at $13,000 \times g$ to remove remaining TCA or acetone (*see* **Note 13**).

4. Resuspend pellets in 250 μL of 1 M Tris–HCl, pH 8. The resulting resuspension is the extracellular fraction.

3.3.3 Preparation of Membrane Fraction

1. Centrifuge 10 mL of an *E. coli* cell suspension for 10 min at 4 °C and $5,000 \times g$ to collect the cells. Discard the supernatant.

2. Resuspend cells in 250 μL of lysis buffer. Sonicate three times for 30 s each, keeping the tube on ice between each sonication. Centrifuge for 10 min at 4 °C and $3,000 \times g$ to collect unlysed cells. Retain the supernatant as cleared cell lysate and dilute using 9.75 mL lysis buffer; dilution is necessary to reach a volume required for ultracentrifugation.

3. Centrifuge cleared lysate at 4 °C and $140,000 \times g$ in ultracentrifuge for 90 min to collect the membrane fraction. Discard the supernatant.

4. Resuspend pellets with 225 μL ice-cold lysis buffer by douncing. This process is carried out until the pellet has been resuspended completely.

5. Add 25 μL of 10 % w/v cymal-7 detergent to a final concentration of 1 % (*see* **Note 3**). Rotate at 4 °C for 1 h to complete membrane fraction extraction.

6. Centrifuge at $10,000 \times g$ for 15 min to remove insoluble material. The supernatant is retained as the membrane fraction.

3.3.4 Preparation of Whole-Cell Fraction

1. Centrifuge 1 mL of an *E. coli* cell suspension for 10 min at 4 °C and $5,000 \times g$ to collect the cells. Discard the supernatant.

2. Resuspend cells in 25 μL of PBS. The resulting resuspension is the whole-cell fraction.

3.3.5 Preparation of Soluble and Insoluble Fractions

1. Centrifuge 10 mL of an *E. coli* cell suspension for 10 min at 4 °C and $5,000 \times g$ to collect the cells. Discard the supernatant.

2. Resuspend cells in 250 μL of PBS. Sonicate three times for 30 s each, keeping the tube on ice between each sonication. Centrifuge for 10 min at 4 °C and $16,000 \times g$ to collect the insoluble material. The supernatant is retained as the soluble fraction.

3. Resuspend the pellet in 1 mL of 50 mM Tris–HCl, pH 8, and 1 mM EDTA to wash the pellet. Centrifuge for 10 min at 4 °C and $16,000 \times g$ and discard the supernatant. Repeat this step twice.

4. Resuspend the pellet in 250 µL of PBS with 2 % SDS. Boil at 100 °C for 10 min. Centrifuge for 10 min at $16,000 \times g$. The supernatant is retained as the insoluble fraction.

3.3.6 Preparation of Fractions for SDS-PAGE

1. Add 50 µL of β-mercaptoethanol to 1 mL of 2× SDS loading buffer (*see* **Note 14**).

2. Mix fractionations with an equal volume of reducing 2× SDS-PAGE loading buffer. Heat samples to 95 °C for 15 min.

3. Vortex samples briefly prior to adding 30 µL of sample on a 4–20 % protein gel (*see* **Note 15**).

4 Notes

1. When using Lemo21(DE3) avoid adding glucose to the medium since it affects lysozyme expression from PrhaBAD.

2. Amp is light and heat sensitive; we recommend Carb to decrease the formation of satellite colonies.

3. If necessary, detergent screening can be performed to identify the detergent(s) suitable for solubilization of a given membrane POI [43]. Briefly, 100 µL of the resuspended membranes are transferred into 1.5 mL tubes and to the following detergents at a final concentration of 1 % unless otherwise noted: octylglucoside (2 %), dodecylmaltoside, 6-cyclohexyl-1-hexyl-β-D-maltoside (Cymal-6), Cymal-7, 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LMPC), Triton X-100, and CHAPSO.

4. Natural *E. coli* proteins such as maltose-binding protein (MBP) can be used as fusion partners to increase the production of heterologous POIs that are prone to misfolding. MBP is a Sec-targeted protein and can be used with its native signal peptide to boost production of its fusion partner in the periplasm [59]. The plasmid pMAL-p5X is commercially available from New England Biolabs for this purpose.

5. In addition to the four periplasmic-targeting signal peptides listed in this work, many other Sec [11] or Tat [60] signal peptides are available. Online signal peptide predictors available from CBS Prediction Servers (http://www.cbs.dtu.dk/services/) are useful in determining the presence of signal peptides as well as cleavage sites [61].

6. Alternative N-terminal fusion tags for the extracellular expression of POIs include *E. coli* outer membrane protein A (OmpA), outer membrane protein F (OmpF), and osmotically inducible protein Y (OsmY) [33]. However, in our hands, YebF is the most efficient of these single-domain carrier proteins [36].

7. In addition to GlpF and MstX, both full-length MBP and *E. coli* thioredoxin (TxrA) can enhance production of membrane POIs [62]. If a fusion protein is undesirable for a given membrane POI, the following plasmids have been used successfully to produce membrane proteins in *E. coli* due to tight regulation of their promoters: pASK75, regulated by anhydrous tetracycline [43], and pRHA-67 (Xbrane Bioscience), tightly regulated by L-rhamnose [63].

8. Cells are typically cultivated at 37 °C; however, lower temperatures (15–30 °C) may improve the ON growth for some strain/plasmid combinations.

9. This corresponds approximately to a 1:100 dilution of ON cultures.

10. IPTG concentrations can be varied from 0.1 μM to 1 mM to control the *lac* promoter. Generally, 0.1–1.0 mM IPTG will result in full induction of the *lac* promoter. Lower IPTG concentrations are often used to decrease expression and control the amount of protein sent to the insoluble fraction. Typically, protein overexpression is better regulated from a pBAD plasmid than from a pET plasmid. We recommend an L-arabinose concentration of 0.2 % for full induction since higher concentration of L-arabinose (>1 %) is toxic to bacteria. Concentrations down to 0.002 % L-arabinose can be used to control production [64].

11. Protein expression is often carried out at a slightly lower temperature than used for cell growth prior to induction. Typically, 3–6 h at 30 °C is adequate; however, longer induction times (i.e., 6–48 h) and lower temperatures (i.e., 16–22 °C) can be used for difficult-to-express POIs. It is often desired to perform small-scale inductions at a variety of temperatures (e.g., 25, 30, and 37 °C) as well as collect a variety of induction time points (e.g., 3, 8, and 24 h) to find the conditions that maximize protein accumulation.

12. It is often desired to compare multiple fractions to determine the efficiency of translocation, secretion, or membrane insertion as well as to determine any potential bottlenecks in these processes. For example, to analyze secretion of YebF fusion proteins, the cytoplasmic, periplasmic, whole-cell, and extracellular fractions can be compared to determine if the protein is accumulating inside cells instead of being excreted. Additionally, fusion proteins that lack a signal peptide and/or genetic knockout strains that are blocked for translocation can be used as negative controls for secretion, translocation, or membrane insertion.

13. If TCA remains in solution, the resulting resuspension will be acidic and turn SDS loading dye from purple to orange-yellow. An additional aliquot of 1 M Tris–HCl, pH 8 is added to adjust the pH to be more basic. TCA precipitation results in

the unfolding of proteins and is not prescribed for the collection of functional supernatant proteins. Centrifuging molecular-weight-cutoff columns is a facile method to concentrate supernatant proteins. Cutoff columns should be used following the manufacturer's instructions.

14. If it is desired to keep disulfide bonds intact, non-reduced samples are prepared with 2× SDS-PAGE loading buffer without β-mercaptoethanol. To ensure that proteins remain oxidized, the entire gel must be prepared with non-reduced samples and fresh running buffer should be used.

15. Due to wide range of production levels for different heterologous proteins in *E. coli*, it is often necessary to adjust the load volume of protein to achieve a satisfactory signal via Western blot or other assay. All of the fractions collected via these methods are a 40-fold concentration of the original cell culture. Other normalizations such as cell number (OD_{600nm}) or total protein (Bradford assay) should be considered if normalizing by culture volume is undesirable.

Acknowledgment

References

1. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22:1399–1408

2. Baba T, Ara T, Hasegawa M et al (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2(2006): 0008

3. Makino T, Skretas G, Georgiou G (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microb Cell Fact 10:32

4. Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12:32–42

5. Carrier TA, Keasling JD (1997) Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. Biotechnol Prog 13:699–708

6. de Marco A (2007) Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. Nat Protoc 2:2632–2639

7. Berkmen M (2012) Production of disulfide-bonded proteins in Escherichia coli. Protein Expr Purif 82:240–251

8. Mansell TJ, Fisher AC, DeLisa MP (2008) Engineering the protein folding landscape in gram-negative bacteria. Curr Protein Pept Sci 9:138–149

9. Weiner JH, Li L (2008) Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. Biochim Biophys Acta 1778:1698–1713

10. Georgiou G, Segatori L (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. Curr Opin Biotechnol 16:538–545

11. Choi JH, Lee SY (2004) Secretory and extra-cellular production of recombinant proteins using Escherichia coli. Appl Microbiol Biotechnol 64:625–635

12. Wagner S, Bader ML, Drew D et al (2006) Rationalizing membrane protein overexpression. Trends Biotechnol 24:364–371

13. Schlapschy M, Grimm S, Skerra A (2006) A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in Escherichia coli. Protein Eng Des Sel 19:385–390

14. Kadokura H, Katzen F, Beckwith J (2003) Protein disulfide bond formation in prokaryotes. Annu Rev Biochem 72:111–135

15. Berlec A, Strukelj B (2013) Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. J Ind Microbiol Biotechnol 40:257–274

16. Lee PA, Tullman-Ercek D, Georgiou G (2006) The bacterial twin-arginine translocation pathway. Annu Rev Microbiol 60:373–395

17. Fisher AC, DeLisa MP (2004) A little help from my friends: quality control of presecretory proteins in bacteria. J Bacteriol 186:7467–7473

18. Fisher AC, Kim JY, Perez-Rodriguez R et al (2008) Exploration of twin-arginine translocation for the expression and purification of correctly folded proteins in Escherichia coli. Microb Biotechnol 1:403–415

19. Bruser T (2007) The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production. Appl Microbiol Biotechnol 76:35–45

20. Mansell TJ, Linderman SW, Fisher AC et al (2010) A rapid protein folding assay for the bacterial periplasm. Protein Sci 19:1079–1090

21. DeLisa MP, Samuelson P, Palmer T et al (2002) Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J Biol Chem 277:29825–29831

22. Fisher AC, Kim W, DeLisa MP (2006) Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein Sci 15:449–458

23. Marrichi M, Camacho L, Russell DG et al (2008) Genetic toggling of alkaline phosphatase folding reveals signal peptides for all major modes of transport across the inner membrane of bacteria. J Biol Chem 283:35223–35235

24. Baker JL, Celik E, DeLisa MP (2013) Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. Trends Biotechnol 31:313–323

25. Merritt JH, Ollis AA, Fisher AC et al (2013) Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. Biotechnol Bioeng 110:1550–1564

26. Valderrama-Rincon JD, Fisher AC, Merritt JH et al (2012) An engineered eukaryotic protein glycosylation pathway in Escherichia coli. Nat Chem Biol 8:434–436

27. Wacker M, Linton D, Hitchen PG et al (2002) N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science 298:1790–1793

28. Papanikou E, Karamanou S, Economou A (2007) Bacterial protein secretion through the translocase nanomachine. Nat Rev Microbiol 5:839–851

29. Pugsley AP, Francetic O (1998) Protein secretion in Escherichia coli K-12: dead or alive? Cell Mol Life Sci 54:347–352

30. Zhang G, Brokx S, Weiner JH (2006) Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in Escherichia coli. Nat Biotechnol 24:100–104

31. Qian ZG, Xia XX, Choi JH et al (2008) Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in Escherichia coli. Biotechnol Bioeng 101:587–601

32. Prehna G, Zhang G, Gong X et al (2012) A protein export pathway involving Escherichia coli porins. Structure 20:1154–1166

33. Kotzsch A, Vernet E, Hammarstrom M et al (2011) A secretory system for bacterial production of high-profile protein targets. Protein Sci 20:597–609

34. Steen EJ, Kang Y, Bokinsky G et al (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463:559–562

35. Bokinsky G, Peralta-Yahya PP, George A et al (2011) Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. Proc Natl Acad Sci U S A 108:19949–19954

36. Haitjema C, Boock JT, Natarajan A et al (2014) A universal genetic assay for engineering extracellular protein expression. ACS Synth Biol 3(2):74–82

37. He SY, Lindeberg M, Chatterjee AK et al (1991) Cloned Erwinia chrysanthemi out genes enable Escherichia coli to selectively secrete a diverse family of heterologous proteins to its milieu. Proc Natl Acad Sci U S A 88:1079–1083

38. Ham JH, Bauer DW, Fouts DE et al (1998) A cloned Erwinia chrysanthemi Hrp (type III protein secretion) system functions in Escherichia coli to deliver Pseudomonas syringae Avr signals to plant cells and to secrete Avr proteins in culture. Proc Natl Acad Sci U S A 95:10206–10211

39. Ni Y, Chen R (2009) Extracellular recombinant protein production from Escherichia coli. Biotechnol Lett 31:1661–1670

40. Luirink J, von Heijne G, Houben E et al (2005) Biogenesis of inner membrane proteins in Escherichia coli. Annu Rev Microbiol 59: 329–355

41. Freigassner M, Pichler H, Glieder A (2009) Tuning microbial hosts for membrane protein production. Microb Cell Fact 8:69

42. Sarkar CA, Dodevski I, Kenig M et al (2008) Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. Proc Natl Acad Sci U S A 105: 14808–14813

43. Link AJ, Skretas G, Strauch EM et al (2008) Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli. Protein Sci 17:1857–1863

44. Nannenga BL, Baneyx F (2011) Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. Protein Sci 20:1411–1420

45. Chen Y, Song J, Sui SF et al (2003) DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in Escherichia coli. Protein Expr Purif 32:221–231

46. Drew D, Lerch M, Kunji E et al (2006) Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods 3:303–313

47. Drew D, Slotboom DJ, Friso G et al (2005) A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci 14:2011–2017

48. Massey-Gendel E, Zhao A, Boulting G et al (2009) Genetic selection system for improving recombinant membrane protein expression in E. coli. Protein Sci 18:372–383

49. Dodevski I, Pluckthun A (2011) Evolution of three human GPCRs for higher expression and stability. J Mol Biol 408:599–615

50. Eshaghi S, Hedren M, Nasser MI et al (2005) An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. Protein Sci 14:676–683

51. Skretas G, Georgiou G (2010) Simple genetic selection protocol for isolation of overexpressed genes that enhance accumulation of membrane-integrated human G protein-coupled receptors in Escherichia coli. Appl Environ Microbiol 76:5852–5859

52. Waraho D, DeLisa MP (2009) Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. Proc Natl Acad Sci U S A 106:3692–3697

53. Perez-Rodriguez R, Fisher AC, Perlmutter JD et al (2007) An essential role for the DnaK molecular chaperone in stabilizing overexpressed substrate proteins of the bacterial twin-arginine translocation pathway. J Mol Biol 367:715–730

54. DeLisa MP, Lee P, Palmer T et al (2004) Phage shock protein PspA of Escherichia coli relieves saturation of protein export via the Tat pathway. J Bacteriol 186:366–373

55. Skretas G, Georgiou G (2009) Genetic analysis of G protein-coupled receptor expression in Escherichia coli: inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. Biotechnol Bioeng 102:357–367

56. Miroux B, Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol 260:289–298

57. Neophytou I, Harvey R, Lawrence J et al (2007) Eukaryotic integral membrane protein expression utilizing the Escherichia coli glycerol-conducting channel protein (GlpF). Appl Microbiol Biotechnol 77:375–381

58. Roosild TP, Greenwald J, Vega M et al (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science 307:1317–1321

59. Nallamsetty S, Austin BP, Penrose KJ et al (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. Protein Sci 14:2964–2971

60. Tullman-Ercek D, DeLisa MP, Kawarasaki Y et al (2007) Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. J Biol Chem 282:8309–8316

61. Emanuelsson O, Brunak S, von Heijne G et al (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953–971

62. Tucker J, Grisshammer R (1996) Purification of a rat neurotensin receptor expressed in Escherichia coli. Biochem J 317:891–899

63. Haldimann A, Daniels LL, Wanner BL (1998) Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. J Bacteriol 180:1277–1286

64. Guzman LM, Belin D, Carson MJ et al (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177:4121–4130

65. Fisher AC, DeLisa MP (2008) Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. PLoS One 3:e2351

Chapter 6

Characterization of Amyloid-Like Properties in Bacterial Intracellular Aggregates

Anna Villar-Pique, Susanna Navarro, and Salvador Ventura

Abstract

Protein aggregation into amyloid conformations is associated with more than 50 different human disorders. Recent studies demonstrate that the expression in bacteria of amyloid proteins results in the formation of intracellular aggregates structurally related to those underlying human diseases. The ease with which prokaryotic organisms can be genetically and biochemically manipulated makes them useful systems for studying how and why protein aggregates inside the cell, providing a tractable environment to rationally model in vivo amyloid formation. In this chapter we present an overview of the methods used to characterize the kinetic, structural, and functional properties of amyloid-like bacterial intracellular aggregates and how they can be employed to screen for lead compounds that might modulate amyloid deposition.

Key words Protein aggregation, Inclusion bodies, Amyloid, Bacteria

1 Introduction

Protein misfolding and aggregation has become a highly active research area due to the recurrent link between the presence of protein deposits in human tissues and the development of dozens of different pathologies [1, 2]. In many cases, these protein aggregates consist of β-sheet-enriched fibrillar structures known as amyloids [3, 4]. In bacteria, the formation of protein aggregates, known as inclusion bodies (IBs), is commonly seen during high-level production of recombinant proteins such as biopharmaceuticals and enzymes of biotechnological importance, precluding their cost-effective commercialization [5]. Although significant effort has been devoted to the characterization of the protein conformations and molecular mechanisms underlying amyloid fibril formation in eukaryotic cells, little is known about the process of protein aggregation inside bacteria and its effect on cellular physiology. However, in the last few years we have witnessed how the adoption

of experimental strategies similar to those previously used to characterize in vitro and in vivo formation in eukaryotic backgrounds to the process of aggregates formation in bacteria has highlighted a high similarity between those molecular reactions [6–8]. The bacterial aggregates formed by these proteins share structural properties with amyloids [9, 10], they are cytotoxic for eukaryotic cells [11], and in the case of prion proteins they might become infective. In fact, this resemblance responds to biophysical constraints since, independent of the organism, the competition between folded and aggregated states inside the cell cannot be avoided, because many of the physicochemical traits that determine the folding into native structures also tend to favor the establishment of intermolecular interactions resulting in the formation of the cross-β motif recurrently observed in the core of different aggregated structures. The study of protein aggregation in bacteria has allowed characterizing intracellular protein aggregation rates [12, 13], assessing the specificity of intracellular protein aggregation [10], screening for aggregation modulators [14], or dissecting the impact of protein aggregation for cell fitness [15]. Here we provide the readers with a detailed list of the different methods our group has employed to provide insights into the amyloid-like nature of bacterial IBs and how they can be exploited to model intracellular misfolding and aggregation as well as to identify inhibitors of these deleterious pathways.

2   Materials

2.1   Preparation of Inclusion Bodies

1. Lysis buffer: 50 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA, 15 mM PMSF, 300 μg/mL lysozyme, pH 8.0.
2. Phosphate buffer saline (PBS).

2.2   Studying Intracellular Protein Aggregation Rates

1. TCS-SP5 AOBS confocal laser scanning microscope (Leica Microsystems, Germany).
2. Leica DMBR microscope equipped with a Leica DFC 500 camera (Leica Microsystems, Germany).
3. LAS AF Lite Software (Leica Microsystems CMS GmbH, Germany).
4. Perkin-Elmer 650–40 spectrofluorimeter (Perkin-Elmer, MA).

2.3   Monitoring the Presence of Amyloid-Like Aggregates Inside Living Bacterial Cells

1. Thioflavin-S.
2. PBS buffer.
3. FacsAria SORP, flow cytometer (BD Biosciences, USA) equipped with a 335 nm UV laser.

2.4 Assessing the Specificity of Intracellular Protein Aggregation

1. Perkin-Elmer 650–40 spectrofluorimeter (Perkin-Elmer, MA).
2. TCS SP2 confocal laser scanning microscope (Leica Microsystems, Germany).
3. LAS AF Lite software (Leica Microsystems CMS GmbH, Germany).

2.5 Exploiting the Competition Between Folding and Aggregation to Screen for Aggregation Modulators

1. Chelex 100 chelating resin.
2. Guanidine hydrochloride (Gnd·HCl).
3. Victor 3 plate reader (Perkin-Elmer, MA).

2.6 Visualizing Amyloid Structures Inside Intracellular Protein Aggregates

1. Proteinase-K.
2. Coomassie Blue staining: 0.1 % Coomassie Blue, 10 % acetic acid, 40 % methanol.
3. 3,5-Dimethoxy-4-hydroxy-cinnamic (sinapinic acid) and a-cyano-4-hydroxy-cinnamic acid.
4. Uranyl acetate.
5. Destaining solution: 25 mM ammonium bicarbonate in 50 % acetonitrile.
6. Matrix solution: 10 mg/mL sinapinic acid dissolved in aqueous 30 % acetonitrile with 0.1 % trifluoroacetic acid.
7. For electron microscopy: Hitachi H-7000 transmission electron microscope (Hitachi, Japan).
8. For atomic force microscopy: Multimode atomic force microscope (Veeco Instruments, Inc., USA), highly oriented pyrolytic graphite (HOPG) (NT MDT Co., Russia).
9. Multimode atomic microscope equipped (Veeco Instruments, Inc., USA).
10. In tapping mode, Veeco NP-S probes from Bruker Optics Inc. (Karlsruhe, Germany).
11. For mass spectrometry: Ultraflex MALDI-TOF mass spectrometer (Bruker Daltonics, Karlsruhe, Germany).
12. For Edman N-terminal sequencing: ABI Procise Model 492 Edman Micro Sequencer connected to an ABI Model 140 °C PTH Amino Acid Analyzer (Perkin Elmer Applied Biosystems, USA).

2.7 Deciphering the Molecular Contacts That Sustain Intracellular Aggregates

1. Bruker Tensor 27 FT-IR Spectrometer (Bruker Optics Inc., Germany) with a Golden Gate MKII ATR accessory.
2. The PeakFit package for nonlinear peak-fitting (Systat Software, USA).

2.8 Characterizing the Toxic Properties of Intracellular Bacterial Aggregates

1. 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT).

2. Dimethylsulfoxide (DMSO).

3. Human cervical adenocarcinoma (HeLa) cell line (American Type Culture Collection, USA).

4. Fetal bovine serum (FBS).

5. Dulbecco's Modified Eagle's medium (DMEM) culture medium.

6. Penicillin–streptomycin antiobiotics.

7. Microtiter plates.

8. Propidium iodide solution, at 1.3 mg/mL in water.

9. Staining buffer: Phosphate-buffered saline, 1 mM EDTA, 0.2 % Pluronic™ F-68, 0.1 % sodium azide, pH 7.4. Tween-20 at 0.01 % can be substituted for Pluronic F-68.

10. Victor 3 plate reader (Perkin-Elmer, MA).

11. BD FACS™ brand flow cytometer (BD FACSCalibur™ flow cytometer or equivalent) equipped with 488 nm laser excitation.

2.9 Characterizing the Infectious Properties of Intracellular Prionic Aggregates

1. D-Sorbitol.

2. Lyticase.

3. YPD medium: 2 % peptone, 1 % yeast extract, and 2 % glucose.

4. SCE buffer: 1 M sorbitol, 10 mM EDTA, 10 mM dithiothreitol, and 100 mM sodium citrate, pH 5.8.

5. STC buffer: 1 M sorbitol, 10 mM $CaCl_2$, and 10 mM Tris–HCl, pH 7.4.

6. PEG buffer: 20 % PEG 8000, 10 mM $CaCl_2$, and 10 mM Tris–HCl, pH 7.5.

7. SOS medium: 1 M sorbitol, 7 mM $CaCl_2$, 0.25 % yeast extract, and 0.5 % peptone.

3 Methods

3.1 Preparation of Inclusion Bodies

IBs are insoluble protein aggregates usually found in recombinant bacteria when they are forced to produce heterologous protein species. These particles are formed by polypeptides that cross-interact through sterospecific contacts and that are steadily deposited in the cytoplasm or the periplasm. Following we detail a protocol to obtain and purify IBs.

1. Protein expression is induced for a minimum of 4 h (*see* **Note 1**).

2. Cells are harvested by centrifugation at $1,500 \times g$ for 20 min at 4 °C and resuspended in lysis buffer (30 mL buffer/L culture).

3. After 30 min of incubation at 37 °C under gentle agitation, detergent NP-40 is added at 1 % (v/v) and cells are incubated at 4 °C for 50 min under mild agitation.

4. Then, 15 μg/mL of DNase I and RNase and 15 μM MgSO₄ are added to mixtures and the resulting mixture is further incubated at 37 °C for 30 min to remove nucleic acids.

5. Protein aggregates are collected by centrifugation at $12,000 \times g$ for 15 min at 4 °C.

6. Finally, IBs are washed once in lysis buffer containing 0.5 % Triton X-100 and three times with sterile PBS. After a final centrifugation at $12,000 \times g$ for 15 min, pellets are stored at –20 °C until analysis (*see* **Note 2**).

3.2 Studying Intracellular Protein Aggregation Rates

In the crowded cytoplasmic space, protein folding and aggregation are competing processes directed by native intramolecular contacts and nonnative intermolecular ones, respectively [16]. This kinetic competition determines not only the balance between soluble and insoluble protein, but also the amount of native-like species in both fractions. In this sense, studies in bacterial IBs carried out during the last decade have demonstrated that these aggregates are not amorphous and unstructured assemblies as traditionally considered, rather they contain a wide range of structures, including native-like conformations, thus becoming partially functional particles [17]. This relevant discovery enables to use the activity of the aggregates as an assessment of their formation rate, since the latter determines the ratio between native and nonnative contacts [12, 15]. In this section, we describe the tagging to a fluorescent protein as a reporter strategy to study protein aggregation.

3.2.1 Fusion to Fluorescent Reporters

The green fluorescent protein (GFP), was first discovered in 1962 in the jellyfish *Aequorea victoria* by Shimomura and coworkers [18], together with its derivatives, has become a powerful tool in cell biology studies. Among other applications, its use as a protein tag has traditionally enabled to monitor gene expression levels and subcellular localization [19]. In this sense, a relevant advance came along with the implementation of GFP as a reporter of protein solubility. Waldo and coworkers developed a protein folding assay based on the N-terminal fusion of a target protein to the GFP. They found that productive folding of the GFP domain was determined by the solubility of the tagged protein of interest and, therefore, that the fluorescence of *Escherichia coli* cells expressing that fusion protein was a direct measure of its insolubility [20]. Later on, it was found that the expression of GFP-tagged amyloidogenic peptides resulted in the formation of active intracellular aggregates, whose fluorescence correlated with the aggregation propensity of the peptides, indicating that GFP can be employed as a reporter of aggregation rather than solubility [12, 17].

Here, we detail a method for in vivo imaging of protein aggregate formation in the cytoplasm of bacteria. Although the protocol is based on the enhanced GFP, any of its fluorescent derivatives may be, in principle, used as long as microscopy equipment is provided with the suitable lasers and settings. A linker between the fluorescent protein and the protein of interest is indispensable to facilitate the proper folding of the former (*see* **Note 3**).

Time-Lapse Imaging of In Vivo Formation of Bacterial Aggregates

1. Bacterial cultures, transformed with the plasmid containing the construct of interest, are grown in standard conditions. When they reach an OD_{600nm} of 0.4–0.6, protein expression is induced (*see* **Note 4**).

2. Under sterile conditions, a small drop of culture is placed on a glass slide covered with a thin layer of solidified medium (composed by liquid medium, containing the appropriate antibiotic and the expression inductor, with 2 % of agarose). Microscope slides with living cells must be covered with a sterile cover slip.

3. Time-lapse experiment is carried out in a confocal laser scanning microscope supplemented with an acclimatized incubation chamber to maintain a fixed growing temperature.

4. GFP is excited using a 458 nm argon laser, and the fluorescence emission is collected within a bandwidth between 500 and 600 nm. Images are digitally captured at specific time intervals (*see* **Note 5**).

3.2.2 Correlation Between Fluorescence and Aggregation

As aforementioned, the fluorescence of a GFP-tagged aggregate depends on the formation rate. Thus, in this section we describe the measurement of the aggregate activity as a reporter of the self-assembly kinetics and aggregation propensity of the polypeptide. This propensity can be assessed by several bioinformatic algorithms [21–23], whose theoretical predictions should, in principle, correlate with experimental fluorescence obtained.

Although GFP fluorescence can be quantified in whole cells, the presence of soluble protein strongly interferes in the measurement of aggregate activity. Hence, intracellular protein aggregates must be previously isolated as described in Subheading 3.1 and adjusted to a selected optical density. The turbidity of a solution containing proteinaceous aggregates is proportional to the amount of protein embedded, and thus, in the protocols here described, the (OD_{360nm}) of purified aggregates will be taken into consideration as a measure of relative concentration. Before adjusting the OD_{360nm} of the purified aggregates, it is indispensable to homogenize the solution by passing through a 25-gauge needle in order to fragment the largest protein deposits, which can alter the turbidity determination.

The activity of GFP-tagged aggregates can be imaged under UV light in an optical microscope as well as quantified in a spectrofluorimeter for an exact determination of the relative GFP emission.
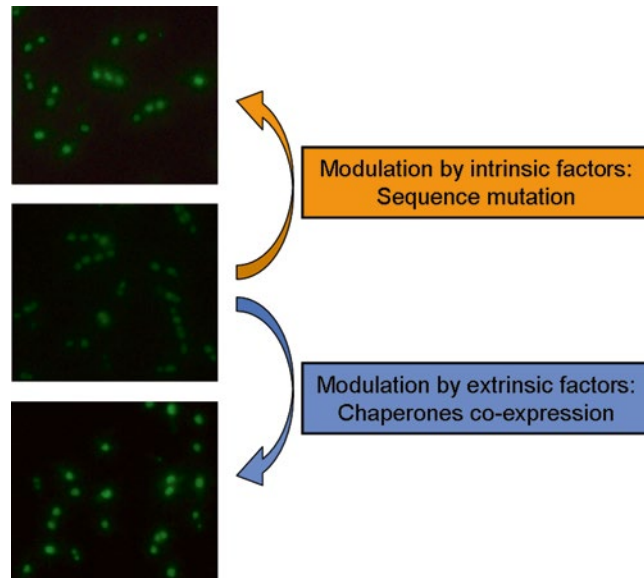
Fig. 1 Monitoring the modulation of the conformational quality of Aβ-42 Alzheimer-related peptide by genetic mutations and molecular chaperones using fluorescence microscopy. An Aβ-42-GFP fusion is expressed recombinantly in bacteria resulting in the formation of IBs. Mutations that decrease the aggregation propensity of this fusion protein or the overexpression of the DnaK/DnaJ and GroELS chaperones decrease their in vivo aggregation rate, resulting in a higher proportion of folded GFP in the aggregates and thus in brighter IBs, when observed under UV light

Microscopy Imaging of Protein Aggregates

1. Purified aggregates are three times washed in PBS and the OD_{360nm} must be adjusted between 0.5 and 1 (*see* **Note 6**).

2. Aggregates are placed on top of a microscope glass slide. Clear nail polish can be employed to seal the cover slip.

3. Images are obtained under UV light using a fluorescence microscope (Fig. 1). The emission filter must be appropriate for the fluorescent tag used. Magnification should be at least 100-fold due to the reduced size of intracellular aggregates.

Quantification of the Aggregate Fluorescence

1. Purified aggregates are resuspended in PBS at pH 7.4. Tris–HCl-based buffers can also be used; however, the pH must be kept between pH 7 and 8 (*see* **Note 7**).

2. The fluorescence of aggregates is quantified on a spectro-fluorimeter using 1 mL quartz cuvettes. The emission is recorded from 500 to 600 nm (the GFP emission peak appears around 510 nm) using an excitation wavelength of 488 nm (*see* **Note 8**).

3. Correlation of GFP activity and aggregation can be explored by means of theoretical aggregation propensity values previously calculated. Among relevant aggregation predictors, we recommend AGGRESCAN (http://bioinf.uab.es/aggrescan) since this algorithm was developed from empiric results obtained in similar experiments with fluorescent aggregates [24] (*see* **Note 9**).

3.3 Monitoring the Presence of Amyloid-Like Aggregates Inside Living Bacterial Cells

Bacterial cells are valuable systems to understand the integration of metabolic, regulatory, and structural features in living cells. The similarities between bacterial aggregates and the deposits formed in higher organisms in pathological processes like amyloid fibrils, nuclear inclusions, and aggresomes [25, 26] provide a unique opportunity to dissect the molecular pathways triggering these disorders in a simple, yet physiologically relevant, organism.

Accordingly, *E. coli* has been the model used to study the link between protein aggregation and ageing [27], the effect of anti-aggregational drugs [28], the role of the highly conserved protein quality machinery on the conformational properties of aggregated states [25, 29], the effect of the protein sequence on in vivo aggregation kinetics [12], or the influence of extrinsic factors like temperature on protein aggregates properties [30, 31].

The coupling of the amyloid-specific dye thioflavin S (Th-S) and flow cytometry provides a very fast, high-throughput, quantitative, and noninvasive technique to monitor the in vivo intracellular aggregation of amyloidogenic proteins [32].

3.3.1 Th-S Staining Imaging and Coupled to Flow Cytometry

Th-S is a fluorescent cell-permeable amyloid-binding benzothiazol salt dye that binds to amyloid fibrils but not monomers. This binding induces a shift in excitation (385–450 nm) and emission (445–482 nm) spectra of the dye. Th-S has the ability to penetrate biological membranes and accumulate in amyloid deposits, thus detecting the presence of amyloid-like protein conformations inside living bacteria cells.

1. Bacterial cells are washed with PBS and diluted at an OD_{600nm} of 0.05. Cells are incubated for 1 h in the presence of 125 μM of Th-S diluted in PBS.

2. Cells are pelleted by centrifugation at $1,500 \times g$ for 5 min and resuspended in PBS.

3. The acquisition of prepared bacteria samples on a BD FACS brand flow cytometer at low speed initial instrument settings should be performed as following:

   • Threshold parameter—SSC (side scatter), FSC (forward scatter).

   • FSC—E01, logarithmic amplification.

- SSC—375 V, logarithmic amplification.
- FL1—600 V, logarithmic amplification.
- Compensation—none used.

A minimum of 10,000 events should be recorded.

4. Data analysis is performed gating first (R1) in the dot plot of (FSC) and (SCC) signals obtained of unstained bacteria.

5. Gated cells in P1 are analyzed for Th-S fluorescence emission measured on an FL1 detector using 355 nm excitation and 530/30 nm long-pass filter (*see* **Note 10**).

3.4 Assessing the Specificity of Intracellular Protein Aggregation

In vivo protein aggregation is a selective process, where intermolecular contacts are established between specific stretches of the involved polypeptides [10, 33]. A simple strategy to assess this sequential specificity in bacterial cells is the simultaneous coexpression of two distinct proteins. In an early work, this resulted in the formation of two different aggregates in a single cell with different morphology and composition, demonstrated by biochemical analyses including detergent-based protein extractions, differential centrifugation, protein electrophoresis, and electron microscopy [34]. In a more recent study, our group addressed this issue by means of fluorescence resonance energy transfer (FRET) between fluorescent proteins [10]. FRET only occurs if the distance between both fluorophores ranges between 10 and 100 Å, resulting in a high-resolution approach to explore colocalization.

   Although FRET can be quantified using different techniques, in this section, sensitized acceptor emission and acceptor photobleaching approaches are explained, being simple and fast assessments for FRET efficiency. In addition, the protocols described below take BFP (the donor) and GFP (the acceptor) as an example of a FRET couple and, thus, the detailed settings for microscopy imaging and spectrofluorometric measurements are only valid for this specific case. However, other fluorophore FRET couples are possible as long as the emission spectrum of the donor molecule overlaps the excitation spectrum of the acceptor one.

3.4.1 Double-Fluorescent Tagging in Colocalization and FRET Experiments

Preparation of Cells Expressing the Fusion Proteins

1. Each of the target proteins is fused to GFP or BFP (the fluorescent tagging method is described in Subheading 3.2.1). Plasmids encoding both protein chimeras must be compatible.

2. Competent bacterial cells are transformed with both plasmids and are grown in standard conditions with the appropriated antibiotics (*see* **Note 11**).

3. Protein expression is induced at an OD_{600nm} of 0.6. Cells are grown for 16 h in standard conditions (*see* **Note 12**).

4. Cells are harvested by centrifugation and washed with 0.22 μm filtered PBS.

FRET Quantification by
Fluorescence Spectroscopy

1. Resuspended cells in PBS are diluted to an OD_{600nm} of 0.01 (*see* **Note 13**).

2. The relative fluorescence of the cultures is recorded in a spectrofluorimeter using 1 mL quartz cuvettes. The excitation wavelength used corresponds to the excitation peak of the BFP (389 nm) and the fluorescence emission is recorded from 420 to 560 nm, where both BFP and GFP present emission peaks (*see* **Note 14**).

3. The ratio between emission at 510 nm (GFP emission peak) and 445 nm (BFP emission peak) is determined to calculate the acceptor-sensitized emission. In addition, this ratio is also indicative of the quenching of the donor fluorescence.

FRET Quantification
and Colocalization Analysis
by Confocal Microscopy

1. Cells expressing the fusion proteins are resuspended with PBS containing 0.1 % of formaldehyde to fix them (*see* **Note 15**).

2. 5 μL of fixed cells are placed on top of a microscope slide and covered with cover slip. Photographs are taken with a confocal microscope using suitable excitation laser lines: 488 nm for GFP and 351 or 364 nm for BFP (*see* **Note 16**).

3. Images of the same region of interest (ROI) are taken simultaneously using emission channels for GFP (around 510 nm) and BFP (445 nm). The merged image of both channels reveals the degree of colocalization (*see* **Note 17**).

4. For acceptor photobleaching approach, a ROI is selected and BFP emission is recorded from 409 to 468 nm (D_{pre}). Afterwards, cells are photobleached using the 488 nm laser line and the BFP emission is recorded again using the same emission wavelength window (D_{post}). FRET efficiency ($FRET_{eff}$) is calculated using the following formula (*see* **Note 18**):

$$FRET_{eff} = \left(D_{post} - D_{pre} \right) / D_{post}$$

3.5 Exploiting the Competition Between Folding and Aggregation to Screen Aggregation Modulators

The kinetic competition between folding and aggregation during protein deposition in the cell (*see* Subheading 1) can be also exploited in vitro to develop a screening assay for aggregation modulator compounds. The result of this competition depends on the intrinsic aggregation propensity of the protein but also on external factors. Thus, the aggregation rate might be affected by the presence of the tested compounds.

On one side, the assay described below is performed with the GFP-tagged amyloid β-peptide, allowing the use of the GFP fluorescence to measure the conformational quality of the final aggregates (*see* **Note 19**) [14]. On the other side, it is based on the refolding of aggregates after a denaturation step, which is essential in order

to start the refolding process from totally unstructured species. The refolding step is performed with PBS-based refolding buffers. Although other standard laboratory buffers may also be employed, it is highly recommended to previously test them in control samples since the composition might affect the refolding process.

The amyloid structure contained in the bacterial protein deposits [9, 35] converts the method into a straightforward approach to detect compounds with inhibitor or enhancer activity over amyloid deposition. Since traditional screening assays for amyloid inhibitors are generally performed with synthetic peptides, the herein proposed system constitutes a major advancement in the field, as the production of bacterial aggregates is fast, easy, and cheap and the method can be performed using 96-well plates, thus becoming a high-throughput cost-effective assay [14].

3.5.1 Evaluation of Fluorescence Recovery in 96-Well Plates

Denaturation of Purified Aggregates

1. 150 µL of purified aggregates at an OD_{360nm} of 1 are centrifuged at maximum speed for at least 10 min and the supernatant is removed. This step must be repeated in separated tubs, one for each refolding experiment (*see* **Note 20**).

2. Pellets are resuspended in 10 µL of 8 M Gnd·HCl to denature the aggregates. Samples must be incubated at room temperature for 4 h.

Refolding Assay

Since refolding in PBS favors a slow establishment of native contacts the screening for amyloid aggregation promoter compounds can be performed straightforwardly by adding them to the refolding buffer. If they act as aggregation enhancers, the GFP fluorescence recovery will be lower. On the contrary, if the assay aims to detect amyloid inhibitor molecules, the refolding buffer must favor the amyloid aggregation, which can be easily achieved by adding the aggregation promoter metals Cu^{2+} and Zn^{2+}, as previously described [14]. Hence, the addition of compounds with aggregation inhibitor activity will increase the recovered fluorescence.

1. Denatured aggregates are diluted in 990 µL of refolding buffers, which are obtained by adding the tested compounds to PBS buffer. PBS must be previously treated with Chelex 100 chelating resin to remove undesired ions that might alter the refolding process (*see* **Note 21**).

2. Samples are incubated overnight (ON) at 4 °C. However, it is recommended to previously perform an assessment of the kinetics since less refolding time may be enough.

3. GFP fluorescence of the solutions containing refolded aggregates is measured in a 96-well plate in a plate reader with the corresponding emission/excitation settings.

3.6 Visualizing Amyloid Structures Inside Intracellular Protein Aggregates

Although IBs have amorphous macroscopic appearance and are conventionally described as disordered aggregates being formed by nonspecific interactions of exposed hydrophobic surfaces, they are highly ordered protein aggregates formed through a process similar to that observed during amyloid deposition [6, 9, 35].

Due to their partial amyloid nature, IBs display regions with high resistance against controlled proteolytic digestion with proteinase K, corresponding to a protected β-sheet core. Thus, the presence of fibrillar structures with amyloid-like morphology in IBs could be observed directly or after controlled proteolysis by transmission electronic microscopy, and atomic force microscopy [11, 36].

3.6.1 Limited Proteolysis Followed by SDS-PAGE Electrophoresis and Mass Spectrometry, Transmission Electron Microscopy, and Atomic Force Microscopy

Proteinase K (pK) is a serine protease that exhibits a broad cleavage specificity. It cleaves peptide bonds adjacent to the carboxylic group of aliphatic and aromatic amino acids and it has been commonly used for general digestion of protein in biological samples as in molecular biology to remove contamination from preparations of nucleic acid [37]. While α-helix, random-coil, and β-turn region can be easily digested by pK, the β-sheet regions and specially cross-β-sheet characteristic of amyloid fibrils are highly resistant to pK activity this way, the limited proteolysis assay allows identifying the core of amyloid-like aggregates. Two approaches can be considered to perform the assay: i) to determine the pattern of digestion at different pK concentrations in a fixed period of time; or ii) to analyze the time course of the digestion for a selected pK concentration. The result of digestion experiments can be resolved by various techniques as SDS-PAGE electrophoresis, mass spectrometry or microscopy.

Limited Proteolysis Followed by SDS-PAGE Electrophoresis and Mass Spectrometry

1. Protein aggregates are prepared at OD 360 nm1 in PBS.
2. The required pK:protein ratio varies from 1:50 to 1:5,000 depending on the sample.
3. The digestion is performed at 37 °C and can be stopped by the addition of 1 volume of electrophoresis loading buffer followed by incubation at 100 °C for 5 min (*see* **Note 22**).
4. The resulting samples can be resolved in 12 % SDS-PAGE gels stained with Coomassie blue or by silver staining protocol.
5. The identity of the pK-resistant bands is analyzed on gel. The bands of interest are extracted cutting out a gel slice and placed in a microcentrifuge tube previously rinsed with 60 % acetonitrile.
6. The gel slice is destained in 100 mL of destaining solution for 20–30 min. This step should be repeated three to four times until the gel slice becomes completely destained.
7. The gel slice is dehydrated in 100 mL of 100 % acetonitrile for 5–10 min and dried at RT.

8. 30 mL of 2 % acetonitrile in 0.1 % formic acid are added to the samples, which are subsequently incubated at RT for 15 min.

9. The samples are further vortexed and sonicated for 1 min.

10. Eluted peptides are vacuum dried in a vacuum centrifuge for 45–60 min, being ready for analysis by mass spectrometry.

11. Equal volumes of the sample and a matrix solution mixed and dried by doplet method.

12. The molecular masses of the pK-resistant fragments are determined by MALDI-TOF spectroscopy. Analysis of the pK-resistant peptides allows a precise identification of the amyloid core mass of the aggregates.

Transmission Electron Microscopy

Transmission electron microscopy is the default technique for the visualization of fibrils and highly dense amyloid bundles overhanging from digested IBs. Negative staining is required to obtain image contrast..

1. IBs are diluted in their own buffer (*see* **Note 23**).

2. 10 µL of sample is placed on a carbon-coated copper grid.

3. The IB samples are allowed to settle to the grid for 5 min. Liquid excess can be removed with absorbent paper.

4. The grid is rinsed with ddH$_2$O, and stained with 2 % (w/v) uranyl acetate for 1 min.

5. The excess of uranyl acetate is discarded with absorbent paper and the grids are let to completely dry on air under sterile conditions.

6. Electron micrograph analysis is performed using a Hitachi H-7000 transmission electron microscope operating at an accelerating voltage of 75 kV.

Atomic Force Microscopy

Tapping-mode atomic force microscopy is a high-resolution type of scanning probe microscopy, which can be employed when a detailed study of amyloid formation in solution is needed.

1. Protein aggregates are pelleted and resuspended in ddH$_2$O (*see* **Note 24**).

2. 50 µL of sample is deposited on cleaved oriented pyrolytic graphite (HOPG) and allowed to adsorb for 20 min before starting the measurements.

3. Images of protein aggregates are obtained with a multimode atomic force microscope equipped with a 12 µm scanner (E-scanner). The images are taken in liquid medium using a liquid cell without the O-ring seal.

4. Veeco NP-S probes are used to scan the samples in tapping mode at a scan rate of 0.5 or 1 Hz.

3.7 Deciphering the Molecular Contacts That Sustain Intracellular Aggregates

3.7.1 Conformational Analysis by ATR-FTIR Spectroscopy

Fourier transformed infrared spectroscopy (FTIR) is a well-established low-resolution technique for studying the secondary structural composition and structural dynamics of proteins. The protein repeat units give rise to nine characteristic infrared absorption bands, namely as amide A, B, and I–VII [38]. Among them, the amide I and II bands are the two most prominent vibrational bands of the protein backbone. In particular, amyloid fibrils display a characteristic band at 1,620–1,630 cm^{-1} in the amide I region of the infrared spectra that is attributed to the tightly bound intermolecular β-strands of the amyloid core. In addition, a secondary band at 1,692 cm^{-1} has been assigned to antiparallel β-sheet conformation. To analyze the amide I band component, second derivative spectra need to be curve fitted. Since, IBs and amyloid fibrils tend to precipitate, attenuated total reflectance (ATR)-FTIR in which the aggregates can be deposited and analyzed in the solid state, is more convenient.

FTIR spectroscopy is used combined with H/D exchange in protein conformational analysis as the intensity changes of the amide I and II bands and the intensity change of the secondary structural elements α-helix and β-sheet can be determined.

1. For ATR-FTIR spectroscopy analysis, protein aggregates and IBs are purified as described in Subheading 3.1.

2. 5–10 μL of sample is placed in an FTIR spectrometer with a Golden ATR accessory. Samples are dried under N_2 atmosphere.

3. Each spectrum comprises 20 independent scans, measured at a spectral resolution of 1 cm^{-1} in the 1,700–1,600 cm^{-1} range.

4. Spectral data are acquired with OPUS MIR Tensor 27 software. All the absorbance spectra are normalized to avoid concentration-dependent effects. Additionally, buffer spectra have to be subtracted from each single spectrum (*see* **Note 25**).

5. Second derivatives of the amide I band spectra are used to determine the frequencies at which the different spectral components are located.

6. Infrared spectra can be fitted through overlapping Gaussian curves and the amplitude, center, and area of each Gaussian function calculated with a nonlinear peak-fitting program. Amyloid fibrils and native β-sheet proteins display maxima within two characteristic, although partially overlapping, spectral regions. The range of amyloid fibrils usually extends from 1,615 to 1,630 cm^{-1}, whereas native β-sheet proteins produce amide I′ peaks clustering between 1,630 and 1,643 cm^{-1}. The signals of amyloid-like structures inside bacterial aggregates overlap with those of *in vitro*-formed regular amyloid structures.

3.7.2 Residue-Level Analysis by NMR-Based H/D Experiments

The structural dynamics of a given conformation is likely to influence the activity of the protein. It has been established that the rates at which the amide proton exchange with solvent deuterium reflect the structural dynamics of proteins and they are sensitive to the secondary structural composition and experimental conditions such as pH, temperature, and pressure. Considering this, amyloid fibrils can be modeled at atomic resolution by Nuclear magnetic resonance (NMR) based H/D exchange [39].

1. Homogenously ^{15}N-labeled IBs are used for H/D-exchange experiments (*see* **Note 26**).

2. 0.5 mL of 200 μM IB solution is pelleted at $13,000 \times g$ for 3 min and washed with D_2O twice.

3. Pellet is resuspended in 0.5 mL of D_2O and kept at 4 °C for H/D exchange.

4. Immediately prior to each NMR measurement, the H/D-exchanged IBs are sedimented at $13,000 \times g$ for 3 min and dissolved in DMSO containing 0.05 % trifluoroacetic acid and 25 mM dithiothreitol.

5. The [^{15}N, ^{1}H]-correlation spectra (HMQC) are measured for 5 min, immediately after resuspending the IBs.

6. Residues that display high intrinsic exchange rates in DMSO are determined by the addition of H_2O followed by the measurement of a series of 2-D spectra. This control measurement let to exclude some residues from the H/D-exchange data analysis.

3.8 Characterizing the Toxic Properties of Intracellular Bacterial Aggregates

3.8.1 Metabolic Assay with MTT in Cultured Mammalian Cells

Proliferation assays are widely used in cell biology for the study of growth factors, cytokines, and nutrients and for the screening of cytotoxic or chemotherapeutic agents. In 1956, the first paper was published on the use of tetrazolium salts as indicators of cell viability. The method was based on the finding that living cells are capable to reduce slightly or uncolored tetrazolium salts into intensely colored formazan derivatives [40]. This reduction process requires functional mitochondria, which are inactivated within a few minutes after cell death. This method provides an excellent tool for the discrimination of living and death cells.

Growth Inhibition Assay

The effect of bacterial IBs on cell growth is determined by measuring 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide dye (MTT) absorbance.

1. Human cervical adenocarcinoma (HeLa) cells are cultured in DMEM supplemented with 10 % FBS and 1 % penicillin–streptomycin. HeLa cells are maintained in a humidified incubator containing 5 % CO2 at 37 °C.

2. Cells are routinely cultured in a logarithmic phase of growth.

3. 2.5×10^3 cells/well are seeded in 96-well microtiter plates in 200 μL of completed DMEM medium (*see* **Note 27**).

4. Cells are exposed to selected concentrations of IBs for 72 h and, afterwards, 20 μL of (MTT) solution (2 mg/ml in PBS) is added into each well. The plate should be gently mixed and the absorbance recorded.

5. The plates are incubated at 37 °C for 4 additional h (*see* **Note 28**).

6. After carefully removing the medium, 200 μL of DMSO is added to each well to solubilize the formazan crystals.

7. The absorbance is measured by a microplate reader set at Abs_{450nm} or Abs_{492nm} (*see* **Note 29**).

3.8.2 Viability Staining with Propidium Iodide

Accurate determination of live, dead, and total bacteria is important in many microbiology applications. Traditionally, viability in bacteria is measured as the ability to form colonies on solid growth medium or to proliferate in liquid nutrient broths. These culture-based tests are time consuming and can work poorly with slow-growing or viable. Furthermore, they do not provide real-time results or timely information.

Live cells have intact membranes and are impermeable to dyes such as propidium iodide (PI). PI is a fluorescent intercalating agent, which is not able to penetrate biological membranes and, thus, and generally excluded from viable cells. Therefore, PI can only leak into cells with compromised/permeabilized membranes. Upon entering cells, PI will bind to DNA and RNA and its fluorescence will be enhanced 20- to 30-fold, with the following settings: Excitation max = 536 nm/emission max = 617 nm.

1. Bacterial cultures should be diluted to a concentration range of 5×10^5 to 9×10^6 bacteria/mL in staining buffer. Importantly, the staining buffer should be previously 0.22 μm-filtered.

2. Bacterial suspensions must be vortexed and diluted at least 1:10 in staining buffer.

3. 5.0 μL of dye solution is added to 200 μL of bacterial suspension, in staining buffer. The final PI concentration is 48 μM PI.

4. Samples are vortexed and incubated for 5 min at RT.

5. Positive control consists of dead bacteria, which are prepared by heating at 95 °C for 10 min 200 μL of diluted bacterial suspension, and staining it as described above.

6. Acquisition of samples on a BD FACS brand flow cytometer (BD FACSCalibur flow cytometer or equivalent) is carried out as following: (*see* **Note 30**):

- Set initial instrument to logarithmic amplification. Use SSC and FSC to set threshold parameter. With an unstained sample of bacteria, set the amplification signals so that the bacteria are in the middle of the SSC-FSC dot plot. Gate this population as R1.

- Acquire a minimum of 10,000 events and maintain flow rate ≤1,000 event/second to avoid coincidence error.

- On the red fluorescence histogram (ex/488 nm, em/670 nm long-pass filter) used for PI analysis, set the fluorescence amplification photomultiplier (PMT), so that the unstained bacteria are in the lowest decade of the plot.

- Verify that the positive control stained with PI appears above the level of the unstained bacteria.

- Acquire experimental samples containing stained bacteria gated from R1. Compensation settings are not necessary.

3.9   Characterizing the Infectious Properties of Intracellular Prionic Aggregates

The amyloid nature of bacterial protein deposits and the sequential specificity accounted in their formation have been demonstrated by a wide set of physicochemical techniques together with seeding and cross-seeding experiments [9, 11]. However, the most conclusive proof of those properties has come along with the demonstration that they contain infectious material. On this regard, the bacterial expression of the *Podospora anserina* prion, Het-S, and the baker's yeast prion, Sup35, resulted in the formation of intracellular aggregates with amyloid properties able to specifically seed the in vitro amyloidogenesis reactions from homologous monomers. Moreover, the transfection of these aggregates into prion-free strains induced the prion conformation phenotype [41–43].

Here, we described a methodology that enables to explore the infectious properties of bacterial aggregates based on the yeast prion Sup35, which acts as a translation termination factor in its soluble and active form ([*psi–*]). The conversion into the prion form ([*PSI+*]) impairs its functionality and thus prion-infected strains exhibit a nonsense suppressor phenotype [44]. The protocol is based on the spheroplast transformation methodology previously described by Tanaka and Weissman [45, 46]. One of the main advantages of the method resides in the easy and straightforward analysis of the prion phenotype. A nonsense mutation in the *ade1* gene provokes that [*psi–*] strains appear as red colonies, while [*PSI+*] cells exhibit a pink or white color depending on if the prion phenotype is weak or strong, respectively [47].

The methodology detailed below also includes a curing assay, where prion-infected strains are incubated in the presence of a denaturing agent that solubilizes the prion aggregates and thus, cells can recover the [*psi–*] phenotype. The efficiency of the phenotype

conversion is indicative of the aggregate resistance to chemical denaturation [43].

3.9.1 Infecting Yeast with Bacterial Intracellular Aggregates

Spheroplast Preparation (*See* **Note 31**)

1. Cells are grown in standard YPD plates at 30 °C until colonies are visible (2–3 days).

2. Both [*psi–*] and [*PSI+*] yeast strains are required, also for control experiments. Background genotype must include a specific nonsense mutation in *ade1* gene (*see* **Note 32**).

3. Isolated colonies are grown ON in 10 mL of YPD liquid medium at 30 °C.

4. 25 mL of fresh YPD medium is inoculated with 2.5 mL of ON culture. New cultures are grown at 30 °C until an OD_{600nm} of 0.5 and afterwards, they are centrifuged at RT for 10 min at $1,500 \times g$.

5. Recovered cells are washed with 10 mL of ddH_2O, and subsequently with 1 M sorbitol, and centrifuged at RT for 5 min at $1,500 \times g$.

6. Pelleted cells are resuspended in 10 mL of SCE buffer.

7. Cells are digested with lyticase at a final concentration of 10 units/mL (*see* **Note 33**). Digestion must be carried out at 30 °C and stopped when 85–90 % of spheroplasts are reached. Concentration of spheroplasts is determined by measuring the OD_{800nm} of 200 μL of cells mixed with 800 μL of 5 % SDS with the following formula:

$$\% \text{ spheroplasts at time } t_x = 100 - \left[\left(OD_{800nm} t_x / OD_{800nm} t_0 \right) \times 100 \right]$$

8. Spheroplasts are centrifuged at RT for 10 min at $750 \times g$ (*see* **Note 34**).

9. Recovered spheroplasts are washed with 10 mL of 1 M sorbitol, and subsequently with STC buffer, and finally centrifuged at RT for 10 min at $750 \times g$.

10. Immediately before using, spheroplasts are resuspended in 100 μL of STC buffer, suitable for carrying out four transformations with 25 μL of spheroplast solution (control transformation experiments must be simultaneously carried out).

Spheroplast Transformation

1. 25 μL of spheroplasts are mixed with 3 μL of purified bacterial aggregates, 20 μg/mL URA3-marked plasmid, and 100 μg/mL salmon sperm DNA. The mixtures are incubated in a sonicator bath for 5 min (*see* **Note 35**).

2. Samples are incubated for 30 min at RT after adding 9 volumes of PEG buffer.

3. Yeast cells are centrifuged at RT for 10 min at $750 \times g$ and resuspended in SOS medium.

4. Cells are incubated at 30 °C for 30 min and plated on synthetic medium lacking uracil (SC-URA) overlaid with top agar at 2.5 %. It is indispensable that the medium contains adenine at 20 mg/mL since it is essential for the survival of [*psi–*] colonies.

Yeast Phenotype Analysis

1. Transformed yeast cells are grown on SC-URA plates for at least 5 days at 30 °C.

2. Isolated transformed colonies are streaked onto ¼ YPD plates (*see* **Note 36**).

3. After growing at 30 °C for 3 days, prion phenotypes are revealed by colony color. Strong [*PSI+*], weak [*PSI+*], and prion-free phenotype [*psi–*] appear as white, pink, and red cells, respectively.

Recovery of the Prion-Free Phenotype [*psi–*]

1. Selected [*PSI+*] colonies are used to inoculate 10 mL of fresh YPD medium containing 3 mM Gnd·HCl.

2. Cultures are grown for 48 h at 30 °C under gentle agitation.

3. 5 µL of each culture is spotted onto ¼ YPD. Cells recovering the [*psi–*] phenotype appear as red colonies.

4   Notes

1. In order to improve the yield of IBs recovery, protein expression is generally induced when bacterial culture reaches an OD_{600nm} of about 0.5.

2. Troubleshooting for soluble protein expression can include the following: (1) expression at RT or at 16 °C rather than 37 °C and (2) expression from a freshly transfected strain rather than frozen glycerol stock.

3. In order not to disrupt the fusion between GFP and the target protein, the linker must be short and lacking of large bulky hydrophobic residues [20].

4. Protein expression induction must be performed immediately before starting the time-lapse imaging in order to acquire microscopy images from the beginning of the experiment. It is also recommended to include a non-induced culture as control.

5. It is important to perform a z-stack capture at each interval, since the medium on the glass slide melts upon the laser incidence and images might be out of focus.

6. PBS must be 0.22 µm filtered to avoid contamination particles in microscopy images.

7. The OD_{360nm} of aggregates for spectrofluorometry measurements must be adjusted depending on the fluorescence activity of the samples.

8. *E. coli* and other microorganisms emit basal fluorescence; thus it is convenient to record the fluorescence emission of a non-induced control.

9. It is important to include a reference sample since fluorescence measurements obtained are relative values.

10. Settings must be adjusted depending on the flow cytometer and the appropiate controls must be included unstained and single-color controls should be used to locate populations and to confirm that PMT voltages for FL-1 are suitable.

   – Set up an FSC vs. SSC plot to gate live bacterial population and to discard debris in the sample.

   – When convenient, analyze an aliquote of the stained bacteria in suspension by fluorescence microscopy.

   – Filter previously all buffers and solutions through 0.22 μm sterile syringe filter to avoid any bacterial contamination.

11. It is important to include negative controls of cells expressing only the GFP-tagged protein and cells expressing only the BFP-tagged protein.

12. Protein expression time must be adjusted in each case. It is recommended to perform a previous experiment to follow the aggregation kinetics, since cells should be harvested when aggregate formation reaches the equilibrium phase.

13. The OD_{600nm} must be adjusted depending on the fluorescence emission of the cells.

14. Mixed equal amounts of control cells expressing the fusion proteins separately can be used as negative control in this experiment.

15. At this stage fixed cells with formaldehyde can be stored at 4 °C until observed.

16. It is important to make sure that the confocal microscope employed is provided with laser lines suitable for the independent excitation of the fluorophores used in the experiment.

17. For a more detailed quantification, the number of colocalized pixels can be determined by plotting the pixels of each channel in a scatter diagram.

18. For an accurate and easy analysis of FRET measurements, it is highly recommendable to use specific software for confocal microscopy images.

19. The method described here is based on the amyloid β-peptide fused to the GFP and thus settings are adjusted for this specific case. However other fusion proteins can be employed.

20. In this step, it is important to add reaction tubes with the same amount of aggregates for further control experiments.

21. The concentration of the tested compounds in the refolding buffers may vary. Aggregation enhancer metals might require a minimum concentration of 10 μM and, as a reference, 25 μM has been enough to detect inhibitor activities [14].

22. Variation in the temperature or the pH of the pK reaction buffer could affect enzyme efficiency.

23. If aggregates are too big or dense, samples can be previously sonicated in an ultrasonic bath for 10 min.

24. This process must be carried out three times to completely eliminate organic constituents such as DMSO from the incubation buffer as they are absorbed on HOPG.

25. When the buffer spectrum interferes strongly with the measurements, the buffer of the samples should be exchanged for dd H_2O.

26. To ensure that, during H/D-exchange measurement, IBs preserve their structural properties, thioflavin T binding and electron microscopy visualization can be performed.

27. It is advisable to use as few cells as possible, otherwise the occurrence of a nonlinear titration curve may be possible.

28. Incubation time with MTT substrate should be adjusted depending on the cell mitochondrial activity rate.

29. It is recommended to measure the reference absorbance at 620 nm (or any wavelength between 620 and 690 nm) to correct measures for nonspecific background values, caused by cell debris, fingerprints, or other potential interferences. Absorbance values can show significant differences in the metabolic activity depending on the cell line used and its metabolic rate.

30. The following items must be considered during the flow cytometry acquisition data:

    – Cellular debris are excluded from analysis by raising the FSC threshold. Use an unstained bacterial sample control to confirm that PMT voltages are set appropriately.

    – Use a mixture of live and dead bacteria to confirm that stained live and dead populations are sufficiently resolved.

    – Since not all bacteria populations display the same PI-uptake ability, it is recommended to previously assess the staining procedure by epifluorescence microscopy.

    – The combination of the dyes SYTO green (Molecular Probes) or Thiazole orange (BD Biosciences) with PI provides a rapid and reliable method for discriminating live and dead bacteria.

31. All the required buffers must be freshly prepared and sterilized by 0.22 μm filtration.

32. Isogenic yeast [psi–] and [PSI+] derivatives of 74D-694 [MATa, his3, leu2, trp1, ura3; suppressible marker ade1-14(UGA)] are recommended [46].

33. Lyticase stock solution might be previously prepared in phosphate buffer pH 7.4 with 50 % glycerol at a final concentration of 10,000 units/mL and stored at –80 °C for further uses.

34. The spheroplast frailty requires gentle handling.

35. Spheroplast transformation must be performed with the following considerations: (1) bacterial aggregates must be prepared at high concentration and briefly sonicated previous to their use in transformation experiments; (2) the URA3-marked plasmid is strongly recommended for a first selection of cells that have introduced external material; pRS316 is stated as an example; and (3) salmon sperm DNA must be previously heated at 95 °C for 10 min and afterwards, incubated on ice for 20 min.

36. Standard YPD medium must be fourfold diluted (¼ YPD) to permit an easy visualization of the color phenotype.

Acknowledgments

References

1. Invernizzi G, Papaleo E, Sabate R et al (2012) Protein aggregation: mechanisms and functional consequences. Int J Biochem Cell Biol 44:1541–1554

2. Calamai M, Kumita JR, Mifsud J et al (2006) Nature and significance of the interactions between amyloid fibrils and biological polyelectrolytes. Biochemistry 45:12806–12815

3. Fernandez-Busquets X, de Groot NS, Fernandez D et al (2008) Recent structural and computational insights into conformational diseases. Curr Med Chem 15:1336–1349

4. Nelson R, Eisenberg D (2006) Recent atomic models of amyloid fibril structure. Curr Opin Struct Biol 16:260–265

5. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24:179–185

6. de Groot NS, Sabate R, Ventura S (2009) Amyloids in bacterial inclusion bodies. Trends Biochem Sci 34:408–416

7. Sabate R, de Groot NS, Ventura S (2010) Protein folding and aggregation in bacteria. Cell Mol Life Sci 67:2695–2715

8. Garcia-Fruitos E, Sabate R, de Groot NS et al (2011) Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J 278:2419–2427

9. Carrio M, Gonzalez-Montalban N, Vera A et al (2005) Amyloid-like properties of bacterial inclusion bodies. J Mol Biol 347:1025–1037

10. Morell M, Bravo R, Espargaro A et al (2008) Inclusion bodies: specificity in their aggregation process and amyloid-like structure. Biochim Biophys Acta 1783:1815–1825

11. Dasari M, Espargaro A, Sabate R et al (2011) Bacterial inclusion bodies of Alzheimer's disease beta-amyloid peptides can be employed to study native-like aggregation intermediate states. Chembiochem 12:407–423

12. de Groot NS, Ventura S (2006) Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. J Biotechnol 125:110–113

13. de Groot NS, Aviles FX, Vendrell J et al (2006) Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. FEBS J 273:658–668

14. Villar-Pique A, Espargaro A, Sabate R et al (2012) Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. Microb Cell Fact 11:55

15. Villar-Pique A, de Groot NS, Sabate R et al (2012) The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. J Mol Biol 421:270–281

16. Jahn TR, Radford SE (2008) Folding versus aggregation: polypeptide conformations on competing pathways. Arch Biochem Biophys 469:100–117

17. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27

18. Shimomura O, Johnson FH, Saiga Y (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol 59:223–239

19. Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67:509–544

20. Waldo GS, Standish BM, Berendzen J et al (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17:691–695

21. Belli M, Ramazzotti M, Chiti F (2011) Prediction of amyloid aggregation in vivo. EMBO Rep 12:657–663

22. Castillo V, Grana-Montes R, Sabate R et al (2011) Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. Biotechnol J 6:674–685

23. Guidolin D, Agnati LF, Albertin G et al (2012) Bioinformatics aggregation predictors in the study of protein conformational diseases of the human nervous system. Electrophoresis 33:3669–3679

24. Conchillo-Sole O, de Groot NS, Aviles FX et al (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 8:65

25. Woulfe J (2008) Nuclear bodies in neurodegenerative disease. Biochim Biophys Acta 1783:2195–2206

26. Kopito RR (2000) Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol 10:524–530

27. Lindner AB, Madden R, Demarez A et al (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc Natl Acad Sci U S A 105:3076–3081

28. Kim W, Kim Y, Min J et al (2006) A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. ACS Chem Biol 1:461–469

29. Martinez-Alonso M, Vera A, Villaverde A (2007) Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells. FEMS Microbiol Lett 273:187–195

30. Vera A, Gonzalez-Montalban N, Aris A et al (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng 96:1101–1106

31. de Groot NS, Ventura S (2006) Effect of temperature on protein quality in bacterial inclusion bodies. FEBS Lett 580:6471–6476

32. Espargaro A, Sabate R, Ventura S (2012) Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation. Mol Biosyst 8:2839–2844

33. Rajan RS, Illing ME, Bence NF et al (2001) Specificity in intracellular protein aggregation and inclusion body formation. Proc Natl Acad Sci U S A 98:13060–13065

34. Hart RA, Rinas U, Bailey JE (1990) Protein composition of Vitreoscilla hemoglobin inclusion bodies produced in Escherichia coli. J Biol Chem 265:12728–12733

35. Wang L, Maji SK, Sawaya MR et al (2008) Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol 6:e195

36. Cano-Garrido O, Rodriguez-Carmona E, Diez-Gil C et al (2013) Supramolecular organization of protein-releasing functional amy-

loids solved in bacterial inclusion bodies. Acta Biomater 9:6134–6142

37. Hubbard SJ (1998) The structural aspects of limited proteolysis of native proteins. Biochim Biophys Acta 1382:191–206

38. Kong J, Yu S (2007) Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim Biophys Sin (Shanghai) 39:549–559

39. Tycko R (2006) Solid-state NMR as a probe of amyloid structure. Protein Pept Lett 13: 229–234

40. Denizot F, Lang R (1986) Rapid colorimetric assay for cell growth and survival. Modifications to the tetrazolium dye procedure giving improved sensitivity and reliability. J Immunol Methods 89:271–277

41. Wasmer C, Benkemoun L, Sabate R et al (2009) Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218–289) are amyloids. Angew Chem Int Ed Engl 48:4858–4860

42. Garrity SJ, Sivanathan V, Dong J et al (2010) Conversion of a yeast prion protein to an infectious form in bacteria. Proc Natl Acad Sci U S A 107:10596–10601

43. Espargaro A, Villar-Pique A, Sabate R et al (2012) Yeast prions form infectious amyloid inclusion bodies in bacteria. Microb Cell Fact 11:89

44. Liebman SW, Derkatch IL (1999) The yeast [PSI+] prion: making sense of nonsense. J Biol Chem 274:1181–1184

45. Tanaka M, Weissman JS (2006) An efficient protein transformation protocol for introducing prions into yeast. Methods Enzymol 412:185–200

46. Tanaka M (2010) A protein transformation protocol for introducing yeast prion particles into yeast. Methods Enzymol 470:681–693

47. Chernoff YO, Lindquist SL, Ono B et al (1995) Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science 268:880–884

Part II

Strategies to Produce Insoluble Proteins in Cell-Free Expression Systems

Chapter 7

Co-translational Stabilization of Insoluble Proteins in Cell-Free Expression Systems

Lei Kai, Erika Orbán, Erik Henrich, Davide Proverbio, Volker Dötsch, and Frank Bernhard

Abstract

Precipitation, aggregation, and inclusion body (IB) formation are frequently observed problems upon overexpression of recombinant proteins. The open accessibility of cell-free reactions allows addressing such critical steps by the addition of protein stabilizers such as chemical chaperones or detergents directly into the expression reactions. This approach could therefore reduce or even prevent initial protein precipitation already in the translation environment. The strategy might be considered to generally improve protein sample quality and to rescue proteins that are difficult to refold from IBs or from aggregated precipitates. We describe a protocol for the co-translational stabilization of difficult proteins by their expression in the presence of supplements such as alcohols, poly-ions, or detergents. We compile potentially useful compounds together with their recommended stock and working concentrations. Examples of screening experiments in order to systematically identify compounds or compound mixtures that stabilize particular proteins of interest are given. The method can primarily be considered for the production of unstable soluble proteins or of membrane proteins containing larger soluble domains.

Key words Chemical chaperone, Co-translational stabilization, Cell-free expression, Insoluble proteins, Linear and correlated screening, Inclusion bodies, Protein aggregation

1 Introduction

Overexpression of recombinant proteins often causes formation of aggregates or inclusion bodies (IBs) [1]. Stress signals or other unfavorable environmental conditions can induce similar effects in living cells. Many organisms are therefore able to synthesize organic substances, known as chemical chaperones, that are able to co-translationally stabilize proteins at suboptimal conditions [2]. While chemical chaperones have significant potentials in biotechnology, the restricted access to the inner cell compartment in most cases prevents their usage for the co-translational protein stabilization in conventional cellular overexpression systems. Stabilization strategies implementing chemical chaperones are therefore usually

restricted to manipulations of growth conditions or to posttranslational refolding approaches.

Cell-free (CF) expression systems became routine techniques for the production of difficult proteins, e.g., membrane proteins [3, 4], toxins [5], antibodies [6], or other problematic proteins [7]. A unique characteristic is the open accessibility of CF expression reactions. A variety of additives such as chemical chaperones, detergents, surfactants, or other compounds can be added at any time point of the reaction and can act co-translationally at the nascent peptide chains [8]. Already initial precipitation or unfolding of protein could therefore be reduced or even prevented.

In this chapter, we focus on the co-translational stabilization of difficult proteins via expression in CF system. We describe the implementation of potential chemical chaperones and stabilizers, including poly-ions, alcohols, and amino acids, as supplements for CF reactions. Each new compound with potential stabilizing effect must first be tested for its general compatibility with the CF expression system. We therefore discuss the compatibility screening of interesting compounds and how to define suitable working concentrations in CF expression reactions. For analytical scale screening in throughput approaches, we describe the single compartment CF batch configuration operated in microplate format. Linear screening of individual compound concentrations as well as the correlated screening of compound mixtures can be performed. We recommend fusions with green fluorescent protein (GFP) derivatives for the fast monitoring of additive effects and for the generation of initial short lists of effective stabilizers [9, 10]. If a stabilization protocol has been established, preparative scale CF reactions intended for further processing of the synthesized proteins such as purification or enzymatic characterization can be set up in the more efficient two-compartment continuous exchange cell-free (CECF) configuration.

2   Materials

All stock solutions should be prepared with ultrapure water and stored at –20 °C if not otherwise stated.

2.1   General Materials

1. Fermenter for 5–10 L of culture volume.
2. French press or other high-pressure cell-disruption equipment.
3. Photometer.
4. Standard centrifuges and set of rotors.
5. Ultrasonic water bath.
6. Thermo shaker for incubation.
7. Chromatographic system (e.g., Äkta purifier, GE Healthcare).

8. Q-Sepharose column (GE Healthcare).

9. Centriprep filter devices, 10 kDa MWCO (Millipore).

10. Plasmid and PCR product purification kits.

11. Dark microplate for fluorescence measurement.

12. GFP assay buffer: 20 mM Tris–HCl, pH 7.8, 150 mM NaCl.

13. Labsonic homogenizer (B. Braun Biotech International).

14. 1.8 mL Nunc cryotubes (ThermoScientific).

15. L-[^{35}S]Methionine.

16. 20 mL scintillation vials.

17. Rotiszint®eco scintillation cocktail (Roth).

18. LS6500 multipurpose scintillation counter (Beckmann).

2.2 Chemical Chaperones and Hydrophobic Compounds	Chemical chaperones useful for CF expression are listed in Table 1. Chemicals should be obtained with the highest purity. Detergents for the stabilization of hydrophobic proteins or of membrane proteins are listed in Table 2.
2.3 Materials for CF Expression Reactions	CF reactions can be performed either in the batch configuration consisting of only the reaction mix (RM) or in the continuous exchange (CECF) configuration consisting of RM and feeding mix (FM). RM and FM are separated in the reaction containers by a semipermeable membrane (12–14 kDa MWCO).

2.3.1 CF Batch Configuration

1. V-shaped 96-well microplates.

2. Stock solutions required for batch reactions are listed in Table 3. A number of compounds are combined in a 10× premix before pipetting into the microplate.

2.3.2 CECF Configuration

1. 24-well microplates.

2. Dialysis tubes, 12–14 kDa MWCO.

3. Reaction containers: analytical scale Mini-CECF-Reactors and preparative scale Maxi-CECF-Reactors [11] (*see* **Note 1**). D-tube™ dialyzer, 12–14 kDa MWCO (Merck Biosciences); Slide-A-Lyzer, 10 kDa MWCO (Pierce).

4. Stock solutions required for CECF reactions are listed in Table 3. Compounds common to the RM and FM are first combined into an RFM-mix.

2.4 Materials for S30 Extract and T7RNAP Preparation

1. 40× S30-A/B buffer: 400 mM Tris–acetate, pH 8.2, 560 mM Mg(OAc)$_2$, 2.4 M KCl. Supplement 1× S30-A buffer with 6 mM ß-mercaptoethanol. Supplement 1× S30-B buffer with 1 mM DTT and 1 mM phenylmethanesulfonyl fluoride (PMSF).

Table 1
Chemical chaperones for the co-translational stabilization of unstable proteins [11]

Compound	Stock concentration	Working concentration limit[a]
Betaine	1 M	>250 mM
Choline	100 mM	>20 mM
Ectoine	1 M	≤150 mM
Sucrose	40 % (w/v)	≤10 %
D-Trehalose	40 % (w/v)	≤4 %
D-Mannose	40 % (w/v)	≤2 %
D-Sorbitol	40 % (w/v)	≤4 %
Glycerol	40 % (w/v)	≤8 %
L-OH- proline	50 mM	>10 mM
N-Acetyl-L-lysine	500 mM	≤100 mM
L-Carnitine	100 mM	≤10 mM
L-Arginine	100 mM	>20 mM
Sarcosine	200 mM	>40 mM
L-Glutamic acid	2 M	>400 mM
Methanol	20 % (v/v)	≤5 %
Ethanol	20 % (v/v)	≤8 %
Isopropanol	20 % (v/v)	≤5 %
Butanol	20 % (v/v)	≤3 %
Pentanol	20 % (v/v)	n.t. (<1 %)
Hexanol	20 % (v/v)	n.t. (<1 %)
PEG 200	40 % (w/v)	>6 %
PEG 400	40 % (w/v)	≤4 %
PEG 1,000	40 % (w/v)	≤6 %
PEG 6,000	40 % (w/v)	≤4.8 %
PEG 8,000	40 % (w/v)	≤4.8 %
PEG 10,000	20 % (w/v)	n.t. (<1 %)

[a]> upper concentration limit not defined yet; n.t., not tolerated by CF systems

2. 40× S30-C buffer: 400 mM Tris–acetate, pH 8.2, 560 mM $Mg(OAc)_2$, 2.4 M KOAc. Supplement $1 \times S30$-C buffer with 0.5 mM DTT.

3. 2×YTPG medium: 22 mM KH_2PO_4, 40 mM K_2HPO_4, 100 mM glucose, 16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl.

Table 2
Detergents for the co-translational stabilization of hydrophobic proteins and membrane proteins

Name	Stock	Working concentration[a]	Reference
Concentration			
Brij-35	5 % (w/v)	≤0.1 %	[17–19]
Brij-58	15 % (w/v)	≤1.5 %	[17–19]
Brij-78	15 % (w/v)	≤1.0 %	[12, 23]
Brij-98	5 % (w/v)	≤0.2 %	[12, 23]
Digitonin	4 % (w/v)	≤0.4 %	[14, 23]
Triton X100	5 % (w/v)	≤0.1 %	[12, 14, 23]
DDM	2 % (w/v)	≤0.1 %	[12, 23]
CHAPS	10 % (w/v)	≤1 %	[20, 21]
Nvoy10	5 mM	≤0.5 mM	[22]
Tween 20	10 % (w/v)	≤1 %	[23, 17]
DHPC	2 % (w/v)	≤0.2 %	[23]
DM	5 % (w/v)	≤0.2 %	[23]
Cholate	10 % (w/v)	≤1 %	[20]

[a]Tolerance might change if compounds are used as mixtures
Brij-35, polyoxyethylene-(23)-lauryl-ether; Brij-58, polyoxyethylene-(20)-cetyl-ether; Brij-78, polyoxyethylene-(20)-stearyl-ether; Brij-98, polyoxyethylene-(20)-oleyl-ether; Triton X100, polyethylene glycol *tert*-octylphenyl ether; DDM, *n*-dodecyl-β-D-maltoside; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; Nvoy, NV10 polymer; Tween 20, polyethylene glycol sorbitan monolaurate; DHPC, 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine; DM, *n*-decyl-β-maltoside; cholate, cholalic acid sodium salt

4. LB medium: 10 g/L peptone, 5 g/L yeast extract, 5 g/L NaCl.

5. Buffer-T7RNAP-A: 30 mM Tris–HCl, pH 8.0, 50 mM NaCl, 10 mM EDTA, 10 mM ß-mercaptoethanol, 5 % glycerol.

6. Buffer-T7RNAP-B: 10 mM K_2HPO_4/KH_2PO_4, pH 8.0, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5 % glycerol.

7. 20 % streptomycin sulfate in H_2O.

8. *E. coli* strain for extract preparation, e.g., A19 (*E. coli* Genetic Stock Center, New Haven, CT) or BL21 (Merck Biosciences).

9. BL21(DE3) Star × pAR1219 for T7RNAP preparation [12].

2.5 DNA Template Preparation

1. PCR purification kit.

2. Plasmid DNA purification kit.

3. DNA polymerase.

4. Agarose.

Table 3
CF reaction preparation for batch and CECF configurations

CECF configuration			Batch configuration		
Compound	Stock conc.	Final conc.	Compound	Stock conc.	Final conc.
RFM-Mix			RM		
RCWMDE amino acid mix	16.7 mM	1 mM	20 Amino acid mix	25 mM	2 mM
20 Amino acid mix	25 mM	0.5 mM	Phospho(enol)pyruvic acid (K[+]), pH 7.0[a]	1 M	30 mM
Acetyl phosphate (Li[+], K[+]), pH 7.0[a]	1 M	20 mM	75×NTP mix, pH 7.0[b]	90 mM ATP	2.5 mM
Phospho(enol)pyruvic acid (K[+]), pH 7.0[a]	1 M	20 mM		60 mM G/C/UTP	1.7 mM
75×NTP mix, pH 7.0[b]	90 mM ATP 60 mM G/C/UTP	1.2 mM 0.8 mM	t-RNA (*E. coli*) T7RNAP	40 mg/mL 0.7 mg/mL	0.17 mg/mL 0.01 mg/mL
DTT	500 mM	2 mM	DNA template	0.2–0.5 µg/µL	2–20 ng/µL
Folinic acid (Ca[2+])	10 mg/mL	0.1 mg/mL	*E. coli* S30 extract		
Complete cocktail (Roche Diagnostics)	50×	1×	10× Premix	1×	0.24×
Hepes/EDTA, pH8.0[a]	24×	1×			
Mg(OAc)$_2$	1 M	11.1 mM[c]	Putrescine	15 mM	1.5 mM

KOAc	4 M	110 mM[c]	Spermidine	15 mM	1.5 mM
PEG 8000	40 %	2 %	K$^+$-glutamate	2500 mM	250 mM
NaN$_3$	10 %	0.05 %	NH$_4$$^+$-glutamate	100 mM	10 mM
			Mg^{2+}-glutamate	100 mM	26 mM[d]
RM-Mix	10 mg/mL		Na$^+$-oxalate	40 mM	4 mM
Pyruvate kinase		0.04 mg/mL	Na$^+$-pyruvate	330 mM	33 mM
t-RNA (*E. coli*)	40 mg/mL	0.5 mg/mL	Folinic acid	340 µg/mL	34 µg/mL
T7RNAP	1.4 mg/mL	0.05 mg/mL	DTT	10 mM	1 mM
RiboLock	40 U/µL	0.3 U/µL	NAD$^+$	5.3 mM	0.53 mM
DNA template	0.2–0.5 µg/µL	2–20 ng/µL			
E. coli S30 extract	1×	0.35×			
FM-Mix					
S30-C buffer	1×	0.35×			
20 Amino acid mix	25 mM	0.5 mM			

[a]Adjusted with KOH
[b]Adjusted with NaOH
[c]Final total concentrations are 16 mM Mg^{2+} and 270 mM K$^+$ as additional ions are contributed by other compounds
[d]If not used as screening compound, the total final Mg^{2+} concentration was adjusted to 26 mM

3   Methods

3.1   Preparation of S30 Extract

The *E. coli* strain A19 is one of the most frequently recommended extract sources as it is low of internal RNases. The S30 extract preparation is a standard procedure which can be performed in 1 day (Fig. 1) [11–13]. The major steps of the S30 preparation
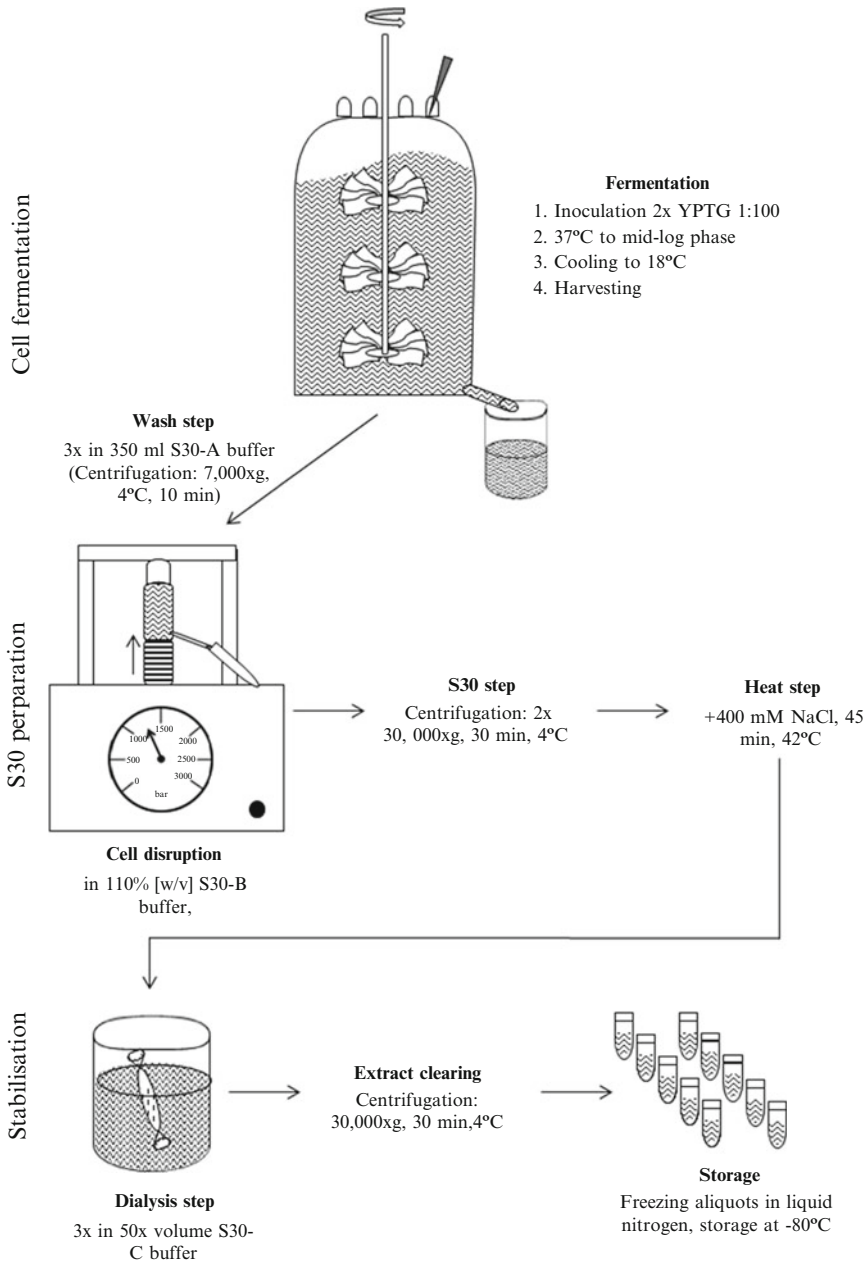
Fig. 1 Flow chart of S30 extract preparation. With a 10 L fermenter, cell fermentation and S30 preparation are performed within 1 day, while the stabilization part takes ON dialysis and harvesting of the extract within the next morning

protocol are (1) fermentation of the cells at 37 °C at good aeration to mid-log growth phase followed by subsequent chilling to 18 °C (*see* **Note 2**), (2) cell washing (*see* **Note 3**), (3) cell disruption by French press or similar and "S30" (30,000 × g) centrifugation, and (4) a high-salt heat step in order to remove endogenous mRNA and undesired proteins (*see* **Note 4**). Subsequent dialysis (14 kDa cutoff) can be performed ON and precipitates are removed by another S30 centrifugation step before aliquoting and storage of the extract (*see* **Note 5**). Starting with a 10 L fermentation yields approximately 60 mL of S30 extract.

3.2 T7 RNA Polymerase (T7RNAP) Preparation

T7RNAP is produced from the *E. coli* strain BL21(DE3) Star (pAR1219) by conventional cultivation in Erlenmeyer flasks with LB medium (Fig. 2) [12]. T7RNAP is heavily overexpressed and visible as prominent band of approximately 90 kDa upon Coomassie
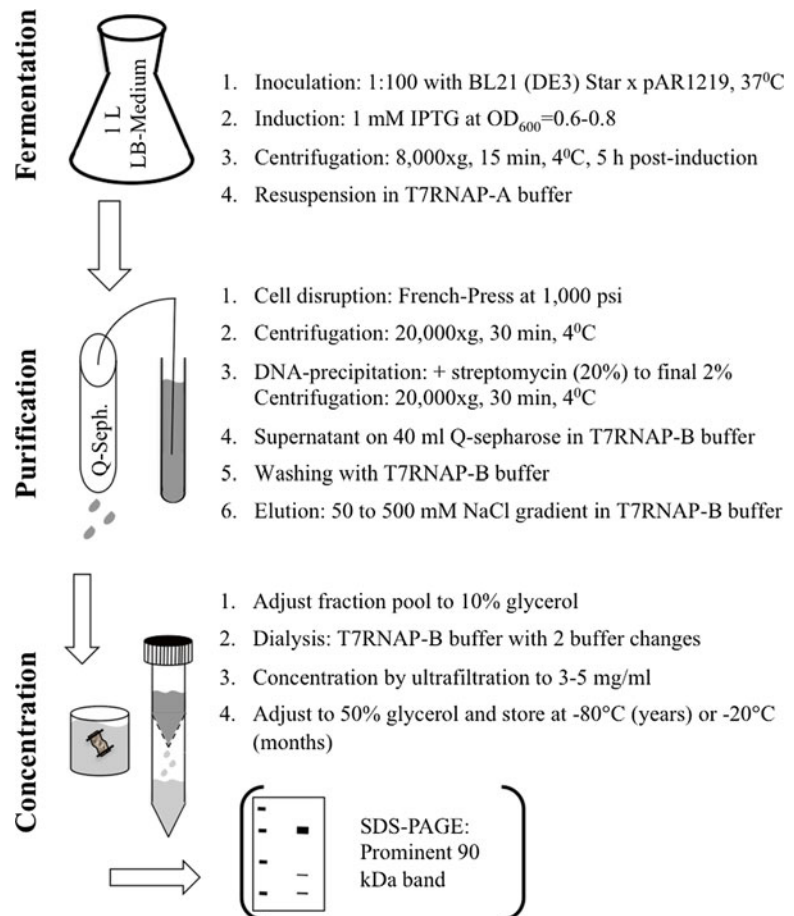
Fig. 2 Flow chart of T7RNAP preparation. The protein is expressed by standard *E. coli* fermentation in Erlenmeyer flasks

Blue staining. On average, approximately 20,000–40,000 T7RNAP units can be isolated out of 1 L culture. The critical steps of T7RNAP preparation are: (1) purification by anion exchange chromatography (*see* **Note 6**) and (2) concentration and storage of the isolated enzyme (*see* **Note 7**).

3.3 DNA Template Preparation

The reading frame to be expressed must be under control of the T7 promoter and T7 terminator. Commonly used vectors are the pET (Merck Biosciences) or pIVEX (Roche Diagnostic) series. High quality and purity of DNA templates is crucial for efficient CF expression. The optimal template concentration should be determined for each new target with an initial concentration screen in the range of 0.1–20 ng/μL of RM.

1. Plasmid templates should be prepared with commercial standard kits such as "Midi" or "Maxi" DNA purification kits. "Mini" kit preparations are not suitable due to the low quality of the purified DNA. The DNA should be dissolved in pure MQ H_2O and optimal stock concentrations are in between 0.2 and 0.5 mg/mL.

2. If the target gene is already present in a suitable vector under control of T7 promoter elements, fragments containing the T7 regulatory sequence and the target gene can be amplified by standard PCR and directly used in the CF reaction. In case the target gene is under control of a different promoter, the appropriate T7 regulatory elements might be attached by a multistep overlap PCR strategy [14].

3.4 GFP and Translational GFP Fusions as Monitoring System

Expression of translational fusions of the target protein with C-terminally attached GFP could allow a fast monitoring of soluble protein expression directly in the RM. In addition, in some cases the functional folding of the C-terminal GFP moiety may correlate with the folding of the N-terminal target protein [15].

Besides wild-type GFP [$\lambda_{Ex} = 395/\lambda_{Em} = 510$], the most frequently used derivatives are shifted GFP (sGFP) [$\lambda_{Ex} = 484/\lambda_{Em} = 510$] and superfolder GFP [$\lambda_{Ex} = 484/\lambda_{Em} = 510$] [16]. GFP expression alone could further be used as control for establishing and optimization of CF expression protocols as well as for testing the quality of freshly prepared stock solutions and extract batches. It further serves as monitor in compatibility screens for determining the tolerance of the CF expression system for new additives such as potential chemical chaperones.

GFP production can be quantified by fluorescence measurement.

1. After expression, keep GFP samples at 4 °C for 12 h to allow complete GFP folding.

2. Add 3 μL of sample into 297 μL GFP assay buffer in a 96-well dark microplate. The dilution of the sample may be adjusted in order to stay within the range of the calibration curve.
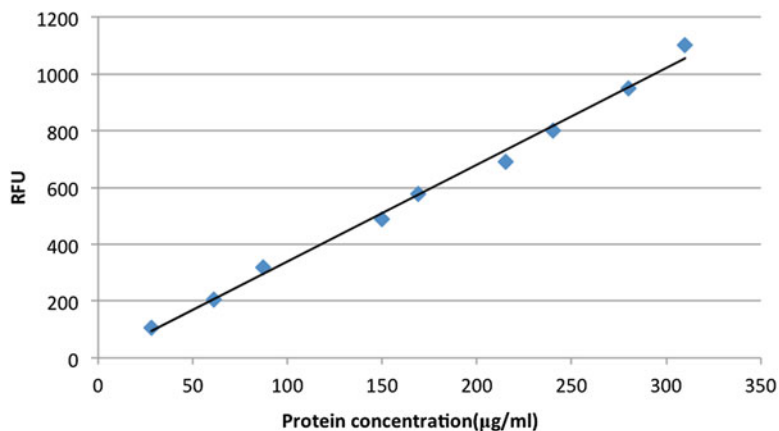
Fig. 3 Calibration curve of sGFP. RFU: relative fluorescent unit

Dependent on the GFP concentration in the sample, either dilutions should be made or increased sample volumes should be measured, if appropriate.

3. Incubate the plate with shaking for 5 min at 22 °C.

4. Perform fluorescence measurement at appropriate wavelengths.

5. Use a suitable calibration curve with purified GFP to quantify the measured fluorescence (Fig. 3).

3.5 CF Expression Reactions in Batch Configuration

One-compartment batch reactions performed in standard V-shaped 96-well microplates are suitable for initial throughput screens. Total volumes of ≥25 μL in the individual microplate cavities are recommended for batch reactions.

1. Prepare each individual compound as stock solution (Table 3).

2. In order to reduce pipetting time, a set of compounds can be combined in a 10× premix (Table 3). The premix can be stored at –20 °C (*see* **Note 8**).

3. Calculate individual compound volumes and prepare a pipetting scheme for the intended set of reactions (*see* **Note 9**). The volume of master mixes should be 110 % of the calculated volumes in order to compensate for volume losses during mixing and pipetting.

4. Thaw the premix and other stock solutions on ice.

5. Combine all common compounds (except DNA template) including the premix into a master mix. The premix and amino acid mix need to be resuspended before pipetting as some compounds might not be dissolved completely.

6. Transfer appropriate aliquots of the master mix into the cavities of 96-well microplates.

7. Fill up to the desired final volume (minus volume for the DNA template) with MQ H$_2$O, e.g., 25 μL. In screening experiments, this free water volume can be replaced by the selected additives.

8. Start the reactions by addition of the DNA template.

9. Seal the microplates by Parafilm. Incubate at 30 °C (*see* **Note 10**) for 2–4 h with slight shaking.

**3.6   Co-translational Linear and Correlated Screening of Chemical Chaperones in the Batch Configuration**

Screening experiments can be performed in linear concentration screens with one compound or in correlated screens with two or more compounds (Fig. 4). Linear screens are suitable for initial compatibility tests and for analyzing the general effect of a new compound on the quality of a target protein. Compatibility screens are mandatory as first experiments in order to define the appropriate working range of a newly selected additive. If the additive is tolerated, the effect of the compound on the co-translational stabilization of proteins of interest is analyzed in subsequent linear concentration screens within the defined working range. Once a number of beneficial compounds have been identified,
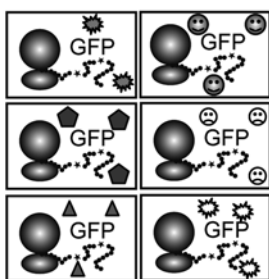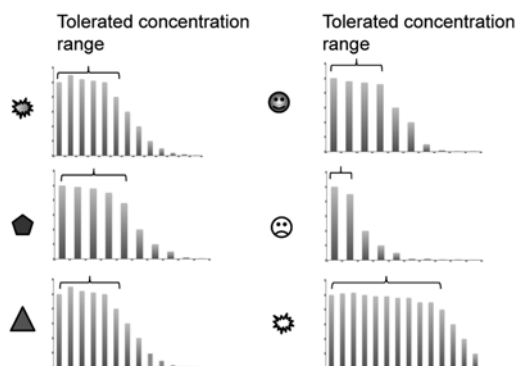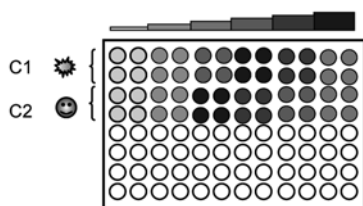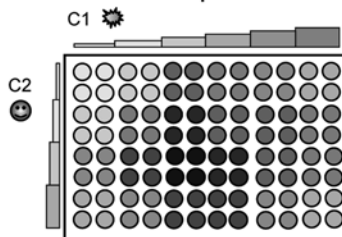

Fig. 4 Overview of the chemical chaperone screening strategy

further correlated concentration screens can be approached for analyzing synergistic effects of the individual compounds on the protein of interest (*see* **Note 11**). Prepare all additive stocks according to the recommended concentrations listed in Tables 1 and 2. Number and maximal concentration of tested compounds must be adjusted according to the available water volume in the RM.

1. Prepare stock solutions and a master mix out of the individual RM compounds and the premix (Table 3). Leave out the volume of the template DNA.

2. Transfer suitable aliquots of the master mix to the 96-well microplates. Reactions should be performed in triplicates in order to obtain representative results. All solutions and the microplate should be kept on ice or on a cooled carrier during pipetting.

3. Add the calculated volumes of the desired screening compound(s). For new compounds, an initial evaluation of the tolerated concentration range with GFP as monitor is recommended. The combined volume of the additives must not exceed the free water volume of the reaction.

4. If necessary, fill up with MQ H_2O to the final reaction volume (minus volume of the DNA template).

5. Start the reactions by addition of template DNA.

6. Seal the microplate with Parafilm in order to prevent evaporation.

7. Incubate the microplate with gentle shaking at 30 °C for 2–4 h.

8. Analyze the reactions by suitable techniques (*see* **Note 12**).

3.7 Preparative Scale CF Expression in the CECF Configuration

When larger amounts of protein are required, e.g., if expression rates in the CF batch configuration are too low or if assays are not sensitive enough, the more efficient CECF configuration is recommended for protein expression. The CECF configuration can be operated in analytical scales (50–100 µL RM) or in preparative scales (several mL RM). RM–FM ratios in between 1:14 and 1:20 are recommended (*see* **Note 13**).

1. Calculate the individual compound volumes according to the desired number of reactions.

2. Prepare a common master RFM-mix for the RM and FM (Table 3). Determine the total volume of RM and FM (i.e., 1 mL RM and 17 mL FM, 18 mL in total). Calculate the volume needed for each compound listed in the RFM-mix according to the final concentration and combine them into one reaction tube.

3. Reconstitute RM and FM (Table 3). The RFM-mix is vortexed briefly and an appropriate aliquot is transferred to a new

reaction tube for reconstitution of the RM. The remaining RFM-mix is used for the FM. The RM and FM are then completed with the other compounds listed in Table 3.

4. RM and FM are completed by addition of MQ H_2O. The FM can be vortexed briefly; the RM should only be mixed by inverting or by pipetting up and down.

5. For screening reactions, common RM and FM master mixes may be prepared and then aliquoted into the desired number of reaction containers. For compound screens comprising a series of analytical CECF reactions, master mixtures are prepared with the lowest concentration of the screening compound. The RM and FM master mixtures are then aliquoted according to the number of reactions and adjusted to the desired screening compound concentrations (*see* **Note 14**).

6. Fill the RM and FM aliquots into reaction containers. There are currently no commercial reaction containers available specifically designed for CF expression. However, commercial D-tube dialyzers (Novagen) are well suited for analytical scale and for preparative scale CF reactions. The D-tubes are available in different sizes and with different membrane MWCOs and can be used for analytical and preparative scale CF expressions. The D-tube dialyzer holds the RM and need to be placed in a suitable tube holding the appropriate volume of FM. We, e.g., recommend 2 mL Eppendorf tubes for the small 10–250 μL analytical scale D-tube dialyzer and 15–50 mL Falcon tubes for larger preparative scale D-tube dialyzer (*see* **Note 15**).

We have further described customized containers made out of Plexiglas [11]. Those Mini-CECF-Reactors are designed for RM volumes of 30–100 μL and can be used in combination with standard 24-well microplates with FM volumes of up to 1.5 mL. The Mini-CECF-Reactors hold the RM and are placed in the cavities of a 24-well plate holding appropriate volumes of FM. A piece of dialysis membrane is fixed to the Mini-CECF-Reactors with a Teflon ring (*see* **Note 16**). The dialysis membrane should be replaced for each new reaction, while the Mini-CECF-Reactors are reusable (*see* **Note 17**). For preparative scale CF reactions, commercial Slide-A-Lyzer devices (Pierce) that can hold up to 3 mL RM volumes may be used. We have designed Plexiglas Maxi-CECF-Reactors that perfectly combine with Slide-A-Lyzer devices as FM container (*see* **Note 18**). Exact descriptions of the Maxi-CECF-Reactors as well as of the Mini-CECF-Reactors have been published [11].

Small plastic boxes or beakers may further be used as FM container for Slide-A-Lyzers. Appropriate pieces of dialysis tubes sealed at both ends by knots can also be used as RM container for preparative scale reactions. The tubes can then be placed into suitable plastic vials, e.g., Falcon tubes, holding the FM.

7. CECF reactions are incubated overnight (ON) at 30 °C. Continuous agitation by shaking or rolling depending on the reaction container setup is necessary in order to ensure efficient substance exchange between RM and FM through the membrane. Shaking water bathes or thermo-controlled cabinets with shaking plates at approx. 150–200 rpm may be used.

The CECF configuration gives higher protein production yields if compared with the batch configurations. However, the optimal final concentration of chemical chaperones might differ in between the two reaction configurations.

3.8 Determination of Total Protein Production

Increased activity of target proteins in the presence of additives could be based on stabilization ligands effects, but also the possible general stabilization of the translation machinery resulting into higher translation efficiency should be considered. It is therefore recommended to determine the total protein production by ^{35}S-methionine incorporation in the presence and in the absence of an additive. This will help to distinguish between specific target-stabilizing effects and more general expression-enhancing effects.

1. ^{35}S-methionine (>1,000 TBq/mmol) is added to the non-labeled amino acid mix in a ratio of 1:40,000.

2. After incubation, the reaction mix is collected and transferred into fresh reaction tubes.

3. Ice-cold TCA (15 %, w/v) with a final concentration of 10 % is added. Precipitated protein is pelleted by centrifugation at $22,000 \times g$ for 10 min at 4 °C.

4. The pellet is washed twice with 10 % ice-cold TCA and once with 95 % ethanol.

5. Carefully remove the supernatant and dry the pellet. The reaction tube containing the dried pellet is dropped into a 20 mL scintillation vial filled with 5 mL scintillation cocktail. Turn the vial upside down to allow efficient contact between the scintillation liquid and the pellet.

6. After 1–2 h, the scintillation is counted for 1 min using a liquid scintillation counter.

3.9 Perspectives

The listed chemical chaperones and stabilizing compounds have been tested for their compatibility with CF expression systems based on at least *E. coli* extracts. However, the list cannot be completed and it shows only a selection of compounds with known stabilizing effects on some proteins. The library of protein stabilizer is certainly much more diverse including even other compound classes such as metal ions, salts, or nonvolatile organic compounds. Individual protein targets might further be stabilized by the addition of specific ligands including inhibitors, cofactors, or substrates. In particular for membrane proteins, a number of newly

synthesized hydrophobic compounds suitable for co-translational stabilization are emerging [13, 14, 20, 21]. The compatibility of many promising compounds with CF expression systems still has to be analyzed, and working concentration ranges have to be defined. In addition, some compounds might be better tolerated in mixtures with others and also synergies of beneficial effects could be highly interesting [14, 22]. It can therefore be assumed that the list of useful additives for CF expression reactions is still not exhausted, and new compounds will continuously be added.

4   Notes

1. Commercial or custom-made reaction container as referenced may be used with similar efficiencies.

2. Media other than 2×YPTG may be used as well. Entering the stationary phase of growth should be avoided and it should be considered that cells might continue to grow during the chilling process. A growth curve of the selected strain should first be determined in a pilot experiment and a strategy for the efficient chilling of the broth media should be established.

3. Suspend pellets thoroughly upon washing. The washed cell pellet might be stored at −80 °C for several months.

4. The supernatant of the S30 step is adjusted to a final concentration of 400 mM NaCl. A significant amount of proteins precipitate during the heat step. This precipitation is necessary in order to improve extract efficiency.

5. Aliquots should not be repeatedly refrozen. The final total protein concentration in the S30 extract should be in between 25 and 35 mg/mL.

6. Significant impurities will still be present in the elution fraction. T7RNAP may smear over several elution fractions. Combine only peak fractions.

7. The total protein concentration should finally be 3–4 mg/mL. T7RNAP may start to precipitate at higher concentrations.

8. Premix composition might be variable, e.g., amino acids could be included if their final concentrations are not subject of evaluation. The proposed premix can be refrozen several times.

9. Calculation templates by standard programs such as Excel may be generated.

10. The reaction is optimized for 30 °C incubation. Higher or lower temperatures rapidly result into reduced expression levels. However, folding and quality of synthesized proteins may be modulated.

11. Some compounds might have different effects if provided in combination with others. Besides synergies of two positively

acting chaperones, also shifts in the tolerated concentration ranges might be observed. Some compounds might therefore be better tolerated if provided in combination with others.

12. The supplied chemical chaperones can have a number of different effects on protein expression. Higher activity of the protein of interest could result from improved folding, but also increased expression rates due to stabilization of the translation machinery by the supplied additive could contribute to this observation. Increased fluorescence of GFP fusions could further result from either better folding or stabilization of the complete fusion protein or just of the GFP moiety. For a reliable interpretation, the totally expressed protein in addition to the amount of functionally expressed protein must always be determined [11].

13. Expression efficiency is neither linear with the RM volume nor with the RM–FM ratio. The indicated volumes and ratios are good economical compromises but may be modified if desired.

14. The volume of screening compounds needs initially to be subtracted from the water volume of the RM and FM mixtures.

15. D-tube dialyzer may be reused few times after extensive washing with water and storage in water with 0.1 % NaN_3. The water must be removed completely before filling with the RM.

16. For the assembly of the Mini-CECF-Reactors, the Teflon ring is placed on a sheet of Parafilm, and then a suitable piece of dialysis membrane (2×2 cm) is placed on top of the Teflon ring. Finally the container is pushed through the Teflon ring which tightly fixes the dialysis membrane between the ring and container.

17. The Mini-CECF-Reactor is filled from the top by touching carefully the membrane with the pipette tip and releasing the RM. For harvesting the RM after incubation, the membrane is perforated from the bottom with a pipette tip and the RM is removed. After the reaction, the Mini-CECF-Reactor is disassembled and the membrane is disposed. The container and the Teflon ring are cleaned by extensively washing with MQ H_2O. Prior to the next usage, the container and Teflon ring should be dried thoroughly. Microplates with the Mini-CECF-Reactors should be sealed with Parafilm to prevent evaporation during incubation.

18. Slide-A-Lyzers are filled with a syringe at one of the preformed openings. This opening should be placed upwards if Maxi-CECF-Reactors are used. It must be sealed if another FM container is used. Care must be taken not to damage the membrane upon filling the RM. Slide-A-Lyzers can be reused few times after extensive washing with water and stored in water with 0.1 % NaN_3. We recommend reuse only for the same protein target.

Acknowledgement

References

1. Carrio M, Gonzalez-Montalban N, Vera A et al (2005) Amyloid-like properties of bacterial inclusion bodies. J Mol Biol 347: 1025–1037

2. Bolen DW (2004) Effects of naturally occurring osmolytes on protein stability and solubility: Issues important in protein crystallization. Methods 34:312–322

3. Junge F, Haberstock S, Roos C et al (2011) Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. N Biotechnol 28:262–271

4. Roos C, Kai L, Proverbio D et al (2013) Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers. Mol Membr Biol 30:75–89

5. Patel KG, Ng PP, Levy S et al (2011) Escherichia coli-based production of a tumor idiotype antibody fragment–tetanus toxin fragment c fusion protein vaccine for b cell lymphoma. Protein Expr Purif 75:15–20

6. Oh IS, Lee JC, Lee MS et al (2010) Cell-free production of functional antibody fragments. Bioprocess Biosyst Eng 33:127–132

7. Bundy BC, Franciszkowicz MJ, Swartz JR (2008) Escherichia coli-based cell-free synthesis of virus-like particles. Biotechnol Bioeng 100:28–37

8. Kai L, Dotsch V, Kaldenhoff R et al (2013) Artificial environments for the co-translational stabilization of cell-free expressed proteins. PLoS One 8:e56637

9. Drew D, Slotboom DJ, Friso G et al (2005) A scalable, gfp-based pipeline for membrane protein overexpression screening and purification. Protein Sci 14:2011–2017

10. Drew D, Newstead S, Sonoda Y et al (2008) Gfp-based optimization scheme for the over-expression and purification of eukaryotic membrane proteins in saccharomyces cerevisiae. Nat Protoc 3:784–798

11. Schneider B, Junge F, Shirokov VA et al (2010) Membrane protein expression in cell-free systems. Methods Mol Biol 601:165–186

12. Li Y, Wang E, Wang Y (1999) A modified procedure for fast purification of t7 rna polymerase. Protein Expr Purif 16:355–358

13. Kai L, Roos C, Haberstock S et al (2012) Systems for the cell-free synthesis of proteins. Methods Mol Biol 800:201–225

14. Haberstock S, Roos C, Hoevels Y et al (2012) A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. Protein Expr Purif 82:308–316

15. Drew D, Lerch M, Kunji E et al (2006) Optimization of membrane protein overexpression and purification using gfp fusions. Nat Methods 3:303–313

16. Pedelacq JD, Cabantous S, Tran T et al (2006) Engineering and characterization of a super-folder green fluorescent protein. Nat Biotechnol 24:79–88

17. Lyukmanova EN, Shenkarev ZO, Khabibullina NF et al (2012) Lipid-protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: Comparison with detergent micelles, bicelles and liposomes. Biochim Biophys Acta 1818:349–358

18. Blesneac I, Ravaud S, Juillan-Binard C et al (2012) Production of ucp1 a membrane protein from the inner mitochondrial membrane using the cell free expression system in the presence of a fluorinated surfactant. Biochim Biophys Acta 1818:798–805

19. Klammt C, Schwarz D, Fendler K et al (2005) Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. FEBS J 272:6024–6038

20. Shimono K, Goto M, Kikukawa T et al (2009) Production of functional bacteriorhodopsin by an escherichia coli cell-free protein synthesis system supplemented with steroid detergent and lipid. Protein Sci 18:2160–2171

21. Genji T, Nozawa A, Tozawa Y (2010) Efficient production and purification of functional

bacteriorhodopsin with a wheat-germ cell-free system and a combination of fos-choline and chaps detergents. Biochem Biophys Res Commun 400:638–642

22. Klammt C, Perrin MH, Maslennikov I et al (2011) Polymer-based cell-free expression of ligand-binding family b g-protein coupled receptors without detergents. Protein Sci 20: 1030–1041

23. Ishihara G, Goto M, Saeki M et al (2005) Expression of g protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. Protein Expr Purif 41:27–37

Part III

Recombinant Protein Production in Lactic Acid Bacteria (LAB)

Chapter 8

Functional Expression of Plant Membrane Proteins in *Lactococcus lactis*

Sylvain Boutigny, Emeline Sautron, Annie Frelet-Barrand, Lucas Moyet, Daniel Salvi, Norbert Rolland, and Daphné Seigneurin-Berny

Abstract

The study of most membrane proteins remains challenging due to their hydrophobicity and their low natural abundance in cells. *Lactococcus lactis*, a Gram-positive lactic bacterium, has been traditionally used in food fermentations and is nowadays widely used in biotechnology for large-scale production of heterologous proteins. This system has been successfully used for the production of prokaryotic and eukaryotic membrane proteins. The purpose of this chapter is to provide detailed protocols for (1) the expression of plant peripheral or intrinsic membrane proteins and then for (2) their solubilization, from *Lactococcus* membranes, for further purification steps and biochemical characterization.

Key words *Lactococcus*, Nisin, Plant membrane proteins, Expression, Solubilization

1 Introduction

Cell proteins can be divided in soluble proteins that are present in aqueous compartments (e.g., cytosol, mitochondrial matrix) and proteins that are embedded in biological membranes (membrane proteins). In both prokaryotic and eukaryotic genomes, 20–30 % of genes code for membrane proteins [1–3]. Membrane proteins have been implicated in many cellular functions. However, despite their functional importance in key processes, the vast majority of them still have no assigned function. Membrane proteins are either peripherally associated with membranes or inserted into the membrane by one or several transmembrane domains. Transmembrane proteins are insoluble in aqueous phases and require detergents to be extracted from their biological membranes. Generally, the low abundance of these proteins in native membranes and their hydrophobic nature are the major bottlenecks for their studies. One approach to overcome this difficulty consists in overexpressing these proteins in heterologous systems for further biochemical

studies. However, their hydrophobicity also strongly limits overexpression of functional proteins even in heterologous host and can also lead to difficulties during their solubilization, purification, functional characterization, and crystallization. Several heterologous prokaryotic and eukaryotic expression systems can be tested to produce these difficult proteins, and the challenge is thus to identify the most suitable expression system for the studied membrane protein [4, 5]. Among prokaryotic systems, the Gram-negative bacterium *Escherichia coli* usually provides an optimal environment for expression of soluble proteins and prokaryotic membrane proteins. However, its capability to host eukaryotic membrane proteins is limited. Indeed, in most cases, expression of such proteins is associated with low expression level, toxicity, and formation of inclusion bodies. The prokaryotic Gram-positive bacterium *Lactococcus lactis* is nowadays widely used for large-scale production of homologous or heterologous proteins. It was shown to be an attractive system for efficient and functional production of eukaryotic membrane proteins (for reviews *see* 6–8). Heterologous expression in *Lactococcus* can be performed using the nisin-controlled gene expression (NICE) system, in which nisin, an antimicrobial peptide, is used to promote the expression of genes cloned under the control of the nisin-inducible promoter *PnisA*. This system is well suited to the expression of integral membrane proteins since (1) *L. lactis* has a single membrane and relatively low proteolytic activity and since (2) the PnisA promoter is a tightly regulated promoter and thus allows controlling the expression level of the recombinant protein according to the inducer concentration [8]. Previous studies have shown that the expression level can differ significantly among eukaryotic membrane proteins, but this expression system is generally adequate to purify enough recombinant protein for further biochemical and biophysical studies.

In this chapter, we describe the protocol used for the expression of plant peripheral or intrinsic membrane proteins that were successfully expressed in an active form using this expression system [5, 9, 10]. In a previous chapter of the book entitled *Heterologous Expression of Membrane Proteins*, the methodologies required for (1) the subcloning of genes into NICE expression vectors, (2) the transformation of *L. lactis*, and (3) the expression of recombinant proteins in this bacterium have already been detailed [11]. In this new chapter, we provide more details on the expression conditions that can be tested to improve the soluble production level of recombinant membrane proteins. Since membrane proteins are produced in the bacterial membrane, we also provide protocols for the solubilization of both peripheral and intrinsic membrane proteins.

2 Materials

2.1 Cloning Strategies

Refer to the previous described chapter in the book entitled *Heterologous Expression of Membrane Proteins* [11] (*see* **Notes 1** and **2**).

1. High-Fidelity Taq DNA Polymerase which is compatible with high GC-containing primers.

2. Purification Kit for extraction of DNA fragments from agarose gels and Plasmid DNA Purification Kit.

3. Restriction endonucleases, T_4 DNA ligase, and provided T_4 DNA ligase buffer 10×.

4. Agarose gels.

5. pNZ8148 plasmid (*see* **Note 3**).

2.2 Transformation of Bacteria

1. *Lactococcus lactis* NZ9000 strain.

2. M17 medium: M17B (broth) or M17A (agar) (*see* **Note 4**).

3. Glucose: Prepare 20 % (w/v) solution in MQ H_2O, and sterilize by filtration through a 0.2 μm membrane under hood bench. Store at 4 °C.

4. Chloramphenicol: 34 mg/mL in absolute ethanol. Store at −20 °C.

5. M17BG [M17 broth, 0.5 % (w/v) glucose] medium or M17BGChl [M17BG, 10 μg/mL chloramphenicol] medium.

6. G-SGM17B medium: M17 Broth, 0.5 M sucrose, 0.33 M glycine. Sterilize the solution by autoclaving. Add glucose 20 % (w/v) to a final concentration of 0.5 %.

7. Medium A: 0.5 M sucrose, glycerol 10 % (v/v) in MQ H_2O. Sterilize the solution by autoclaving.

8. Medium B: 0.5 M sucrose, glycerol 10 % (v/v) and 0.05 M EDTA, pH 8.0, in MQ H_2O. Sterilize the solution by autoclaving.

9. M17AGChl medium: M17 agar, 0.5 % (w/v) glucose, 10 μg/mL chloramphenicol.

10. M17G1%Chl medium: M17 Broth, 1 % (w/v) glucose, 10 μg/mL chloramphenicol.

11. Glycerol RPE.

12. Electroporation cuvettes.

2.3 Expression of Recombinant Proteins and Optimization of Expression Conditions

1. *Lactococcus lactis* NZ9700 strain.

2. Nisin.

3. M17G1% medium (*see* **Note 5**): M17 Broth, 1 % (w/v) glucose.

4. M17G1%Chl medium: M17 Broth, 1 % (w/v) glucose, 10 μg/mL chloramphenicol.

5. Hepes/glycerol solution: 20 mM Hepes, pH 6.0, 10 % (v/v) glycerol.

6. Laboratory glassware bottles (1 L Schott bottles) and Falcon tubes (15 mL and 50 mL).

7. Incubator for cell growth.

8. Lysozyme.

9. DNase I.

10. 20 mM Hepes, pH 6.0.

2.4 Purification of Lactococcus Membranes and Detection of the Recombinant Protein Produced

1. Sonicator.

2. Cell disruption system: one-shot (*see* **Note 6**).

3. Centrifuge and ultracentrifuge.

2.4.1 Preparation of Lactococcus Membranes Using a Cell Disruptor

2.4.2 SDS-PAGE, Transfer and Western Blotting

1. Gel electrophoresis apparatus with the various accessories needed for protein separation by electrophoresis (combs, plates, and casting apparatus).

2. 4× Laemmli stacking gel buffer: 0.5 M Tris–HCl, pH 6.8. Store at 4 °C.

3. 8× Laemmli resolving gel buffer: 3 M Tris–HCl, pH 8.8. Store at 4 °C.

4. Single 12 % acrylamide resolving gels (*see* **Note 7**): 4 mL of acrylamide-bis 30 % solution, 1.25 mL of 8× Laemmli resolving gel buffer, 4.6 mL of H_2O, 50 μL of 20 % (w/v) SDS, 4 μL of TEMED, and 0.1 μL of 10 % (w/v) ammonium persulfate. In each case, the total volume should be ~10 mL (sufficient for two 7-cm-long gels).

5. Stacking 5 % acrylamide gel: 0.83 mL of acrylamide-bis 30 % solution, 1.25 mL of 4× Laemmli stacking gel buffer, 2.8 mL of H_2O, 25 μL of 20 % (w/v) SDS, 5 μL of TEMED, and 50 μL of 10 % (w/v) ammonium persulfate. The total volume will be 4.96 mL (sufficient for two 7-cm-long gels).

6. 4× loading buffer for protein solubilization: 200 mM Tris–HCl, pH 6.8, 40 % (w/v) glycerol, 4 % (w/v) SDS, 0.4 % (w/v) bromophenol blue, and 100 mM dithiothreitol.

7. 10× Laemmli running buffer: 192 mM glycine, 25 mM Tris, 0.1 % SDS. Store at room temperature (RT).

8. Gel-staining medium: acetic acid/isopropanol/water, 10/25/65 (v/v/v), supplemented with 2.5 g/L of Coomassie Brilliant Blue R250. Store in clean and closed bottles.

9. Gel-destaining medium: 30 % (v/v) ethanol.

10. System for protein transfer to nitrocellulose membranes (central core assembly, holder cassette, nitrocellulose filter paper, fibber pads, and cooling unit).

11. Protein transfer buffer: 30.4 mM glycine, 40 mM Tris, 0.08 % (w/v) SDS, 20 % (v/v) ethanol. Dilute gel reservoir buffer with ethanol to obtain 20 % (v/v) final ethanol concentration (prepare about 800 mL).

12. 20× Tris-buffered saline (TBS): 1 M Tris–HCl, pH 7.4, 3 M NaCl. Store at 4 ° C.

13. TBS-T: 1× TBS, 0.05 % (v/v) Triton X-100. Store at RT.

14. Blocking buffer: 3 % (w/v) Bovine Serum Albumin Fraction V in TBS-T.

15. *Strep*-Tactin HRP conjugate.

16. 100 mM Tris–HCl, pH 8.0.

17. Solution A: 90 mM p-Coumaric acid in DMSO. Store at 4 °C and protect from light.

18. Hydrogen peroxide (H_2O_2) 30 % (v/v).

19. Solution B: 250 mM luminol (3-aminophalhydrazin) in DMSO. Store at 4 °C and protect from light.

20. Developer and fixer solutions.

21. Chemiluminescence-adapted films.

2.5 Solubilization of Lactococcus Membrane Containing the Recombinant Protein

2.5.1 Solubilization of Peripheral Membrane Proteins

1. 50 mM MOPS, pH 7.8, containing 0.5 M or 1 M NaCl.

2. 50 mM MOPS, pH 7.8, containing 0.5 M KI or 0.1 M Na_2CO_3, pH 11 or 0.1 M NaOH.

3. 50 mM Tris–HCl, pH 7.5, containing 0.1 % or 0.5 % (v/v) Triton X-100 (*see* **Note 8**).

4. 50 mM Tris–HCl, pH 7.5, containing 0.1 % or 0.5 % (w/v) n-dodecyl-β-maltoside (DDM).

5. 10 mM Tris–HCl, pH 6.8.

6. 10 mM Tris–HCl, pH 8.0.

7. PD10 column.

8. Ultracentrifuge.

2.5.2 Solubilization of Intrinsic Membrane Proteins	1. Solubilization buffer: 50 mM Tris–HCl, pH 8.0, 100 mM NaCl.
	2. Solubilization buffer containing 20 mM DDM, 6 mM dodecyl octaethylene glycol monoether ($C_{12}E_8$).
	3. Tris(2-carboxyethyl)phosphine (TCEP).
	4. Stirring wheel.
	5. Tipped sonicator.
	6. High-speed centrifuge and Eppendorf centrifuge.

3 Methods

3.1 Cloning Strategies

The pNZ8148 vector is used for expression in *Lactococcus*. This vector contains an origin of replication (*ORI*), the gene for the resistance to chloramphenicol, two genes for the replication proteins *repA* and *repC*, the nisin-inducible promoter (*PnisA*), and the transcription terminator (*T*). Methods for vector and insert preparations, ligation, and selection of positive recombinant clone have been already well described in a previous chapter of the book entitled *Heterologous Expression of Membrane Proteins* [11].

3.2 Transformation of Bacteria

3.2.1 Preparation of Electrocompetent Cells

1. Inoculate 5 mL of G-SGM17B medium (in 15 mL Falcon tubes) with a glycerol stock at 30 °C without shaking.
2. 24 h later, inoculate 50 mL of G-SGM17B medium with the 5 mL preculture overnight (ON) at 30 °C without shaking.
3. The next morning, inoculate 400 mL of G-SGM17B medium with the 50 mL preculture.
4. Grow until $OD_{600\ nm}$ reaches 0.2–0.3 (approximately 3 h).
5. Centrifuge for 20 min at $5,000 \times g$, 4 °C and keep the pellet.
6. Wash the bacteria with 400 mL of medium A.
7. Centrifuge for 20 min at $5,000 \times g$, 4 °C and keep the pellet.
8. Keep on ice for 15 min in 200 mL of medium B.
9. Centrifuge for 20 min at $5,000 \times g$, 4 °C and keep the pellet.
10. Wash the pellet with 100 mL of medium A.
11. Centrifuge for 20 min at $5,000 \times g$, 4 °C and keep the pellet.
12. Resuspend the bacteria in 4 mL of medium A.
13. Keep the bacteria into small aliquots (40 μL) at –80 °C.

3.2.2 Electroporation

1. Add 1 μL of recombinant plasmid DNA to 40 μL of electrocompetent cells.
2. Store on ice for 1 min and transfer to the electroporation cuvette.
3. Use the following parameters: 2,000 V, –25 μF, –200 Ω. Press on pulse. The time should be between 4.5 and 5 ms.
4. Add 1 mL of M17BG medium.

5. Store the cuvette on ice for 10 min and incubate at 30 °C for 3 h.

6. Spread bacteria on two independent M17AGChl Petri dishes (1/10th and 9/10th of the volume).

7. Incubate for 1–2 days at 30 °C.

3.2.3 Preparation of a Glycerol Stock

1. Inoculate 5 mL of M17GChl medium (in 15 mL Falcon tube) with an independent recombinant clone and incubate ON at 30 °C without shaking.

2. For a normal glycerol stock, add 850 μL of the small 5 mL culture to 150 μL of sterile 100 % glycerol and store at –80 °C.

3. For a concentrated glycerol stock, prepare 25 mL of M17G1% Chl medium and scrape the normal glycerol stock with tip.

4. Incubate ON at 30 °C (90 rpm).

5. Pellet the bacteria by centrifugation at $4,350 \times g$ for 15 min at 4 °C and add sterile 100 % glycerol in M17B medium to a final concentration of 15 % (v/v).

3.3 Expression of Recombinant Proteins and Optimization of Expression Conditions

1. Inoculate 10 mL of M17G1% with a concentrated glycerol stock of the *Lactococcus* NZ9700 strain.

2. Incubate at 30 °C with gentle agitation (around 90 rpm) for 6 h.

3. Inoculate 500 mL of M17G1% in a Schott bottle with the 10 mL preculture.

3.3.1 Preparation of "Homemade" Nisin

4. Incubate at 30 °C with gentle agitation (90 rpm) for 24 h.

5. Centrifuge the bacteria culture at $6,300 \times g$ for 5 min at 4 °C and transfer the supernatant into 15 mL Falcon tubes (*see* **Note 9**). Store the tubes at –80 °C until use.

3.3.2 Preparation of Commercial Nisin

1. Commercial nisin is provided at a 2.5 % (w/v) concentration (*see* **Note 10**). Prepare a 40 mg/mL solution in 0.05 % (v/v) acetic acid to obtain a final concentration of 1 mg/mL pure nisin (e.g., add 500 μL of 0.05 % acetic acid to 20 μg of nisin).

2. Let it dissolve for 10 min at RT.

3. Centrifuge 1 min at $11,000 \times g$ to pellet insoluble material and keep the supernatant.

4. Aliquot the 1 mg/mL nisin solution and store the aliquots at –20 °C (*see* **Note 11**).

5. Prepare nisin dilution from the 1 mg/mL nisin stock in sterile water just before use. A diluted nisin solution is not stable.

3.3.3 Standard Protocol for Expression

1. Inoculate 25 mL of M17G1%Chl in a 50 mL Falcon tube with a 300 μL concentrated glycerol stock of recombinant bacteria (*see* **Note 12**).

2. Incubate ON at 30 °C with gentle agitation (90 rpm).

3. Inoculate 1 L of M17G1%Chl in a Schott bottle with the 25 mL preculture (*see* **Note 13**).

4. Incubate at 30 °C with gentle agitation until $OD_{600\,nm}$ reaches 0.8.

5. Sample 5 mL of culture for further analysis of the recombinant protein expression directly on a crude extract (*see* Subheading 3.3.5).

6. Induce recombinant protein expression by the addition of either 5 mL of extracted nisin from the NZ9700 strain (*see* **Note 14**) or 1 μg of commercially available nisin (10 μL of a 0.1 μg/μL nisin dilution for 1 L of *Lactococcus* culture) (*see* **Note 15**).

7. Incubate at 30 °C with gentle agitation for an additional 4 h (*see* **Note 16**). Measure the $OD_{600\,nm}$ and sample 5 mL of the culture every hour.

8. Harvest the bacteria in buckets and centrifuge at $5,000 \times g$ for 20 min at 4 °C.

9. Discard the supernatant and resuspend the pellet in 40 mL of Hepes/glycerol solution.

10. Transfer the suspension in a 50 mL Falcon tube and store at –80 °C.

3.3.4 Optimization of Expression Conditions

Several parameters can be modified in order to improve the production of the recombinant protein and have to be tested for each protein. These parameters are the following: (1) impact of growth time after induction, (2) impact of cell concentration ($OD_{600\,nm}$) when adding the inducer, and (3) impact of the nisin concentration (homemade or commercial nisin) (*see* **Note 17**).

Impact of Growth Time After Induction

1. Perform **steps 1–6** as described in the Subheading 3.3.3.

2. After addition of nisin, separate the 1 L culture into four 250 mL cultures in 250 mL Schott bottles.

3. Incubate at 30 °C with gentle agitation for different time for each bottle; for example, 2 h for bottle A, 3 h for bottle B, 5 h for bottle C, and 7 h for bottle D.

4. At the end of the determined induction time, sample 5 mL of the culture and harvest the remaining 245 mL culture in buckets.

5. Centrifuge the remaining 245 mL at $5,000 \times g$ for 20 min at 4 °C. Discard the supernatant and resuspend the pellet in 10 mL of Hepes/glycerol solution. Transfer the suspension in a 50 mL Falcon tube and store the pellet at –80 °C.

6. For the 5 mL sample, proceed as described in Subheading 3.3.5.

Impact of Nisin Concentration (Fig. 1)

1. Perform **steps 1–5** as described in the Subheading 3.3.3.

2. Separate the 1 L culture into four 250 mL cultures in 250 mL Schott bottles.

3. For each bottle, add various amounts of nisin, for example, the equivalent of 0.5, 5, 100, and 200 μg/L culture (that means respectively 0.125, 1.25, 25, and 50 μg of nisin for each 250 mL culture) (*see* **Note 18**).

4. Incubate at 30 °C with gentle agitation for 4 h (*see* **Note 19**).

5. Sample 5 mL of each culture and harvest the remaining 245 mL culture in buckets.

6. Centrifuge the 245 mL culture at $5,000 \times g$ for 20 min at 4 °C. Discard the supernatant and resuspend the pellet in 10 mL of Hepes/glycerol solution. Transfer the suspension in a 50 mL Falcon tube and store the pellet at –80 °C.

7. For the 5 mL sample (*see* Subheading 3.3.5).

Impact of Cell Concentration

1. Perform **steps 1–3** as described in the Subheading 3.3.3.

2. Separate the 1 L culture into four 250 mL cultures.

3. Check the $OD_{600 \text{ nm}}$ and add nisin (0.25 μg for each 250 mL culture, *see* **Note 20**), for example, when $OD_{600 \text{ nm}}$ reaches 0.5 for bottle A, 0.8 for bottle B, 2 for bottle C, and 5 for bottle D.

4. Incubate each culture for an additional 4 h after the addition of nisin, at 30 °C, with gentle agitation.

5. Sample 5 mL of each culture and harvest the remaining 245 mL culture in buckets.
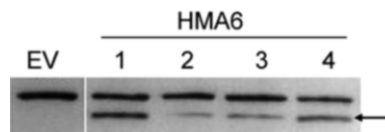
Fig. 1 Impact of nisin concentration on the expression of an intrinsic plant membrane protein. The production of the recombinant protein was induced at $OD_{600 \text{ nm}} = 0.8$ with various amounts of nisin. After 4 h of induction, the cells were harvested and total membrane proteins extracted. The level of the recombinant protein HMA6 was analyzed by SDS-PAGE, and Western blot was performed using an HRP conjugate specific to the *Strep*-tag II. *Arrow* indicates the position of the expressed protein. EV: crude membrane proteins derived from bacteria containing the empty pNZ8148 vector (induction was performed with 20 μg/L of commercially available nisin). (1) Induction with 5 mL/L of extracted nisin from the NZ9700 strain, (2) induction with 1 μg/L of commercially available nisin, (3) induction with 20 μg/L of commercially available nisin, (4) induction with 100 μg/L of commercially available nisin

6. Centrifuge the 245 mL culture at $5,000 \times g$ for 20 min at 4 °C. Discard the supernatant and resuspend the pellet in 10 mL of Hepes/glycerol solution. Transfer the suspension in a 50 mL Falcon tube and store the pellet at –80 °C.

7. For the 5 mL sample (*see* Subheading 3.3.5).

3.3.5 Preparation of Crude Extract to Analyze the Amount of Recombinant Protein

Before performing the purification of *Lactococcus* membranes, the expression of the recombinant protein can be checked on a crude extract if its expression level is sufficient. Crude extracts are obtained from the 5 mL culture samples harvested during expression experiments (*see* **Note 21**).

1. Centrifuge each 5 mL culture at $4,000 \times g$ for 15 min at 4 °C.

2. Discard the supernatant and resuspend the pellet with 100 μL of 20 mM Hepes pH 6.0.

3. Add 10 μL of 10 mg/mL lysozyme and incubate at 37 °C for 20 min.

4. Sonicate the suspension for 3 min (duty cycle 40 %, output control $n = 5$).

5. Add 1 μL of DNAse I and incubate at 37 °C for 20 min.

6. Add SDS 20 % to obtain a final concentration of 6 % (v/v) and incubate for 10 min at RT.

7. Centrifuge at $13,000 \times g$ for 30 s.

8. Keep the supernatant for further SDS-PAGE and Western blot analyses (Fig. 2).
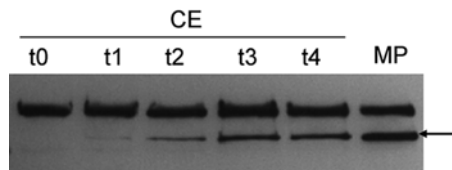
Fig. 2 Impact of growth time after induction on the expression of an intrinsic plant membrane protein. The production of the recombinant protein was induced at $OD_{600\ nm} = 0.8$, and 5 mL of the culture was collected every hour (t0, t1, t2, t3, t4) for preparation of crude extracts. After 4 h of induction, the cells were harvested and total membrane proteins extracted. The level of the recombinant protein HMA6 was analyzed by SDS-PAGE and Western blot performed using an HRP conjugate specific to the *Strep*-tag II. *Arrow* indicates the position of the expressed protein. t0: sample collected just before the addition of nisin. t1, t2, t3, and t4: samples collected after 1, 2, 3, and 4 h of induction. CE: crude extract. MP: membrane proteins extract. 18 μL of each crude extract and 20 μg of membrane proteins were loaded on gel

3.4   Purification of Lactococcus Membranes and Detection of the Recombinant Protein Produced

3.4.1   Preparation of Lactococcus Membranes Using Cell Disruptor

1. Take out the bacterial pellet (*see* Subheadings 3.3.3 or 3.3.4) and let it thaw on ice.

2. Sonicate the suspension for 3 min (duty cycle 40 %, output control *n* = 5).

3. Disrupt cells at 35,000 psi (2.3 kbars) and keep the lysate on ice (*see* **Note 22**).

4. Centrifuge the lysate at $10,000 \times g$ (rotor SS34, Sorvall) for 10 min at 4 °C and transfer the supernatant to ultracentrifuge tubes (*see* **Note 23**).

5. Centrifuge at $150,000 \times g$ (rotor Ti45, Beckman) for 1 h at 4 °C and discard the supernatant.

6. Resuspend the pellet with 1.5 mL of Hepes/glycerol buffer (*see* **Note 24**).

7. Take an aliquot of the membrane protein suspension to quantify the protein concentration and for further analysis by SDS-PAGE and Western blot.

8. Aliquot the remaining membrane proteins in small volumes and store them at –80 °C after freezing in liquid nitrogen.

3.4.2   Analysis of the Expression of Recombinant Protein by SDS-PAGE and Western Blotting

1. Prepare protein samples in loading buffer 1× to have 20 μg of proteins for each sample per lane (*see* **Note 25**) and load the samples on an SDS-polyacrylamide gel (*see* **Note 26**). Load the molecular weight markers in another slot.

2. After electrophoresis, remove the gels from the apparatus; place them in plastic boxes in the presence of gel-staining medium if Western blot is not performed. Shake the box gently for 30 min. Pour off the staining solution and replace it with the gel-destaining medium. Shake the box gently for 15 min. Repeat this step once or twice. If Western blotting is to be conducted, the Coomassie staining step should be omitted and proceed directly as described in Subheading 3.4.2, **step 3** for gel transfer and Western blotting.

3. Western blots should be performed after the separation of proteins by SDS-PAGE to specifically detect a specific protein (here the recombinant protein produced). After gel migration, transfer the gel in plastic boxes containing protein transfer medium and proceed to the transfer of proteins onto nitrocellulose membrane (*see* **Note 26**).

4. After transfer, recover the nitrocellulose membrane. The following incubation and washing steps require agitation on a rocking plate at RT.

5. Rinse the membrane twice with water and then wash the membrane twice for 5 min with TBS-T.

6. Saturate the membrane with BSA-containing TBS-T for 1 h (*see* **Note 27**).

7. Wash the membrane three times for 5 min with TBS-T.

8. Incubate with the *Strep*-Tactin conjugate coupled to HRP diluted at 1/10,000 in TBS-T for 1 h (*see* **Note 28**).

9. Wash the membrane twice for 5 min with TBS-T and then twice with TBS.

10. Prepare 9 mL of 100 mM Tris–HCl, pH 8.0, with 40 μL of solution A and 5 μL of H_2O_2.

11. Prepare 9 mL of 100 mM Tris–HCl, pH 8.0, with 90 μL of solution B.

12. Mix the two above solutions (**steps 10** and **11**) in a dark room and incubate the nitrocellulose membrane for 1 min in this mixture (the chemiluminescence substrate solution).

13. Expose to film for a few seconds and up to several minutes depending on the detected signal.

14. Incubate the film successively in the developer solution (for 1–3 min, depending on the signal to noise ratio), in water (for 10 s), and in the fixer solution (for 2 min). Rinse the film in water and dry it. Typical results are shown in Figs. 1 and 2.

3.5 Solubilization of Lactococcus Membrane Containing the Recombinant Protein

3.5.1 Solubilization of Peripheral Membrane Proteins

According to the nature of the membrane protein (more or less hydrophobic), several solubilization treatments can be performed: saline and alkaline treatments, as well as the use of detergents (*see* **Note 29**). In the following part, we describe three conditions that have been tested for the solubilization of a plant peripheral protein, the ceQORH protein. All the solubilizations are performed with a membrane protein concentration of 1 mg/mL and at 4 °C.

1. Salt treatment: Incubate membrane proteins in 50 mM MOPS, pH 7.8, containing 0.5 or 1 M NaCl or 0.5 M KI for 45 min, at 4 °C. Mix gently the sample every 15 min.

2. Alkaline treatments: Incubate membrane proteins directly in 0.1 M Na_2CO_3, pH 11, or 0.1 M NaOH for 45 min, at 4 °C. Mix the sample every 15 min.

3. Detergents treatments: Incubate membrane proteins in 50 mM Tris–HCl, pH 7.5, containing 0.1 % (v/v) or 0.5 % (v/v) of either Triton X-100 or DDM for 45 min, at 4 °C. Mix the sample every 15 min.

4. For each treatment, centrifuge membranes at $160,000 \times g$, for 1 h, at 4 °C to separate solubilized proteins (in the supernatant) from insoluble membrane proteins (in the pellet). Keep the supernatant for further purification steps (*see* **Note 30**).

5. Resuspend the pellet in the same volume of 10 mM Tris–HCl, pH 6.8, for SDS-PAGE and Western blotting analysis.
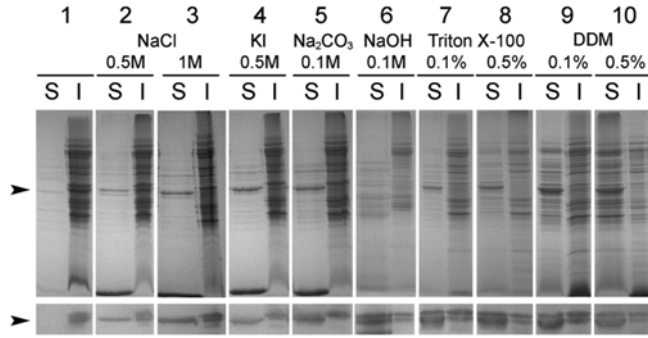
Fig. 3 Impact of salt, pH, and detergents on the solubilization of a peripheral plant membrane protein produced in *Lactococcus lactis*. *Lactococcus* membrane proteins containing the ceQORH protein were incubated in a buffer containing various concentrations of salt (NaCl or KCl), detergent (Triton X-100 or DDM), NaOH, or Na_2CO_3. Solubilized proteins (S) were separated from insoluble membrane proteins (I) by centrifugation. Proteins were analyzed by Coomassie blue-stained SDS-PAGE (*upper panel*) and by Western blot (*lower panel*)

6. Analyze the resulting fractions on SDS-PAGE and by Western blot analysis (*see* Subheading 3.4.2) to determine the optimal condition for solubilization (Fig. 3).

7. For further affinity purification using the *Strep*-Tactin Sepharose matrix (IBA, Goettingen, Germany), desalt the solubilized membrane proteins on a PD10 column (Sephadex G-25 M, GE Healthcare) in 10 mM Tris–HCl, pH 8.0 (*see* **Note 31**).

3.5.2 Solubilization of Intrinsic Membrane Proteins

Intrinsic membrane proteins cannot be solubilized with mild treatments as described above, and their solubilization requires the presence of detergent. We describe here the protocol used for the solubilization of the plant ATPases HMA1 and HMA6 which are intrinsic proteins with 6–8 predicted transmembrane domains. We used a combination of two detergents (DDM and $C_{12}E_8$) that have been already used for the solubilization of such proteins and were shown to preserve activity of the proteins.

1. Incubate the membrane proteins in the solubilization buffer to have a final concentration of 4 mg/mL, for 30 min at 4 °C.

2. Centrifuge at $160,000 \times g$ for 80 min at 4 °C to eliminate soluble proteins.

3. Resuspend the pellet in the solubilization buffer containing 20 mM DDM, 6 mM $C_{12}E_8$, and 100 μM TCEP (*see* **Note 32**).

4. Sonicate the suspension using a tipped sonicator for 3 min on ice (duty cycle 10 %, output control $n = 15$).

5. Incubate the membrane suspension for 1.5 h at 4 °C, under gentle agitation. After incubation, sonicate the suspension once more with the same settings.

6. Centrifuge the suspension at $15,000 \times g$, for 20 min, at 4 °C. The supernatant contains the solubilized membrane proteins and can be used for further affinity purification steps (not described here). Insoluble proteins present in the pellet are resuspended in an equal volume of 50 mM Tris–HCl, pH 8.0, 100 mM NaCl.

7. Analyze the resulting fractions by Western blotting to validate the solubilization of the membrane protein before performing the purification (*see* Subheading 3.4.2 and Fig. 4).

8. Before purification on a *Strep*-Tactin Sepharose matrix, the solubilized membrane proteins are diluted ten times since a high concentration of detergent can impair the interaction between the affinity matrix and the tagged protein (*see* **Note 33**).
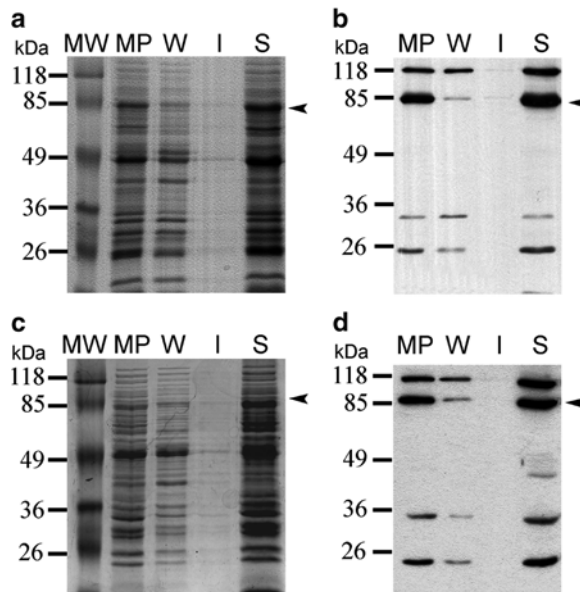
Fig. 4 Solubilization of two intrinsic plant membrane proteins using detergents. *Lactococcus* membrane proteins (MP) containing the transmembrane proteins HMA1 or HMA6 were incubated in 50 mM Tris–HCl, pH 8.0, 100 mM NaCl and subsequently centrifuged to eliminate soluble proteins (W, washing). Membrane pellets were solubilized in the same buffer containing 1 % (w/v) DDM, 0.32 % (w/v) $C_{12}E_8$, and 100 μM TCEP. After incubation for 1.5 h, solubilized membrane proteins (S) were separated from insoluble proteins (I) by centrifugation. Aliquots (15 μg of crude MP and 10 μL of resulting fractions W, I, and S) were loaded on a 10 % SDS-PAGE further stained with Coomassie blue (panels **a**, **c**) and by Western blot (panels **b**, **d**) using the HRP conjugate specific to the *Strep*-tag II

4   Notes

1. All the materials required for the cloning strategies have been already well described in a previous chapter of the book entitled *Heterologous Expression of Membrane Proteins* [11].

2. Other cloning strategies have been described in the literature; see, for example, [12–14].

3. This plasmid carries the chloramphenicol acetyltransferase gene, the PnisA promoter followed by an *Nco*I site for translational fusions at the ATG. It contains a terminator after the MCS.

4. The M17 medium has been adapted for *Lactic Streptococci* [15]. The most commonly used laboratory medium for *Lactococcus* growth is the M17 supplemented with a carbon source such as glucose, lactose, or other sugars and a relevant antibiotic for plasmid selection.

5. Lactose can be used instead of glucose; however, the growth rate of *Lactococcus* is higher in the presence of glucose [16].

6. One-shot disruption system is the most suitable system to disrupt the *Lactococcus* cell wall. It avoids the use of lysozyme which is then recovered in *Lactococcus* membrane preparation [9]. This system also improves the yield of crude membrane preparation compared to the one obtained by lysozyme and French press treatment.

7. According to the apparent molecular weight of the recombinant protein, other concentrations of acrylamide can be used (e.g., higher concentrations for the separation of smaller proteins).

8. Other detergent concentrations can be tested.

9. Aliquot the supernatant in small volumes to avoid freezing thawing of nisin.

10. Nisin can be purchased from Sigma or MoBiTec at a 2.5 % (w/v) concentration.

11. Frozen aliquots are stable for at least 1 year.

12. In some cases, it could be useful to inoculate a culture with a concentrated stock of bacteria containing a nonrecombinant pNZ8148 vector as a negative control.

13. *Lactococcus* is able to grow in anaerobic conditions and thus Schott bottles are suitable for this use and can be filled to the top.

14. This volume has to be optimized for each new protein expressed and also has to be determined for each new preparation of nisin [17].

15. For the induction, concentrations of 0.5–5 ng/mL nisin are often used (typically 1 ng/mL is used). However, other concentrations can be tested (*see*, e.g., 16, 17).

16. Shorter or longer induction time can be applied; this must be optimized for each protein (*see* Subheading 3.3.4).

17. Here, we only describe the optimization of three parameters, but others can also be tested like the pH of the medium or the growth temperature [16, 17].

18. Other concentrations can be tested. However, since nisin is toxic for *Lactococcus lactis*, a high amount of nisin can lead to cell death. This can be followed by monitoring the growth at $OD_{600\ nm}$ during the induction time. For several intrinsic proteins, we have noticed that the best expression is achieved when the growth of the bacteria is arrested (i.e., the $OD_{660\ nm}$ remains constant).

19. Other induction times can be tested.

20. Other amounts of nisin can be used.

21. In the literature, other protocols can be found for the preparation of crude extract [16, 18–20]. Lysozyme is an enzyme that digests the peptidoglycan in cell walls of Gram-positive bacteria. Recombinant membrane proteins with a production yield around 1–3 % of total membrane proteins can be easily detected in a crude extract using Western blotting.

22. In a previous chapter [11], we describe an alternative protocol with the use of lysozyme and French press to obtain membrane extracts.

23. This centrifugation step allows the removal of remaining intact bacteria and cell wall components.

24. Add first 1 mL of Hepes/glycerol buffer to resuspend the pellet and transfer the suspension in a new 1.5 mL-tube. Add then 500 μL more Hepes/glycerol buffer to wash the centrifugation tube and pool this suspension to the first 1 mL. Use a grinder (hand homogenizer like a Potter-Elvehjem Tissue Grinder) to homogenize the membrane suspension. The volume of added Hepes/glycerol buffer can be adapted depending on the size of the pellet.

25. We usually quantify protein amounts using the Bio-Rad protein assay reagent [21]. Crude protein extracts are loaded directly without quantification. SDS-PAGE analyses are performed as described by Chua [22]. We usually heat the samples for 2 min at 95 °C prior to loading on gel. This step should be avoided when using samples from crude extracts of *Lactococcus* (*see* Subheading 3.3.6) since heating seems to enhance the viscosity of the sample.

26. More detailed information on SDS-PAGE analysis and protein transfer onto nitrocellulose membrane are provided in the previous chapter on *Lactococcus* expression system [11].

27. We used a *Strep*-tag II tag for the detection and purification of the recombinant proteins (this tag is fused to the *C*-terminal part of the protein). However, other affinity tags can be used. For more information concerning the detection of *Strep*-tag II proteins, refer to the handbook of IBA (IBA, Goettingen, Germany).

28. Recombinant protein can be detected using specific antibodies. In that case, first incubate the membrane with the primary antibody, wash the membrane and then incubate the membrane with the secondary antibody coupled to HRP. A detailed protocol is described in [23].

29. Membrane proteins either peripherally or intrinsically associated with membranes need to be solubilized to become soluble in aqueous solution before purification steps (for review *see* 24). According to the hydrophobic nature of the protein, various treatments can be performed from mild solubilization to stronger ones. Peripheral membrane proteins can be dissociated from membranes by using high salt or high pH solutions. Detergents that possess amphipathic properties are commonly used to solubilize integral membrane proteins from membranes. The detergent used has to solubilize the protein while preserving its activity. Thus, several detergents should be tested. For the solubilization of the plant transmembrane P-type ATPases HMA1 and HMA6, we used DDM and $C_{12}E_8$ which have been successfully used for the solubilization of the sarcoplasmic Ca-ATPase [25, 26].

30. The ceQORH protein was purified after solubilization in 1 M NaCl. Using the *Strep*-Tactin affinity matrix, the yield of purification of the ceQORH protein was 2–4 mg of purified protein per liter of culture [9].

31. The peripheral ceQORH protein expressed in *Lactococcus*, solubilized in the presence of salt and then purified on *Strep*-*Tactin* matrix, is active [9].

32. Here we used a combination of two detergents, but each detergent can be used alone and other detergents can be tested. Note that the nature of the detergent and its concentration have to be determined for each specific membrane protein.

33. Purification of HMA1 and HMA6 proteins on *Strep*-Tactin matrix was performed in the presence of 0.1 % (w/v) DDM and resulted in a purification yield of around 10–30 % [9].

Acknowledgment

References

1. Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci 7:1029–1038

2. Krogh A, Larsson B, von Heijne G et al (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

3. Wagner S, Bader ML, Drew D et al (2006) Rationalizing membrane protein overexpression. Trends Biotechnol 24:364–371

4. Freigassner M, Pichler H, Glieder A (2009) Tuning microbial hosts for membrane protein production. Microb Cell Fact 8:69

5. Bernaudat F, Frelet-Barrand A, Pochon N et al (2011) Heterologous expression of membrane proteins: choosing the appropriate host. PLoS One 6:e29191

6. Kunji ER, Chan KW, Slotboom DJ et al (2005) Eukaryotic membrane protein overproduction in Lactococcus lactis. Curr Opin Biotechnol 16:546–551

7. Mierau I, Kleerebezem M (2005) 10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis. Appl Microbiol Biotechnol 68:705–717

8. Zhou XX, Li WF, Ma GX et al (2006) The nisin-controlled gene expression system: construction, application and improvements. Biotechnol Adv 24:285–295

9. Frelet-Barrand A, Boutigny S, Moyet L et al (2010) Lactococcus lactis, an alternative system for functional expression of peripheral and intrinsic Arabidopsis membrane proteins. PLoS One 5:e8746

10. Catty P, Boutigny S, Miras R et al (2011) Biochemical characterization of AtHMA6/PAA1, a chloroplast envelope Cu(I)-ATPase. J Biol Chem 286:36188–36197

11. Frelet-Barrand A, Boutigny S, Kunji ER et al (2010) Membrane protein expression in Lactococcus lactis. Methods Mol Biol 601: 67–85

12. Geertsma ER, Poolman B (2010) Production of membrane proteins in Escherichia coli and Lactococcus lactis. Methods Mol Biol 601:17–38

13. Maischberger T, Mierau I, Peterbauer CK et al (2010) High-level expression of Lactobacillus beta-galactosidases in Lactococcus lactis using the food-grade, nisin-controlled expression system NICE. J Agric Food Chem 58:2279–2287

14. Douillard FP, Mahony J, Campanacci V et al (2011) Construction of two Lactococcus lactis expression vectors combining the Gateway and the NIsin Controlled Expression systems. Plasmid 66:129–135

15. Terzaghi BE, Sandine WE (1975) Improved medium for lactic streptococci and their bacteriophages. Appl Microbiol 29:807–813

16. Berlec A, Tompa G, Slapar N et al (2008) Optimization of fermentation conditions for the expression of sweet-tasting protein brazzein in Lactococcus lactis. Lett Appl Microbiol 46:227–231

17. Mierau I, Olieman K, Mond J et al (2005) Optimization of the Lactococcus lactis nisin-controlled gene expression system NICE for industrial applications. Microb Cell Fact 4:16

18. Le Loir Y, Gruss A, Ehrlich SD et al (1998) A nine-residue synthetic propeptide enhances secretion efficiency of heterologous proteins in Lactococcus lactis. J Bacteriol 180: 1895–1903

19. Sirén N, Salonen K, Leisola M et al (2008) A new and efficient phosphate starvation inducible expression system for Lactococcus lactis. Appl Microbiol Biotechnol 79:803–810

20. Abdullah-Al-Mahin, Sugimoto S, Higashi C et al. (2010) Improvement of multiple-stress tolerance and lactic acid production in Lactococcus lactis NZ9000 under conditions

of thermal stress by heterologous expression of Escherichia coli DnaK. Appl Environ Microbiol 76:4277–4285

21. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254

22. Chua NH (1980) Electrophoresis analysis of chloroplast proteins. Methods Enzymol 69: 434–436

23. Salvi D, Moyet L, Seigneurin-Berny D et al (2011) Preparation of envelope membrane fractions from Arabidopsis chloroplasts for proteomic analysis and other studies. Methods Mol Biol 775:189–206

24. Lin SH, Guidotti G (2009) Purification of membrane proteins. Methods Enzymol 463: 619–629

25. Jidenko M, Nielsen RC, Sorensen TL et al (2005) Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 102: 11687–11691

26. Jidenko M, Lenoir G, Fuentes JM et al (2006) Expression in yeast and purification of a membrane protein, SERCA1a, using a biotinylated acceptor domain. Protein Expr Purif 48:32–42

Part IV

Recombinant Protein Production in Yeast

Chapter 9

High Cell-Density Expression System: Yeast Cells in a Phalanx Efficiently Produce a Certain Range of "Difficult-to-Express" Secretory Recombinant Proteins

Yasuaki Kawarasaki, Takeshi Kurose, and Keisuke Ito

Abstract

Yeast's extracellular expression provides a cost-efficient means of producing recombinant proteins of academic or commercial interests. However, depending on the protein to be expressed, the production occasionally results in a poor yield, which is frequently accompanied with a deteriorated growth of the host. Here we describe our simple approach, high cell-density expression, to circumvent the cellular toxicity and achieve in a production of a certain range of "difficult-to-express" secretory protein in preparative amount. The system features an ease of performing: (1) precultivate yeast cells to the stationary phase in non-inducing condition, (2) suspend the cells to a small aliquot of inducing medium to form a high cell-density suspension or "a phalanx," and then (3) give a sufficient aeration to the phalanx. Factors and pitfalls that affect the system's performance are also described.

Key words Difficult-to-express secretory proteins, *Saccharomyces cerevisiae*, Heterologous expression, Extracellular protein production, High cell-density expression system

1 Introduction

Interestingly, extracellular production of a foreign protein or enzyme simplifies the downstream purification process, since certain steps including cell lysis and the following extract clarification steps are not necessary. Furthermore, purification of a secreted protein is much simpler than that from a clarified cell lysate, due to the presence of much less contaminating proteins [1–4].

Saccharomyces cerevisiae has commercial advantages over other host organisms regarding extracellular recombinant protein production. For example, it intrinsically releases a limited number and amount of endogenous proteins to the culture medium. Besides, genetic tools including constitutive or regulatable promoters [2] and artificial secretory signal sequences [5, 6] have also been devised to maximize the extracellular production of industrially useful proteins such as insulin [4–6] and enzymes such as

proteases, glycosidases, and lignolytic enzymes including laccases. However, depending on the protein to be expressed, yeast show deteriorated growth upon the expression induction, and this occasionally results in a poor production of the protein of interest or in the formation of insoluble aggregates known as inclusion bodies. Regardless of the molecular mechanisms that interfere with the production of biologically active proteins, those recombinant proteins are collectively called "difficult-to-express" proteins.

When our target protein is a "difficult-to-express" protein, it is frequently necessary to spend much time on, for instance, optimizing the induction condition, redesigning expression constructs including promoter and terminator sequences, trying an alternative expression vector with reduced gene copy number, and/or changing the host vector system to redeem the poor yield. Thus, in this chapter a detailed protocol for the expression of "difficult-to-express" proteins in yeast is introduced. Specifically, this chapter describes the different attempts performed to produce *Lentinula edodes* laccases (Lcc1 [7] and Lcc4 [8]) in *S. cerevisiae*. When using *S. cerevisiae* strains in classical induction manners (i.e., growth-associated induction), only a trace amount (<0.01 μg/L) of laccase was obtained in the culture supernatant. The poor production rate was accompanied with poor growth of the host as a result of loss of the expression plasmid [9]. The subsequent optimization of the expression condition fortuitously led us to establish a novel expression system that was capable of producing a certain range of "difficult-to-express" proteins [9]. With this system, we have successfully produced several "difficult-to-express" proteins including extracellular catalytic domain of hDPPIV [10], and miracle fruit miraculin [11], as well as the two *L. edodes* laccases [9]. The novel system features (1) an ease of performing, (2) a unique induction manner using a high-density suspension of yeast cells, and (3) a significant increase in the protein yield. Although the high cell-density system was not amenable to simple scale-up due to its increased oxygen requirement, we have overcome this problem by applying a benchtop jar fermenter to the high cell-density system [12]. As a result, the recombinant yeast cells in the "phalanx" produced as much amount of Lcc4 as 0.6 mg/L in 7 days of induction.

2   Materials

All chemicals are reagent-grade purity. We purchased those compounds from Sigma-Aldrich or Wako Pure Chemical Industries. Permission from your institute is required in advance to perform experiments using genetically modified organisms (GMOs). GMOs should be handled in accordance with the institute's safety guideline in any cases. Follow all waste-disposal regulations when disposing waste materials and genetically modified organisms.

2.1 Hosts and Vectors and Constructed Plasmids

Any lab-stock *S. cerevisiae* strains can be used as a host for the expression; however, in our hands, those lacking *pep4* gene (*see* **Note 1**) show frequently better performance. cDNA or DNA fragments amplified by PCR that encodes protein of interest should be cloned in YEp- or YRp-type plasmid under regulatable promoters such as P_{GAL1} (galactose-inducible promoter) or P_{CUP1} (copper-inducible promoter). We use pBG13 [13], a derivative of commercially available pYES2 (Invitrogen) that bears P_{GAL1} promoter and URA3 gene as a selection marker. Alternatively, other YEp-type plasmids can also be used. The choice of the signal sequence is occasionally critical regarding the final protein yield obtained. Although the signal sequence from mating factor MF1 is popularly used, some literatures report the optimization of the signal sequence improves the yield [5].

2.2 Media

1. Non-inducible synthetic dropout-dextrose medium lacking uracil (SD-U): 6.7 g/L yeast nitrogen base without amino acids (Sigma) and 20 g/L glucose supplemented with appropriate synthetic dropout (e.g., without uracil). This media is used for strains with pBG13-derived plasmids.

   The pH of the SD-U medium is usually between 5 and 6, being not necessary to adjust it to any specific pH. Synthetic dropout-dextrose (SD) medium lacking appropriate nutrients as well as inducer is used for precultivation and routine strain maintenance.

2. 2 % galactose-inducing medium (SG-U): Galactose is added to the SD medium instead of the glucose. The pH of the inducing SG medium could be adjusted to the range between 3 and 7 with concentrated HCl or NaOH if necessary, according to the pH stability of the product. Autoclaved media for the cultivation can be kept at room temperature (RT) until use.

3. K medium (SG-UCY) supplemented with 0.5-mM $CuSO_4$: The Lcc4-inducing medium that lacks cysteine (C) and tyrosine (Y) as well as uracil ([9], *see* **Note 2**). The copper sulfate is supplemented by adding 0.5 mL of 1 M $CuSO_4$ stock solution to the autoclaved K medium (*see* **Note 3**).

2.3 Cultivation Vessels and Shakers

Depending on the amount of the product, an appropriate production scale can be chosen (analytical/screening scale, flask scale, or preparative scale). It should be noted that, in any production scale, the aeration of the cell suspension is a critical factor that affects the final protein production yield [9]. Besides, the following general material is needed:

1. Turbidimeter to measure the optical density of the preculture (*see* **Note 4**).

2. Sterile test tube ($\Phi = 18$ mm) or 50-mL conical centrifugation tube.

3. Centrifuge (*see* **Note 5**).

4. Vortex mixer.

2.3.1 Analytical/
Screening Scale (–0.2 mL)

1. 96-deep-well plastic plates (*see* **Note 6**).

2. Sealing film for cultivation (gas-permeable seals such as Axygen's breathable sealing film for 96-well plate).

2.3.2 Test-Tube Scale
(–4 mL)

Tubes with 18 mm in diameter are suitable. Aluminum caps or gas-permeable sterile plugs for the test tubes are also needed. Alternatively, conical 50-mL centrifugation tubes (e.g., CELLSTAR filter-top tubes from Greiner Bio-One) are used instead of the sterile test tubes.

2.3.3 Flask-Scale
Expression (–200 mL)

1. Cultivation vessels with multiple baffles (e.g., 2.5-L Ultra Yield Flask from Thomson Instrument Company with ventilation top seal). Baffles are essential for the production in acceptable yield. It should be noted that the yield obtained in flask-scale expression is decreased to 1/5 to 1/10 of that obtained in a small-scale production [9].

2. Medium-size bioshakers capable of agitating the cell suspension at a high (>140 rpm, vibration stroke = 25 mm) rotation rate is required (*see* **Note** 7).

2.3.4 Preparative Scale

1. Benchtop fermenter equipped with a chilling-water circulation device is strongly recommended for the preparative-scale production (*see* **Notes 8** and **9**).

3 Methods

3.1 Analytical-Scale/
Screening-Scale
Expression

1. Inoculate 4 mL of SD-U medium in a sterile test tube with 2–3 yeast colonies containing the expression plasmid. Incubate the preculture at 30 °C with a reciprocal shaking at 100–120 rpm. The culture will reach the stationary phase ($OD_{660\ nm} = 3$–4) in 24 h.

2. Transfer the preculture to a sterile conical tube or centrifugation tube with appropriate size.

3. Harvest the cells by a brief spin at $500 \times g$ at RT.

4. Remove the culture supernatant.

5. Add the inducing medium (SG-U or the K medium for the laccase expression) to give a high cell-density suspension with $OD_{660\ nm} = 15$. In most cases, the volume of the inducing medium is nearly 20 % of that of the preculture. Swirl vigorously with a vortex mixer to suspend the yeast cell in the pellet.
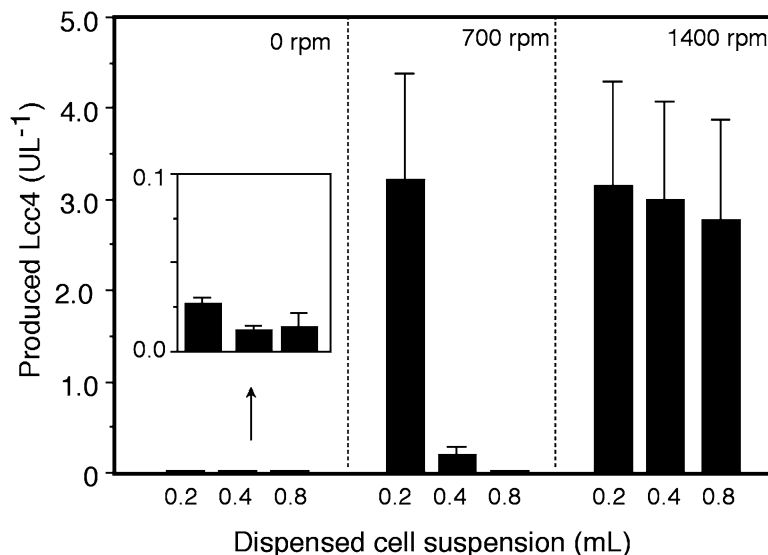
Fig. 1 Analytical-scale expression. Various amounts of a condensed cell suspension were dispensed to wells in a deep-well cultivation plate. Laccase expression was induced at 25 °C for 24 h at the indicated agitation rate. The amounts of the produced laccase from FGY(pBGlcc4) in the 24 h of agitation are shown

6. Dispense the cell suspension to wells in a sterile deep-well plate (*see* **Note 6**). The appropriate amount of the cell suspension per well is 0.2 mL. Wells with 0.4 mL or more cell suspension cannot give sufficient relative surface areas, thus leading to dearth of aeration (Fig. 1).

7. (Optional) Dispense sterile H_2O to unused wells to keep the chamber humid during the induction. Otherwise, the culture volume would decrease particularly in long-term induction by evaporation.

8. Seal the deep-well plate with a sterile, gas-permeable film. Induce the gene expression by aerobically incubating the deep-well plate at 20–30 °C (*see* **Note 10**) with a bioshaker with vigorous agitation at 1,400 rpm (vibration stroke = 2 mm) for a day (*see* **Note 11**). The incubation can be extended to another days for further increase in yield.

3.2 Test-Tube-Scale Expression

1. Inoculate 4 mL SD-U in a sterile test tube with 2–3 yeast colonies containing the expression plasmid. Incubate the culture at 30 °C with a reciprocal shaking at 100–120 rpm for 24 h. The cells in the culture will enter stationary phase ($OD_{660\,nm} = 3.0$).

2. Transfer the preculture to 40 mL of a fresh SD-U in a sterile flask. Cultivate aerobically at 30 °C for 12–24 h, until the culture turbidity increases up to 3.0 $OD_{660\,nm}$ or more.
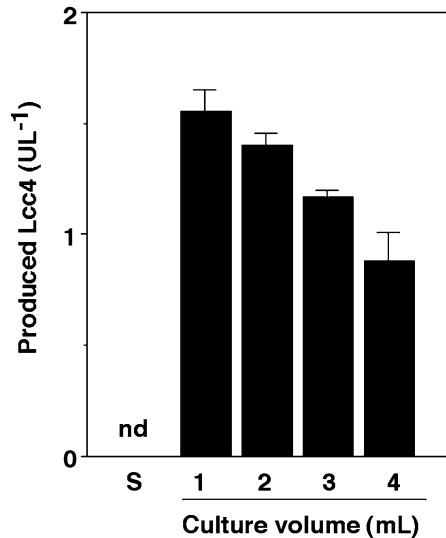
Fig. 2 Test-tube-scale expression. Indicated amounts of a condensed cell suspension ($OD_{660 nm} = 15$) were poured into test tubes ($\Phi = 18$ mm) and reciprocally agitated at 150 rpm for 24 h at 25 °C. Sample depicted as S represents the culture supernatant obtained in static (without agitation) cultivation

3. Transfer the preculture to a centrifugation tube.

4. Harvest the cells by a spin at $500 \times g$ at RT for 15 min.

5. Remove the culture supernatant.

6. Add the inducing medium (SG-U or the K medium for the laccase expression) to give a high cell-density suspension with $OD_{660 nm} = 15$. Swirl vigorously with a vortex mixer to suspend the yeast cells.

7. Pour 2 mL of the cell suspension into a sterile test tube ($\Phi = 18$ mm, *see* **Note 12**). Further addition of the cell suspension to a tube causes decrease in production yield (Fig. 2, [9]).

8. Place the test tube on an angled (30°, *see* **Note 13**) tube rack in a bioshaker (*see* **Note 7**).

9. Induce the gene expression at 20–30 °C (*see* **Note 10**) for 2–3 days. The culture should be reciprocally agitated at 150 rpm (vibration stroke = 22 mm).

3.3 Flask-Scale Expression

1. Prepare 1 L of preculture (*see* **Note 14**) (*see* Subheadings 3.1 and 3.2).

2. Withdraw a small aliquot of the preculture and measure the $OD_{660 nm}$.

3. Transfer the preculture to sterile 500-mL centrifugation tubes.

4. Pellet the cells by a centrifuge at $500 \times g$ for 15 min (*see* **Note 5**).

5. Remove the culture supernatant.
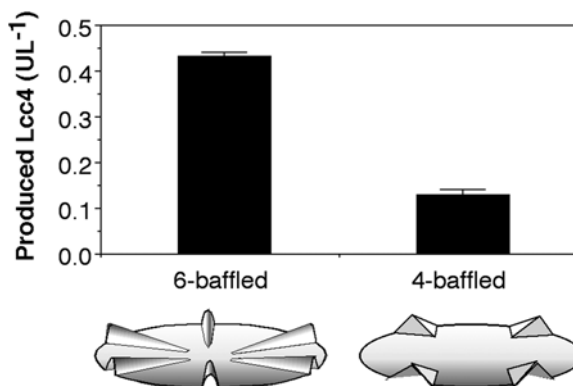
Fig. 3 Flask-scale expression: effect of flask design on high cell-density expression. A dense suspension of FGY217(pBGlcc4) cells ($OD_{660\,nm} = 12$) was divided to 500 mL and dispensed to baffled flasks with different baffling designs. The cultivation was carried out at 25 °C for 24 h at 220 rpm with a presence of antifoam (Sigma). Drawings represent the bottom shapes of the cultivation flasks; Ultra Yield Flask by Thomson Instruments (6-baffle flask) and Nalge-Nunc's baffled flask (4-baffle flask)

6. Dispense about 20–30 mL of the inducing medium (SG-U or the K medium for the laccase expression) to each of the tubes. Swirl vigorously to suspend the yeast cells.

7. Pool the cell suspension into a single cylinder (it should be sterilized in advance), then adjust the optical density to $OD_{660\,nm} = 15$ with an appropriate amount of the inducing medium.

8. Transfer the cell suspension to a 2.5-L sterile baffled flask. Note that the design of the flask (i.e., size of the base area and shape, as well as the number of baffles, etc.) as well as the volume of the cell suspension significantly influences the yield and reproducibility (Fig. 3).

9. Induce the gene expression at 20–30 °C for 2–3 days at a high rotation rate (150 rpm, vibration stroke = 22 mm). It should be emphasized again that the yield in this production scale is easily dropped to 10–20 % of that obtained in a small-scale production (compare the results shown in Fig. 1 with those in Fig. 3).

3.4 Preparative-Scale Production Using a Benchtop Jar Fermenter

1. Pour 1 L of deionized H_2O to a jar-fermenter (*see* **Note 8**) vessel (vessel size = 2 L). Wrap the tubes, air-filter, and condenser for the fermenter with sheets of aluminum foils. Autoclave them at 120 °C for 20 min. Discard the deionized H_2O in the fermenter after the autoclave.

2. Prepare 4 L of preculture (*see* Subheadings 3.1 and 3.2).

3. Withdraw a small aliquot of the preculture and measure the optical density at 660 nm.
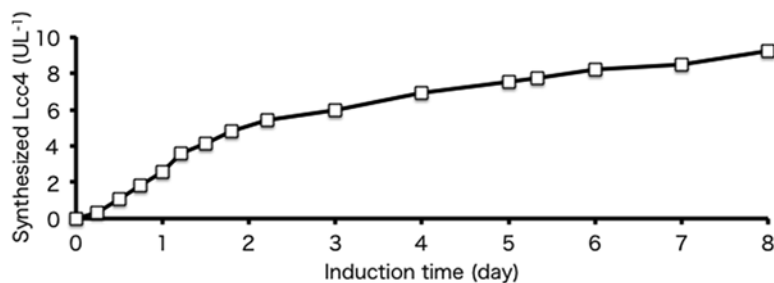
Fig. 4 Preparative-scale expression. A typical Lcc4 production profile in the jar-fermenter-based high cell-density expression is shown. FGY217(pBGlcc4) cells precultured in non-inducing medium (4 L in total) were harvested, and suspended to 1 L of galactose-enriched K medium to give a dense cell suspension (i.e., 15 $OD_{660 \, nm}$), then poured into the Marubishi fermenter vessel (2 L in volume). The cell suspension was agitated at 600 rpm at 20 °C, with a vigorous (2-L filtered air min^{-1}, 2 vvm) aeration. The amount of Lcc4 in the culture supernatant was analyzed at the indicated time point

4. Harvest the cells by multiple runs of centrifugation as in Subheading 3.3.

5. Suspend the cells to the inducing medium (SG-U or the K medium for the laccase expression). Note that the inducing medium with enriched (4 %) galactose often gives increased yield in longer induction periods [12].

6. Pool the cell suspension into a sterile 1-L measuring cylinder, and then adjust the optical density to $OD_{660 \, nm} = 15$ with an appropriate amount of the inducing medium. Measure the total volume of the cell suspension.

7. Transfer the entire (up to 1 L) cell suspension to the fermenter vessel (*see* **Note 8**).

8. Run the jar-fermenter system at 20–30 °C. The aeration rate should be higher than 1.0 vvm. If available, a chilling-water circulation unit for a rotary evaporation should be used to stabilize the induction temperature (*see* **Note 9**). The production of the heterologous protein will last for a week (*see* **Note 15**) when the galactose-enriched induction medium is used (Fig. 4).

4   Notes

1. We have tested various yeast strains including INVSc1 (*MATa, his3Δ1, leu2, trp1-289, ura3-52/MATα, his3Δ1, leu2, trp1-289, ura3-52*; Invitrogen, CA, USA), BY2777 (*MATa, prb1-1122, prc1-407, pep4-3 leu2, trp1, ura3-52*; laboratory stock), FGY217 (*MATa, pep4Δ, ura3-52, lys2Δ201*; laboratory stock),

and BY4741 (*MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*). Among them, FGY217 and BY2777 strains are preferably used in high cell-density expression. The strain used in Figs. 1, 2, 3, and 4 for the Lc7c4 expression is FGY217.

2. Both cysteine and tyrosine can be substrates of Lcc4. The presence of those substrates in the culture medium significantly disturbs the enzymatic assay using chromogenic substrate.

3. The Cu supplementation is required for the active laccase expression, because laccases are multi-copper enzyme. Addition of $CuSO_4$ to the culture medium up to 0.2 mM does not significantly affect cell growth. The 1 M $CuSO_4$ stock solution does not have to be autoclaved.

4. We usually use TAITEC's miniphoto 518R with 660-nm filter. This turbidimeter allows us to measure the turbidity of the culture in test tube without withdrawing it.

5. We usually use TOMY MX-300 centrifuge with a rotor for 50-mL tubes (AR510-04). A larger centrifuge (e.g., CR22N from HIMAC (Hitachi)) can also be used when a preculture with larger cultivation volume is needed.

6. We have used Nunc's round-well plate (#260251/260252; the well size is 8.4 mm in diameter and 29 mm in depth) to get rid of splash of the culture medium during agitation.

7. We use TAITEC BR-43FL.

8. We use B.E. Marubishi's MDL-200.

9. We use EYELA's CCA-1111. The temperature of the circulating chilling water is set to 10 °C. It is difficult to keep the cultivation temperature below 25 °C with conventional tap water-based chilling system.

10. Induction temperature is another important factor that affects the yield. In the case of laccase expression, the lowest temperature (20 °C) is the best.

11. We use TAITEC deep-well maximizer (M•BR-022UP).

12. Test tubes with wider diameters provide larger relative surface area.

13. Tube rack placed in more acute angle provides good aeration. However, the culture, which involves GMO, would be leaked during the induction.

14. The aeration required for the preculture (final $OD_{660\,nm} = 3\text{–}4$) is much lower than that for the high cell-density suspension ($OD_{660\,nm} = 15$). Therefore, the preculture volume can be raised to 0.5 L when the Thomson Instrument Company's 2.5-L baffle flask is used.

15. Antifoam (e.g., Antifoam A from Sigma-Aldrich) is efficiently used for such long-term cultivation.

Acknowledgment

References

1. Sudbery PE (1996) The expression of recombinant proteins in yeasts. Curr Opin Biotechnol 7:517–524

2. Porro D, Sauer M, Branduardi P et al (2005) Recombinant protein production in yeasts. Mol Biotechnol 31:245–259

3. Graf A, Dragosits M, Gasser B et al (2009) Yeast systems biotechnology for the production of heterologous proteins. FEMS Yeast Res 9:335–348

4. Gerngross TU (2004) Advances in production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 22: 1409–1414

5. Kjeldsen T, Pettersson AF, Hach M et al (1997) Synthetic leaders with potential BiP binding mediate high-yield secretion of correctly folded insulin precursors from *Saccharomyces cerevisiae*. Protein Expr Purif 9:331–336

6. Kjeldsen T, Ludvingsen S, Diers I et al (2002) Engineering-enhanced protein secretory expression in yeast with application to insulin. J Biol Chem 277:18245–18248

7. Nagai M, Sato T, Watanabe H et al (2002) Purification and characterization of an extracellular laccase from the edible mushroom *Lentinula edodes*, and decolorization of chemically different dyes. Appl Microbiol Biotechnol 60:327–335

8. Nagai M, Kawata M, Watanabe H et al (2003) Important role of fungal intracellular laccase for melanin synthesis: purification and characterization of an intracellular laccase from *Lentinula edodes* fruit bodies. Microbiology 149:2455–2462

9. Kimata K, Yamaguchi M, Saito Y et al (2012) High cell-density expression system: A novel method for extracellular production of difficult-to-express proteins. J Biosci Bioeng 113:154–159

10. Hikida A, Ito K, Motoyama T et al (2013) Systematic analysis of a dipeptide library for inhibitor development using human dipeptidyl peptidase IV produced by *Saccharomyces cerevisiae* expression system. Biochem Biophys Res Commun 430:1217–1222

11. Ito K, Sugawara T, Koizumi A et al (2010) Bulky high-mannose-type N-glycan blocks the taste-modifying activity of miraculin. Biochim Biophys Acta 1800:986–992

12. Kurose T, Saito Y, Kimata K et al (2014) Secretory expression of *Lentinula edodes* intracellular laccase by yeast high-cell-density system: a sub-milligram production of "difficult-to-express" secretory protein. J Biosci Bioeng. 117:659–663

13. Kamiya T, Ojima T, Sugimoto K et al (2010) Quantitative Y2H screening: Cloning and signal peptide engineering of a fungal secretory LacA gene and its application to yeast two-hybrid system as a quantitative reporter. J Biotechnol 146:151–159

Part V

Recombinant Protein Production in Insect Cells-Baculovirus

Chapter 10

Insect Cells–Baculovirus System for the Production of Difficult to Express Proteins

Judit Osz-Papai, Laura Radu, Wassim Abdulrahman, Isabelle Kolb-Cheynel, Nathalie Troffer-Charlier, Catherine Birck, and Arnaud Poterszman

Abstract

The production of sufficient quantities of homogenous protein not only is an essential prelude for structural investigations but also represents a rate-limiting step for many human functional studies. Although technologies for expression of recombinant proteins and complexes have been improved tremendously, in many cases, protein production remains a challenge and can be associated with considerable investment. This chapter describes simple and efficient protocols for expression screening and optimization of protein production in insect cells using the baculovirus expression system. We describe the procedure, starting from the cloning of a gene of interest into an expression transfer baculovirus vector, followed by generation of the recombinant virus by homologous recombination, evaluation of protein expression, and scale-up. Handling of insect cell cultures and preparation of bacmid for co-transfection are also detailed.

Key words Baculovirus, Insect cells, Homologous recombination, Protein expression and production, Solubility screen

1 Introduction

The production of sufficient quantities of homogenous recombinant protein samples not only is an essential prelude for structural investigations but also represents a rate-limiting step for many functional studies. *Escherichia coli* is a robust and inexpensive expression host for the production of recombinant proteins, but there are serious limitations in using bacteria for synthesis of eukaryotic protein [1, 2]. In particular, bacteria are unable to accomplish posttranslational modifications and folding events required for the generation of fully functional eukaryotic proteins. Many eukaryotic proteins expressed in bacteria are often synthetized as truncated polypeptides or become insoluble as inclusion bodies that are very difficult to recover without harsh denaturants and

subsequent cumbersome protein-refolding procedures. In contrast to *E. coli*, the eukaryotic baculovirus/insect cell, and mammalian systems promote good protein folding and many posttranslational modifications [3, 4]. Advances in vector design and process optimization have resulted in user-friendly and efficient technologies for expression screening and large-/medium-scale production of complexes [5–8].

Baculoviruses are rod-shaped, double-stranded, DNA viruses which infect and kill a large number of different invertebrate species especially insects [9]. The most common baculovirus used for expression studies is *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) [10, 11], which infects the lepidopteran species *Spodoptera frugiperda* as host insects. AcMNPV particles are surrounded with a protective matrix consisting of the protein polyhedron [12], which permits survival in the environment and efficient dissemination to new hosts. Under the control of the extremely strong promoter *pPolh*, polyhedrin is expressed at considerable levels and can represent up to 50 % of total cellular proteins. In cell culture, the polyhedrin coat is not essential for virus propagation and thus heterologous proteins can be expressed under the control of the *pPolh* promoter [10, 13]. Insect cells infected by recombinant baculoviruses can overexpress target proteins with yields which can reach several hundreds of mgs of protein for 1 L of culture. More importantly, the cytoplasmic environment of the insect cells allows for proper protein folding and for most posttranslational modifications which are very often crucial for the folding/function of the target protein [3].

Original methods for recombinant protein expression using the baculovirus expression system (BEVS) were time-consuming and incompatible with parallel processing of multiple targets, for example when screening of mutants or deletion variants is required to identify constructs suitable for structural studies. A remarkable improvement was the incorporation of a bacterial artificial chromosome (BAC) into the viral genome which allows modification of the viral DNA and results in the elaboration of efficient strategies to generate recombinant viruses at frequencies close to 100 %, removing the need to plaque-purify recombinant viruses from parental [14]. One of these strategies, patented by Invitrogen (bac-to-bac), is based on the Tn7-mediated transposition of an expression cassette containing the target gene into the bacmid within *E. coli* [15]. More recently, new bacmids, which consist of a recombinant baculoviral genome with a bacterial replicon at the polyhedrin locus and deletion of a downstream essential gene (ORF 1629), were engineered [16, 17]. The bacmid DNAs which can be produced in *E. coli* do not replicate in insect cells. Homologous recombination between the viral and appropriate transfer plasmid will restore replication, eliminate the bacterial replicon in the polyhedrin locus, and knock-in the gene of interest.

Healthy Sf9, Sf21 and High-five cells (Section 3.4)

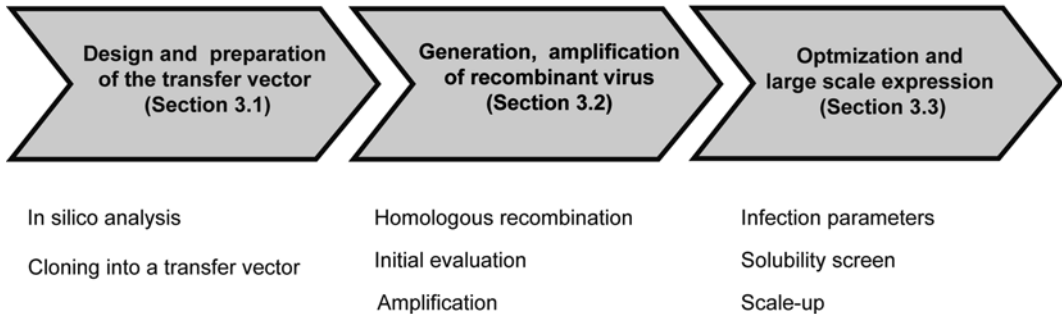Linearized viral DNA for homologous recombination (Section 3.5)

```
┌──────────────────────┐   ┌──────────────────────┐   ┌──────────────────────┐
│ Design and preparation│──▶│ Generation, amplification│─▶│ Optmization and       │
│  of the transfer vector│  │ of recombinant virus   │  │ large scale expression │
│     (Section 3.1)     │   │     (Section 3.2)     │   │     (Section 3.3)     │
└──────────────────────┘   └──────────────────────┘   └──────────────────────┘
```

In silico analysis   Homologous recombination   Infection parameters

Cloning into a transfer vector   Initial evaluation   Solubility screen

                                  Amplification   Scale-up

Fig. 1 Flowchart illustrating the entire process from cloning to protein production in insect cells using the baculovirus expression system

As it is not possible for non-recombinant virus to replicate, there is no need for any selection system which considerably simplifies and reduces production of recombinant virus to a one-step procedure in insect cells.

In this chapter, we provide simple and efficient protocols for expression screening as well as optimization and scale up production. We describe the procedure, starting from the cloning of a gene of interest into an expression transfer baculovirus vector, followed by generation of the recombinant virus by homologous recombination, evaluation of protein expression, and scale-up (Fig. 1). We detail handling of insect cell culture as well as preparation of bacmid for co-transfection. These protocols do not require extensive background in cell biology and can be applied in a lab equipped for basic cell culture and biochemical work.

2 Materials

2.1 Working Environment and Instruments

1. Temperature controlled room or incubator set at 27 °C.

2. Stirring platform for spinner flask operating at 27 °C and up to 150 rpm.

3. Orbital shaker fitted for 250 mL to 2 L Erlenmeyer flasks, with shaking speed of up to 150 rpm (125 mm orbital).

4. Inverted phase-contrast microscope or optionally fluorescence microscope.

5. Cell counting chamber or optionally automated cell counter.

6. Centrifuge with adaptors for 1 L, 250 mL, 50 mL, and 15 mL tubes.

7. Access to liquid nitrogen storage.

8. Thermocycler.

9. Refrigerated benchtop centrifuge.

10. Sonicator with 3 mm probe (optionally four head sonicator).

11. Devices for DNA and protein gel electrophoresis.

12. Western blot transfer system.

13. Multichannel pipette or optionally pipetting robot with a vacuum chamber and gripper.

2.2   Chemicals, Kits, and Strains for Molecular Biology

1. MB grade Ethanol 100 %, Isopropanol 100 %, Na/Acetate, and agarose.

2. Ampicillin (100 mg/mL), Kanamycin (50 mg/mL), Chloramphenicol (34 mg/mL).

3. Luria-Bertani (LB): 10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, and 950 mL of ddH$_2$O. Stir to dissolve all solids and bring the final volume to 1 L with ddH2O before autoclaving.

4. LB-agar plates: add 15 g of agar per L of LB broth prepared as above before autoclaving

5. DH5α competent cells.

6. pAC8 vectors [18], BAC1(ACCATCTCGCAAATAAATAA) and BAC2(ACAACGCACAGAATCTAGCG) primers.

7. PCR reaction kit with high-fidelity DNA polymerase and PCR clean-up kit.

8. Restriction enzymes NdeI, BamHI, Bsu36I.

9. Shrimp Alkaline Phosphatase (SAP).

10. T4 DNA ligase.

11. BA10:KO1629 in *E. coli* DH10B strain [17].

12. DNA plasmid purification kits suitable for isolation of plasmids and BACs.

2.3   Cell Culture Dishes

1. T75 and T175 tissue culture flasks.

2. 6-Well plate, flat bottom, low evaporation lid.

3. 2 L Spinner flasks.

4. 250 mL and 2 L glass or disposable Erlenmeyer flasks.

5. Plate sealer, breathable, gas permeable, 80 × 150 mm.

6. Sterile cryogenic tubes, 1.5 mL.

7. Controlled Rate Freezer System for 12 tubes (freezing rate –1 °C/min).

8. Disposable sterile conical tubes for 15, 50, and 250 mL.

2.4   Insect Cell Lines and Media

1. Trypan blue solution.

2. TNM-FH and serum-free insect cell medium

3. Fetal Bovine Serum (FBS).

4. Sf9 cells adapted in TFNM-TH, 10 % FBS.

5. Sf9, Sf21, and Hi5 cells adapted for suspension growth in serum-free medium.

6. Cell culture grade DMSO.

7. FectoFly™ (Polyplus) or equivalent transfection agent.

2.5 Chemicals and General Stock Solutions

1. Phosphate-buffered saline 10× stock solution (PBS 10×).

2. DNAse I Stock Solutions (40,000 U/mL).

3. Protease inhibitor cocktail.

4. Reducing agent such as β-mercaptoethanol, tris[2-carboxyethyl] phosphine hydrochloride (TCEP), or dithiothreitol (DTT).

2.6 Buffers and Reagents for Protein Purification

1. 24 and 96 deep-well blocks and filters plates.

2. Affinity resins: Ni-NTA agarose; Glutathione sepharose 4B; Strep-Tactin sepharose; Anti-Flag M2 affinity gel. Ni-NTA and GST resins are available in spin column and 96-well spin plate format.

3. Lysis buffer: 30 mM Na phosphate (Tris/HCl or HEPES can be used as well), pH 7.8, 0.5 M NaCl, with or without 6 M urea and detergents as appropriate.

4. Wash buffer: 30 mM Na phosphate (Tris/HCl or HEPES can be used as well), pH 7.8, 0.5 M NaCl, with or without 6 M urea and detergents as appropriate.

5. Elution buffer: Same as wash buffer plus imidazole (200 mM), reduced glutathione (10 mM), D-desthiobiotin (2.5 mM), or Flag-peptide (DYKDDDDK, 100 μg/mL).

6. 4× SDS loading solution.

3 Methods

3.1 Design and Preparation of the Transfer Vector

Variable yields and poor solubility are major impediments to streamlined production of many recombinant proteins, in particular for proteins classified as difficult-to-express including membrane proteins or large multi-domain proteins, which often do not function as isolated entities but in complex with other macromolecules. As solubility and expression level of constructs cannot be predicted, most strategies to optimize production of recombinant proteins rely on systematic screening.

When a full-length protein fails to express in soluble form, a common strategy for improving production is to modify target genes sequences by PCR and clone constructs encompassing single- or multi-domain fragments. Expression constructs should correspond to structural units, and unless needed for functional

reasons, any flexible, unstructured tails should be removed. In absence of direct structural information on domain boundaries from closely related proteins, design of expression constructs relies on educated guesses from analysis of multiple sequence alignments, from predictions of secondary structure and disordered regions, as well as from domain predicting algorithms.

Meta-analysis servers (Table 1) which collect and display information from prediction algorithms and from databases searches help selection of domain boundaries. Widely accepted guidelines for a priori soluble proteins are (1) to respect the boundaries of predicted globular domains as well as predicted secondary structural elements and (2) to avoid inclusion of low-complexity regions, hydrophobic residues at the termini, as well predicted membrane spanning regions. The optimal step size between the nested primers can be a matter of debate; we commonly make constructs to encode proteins that vary in length by 2–10 amino acids at each end. Ideally, the approximate boundaries of the region of interest might be identified using a functional assay, scanning deletion mutagenesis, as well as limited proteolysis combined with mass spectrometry analysis. It might be worth trying to express multiple domains in some cases, as the neighboring domains can stabilize each other and create functional entities.

Foreign cDNAs cannot be directly inserted into the baculoviral genome. Instead, cDNAs are cloned into a transfer vector or donor vector which is used to manipulate the viral genome. Several technologies are available. The pipeline described here is based on

Table 1
Resource portals providing access to software tools and databases for selection of construct boundaries

Meta-analysis servers		
EXPASY SIB Bioinformatics Resource Portal	http://www.expasy.org/	[33]
MPI Toolkit for protein sequence analysis	http://toolkit.tuebingen.mpg.de/	[34]
Protein CCD: Crystallographic Construct Design	http://xtal.nki.nl/ccd	[35]
Order, disorder prediction tools		
IUPred: prediction of intrinsically disordered protein	http://iupred.enzim.hu/	[36]
GLOBPLOT2: Domain & Globularity Prediction, Intrinsic Protein Disorder prediction	http://globplot.embl.de/	[37]
RONN: Regional Order Neural Network	http://www.strubi.ox.ac.uk/RONN	[38]
FoldIndex@: Will this protein fold?	http://bioportal.weizmann.ac.il/fldbin/findex	[39]

homologous recombination in insect cells between linearized baculovirus DNA and a transfer vector containing the gene(s) of choice cloned under the control of the polyhedrin and/or the p10 promoter. A wide panel of transfer vectors is available for the production of proteins with specific peptide tags that aid subsequent protein purification. Vectors that permit the insertion of multiple genes for co-expression of several proteins also exist (Table 2).

Screening includes optimization of not only expression conditions and constructs variants, as described above, but also fusion tags which can have a positive impact on the yield, solubility, and even the folding of their fusion partners [19, 20]. We detail the use of the PAC8 vector suite, a set of vectors with identical backbones designed to facilitate expression screening and enable consistent comparisons of the impact of fusion partners on expression, solubility, and purification [18]. These vectors contain a polyhedrin promoter, the sequence coding for fusion protein/affinity tag including Protein A, FLAG, GST, Strep, and His6, followed by a protease 3C cleavage site and a poly-linker (NdeI, PmeI, and BamHI) or a gateway cloning cassette to insert the target cDNA. Modified versions include vectors that co-express the DsRed fluorescent protein with the target gene to easily monitor transfection, virus amplification, and optimization of culture conditions (Table 2).

Table 2
Transfer vectors for expression screening

Vector	Promoter	Main feature	Fusion	Reference
pVL1392, pVL1393	pH	General purpose transfer vector	–	[15]
pAcUW21	p10	General purpose transfer vector	–	[40]
pAcUW51, pAcUW31	pH, p10	Co-expression of 2 genes	–	[40]
pACAB3, pACAB4	pH, p10	Co-expression of 3 or 4 genes	–	[41]
pBacPAK8, pBacPAK9	pH	General purpose transfer vector	–	Clontech™
pTri-Ex-1.1, –2, –3	p10	Multi-host expression	Yes	Novagen™
pOET1, 2, 3, 4	pH, p6.9	Small size plasmids	Yes	OET™
pOPIN	pH	Multi-host expression, In-fusion cloning	Yes	[42]
pOmni Bac	pH, p10	Co-expression of n genes, LoxP site	–	[43]
pAC8	pH	N-terminal fusions, C3 cleavage	Yes	[18]
pAC8-DsRed[a]	pH	Co-expression of DsRed as marker	Yes	Unpublished
pAC8-GW-Lox[a]	pH	Gateway cloning cassette, Co-expression of n genes, LoxP site	Yes	Unpublished

[a]Not published but available on request

1. Analyze the protein and DNA sequences of the target gene(s) to plan experiments (construct design and selection of affinity tags) and determine the cloning strategy. When a screening of several constructs for a given gene is planned, we first try restriction-/ligation-based cloning and use the NdeI and BamHI restriction sites of pAC8 poly-linker. If this is not possible, for example when the cDNA contains NdeI and BamHl/ BglII sites and gene synthesis not affordable, we use restriction-independent strategies (*see* **Note 1**).

2. Digest the transfer vector with NdeI and BamHI, treat the digested plasmid with a phosphatase, isolate the linearized vector from the rest of the reaction with purification kit, and quantify. Typically, we prepare a large stock (20 μg) of vector that can be stably stored at –20 °C and used for several rounds of subcloning (*see* **Note 2**).

3. Amplify cDNAs using a forward primer that contains a NdeI restriction site and a reverse primer with a BamHI (or a BglII) site and a stop codon. The PCR product is cleaned with commercial DNA clean-up kit, digested with NdeI and BamHI (or BglII). Typically 1–2 μg of the PCR product is digested in a total volume of 20 μL for 1 h. Run digested DNA in an agarose gel, purify to isolate DNA, and quantify the recovered product using a nano UV spectrophotometer. Gel purification can be replaced by PCR clean up. In this case, digest the PCR reaction with DpnI to remove the matrix and inactivate.

4. Set up a DNA ligation to fuse the digested pAC8 vector and the cDNA fragment. Typically 100 ng of the linear plasmid fragment is ligated with threefold molar excess of the insert in 10 μL reaction volume. Different ratios plasmid/insert can be tested. Overnight (ON) ligation at 16 °C is optimal for T4 DNA ligase activity. Do not forget a negative control to evaluate the background from uncut or self-ligating recipient plasmid.

5. Transform the ligation reaction into competent cells, for example *E. coli* DHα5 and plate onto LB agar plates containing 100 μg/mL ampicillin. Significantly more colonies should be obtained in presence of insert than in the negative control.

6. Pick up 4–5 colonies if the ligation background is good (i.e., minimum ten times more colonies are on the plate than the negative control plate), and 8–10 colonies if the ligation background is high. Grow 1 mL pre-cultures that will be stored at 4 °C and use the BAC1 and BAC2 primers which hybridize on both extremities of the expression cassette as first screen for PCR analysis. Select two positive colonies and grow ON cultures to purify DNA. Perform diagnostic restriction digestion of 200 ng with NdeI/BamHI and sequence it using the BAC1 and BAC2 primers.

7. For the next step, pure and sterile transfer vector is required. Perform a mini- or midi-prep of a sequence-validated plasmid and precipitate 10 μg DNA with 300 mM Na/Acetate pH 5.2 (final concentration) and 3 volumes of ethanol 100 %. Place at −80 °C for more than 1 h and centrifuge at $250,000 \times g$ for 15 min. Carefully remove the supernatant, add 1 mL of cold 70 % ethanol, and centrifuge it again.

   From this point manipulate under a laminar flow hood!

8. Remove ethanol and air-dry the precipitated DNA under the sterile hood. Resuspend DNA in 20 μL sterile ultrapure H_2O. Take an aliquot to measure the DNA concentration and store at −20 °C.

3.2   Generation and Amplification of Recombinant Baculovirus

To generate recombinant baculoviruses a transfer vector suitable for homologous recombination is co-transfected with linearized viral DNA in insect cells. This allows integration of the expression cassette into the viral genome which will be replicated and leads to virus production. The co-transfection supernatant is referred to as the initial virus stock (P0). It can be used for a first evaluation of protein expression and will be amplified to obtain amplifications 1 and 2 (P1 and P2) required for large-scale expression. Handling of insect cells is described in Subheading 3.4 where procedures for thawing, maintenance, and freezing are detailed. Note that all waste cells, viruses, used media, and plasticware are to be treated ON with bleach or autoclaved before discarding.

In this section, we detail the co-transfection of the transfer plasmid with a linearized viral DNA. For efficient transfection, virus amplification, and protein production, healthy cells are absolutely required. Cells should be maintained in exponential growth phase, should not be overgrown, and passages should be limited. A doubling time of 18–24 h and a continuous viability >95 % are prerequisites for successful work. We advise setup of quality control experiments with known cDNA and viruses to monitor infectivity and expression levels. Additionally, we found that use of a transfer vector which, in addition to the target protein, expresses the DsRed protein (pCA8-DsRed, Table 2) can be extremely useful, not only as a positive transfection control but also for protocol optimization.

3.2.1   Co-transfection of Transfer Plasmid and Viral DNA

The described protocol was optimized to maximize the number of infected cells from 5 to 7 days after transfection and therefore the titer of the P0 virus stock (*see* **Notes 3** and **4**). We use a serum-containing medium for generation of viruses (*see* **Note 5**) and linearized viral DNA prepared from BAC10:KO1629, as described in Subheading 3.5. For simplicity, we have chosen a transfection agent which is not affected by serum.

1. Seed a 6-well plate using 1.5×10^6 Sf9 cells per well in 1.6 mL insect cell culture medium (TNM-FH + 10 % FBS) and let the cells adhere for 20 min at 27 °C.

2. Meanwhile, under the sterile hood, mix 4 µg of DNA transfer vector with 1 µg of linearized bacmid in 100 µL of sterile 150 mM NaCl and dilute 5 µL of FectoFly™ transfection reagent into 100 µL of sterile 150 mM NaCl (i.e., use 1 µL of FectoFly™ per µg of DNA). Include a negative control without a transfer plasmid and positive control with transfer vector expressing the DsRed protein.

3. Add the FectoFly™ solution to the DNA solution (respect the addition order), mix well, but gently, and incubate at room temperature (RT) for 30 min. Respect recommended incubation time as extended incubation may lead to formation of large and difficult to transfect DNA/transfection agent complexes.

4. Add the 200 µL DNA/FectoFly™ solution *drop wise* to the cells, homogenize by shaking the plate gently, and incubate at 27 °C. Four hours after the co-transfection, add 2 mL of insect cell medium to the cell layer and return to the incubator for at least 5 days.

5. From the second day, observe cells daily under an inverted microscope and search for infected cells, which should swell, stop dividing, and appear uniformly rounded with enlarged nuclei. Cells expressing the DsRed fluorescent protein should be present in the positive control while confluent cell growth should be seen in the negative control.

6. After 5 days, when more than 50 % of cells in the positive control express the DsRed fluorescent protein, carefully collect the supernatant by centrifugation at $200 \times g$ for 10 min (P0) and store at 4 °C, protected from light. This P0 virus stock should be stable for at least 3–6 months.

3.2.2 Early Evaluation of Protein Expression

P0 can be used for an initial screening to determine if the protein(s) of interest are expressed using Western blot analysis and for small-scale purification. When possible, experiments are performed in semi-denaturing or denaturing conditions to provide a first indication of the expression level independently of protein solubility. The protocol below is used for purification of histidine-tagged proteins under denaturing conditions (6 M urea, *see* **Note 6**). Note that the titer of P0 can be very low and not sufficient for reliable evaluation of expression levels.

1. Add 1.5×10^6 Sf9 cells (e.g. 0.75 mL of Sf9 cells at 2.0×10^6 Sf9 cells/mL) grown in insect cell culture medium (TNM-FH + 10 % FBS) in each well of a 6-well plate, and let the cells adhere for 20 min at 27 °C.

2. Discard the medium, add 300 microliters of fresh medium and 150 µL of P0 to attached cells. After 1 h incubation at 27 °C,

add 3 mL of insect cell medium and return to the incubator for 48 h.

3. Resuspend cells by gently pipetting up and down. Infected insect cells are very sensitive and should be resuspended gently to minimize cell lysis at this step. Centrifuge the resuspended cells at $200 \times g$ for 10 min at 4 °C and discard the supernatant.

4. Wash cell pellet with 1 mL of PBS + glycerol 10 % followed by centrifugation at $200 \times g$ for 10 min. Discard the supernatant and store the pellet at –80 °C or proceed immediately.

5. Resuspend cells in 0.8–1.5 mL of lysis buffer supplemented with EDTA-free protease inhibitor cocktail at recommended concentration and sonicate them for 30 s with 3 mm probe at 20 % intensity. Alternatively, use a lysis buffer containing 1 % Tween 20 and 400 U/mL of DNAse Type I, and shake the mixture for 15 min at RT. Take 15 µL aliquots and add 5 µL of 4× SDS loading dye (total extract).
At this stage, samples can be processed in individual 1.5 mL tubes or transferred into 24 deep-well blocks for parallel processing.

6. Centrifuge the lysate at $6,500 \times g$ for 60 min at 4 °C and optionally filter the supernatant using a 0.2 µm filter plate. Take a 15 µL aliquot and add 5 µL of 4× SDS loading dye (soluble extract).

7. Incubate the soluble extract with equilibrated affinity resin at 4 °C. Use 25 µL of resin for batch purification and incubate for 15–120 min with slow end-over-end mixing. Use 100 µL for spin-column or filter-based chromatography.

8. Wash the resin three times with 800 µL of washing buffer (which can be supplemented with 10 mM imidazole to limit a-specific binding in case of IMAC affinity) and elute with 50 µL of elution buffer for batch purification or with 200 µL of elution buffer for spin-column or filter-based chromatography. Keep all buffers on ice. Take a 15 µL aliquot from each elution and add 5 µL of 4× SDS loading dye (elutions).

9. Analyze the different samples using SDS-PAGE with Coomassie staining and/or with Western blotting in case of low expression levels.

3.2.3   Virus Amplification from suspension cultures

Amplification of the recombinant P0 virus stock is necessary before large-scale recombinant production. Insect cells are infected with a small quantity of virus which will replicate, release new viral particles that will infect more cells, and so on. It is of major importance to ensure that cells are healthy, in exponential growth phase, and that they are infected at a very low multiplicity of infection (MOI) (less than one virus per cell). In these conditions, few cells are infected initially; the virus replicates to release the budded virus,

which then infects more cells and so on. Multiple rounds of replication occur and high virus titers can be obtained. When cells are infected at high MOI, all the cells are initially infected and only a single round of replication will occur, giving a poor virus amplification (*see* **Note 7**). After the first round of amplification a high-titer viral stock called P1 is obtained. P2 and P3 viral stocks correspond to the second and the third round of amplification respectively.

Virus amplification can be performed with either adherent or suspension cultures. The protocol below is used for amplifying viruses harvested from co-transfection in suspension using Sf9 cells grown in serum-containing medium (*see* **Note 5**). For non-experienced users, we recommend to start with viruses obtained from a transfer vector that will co-express your target protein with a fluorescent protein such as DsRed (Table 2) to monitor amplification.

1. Add 250 μL of P0 to a spinner flask containing 250 mL of TNM-FH + 10 % FBS seeded with Sf9 cells at 1.0×10^6 cells/mL in exponential growth phase and incubate at 27 °C with agitation (100 rpm). Ideally, different volumes of P0 can be tested.

2. Observe cells daily for signs of infection under an inverted microscope (or with a fluorescent microscope if appropriate). Infected cells should swell, stop dividing, and appear uniformly rounded with enlarged nuclei. During infection cell size can increase up to 20–30 % (Fig. 2).

3. Incubate at 27 °C for 5–7 days and harvest virus when 50 % of cells or more collapse. Centrifuge at $1,000 \times g$ for 10 min to remove cell debris and store the supernatant at 4 °C, protected from light. This is the P1 virus stock.

3.2.4 Virus Amplification from adherent cultures

P1 is sufficient for initial protein expression studies. If large volumes of virus are required, repeat the co-transfection or amplify P1 to obtain P2 (and eventually P3). Viruses may also be amplified from adherent cultures when smaller volumes are needed.

1. Seed a T175 flask with 20×10^6 Sf9 cells (e.g., 10 mL at 2.0×10^6 cells/mL) grown in TNM-FH + 10 % FBS medium and add 25–100 μL of the P0 viral stock and incubate at 27 °C to favor infection.

2. After 1 h incubation, add TNM-FH + 10 % FBS medium up to 25 mL and incubate the cells for 5–7 days more. Observe cells for signs of infection under an inverted microscope (or with a fluorescent microscope, if appropriate).

3.3 Optimization and Large-Scale Expression of Proteins

The yield of the recombinant protein as well as its quality is affected by a myriad of factors. Thus, once a concentrated baculovirus stock has been amplified, optimization experiments are required before large-scale expression and purification can start.
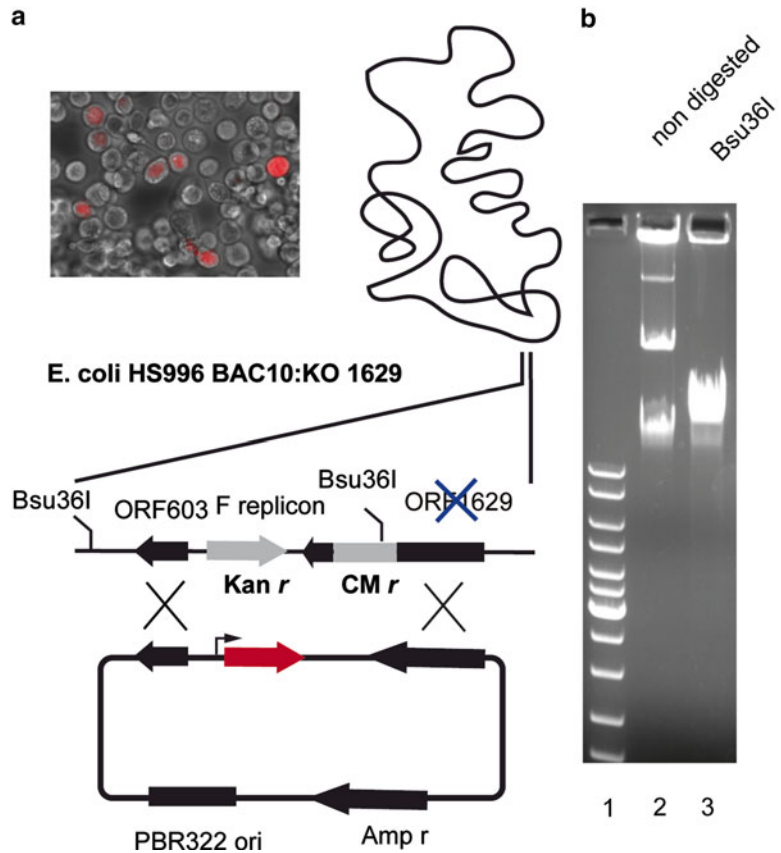
Fig. 2 Construction of a recombinant baculovirus by homologous recombination. (**a**) Linearized viral DNA (purified BAC10:KO1629) is co-transfected with a transfer vector containing the cDNA (DsRed) to be inserted into the viral genome. Homologous recombination replaces the F replicon by the expression cassette for the gene of interest (DsRed) and restores the essential ORF1629, allowing viral replication and expression of the desired protein. 2–3 days after transfection, fluorescence of the DsRed protein should be visible. In this case only a few cells are infected. (**b**) Analysis of purified BAC10:KO1629 on a 0.8 % agarose gel before (*lane 2*) and after digestion with Bsu36I (*lane 3*)

3.3.1 Infection Parameters and First Evaluation of Solubility

Experiments start with Sf9 or Sf21 cells grown in suspension using serum-free medium. Small-scale purifications with two lysis/purification buffers will provide first information on protein solubility and expression yield.

For each virus, the optimization of growth and infection requires a careful analysis. Key parameters of the process are the amount of virus and the time of infection: (1) When infecting cells for protein production, the object is to get all cells infected synchronously and, therefore, we can use more viral particles than cells (*see* **Note** 7). Typically, conditions which correspond to MOIs in the range of 0.5–10 are tested. (2) The best time to harvest depends on the

a

Un-infected cells (Day 0)

b
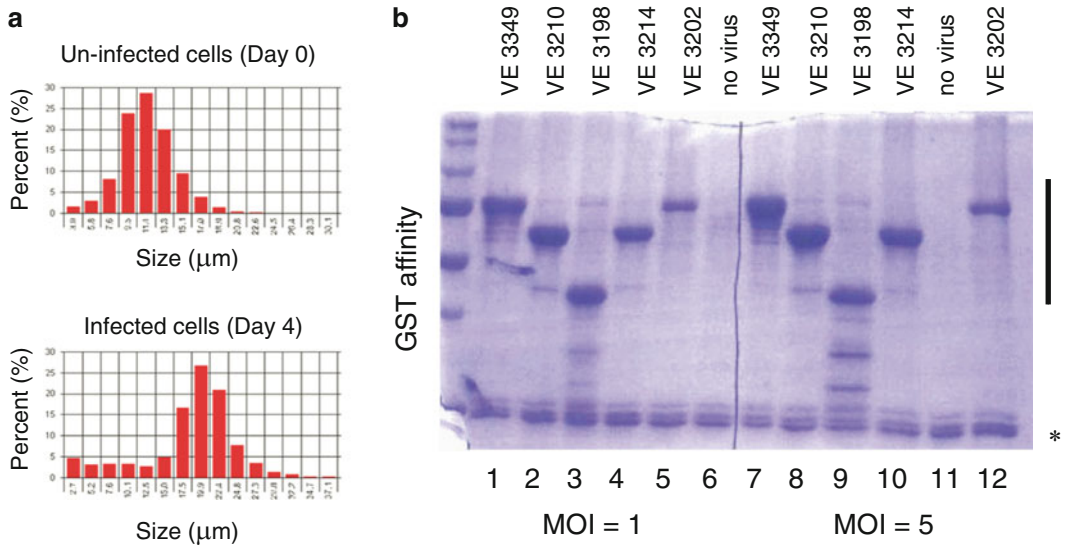
Fig. 3 Amplification and optimization. (**a**) To amplify a P0 virus stock, 250 mL of Sf9 cells seeded at 1×10^6 cells/mL in TNM-FH + 10 % FBS were infected by 250 μL of transfection supernatant and the size distribution of cells daily analyzed using a cell counter. Cell diameter histograms before (un-infected cells, day 0) and after 4 days infection (infected cells, day 4). The average diameter increases from 11.5 (un-infected cells) to 19 nm. (**b**) Five baculoviruses for expression of a target protein in fusion with a GST affinity tag have been generated and the corresponding P1 viral stock prepared. To test expression, Sf21 cells were infected by increasing volumes that correspond to MOIs of 0.5, 2.0, and 10. Aliquots were harvested 48, 72, and 96 h after incubation and affinity purified. A representative SDS-PAGE analysis is shown. *Asterisk* corresponds to endogenous GST

nature of the target protein. Cells are usually analyzed 48, 72, and 96 h postinfection. Some stable proteins might accumulate to high levels 72 or 96 postinfection while others, sensitive to degradation, will need to be collected after 24 or most commonly 48 h (Fig. 3).

Protein expression may also be affected by the cell line and expression obtained using Sf9, Sf21, or High-Five cells can be compared. New cell lines such as superSF9-1™ (OET) or superSF9-2™ (OET), which feature a prolonged expression time (suited for stable proteins) or an intense peak of recombinant protein production (suited for toxic or highly unstable proteins), are worth being tested.

1. Seed 3×250 mL Erlenmeyer flasks with exponentially growing Sf9 or Sf21 cells at 1.0×10^6 cells/mL in 90 mL of appropriate medium. Add 0.45, 1.8, or 9 mL P1 stock to each flask, close the flask with an air pore sheet or with a vented cap, and incubate it at 27 °C. Assuming a titer of 1.0×10^8 pfu/mL, these conditions correspond to MOIs of 0.5, 2, and 10.

2. After 48 h, sample 30 mL of cell suspension and return the flask to the shaker. Centrifuge the cell suspension at $200 \times g$ for 10 min in 50 mL tubes, resuspend cells in 3 mL PBS + 10 %

glycerol, split into 3 aliquots of 1 mL (each aliquot corresponds to 10 mL culture), centrifuge again, and store pellets at –80 °C. When several constructs are tested simultaneously, transfer aliquots into 24 deep-well blocks and use the block for centrifugation and storage.

3. Repeat **step 2** after 72 and 96 h.

4. Lyse and purify the first aliquot under denaturing condition as described in Subheading 3.2.2 (**steps 5–9**). This will provide an indication of the total yield independently of protein solubility.

5. Proceed with the second and third aliquots in native or semi-denaturing conditions. For a histidine-tagged intracellular protein and IMAC affinity purification, we would use a buffer containing 150 mM NaCl, but without detergent, for the purification of the second aliquot and a buffer containing 1 M NaCl and 1 % Tween 20 for the third one (*see* **Note 6**).

This will provide preliminary information on protein solubility and a quantitative indication of the amount of soluble protein that one should be able to purify per liter of culture in native conditions.

3.3.2 Solubility Screen

Preparation of the cell lysate is a critical step which often requires a second round of optimization to identify a suitable lysis buffer. Optimal conditions should maximize solubility and stability of the target protein while minimizing oxidation, unwanted proteolysis, and aggregation. If the protein of interest can be tested in vitro, screening should include the use of a functional assay to control/optimize activity of the recombinant protein.

In absence of prior knowledge, we would systematically vary the pH, test a few detergents (ionic and nonionic detergent) and different salt concentrations (low, medium, and high salt). High ionic strength enhances solubility of many proteins but is not compatible for a number of applications (ion exchange chromatography, native mass spectrometry, in vitro DNA binding assays, etc.). Indeed, a substantial fraction of proteins precipitate when the salt concentration is reduced to physiological levels. For intracellular proteins, care should be taken to maintain a reducing environment (*see* **Note 6**). Inclusion of glycerol (10 %), detergents, and/or addition of specific ligands (if known and available) can help and enhance solubility and stability.

Below, a guide to design a buffer screening with 24 conditions to test protein solubility is detailed (Table 3):

1. Based on previous experience and knowledge of the target protein, identify a set of 24 conditions for a buffer screen and prepare 3 deep well plates of 24 wells with the corresponding solutions. The first plate will contain the solutions used to

Table 3
Example of solubility screen

	–	2 mM CHAPS	0,1 % Triton X-100	0.5 M SB 201	0.4 mM ZW 3–14	10 % glycerol
50 mM NaCl						
150 mM NaCl						
500 mM NaCl						
1,000 mM NaCl						

Salt concentrations and detergents are systematically varied; pH is fixed (Tris/HCl 30 mM, pH 7.8). Triton X-100 is a nonionic detergent, while the CHAPS, Sulfobetaine 201 (SB 201), and zwittergent 3–14 (ZW 3–14) are zwitterionic detergent. The glycerol helps to stabilize the protein in solution

resuspend the pellets. Prepare 1.5 mL per condition and add the EDTA-free protease inhibitor cocktail. The second plate will contain solutions used to equilibrate and wash the beads. Prepare 2.5 mL per condition. The third plate will contain the elution buffer. Prepare 250 µL per condition.

2. Seed a 2 L Erlenmeyer flask with 500 mL medium and infect cells using parameters determined previously in Subheading 3.3.1. Harvest cells, wash them in PBS + 10 % glycerol, and resuspend them in 50 mL of the same buffer. Aliquot the resuspended cells in two 24-deep well plates with 1ml per well, centrifuge at $200 \times g$ for 20 min, and snap freeze the plate after removal of the supernatant.

Resuspend cells in 0.8–1.5 mL of the 24 different lysis buffers. Lyse and purify as described in Subheading 3.2.2 (**steps 5–9**). This should identify suitable condition for the preparation of the cell lysate and provide an indication of the total yield of soluble protein per L of culture. If applicable, test the purified proteins using functional assays. There are no generic recipes to solve expression/solubility problems. Many eukaryotic proteins are components of multiprotein assemblies. They are often insoluble when expressed individually and require the presence of an interacting protein for folding and stability, which can be extremely challenging and require specific technologies to overcome the encountered bottlenecks [7, 21, 22]. Co-expression of the target protein with its partners often helps. Proteins can be expressed either from a single baculovirus carrying multiple foreign genes or from co-infection of multiple baculoviruses each carrying a single foreign gene [23]. The second approach, which requires minimal efforts if partner proteins are known and the corresponding viruses available, is adapted for expression screening. Along the same lines, proteins can be stabilized by the binding of a small molecule ligand. If a sufficiently soluble, cell-permeable ligand is known and available, it can be used to stabilize and promote solubility of recombinant proteins.

Ligands can be directly added to the culture medium or included in the lysis buffer.

Once the culture conditions and composition of the buffer used for the preparation of the clarified lysate have been determined, productions can be scaled up. Large-scale experiments can be performed with 2 L Erlenmeyers containing 500 mL culture and up to 8 flasks can be used in parallel. For larger volumes, devices suitable for production at 5–20 L scale such as wave bags as well as further optimization and standardization are required. This includes, for example, a setup of procedures for virus stock preservation and infection of large volumes of culture [24].

3.4 Handling Insect Cell Cultures

Successful expression work with the BEVS drastically depends on the health of insect cell cultures that should be carefully maintained and tested on a regular basis. The most common cell lines used for BEVS applications are derived from *Spodoptera frugiperda* (Sf9, Sf21) or *Trichoplusia ni* (High-Five™). *Lepidopteran* cells can be cultured between 25 and 30 °C with an optimum at 27 °C in phosphate buffered media, which means that CO_2 incubators are not required. Insect cell and baculovirus work requires a basic familiarity with insect cell physiology [25] and general cell culture methods (*see* **Note 8**). Invertebrate cell cultures are sensitive to environmental factors and we recommend a careful control of growth characteristics to determine the accurate cell density range where mid-log phase of growth falls based on actual culture conditions (i.e., incubator, equipment, technicians).

3.4.1 Thawing Cells

The following protocol describes initiation of cultures from frozen vials.

1. Equilibrate the appropriate growth medium at room temperature.

2. Select a cryovial from liquid nitrogen storage and thaw in a 37 °C water bath until only a small ice crystal remains.

3. Wipe or spray the cryovial with 70 % ethanol and take it under the hood.

4. Transfer cells into a sterile 50 mL tube containing 45 mL fresh medium tube, centrifuge at $50 \times g$ for 15 min at 20 C°, discard the supernatant that contains DMSO, and gently suspend cells in 5 ml of appropriate growth medium.

5. Seed the entire suspension into a T75 flask containing 15 mL of medium or to a 100 mL Erlenmeyer with 20 mL as appropriate (Table 4). The initial viable cell density in the culture should be at least 2.0×10^4 cells/cm² for adherent cells and $3–5 \times 10^6$ viable cells/mL for suspension (*see* **Note 9**).

6. The following information should be recorded in a cell notebook: cell line and batch, medium, passage number and date, density to which the culture has been split (dilution factor or cell density).

Table 4
Culture parameters for Sf9, Sf21, and High-Five™

	Sf9	Sf9	Sf9	Sf21	High-five™
Adherent/ Suspension	Adherent	Suspension	Suspension	Suspension	Suspension
Medium	TNM-FH, 10 % FBS	TNM-FH, 10 % FBS	SF900-II	SF900-II	Express-five
Max cell densities	1.50×10^5 cells/cm² 80 % confluency	$2–3 \times 10^6$ cells/mL	$2–3 \times 10^6$ cells/mL	10×10^6 cells/ mL	3×10^6 cells/ mL
Seed density	$2–4 \times 10^4$ cells/cm² Dilution 1:3 or 1:4	0.50×10^6 cells/mL	0.50×10^6 cells/mL Dilution 1:4	0.50×10^6 cells/mL Dilution 1:10	0.30×10^6 cells/mL Dilution 1:5
Doubling time	24–30 h	24–30 h	24–30 h	18–24 h	18–24 h
Agitation	na	100 rpm	120 rpm	120 rpm	120 rpm
Freezing cells					
Cryopreservation medium	70 % TNM-FH, 20 % FBS, 10 % DMSO	na	60 % SF900-II, 30 % FBS, 10 % DMSO	60 % SF900-II, 30 % FBS, 10 % DMSO	45 % Express Five, 45 % conditioned Express Five, 10 % DMSO
Density(cell/vial)	$>10 \times 10^7$	na	$>10 \times 10^7$	$>10 \times 10^7$	$>0.30 \times 10^7$

7. Place them into the incubator or shaker and proceed as indicated below. Cells should be dividing within 2 days. Passage the cells at least two times before using in other applications.

3.4.2 Maintenance of Cell Cultures

Insect cells can be grown as monolayers or as suspension cultures with spinner flasks or Erlenmeyers (Table 4). Cells adapted to serum-free media are grown in Erlenmeyers with an orbital shaker while cells grown in serum-supplemented media are generally cultivated as monolayers or in spinner flasks with gentle stirring.

We use Sf9 cells adapted to a serum-containing medium for generation of viruses, initial evaluation of protein expression, and virus amplification. Sf9 adherent cells are maintained in T75 flasks and passaged every 2–3 days (*see* **Note 10**).

1. Observe cells using an inverted microscope to verify that they look healthy (no visible contamination, limited number of floaters, etc.) and that the monolayer has reached 90 % confluence.

2. Detach cells by tapping the flask or by sloughing (streaming medium over the monolayer with a pipette to dislodge cells).

3. Seed new flask at a density of $2–4\times10^4$ cells/cm^2. This corresponds to a 1:3 dilution, i.e., seeding 5 ml of cell suspension into a T75 flask containing 10 ml of fresh medium (TNM-FH + 10 % FBS). Once cells have attached, verify that initial cell density is close to 30 %. Depending on needs, nearly confluent monolayer cells can be split at any ratio between 1:2 and 1:20. and/or

4. Seed a spinner flask a density of 5×10^5 cells/mL. For example, add the cell suspension from 6 to 8 T75 flasks into a 2 L spinner flask containing 400–500 mL of fresh medium (TNM-FH + 10 % FBS).

5. Incubate at 27 °C for 48 h with appropriate agitation if relevant. Aliquots can be stored for 1–4 weeks at 4 °C as backup. Don't forget to fill the cell notebook.

Sf9, Sf21, and High-Five™ cells adapted to serum-free medium are maintained in suspension using an orbital shaker (*see* **Note 11**) and are used when expression of the recombinant protein has been validated for optimization and large-scale productions. Cultures are usually maintained in 250 mL disposable, sterile Erlenmeyer flasks containing 50–100 mL of medium and passaged every 2–3 days.

1. Take an aliquot of the stock culture, count cells, and determine their viability. If the cell density is below 2.0×10^6 cells/mL, continue to grow the cells.

2. When the cell density reaches 2.0×10^6 cells/mL, passage the cells at a seeding density of $0.3–0.50\times10^5$ cells/mL into the desired flask and scale up accordingly.

3. Incubate at 27 °C for 48 h with appropriate agitation. Aliquots can be stored for 1–4 weeks at 4 °C as backup. Don't forget to fill the cell notebook.

3.4.3 Freezing Cells

Once a culture is fully adapted to the culture conditions and growth medium, it is essential to establish a frozen master cell seed stock that should be prepared with the lowest possible passage.

1. Prepare appropriate freezing medium for the cell line and cryogenic storage vials (i.e., cryovials) by labeling each vial appropriately with the cell name, density, and date.

2. Measure the cell density and viability of the stock cells culture to cryopreserve, transfer cells from the culture flask into sterile 50 mL conical tubes, and centrifuge at 50 ×*g* for 15 min at 20 °C to pellet.

3. Remove spent medium and resuspend the cells in an appropriate volume of cold freezing medium to achieve the desired cell density to add to each cryovial (Table 4).

4. Dispense aliquots of the cell suspension into labeled cryovials. Frequently and gently mix the cell suspension to maintain a homogeneous solution.

5. Place the ampoules in a small Styrofoam box (freezing rate −1 °C/min) and place it in a −80 °C ultralow freezer ON. Store the cell vials in liquid nitrogen.

6. Qualify frozen cells by thawing one vial and testing the thawed cells for bacterial, fungal, and mycoplasma contamination, normal cell morphology, robust growth, and capacity to express recombinant protein.

3.5 Linearized Viral DNA for Co-transfection

Homologous recombination of a transfer vector with linearized baculoviral DNA is the simplest system to generate recombinant virus but the use of this technology was limited by the fact that isolation of recombinant from parental virus is required. The inactivation of the essential viral gene (ORF1629) lying adjacent to the PH locus used for recombination resulted in a renewal of this technology, which is now labor- and time-effective.

Ready-to-use and genetically optimized linearized baculovirus DNA can be purchased from a number of sources. A number of these contain modifications of viral genome, such as knock-out of viral-nonessential encoded genes that might interfere with expression of the recombinant proteins or contain helper modules for expression of molecular chaperones that can enhance folding of target proteins in the endoplasmic reticulum (Calreticulin and protein disulfide isomerase). They can contain reporter proteins to to follow the infection. For standard applications and/or initial construct evaluation, we mainly use the bacmid BAC10:KO1626 (Zhao et al. 2003) as source of viral DNA. It consists of the wild-type AcMNPV genome with a low copy bacterial replicon flanked by 2 Bsu36I restriction sites. These elements as well as kanamycin and chloramphenicol resistance markers are inserted at the polyhedrin locus. The bacmid is produced in *E. coli*, purified using a plasmid/bacmid isolation kit, and linearized to enhance recombination efficiency. A 400 mL culture typically yields 50 μg of purified bacmid, which is sufficient for 50 transfections in 6-well plate or 200 in 24-well format.

1. Inoculate 10 mL of LB medium containing 50 μg/mL kanamycin and 34 μg/mL chloramphenicol with a single colony DH10B BAC10:KO1629 from a freshly peaked plate or directly with 50 μL of a glycerol stock and incubate at 37 °C for 4–8 h to prepare a fresh pre-culture.

2. Seed 4 mL of the pre-culture into a 2 L Erlenmeyer containing 400 mL LB with the appropriate combination of antibiotics and incubate at 37 °C for 12–16 h. The culture should be grown to an $OD_{600\,nm}$ of 2.0–3.0 which corresponds to 1.2–1.8 g of wet cells and is needed for one bacmid purification (*see* **Note 12**).

3. Extract bacmid DNA using a commercially available purification kit that is suitable for purification of a 135 kbp DNA and follow carefully the manufacturer's instructions. In particular, do not use overgrown cultures, as starvation of cells can lead to degradation of large constructs. Adapt the volumes of buffers used to cell mass in order to optimize purification and treat lysates extremely carefully as bacmid DNA, due to its large size, is sensitive to shearing—i.e., use large orifice pipette tips, and do not vortex.

   Since the bacmid has to be sterile, manipulations should be performed under a sterile hood.

4. The procedures include an isopropanol or an ethanol precipitation to concentrate DNA which is followed by a washing step with 70 % ethanol to remove traces of salts (*see* **Note 13**). After centrifugation, carefully remove ethanol from the tube with a pipette tip, air-dry the pellet at RT for 2–4 h, but not ON as it might be difficult to dissolve it when over-dried. Resuspend the dried bacmid pellet with 200 μL of sterile ultrapure H_2O and incubate it ON at 4 °C. Do not vortex.

5. Check the homogeneity of the bacmid solution by pipetting up and down slowly and if the solution is too viscous add sterile H_2O until a homogenous solution is obtained. Take an aliquot for UV quantification. Adjust the concentration to 125 μg/mL and store the sample at 4 °C in a sterile 1.5 mL tube. As a quality control, we verify that the $OD_{260\ nm}/OD_{280\ nm}$ ratio is close to 1.8 and we analyze the bacmid before and after restriction with Bsu36I and BamHI on a 0.8 % agarose gel. DNA (final concentration 100 μg/mL in the appropriate buffer) is digested using 10 U of restriction enzyme per μg DNA during 2–4 h at 37 °C. Neither high nor low molecular weight nucleic acid should be visible in the gel.

6. As homologous recombination is more efficient with linear than with circular DNA, the purified bacmid is finally linearized using Bsu36I at preparative scale using the same experimental conditions as described in **step 5**.

   For the digestion of 25 μg bacmid, we mix under a cell culture hood 200 μL of bacmid (125 μg/mL), 25 μL 10× NEB3 buffer, and 25 μL Bsu36I (NEB) (10 U/μL). After an incubation of 2–4 h at 37 °C, an aliquot is analyzed by gel electrophoresis to control digestion before heat inactivation of Bus36I (20 min at 72 °C). If digestion is not complete add 12.5 μL Bsu36I (NEB) (10 U/μL) and re-incubate for 2 h.

7. The linearized bacmid can be stored at 4 °C for 1–2 month. Alternatively, prepare aliquots of 6.5 μg (65 μL) which are sufficient for 6 transfections in a 6-well plate format or 24 transfections in a 24-well plate format and freeze them at 20 °C. Once an aliquot was thawed, keep DNA at 4 °C and do not re-freeze again.

4   Notes

1. If restriction/ligation cloning cannot be used or fails, we generally try Sequence and Ligation Independent Cloning (SLIC) as described in [26, 27]. Use an acceptor vector double-digested with NdeI and BamHI and a PCR product amplified with a high-fidelity polymerase and primers with 30 bp of homology to the vector. We recommend the use of web tools such as SODA (http://slic.cgm.cnrs-gif.fr/) or NebBuilder (http://nebuilder.neb.com/) for primer design.

2. It is critical to obtain a maximal yield of the transfer vector double digested. We can first digest the plasmid with each enzyme independently using 5U per µg DNA for 2 h and analyze the result on an agarose gel. The two reaction mixes are pooled and the same amount of the other enzyme is added to the tube which is incubated for another 2 h or ON. The phosphatase (typically SAP) is directly added to the reaction and incubated for 1 h before inactivation. Follow the manufacturer's instructions.

3. Co-transfection is a critical step in the expression pipeline, but this step can easily be optimized using a fluorescent reporter protein. Don't hesitate to run an optimization plate where the cell seeding density, the amount of linearized viral DNA and of transfer vector, and the DNA/transfection agent vary.

4. The virus titer is an estimation of the concentration of active viral particles, which can be determined using plaque assays, end-point dilutions, or Q-PCR [28–31]. For high-titer virus stocks, values between 0.50×10^8 and 5×10^8 pfu/mL are expected.

5. We use a serum-containing medium for generation of viruses, initial evaluation of protein expression, and amplification of high-titer virus stocks. This is not a requirement, but if a serum-free medium is used for amplification, 10 % FBS should be added to stabilize the virus stocks. Virus can be stored at 4 °C and protected from light for 6–12 months or longer. However, after more than 3–4 months, it is recommended to recalibrate experiments before use the stock or to re-amplify. Sf9 or Sf21 but not High Five cells are suitable to produce or amplify virus.

6. IMAC affinity resins are compatible with high urea concentration (up to 6 M). For purifications with GST, Strep, or FLAG affinity resins, nonionic detergents such as Triton-X100 or Tween 20 should be used (up to 1 %, depending on the resin). Some proteins have to be manipulated in presence of reducing agent for stability and/or solubility. In that case, carefully check resin compatibility.

7. The MOI is defined as the average number of viral particles per cell that is equal to the ratio: (number of viral particles)/(number of cells). For efficient amplification one should infect cells using a low MOI (typically 0.05–0.2) which ensures that only a few cells are infected initially and limits accumulation of defective interfering particles (DIPs), i.e., partial genomes packaged by complementation from intact genomes co-infected in the same cell [32]. For protein production, all cells should be infected simultaneously and high MOIs, usually above 1.0 and up to 10.0, are used.

8. Insect cells and viruses are handled in laminar flow hood under aseptic conditions without antibiotic as these can mask low levels of contamination. However, the addition of penicillin (50–100 U/mL) and streptomycin (50 µg/mL) or gentamicin (50 µg/mL) can be useful to face a contamination. Material taken inside the hood should be treated with 70 % ethanol and taken out properly decontaminated (autoclave or bleach). Do not use soap when cleaning vessels. We recommend washing with commercially available cleaners and wash intensively with MQ H_2O prior autoclaving.

9. Cell viability can be evaluated with trypan blue. Mix one volume of cells with one volume of a 0.1 % stock solution of trypan blue (in PBS or other isotonic salt solution). Nonviable cells will take up Trypan blue. Healthy, log-phase cultures should contain more than 97 % unstained viable cells.

10. Supplementation of media with serum promotes cell growth, provides shear force protection, and prolongs stability of virus stocks. However, it has a non-negligible associated cost and leads to excessive foaming with subsequent cell damage. In addition, serum batches can exhibit significant lot to lot variability and should be carefully tested before use.

11. Insect cells maintained under serum-free conditions may attach very tightly to surface and require additional effort to detach. To dislodge the cells, you may need to shake the flask vigorously two to three times using a wrist-snapping motion or use a cell scraper.

12. By rule of thumb, 1 L of *E. coli* culture with an $OD_{600 \, nm}$ of 1 consists of 1.0×10^{12} cells and yields about 1.5 g cell wet weight. We usually grow several cultures in parallel and prepare cell pellets in 1.5 g aliquots which are either processed immediately or stored at −20 °C.

13. To concentrate DNA, do not use membrane-based tools such as Nucleobond Finalizer™ (Macherey Naeglel) which are not recommended for constructs larger than 50 kbp.

Acknowledgments

References

1. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol 10:411–421

2. Brondyk WH (2009) Selecting an appropriate method for expressing a recombinant protein. Methods Enzymol 463:131–147

3. Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nat Biotechnol 23:567–575

4. Rosser MP, Xia W, Hartsell S et al (2005) Transient transfection of CHO-K1-S using serum-free medium in suspension: a rapid mammalian protein expression system. Protein Expr Purif 40:237–243

5. Aricescu AR, Assenberg R, Bill RM et al (2006) Eukaryotic expression: developments for structural proteomics. Acta Crystallogr D Biol Crystallogr 62(Pt 10):1114–1124

6. Nettleship JE, Assenberg R, Diprose JM et al (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. J Struct Biol 172:55–65

7. Vijayachandran LS, Viola C, Garzoni F et al (2011) Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J Struct Biol 175:198–208

8. Assenberg R, Wan PT, Geisse S et al (2013) Advances in recombinant protein expression for use in pharmaceutical research. Curr Opin Struct Biol 23:393–402

9. Harrap KA (1972) The structure of nuclear polyhedrosis viruses. I. The inclusion body. Virology 50:114–123

10. Smith GE, Summers MD, Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol Cell Biol 3:2156–2165

11. Ayres MD, Howard SC, Kuzio J et al (1994) The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. Virology 202:586–605

12. Fraser R, Heslop VR, Murray FE et al (1986) Ultrastructural studies of the portal transport of fat in chickens. Br J Exp Pathol 67:783–791

13. Smith GE, Fraser MJ, Summers MD (1983) Molecular Engineering of the Autographa californica Nuclear Polyhedrosis Virus Genome: Deletion Mutations Within the Polyhedrin Gene. J Virol 46:584–593

14. Roy P, Noad R (2012) Use of bacterial artificial chromosomes in baculovirus research and recombinant protein expression: current trends and future perspectives. ISRN Microbiol 2012:628797

15. Luckow VA, Lee SC, Barry GF et al (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. J Virol 67:4566–4579

16. Hitchman RB, Possee RD, Crombie AT et al (2010) Genetic modification of a baculovirus vector for increased expression in insect cells. Cell Biol Toxicol 26:57–68

17. Zhao Y, Chapman DA, Jones IM (2003) Improving baculovirus recombination. Nucleic Acids Res 31:E6–6

18. Abdulrahman W, Uhring M, Kolb-Cheynel I et al (2009) A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. Anal Biochem 385:383–385

19. Walls D, Loughran ST (2011) Tagging recombinant proteins to enhance solubility and aid purification. Methods Mol Biol 681:151–175

20. Waugh DS (2005) Making the most of affinity tags. Trends Biotechnol 23:316–320

21. Perrakis A, Musacchio A, Cusack S et al (2011) Investigating a macromolecular complex: the toolkit of methods. J Struct Biol 175:106–112

22. Berger I, Blanco AG, Boelens R et al (2011) Structural insights into transcription complexes. J Struct Biol 175:135–146

23. Sokolenko S, George S, Wagner A et al (2012) Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. Biotechnol Adv 30:766–781

24. Wasilko DJ, Lee SE, Stutzman-Engwall KJ et al (2009) The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. Protein Expr Purif 65:122–132

25. Lynn DE (2007) Routine maintenance and storage of lepidopteran insect cell lines and baculoviruses. Methods Mol Biol 388:187–208

26. Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59

27. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256

28. Cha HJ, Gotoh T, Bentley WE (1997) Simplification of titer determination for recombinant baculovirus by green fluorescent protein marker. Biotechniques 23(782–4):786

29. Hopkins R, Esposito D (2009) A rapid method for titrating baculovirus stocks using the Sf-9 Easy Titer cell line. Biotechniques 47:785–788

30. Hitchman RB, Siaterli EA, Nixon CP et al (2007) Quantitative real-time PCR for rapid and accurate titration of recombinant baculovirus particles. Biotechnol Bioeng 96:810–814

31. Roldao A, Oliveira R, Carrondo MJ et al (2009) Error assessment in recombinant baculovirus titration: evaluation of different methods. J Virol Methods 159:69–80

32. Kool M, Voncken JW, van Lier FL et al (1991) Detection and analysis of Autographa californica nuclear polyhedrosis virus mutants with defective interfering properties. Virology 183:739–746

33. Artimo P, Jonnalagedda M, Arnold K et al (2012) *ExPASy: SIB bioinformatics resource portal*, in. Nucleic Acids Res W597–603

34. Biegert A, Mayer C, Remmert M et al (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. Nucleic Acids Res 34(Web Server issue):W335–W339

35. Mooij WT, Mitsiki E, Perrakis A (2009) ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. Nucleic Acids Res 37(Web Server issue):W402–W405

36. Dosztanyi Z, Csizmok V, Tompa P et al (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433–3434

37. Linding R, Russell RB, Neduva V et al (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31:3701–3708

38. Yang ZR, Thomson R, McNeil P et al (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21:3369–3376

39. Prilusky J, Felder CE, Zeev-Ben-Mordehai T et al (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438

40. Weyer U, Possee RD (1991) A baculovirus dual expression vector derived from the Autographa californica nuclear polyhedrosis virus polyhedrin and p10 promoters: co-expression of two influenza virus genes in insect cells. J Gen Virol 72:2967–2974

41. Belyaev AS, Roy P (1993) Development of baculovirus triple and quadruple expression vectors: co-expression of three or four bluetongue virus proteins and the synthesis of bluetongue virus-like particles in insect cells. Nucleic Acids Res 21:1219–1223

42. Berrow NS, Alderton D, Sainsbury S et al (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res 35:e45

43. Vijayachandran LS, Thimiri Govinda Raj DB, Edelweiss E et al (2013) Gene gymnastics: Synthetic biology for baculovirus expression vector system engineering. Bioengineered 4:279–287

Part VI

Recombinant Protein Production in Mammalian Cells

Chapter 11

Transient Expression in HEK 293 Cells: An Alternative to *E. coli* for the Production of Secreted and Intracellular Mammalian Proteins

Joanne E. Nettleship, Peter J. Watson, Nahid Rahman-Huq, Louise Fairall, Mareike G. Posner, Abhishek Upadhyay, Yamini Reddivari, Jonathan M.G. Chamberlain, Simon E. Kolstoe, Stefan Bagby, John W.R. Schwabe, and Raymond J. Owens

Abstract

Transient transfection of human embryonic kidney cells (HEK 293) enables the rapid and affordable lab-scale production of recombinant proteins. In this chapter protocols for the expression and purification of both secreted and intracellular proteins using transient expression in HEK 293 cells are described.

Key words Mammalian, HEK, Transient transfection, Secreted protein, Intracellular protein, Protein complexes, FLAG purification

1 Introduction

High-quality, pure proteins are important reagents for a wide variety of applications such as biochemical assays, protein-based therapeutics, and protein crystallography. *E. coli* remains as the most commonly used expression host for producing recombinant proteins for research purposes, for example, structural studies, due to its ease of use and relatively low cost. However, production of recombinant proteins in high yield from *E. coli* can be challenging due to low expression levels and poor solubility. This is particularly the case for mammalian proteins. Although expression of many human intracellular proteins has been tried in *E. coli*, about 65 % are either not expressed or expressed insolubly [1]. These problems may be overcome by using mammalian cells for protein production as these express the necessary chaperones for correct folding and contain the machinery for adding posttranslational modifications (PTMs). Mammalian cells also contain small

molecules and cofactors which may be required for protein expression or complex formation.

Two mammalian cell lines are routinely used for the production of recombinant proteins, Chinese hamster ovary (CHO) and human embryonic kidney (HEK 293) cells. Of these, HEK 293 cells have become the mammalian cell line of choice for lab-scale protein production due to their ease of culture and high transfection efficiency [2]. A useful variant of HEK cells is the 293T cell line which expresses the SV40 large T antigen. Expression vectors containing the SV40 origin of replication are episomally amplified within the 293T cells, which increases the plasmid copy number per cell and can lead to higher levels of transient expression [3]. A further variant of the HEK cell line is the FreeStyle™ HEK 293F cell line (Life Technologies, UK) in which the HEK 293 cells are adapted to suspension growth in FreeStyle™ 293 expression medium. The medium is designed to support high-density growth and has the advantage of allowing transfection without the need to change medium.

The use of the inexpensive cationic polymer polyethylenimine [4, 5] as the DNA-condensing reagent has meant that large-scale transient transfection of HEK 293 cells has become economically feasible and is routinely used for the production of secreted and cell surface glycoproteins (reviewed by Aricescu and Owens [6]).

In contrast to their use with secreted proteins, mammalian cells have not been used routinely for the production of intracellular proteins due to the relatively low levels of expression compared with insect or bacterial systems. However, by using highly selective purification methods, e.g., FLAG® tag [7] or HaloTag® [8], it is possible to achieve useful yields of intracellular proteins. Again, transient expression in HEK 293 cells offers a way of rapidly assessing the protein yield and quality. Subsequent production of stable cell lines, typically by co-selection, may be required to sustain and improve the production levels of a particular product.

In this chapter, protocols for the production of both secreted and intracellular proteins by transient transfection of HEK 293 cells are described. The methods are exemplified by reference to the production of the secreted protein, human serum amyloid P component (SAP), and the intracellular proteins, human brain-specific protein kinase C isoform protein kinase M zeta (PKMζ) and human histone deacetylase 3 (HDAC3) in complex with its activation domain from the SMRT corepressor (SMRT-DAD).

SAP is a plasma glycoprotein [9] which participates in the innate human immune system but also plays a role in the molecular pathology of diseases such as amyloidosis and amyloid-associated diseases such as Alzheimer's and type II diabetes [10]. SAP is of increasing clinical relevance as radiolabeled SAP is used for identifying sites of amyloid deposition [11], while drug development programs attempting to deplete serum levels (for treatment of amyloidosis) and also administer protein (for treatment of fibrosis) are currently underway [12, 13]. SAP contains an *N*-glycan and a disulfide bridge and is representative of proteins with these modifications.

PKMζ is a neuron-specific isoform of atypical protein kinase C (aPKC) that lacks the normal N-terminal regulatory region and therefore comprises just a kinase catalytic domain [14]. In vivo phosphorylation of PKMζ by PDK1 converts PKMζ into a conformation with high constitutive activity [15]. Although there is controversy as to the extent and nature of its role, PKMζ has been implicated in both memory [16] and pain [17]. PKMζ contains five cysteines (with the potential for disulfide bridge formation) and is activated via phosphorylation.

HDAC3 is a class I histone deacetylase (HDAC) that is involved in transcriptional regulation [18]. Like the other class I HDACs, HDAC3 requires recruitment to its cognate corepressor protein (SMRT) to have full enzymatic activity [19]. HDACs are important therapeutic targets for the treatment of cancer [20] and are involved in other diseases such as Alzheimer's and HIV [21, 22]. HDAC3 and its activation domain from the SMRT corepressor (SMRT-DAD) do not interact when expressed in bacterial cells but require expression in higher eukaryotes to form a complex. The HDAC3–SMRT-DAD complex is phosphorylated in the C-terminal region of HDAC3 and also acetylated (as determined by mass spectrometry). The structure of the HDAC3–SMRT-DAD complex revealed the presence of an Ins(1,4,5,6)P$_4$ molecule at the interface between HDAC3 and SMRT which is required for complex formation and activation [23].

To show the benefit of using mammalian rather than bacterial cells to express human proteins, SAP and PKMζ were tested for expression in both *E. coli* and HEK 293 cells [24–26]. In Fig. 1 it can be seen that SAP in a vector containing a signal sequence is expressed and secreted using HEK cells (Fig. 1, lanes 1 and 2) but
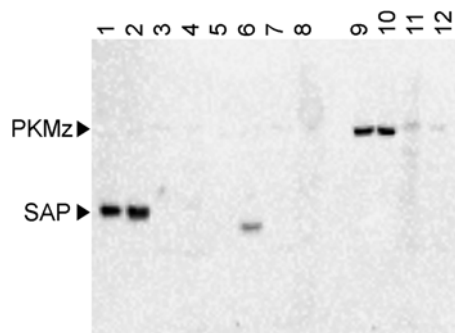
Fig. 1 Anti-His Western blot showing expression of SAP and PKMζ using HEK 293T cells and *E. coli. Lanes 1–4* show expression of SAP using a signal sequence with *lane 1* showing secreted product from HEK cells; *lane 2*, the HEK whole cell extract; *lane 3*, secreted product from *E. coli*; and *lane 4, E. coli* whole cell extract. In a similar way, *lanes 5–8* show expression of SAP without the signal sequence. *Lanes 9–12* show expression of PKMζ with *lane 9* showing whole cell extract from HEKs; *lane 10*, soluble extract from HEK cells; *lane 11, E. coli* whole cell extract; and *lane 12, E. coli* soluble protein extract (For information about the vectors used, *see* **Note 1**)

not in *E. coli* (Fig. 1, lanes 3 and 4). Without the signal sequence, some SAP is accumulated in the cells using HEK cell expression (Fig. 1, lane 6), but there is no expression in *E. coli* (Fig. 1, lanes 7 and 8). The band in lane 6 migrates lower than those in lanes 1 and 2 as no posttranslational modification has taken place. For PKMζ, expression can be seen in the cells and in the soluble extraction for both HEK cells and *E. coli* (Fig. 1, lanes 9–12); however, expression levels are higher using HEK cells (Fig. 1, lanes 9 and 10).

2   Materials

2.1   Protein Expression Using HEK 293T Cells

1. HEK 293T cells (ATCC no. CRL-1573—LGC Standards, UK).
2. Dulbecco's Modified Eagle's Medium (DMEM).
3. Fetal calf serum (FCS).
4. Non-essential amino acids (1:100).
5. L-Glutamine.
6. T175 tissue culture flask.
7. Plasmid DNA: The gene of interest needs to be contained in a vector compatible with mammalian expression systems (*see* **Note 1**).
8. Polyethylenimine (PEI) (25 kDa branched PEI). Prepare a 100 mg/mL stock solution in water before diluting to 1 mg/mL. Neutralize the solution with HCl, filter sterilize, and store at –20 °C in aliquots.
9. Kifunensine (Toronto Research Chemicals, Canada).
10. Expanded surface roller bottles.
11. PBS: 10 mM phosphate buffer, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.4.

2.2   Protein Expression Using HEK 293F Cells

1. FreeStyle™ HEK 293F cells (Life Technologies, UK).
2. Gibco® FreeStyle™ 293 expression medium (Life Technologies, UK).
3. 250 mL Erlenmeyer flask with vent cap.
4. Vent cap roller bottle.
5. Plasmid DNA: The gene of interest needs to be contained in a vector compatible with mammalian expression systems (*see* **Note 1**).
6. Polyethylenimine (PEI) (25 kDa branched PEI). Prepare a 0.5 mg/mL stock solution in water. Neutralize the solution with HCl, filter sterilize, and store at –20 °C in aliquots.
7. Kifunensine (Toronto Research Chemicals, Canada).
8. Dulbecco's phosphate buffered saline.

2.3 Purification of Secreted Proteins Using a His6 Tag	1. ÄKTA purification system such as ÄKTAxpress (GE Healthcare Life Sciences, UK).
	2. HiLoad 16/600 Superdex S75 or S200 (GE Healthcare Life Sciences, UK).
	3. Gel Filtration Buffer: 20 mM Tris–HCl, 200 mM NaCl, pH 8.0.
	4. Nickel Wash Buffer: 50 mM Tris–HCl, 500 mM NaCl, 30 mM imidazole, pH 8.0.
	5. Nickel Elution Buffer: 50 mM Tris–HCl, 500 mM NaCl, 500 mM imidazole, pH 8.0.
	6. 96 deep-well plate.
	7. 5 mL HisTrap FF column (GE Healthcare Life Sciences, UK).
2.4 Purification of Intracellular Proteins Using a 3×FLAG® Tag	1. FLAG Lysis Buffer: 100 mM potassium acetate, 50 mM Tris-HCl, pH 7.5, 5 % (v/v) glycerol, 0.3 % Triton X-100, Roche complete protease inhibitor tablet (Roche, UK).
	2. FLAG Wash Buffer 1: 100 mM potassium acetate, 50 mM Tris-HCl, pH 7.5, 5 % (v/v) glycerol, 0.3 % Triton X-100.
	3. FLAG Wash Buffer 2: 300 mM potassium acetate, 50 mM Tris-HCl, pH 7.5, 5 % (v/v) glycerol.
	4. FLAG Cleavage Buffer: 50 mM potassium acetate, 50 mM Tris-HCl, pH 7.5, 5 % (v/v) glycerol, 0.5 mM Tris(2-carboxyethyl) phosphine (TCEP).
	5. Anti-FLAG® M2 resin (Sigma-Aldrich, UK).
	6. FLAG Equilibration Buffer: 100 mM potassium acetate, 50 mM Tris-HCl, pH 7.5.
	7. His-TEV (tobacco etch virus) protease solution. A plasmid for the expression of His-tagged TEV protease using *E. coli* is available from Addgene, USA (www.addgene.org).
	8. Amicon Ultra Centrifugal filter (Millipore, UK).
	9. HiLoad 10/300 Superdex S75 or S200 (GE Healthcare Life Sciences, UK).
	10. Gel Filtration Buffer: 50 mM potassium acetate, 25 mM Tris-HCl, pH 7.5, 0.5 mM TCEP.

3   Methods

3.1 Protein Expression Using Attached HEK 293T Cells	Depending on the scale of expression required and the equipment available, methods for both T175 static flasks and roller bottles are described below.
3.1.1 Medium Scale Using T175 Flasks	1. All cell manipulations are carried out in a Class 2 laminar flow hood.

2. Seed HEK 293T cells at 7.5×10^5 cells/mL in 5 mL so that the cells are ~80 % confluent after 24 h. Make up to 45 mL with DMEM containing 2 % FCS, 1× non-essential amino acids, and 1 mM glutamine.

3. Incubate the cells at 37 °C in a 5 % CO_2/95 % air atmosphere for 24 h.

4. Mix 87.5 μL plasmid DNA (*see* **Note 2**) with 2.6 mL of DMEM supplemented with 1× non-essential amino acids and 1 mM glutamine.

5. In a separate vessel, mix 154 μL 1 mg/mL PEI with 2.6 mL of DMEM containing 1× non-essential amino acids and 1 mM glutamine. Add this to the DNA cocktail made in **step 4** and mix thoroughly.

6. Incubate at room temperature (RT) for 10 min (*see* **Note 3**).

7. Remove the supernatant from the T175 flask of confluent HEK 293T cells.

8. Add the transfection cocktail made in **steps 4–6** to the cells.

9. Top up the flask with 40 mL of DMEM containing 2 % FCS, 1× non-essential amino acids, and 1 mM glutamine.

10. If control of glycosylation is required, add 45 μL of 1 mg/mL kifunensine to the T175 flask (*see* **Note 4**).

11. Incubate the flask at 37 °C in a 5 % CO_2/95 % air atmosphere for 3 days at which point the phenol red pH indicator in the DMEM should start to change color to orange.

12. To harvest a secreted protein: Collect the supernatant (which contains the protein), centrifuge at $6,000 \times g$ for 15 min to remove any detached cells, and filter through a 0.22 μm bottle top filter before storing at 4 °C.

13. To harvest an intracellular protein: Remove the supernatant and discard before freezing the T175 flask at –80 °C.

3.1.2 Large Scale Using Roller Bottles

1. Each roller bottle contains 250 mL of culture so four roller bottles are needed per L of culture.

2. Seed HEK 293T cells at around 7.5×10^5 cells/mL in 20 mL into each roller bottle (*see* **Note 5**) and add 250 mL DMEM containing 2 % FCS, 1× non-essential amino acids, and 1 mM glutamine.

3. Incubate the roller bottle at 37 °C for 4 days with the bottle rotating at 30 rpm (*see* **Note 6**). After this time, the cells should be ~80 % confluent.

4. Remove the spent medium from the roller bottle and replace with 200 mL DMEM containing 2 % FCS, 1× non-essential

amino acids, and 1 mM glutamine. Return the roller bottle to the incubator.

5. Mix 0.5 mg of plasmid DNA (*see* **Note 2**) with 25 mL of DMEM with 1× non-essential amino acids and 1 mM glutamine.

6. In a separate vessel, mix 875 μL of 1 mg/mL PEI with 25 mL of DMEM containing 1× non-essential amino acids and 1 mM glutamine. Add this to the DNA cocktail from **step 5** and mix thoroughly.

7. Incubate at RT for 10 min (*see* **Note 3**).

8. Add the transfection cocktail made in **steps 5–7** to the roller bottle.

9. If control of glycosylation is required, add 0.25 mL of 1 mg/mL kifunensine to the roller bottle (*see* **Note 4**).

10. Incubate the roller bottle at 37 °C with the bottle rotating at 30 rpm (*see* **Note 6**) for 3–6 days. The point of harvest is determined by the phenol red pH indicator in the DMEM starting to change color to orange.

11. To harvest a secreted protein: Collect the supernatant (which contains the protein), centrifuge at $6,000 \times g$ for 15 min to remove any detached cells, and filter through a 0.22 μm bottle top filter before storing at 4 °C.

12. To harvest an intracellular protein: Remove the supernatant and discard. Detach cells from the roller bottle by shaking and harvest by centrifugation at $6,000 \times g$ for 15 min. Wash the roller bottle in 125 mL PBS and use this solution to resuspend the cell pellet, thus washing the cells to remove any remaining medium. Centrifuge for a further 15 min at $6,000 \times g$ and freeze the resulting pellet at –80 °C.

3.2 Protein Expression Using Suspension HEK 293F Cells

Depending on the scale of expression required and the equipment available, methods for both 250 mL Erlenmeyer flasks and roller bottles are described below. For co-transfections of two or more plasmids, the total amount of DNA used must be as indicated in the protocols below.

3.2.1 Medium Scale Using 250 mL Flasks

1. 250 mL flasks will support between 30 and 100 mL culture. For transfection volumes greater than 30 mL, the protocol can be scaled accordingly.

2. All cell manipulations are carried out in a Class 2 laminar flow hood.

3. Seed cells at 3.5×10^5 cells/mL in to a 250 mL flask with a final volume of 30 mL.

4. Incubate flask at 37 °C in a 5 % CO_2/95 % air atmosphere with the flask rotating at 120 rpm for 3 days until the cells reach a density of $>2 \times 10^6$ cells/mL.

5. Dilute 30 μg plasmid DNA (total) (*see* **Note 2**) in 3 mL PBS and vortex briefly.

6. Add 120 μL of 0.5 mg/mL PEI to the diluted DNA and vortex briefly.

7. Incubate at RT for 20 min.

8. Add the PBS, DNA, and PEI cocktail to 27 mL cells at 1×10^6 cells/mL final concentration.

9. If control of glycosylation is required, add 30 μL of 1 mg/mL kifunensine to the flask (*see* **Note 4**).

10. Incubate flask at 37 °C in a 5 % CO_2/95 % air atmosphere for 48 h.

11. To harvest protein: Centrifuge cells at $6,000 \times g$ for 5 min. For intracellular protein retain cells and store at –80 °C, and for secreted protein retain supernatant and filter through a 0.22 μm bottle top filter before storing at 4 °C.

3.2.2 Large Scale Using Roller Bottles

1. Each roller bottle contains 300 mL of culture so four roller bottles are needed for 1.2 L of culture. Roller bottles will support a minimum volume of 150 mL and a maximum volume of 300 mL, for volumes less than 300 mL the protocol can be scaled accordingly.

2. Seed cells at 3.5×10^5 cells/mL into a roller bottle with a final volume of 300 mL.

3. Incubate the roller bottle at 37 °C in a 5 % CO_2/95 % air atmosphere with the vertically orientated bottle shaking at 120 rpm for 3 days until the cells reach a density of $>2 \times 10^6$ cells/mL.

4. Dilute 300 μg plasmid DNA (total) (*see* **Note 2**) in 30 mL PBS and vortex briefly.

5. Add 1.2 mL of 0.5 mg/mL PEI to the diluted DNA and vortex briefly.

6. Incubate at RT for 20 min.

7. Add the PBS, DNA, and PEI cocktail to 270 mL cells at 1×10^6 cells/mL final concentration.

8. If control of glycosylation is required, add 0.3 mL of 1 mg/mL kifunensine to each roller bottle (*see* **Note 4**).

9. Incubate the roller bottle at 37 °C in a 5 % CO_2/95 % air atmosphere for 48 h.

10. To harvest protein: Centrifuge cells at $6,000 \times g$ for 5 min. For intracellular protein retain cells and store at –80 °C, and for secreted protein retain supernatant and filter through a 0.22 μm bottle top filter before storing at 4 °C.

3.3 Purification of Secreted Proteins Using a His6 Tag

This protocol describes an automated method for purification of secreted proteins from large volumes of medium using an ÄKTAxpress system (GE Healthcare Life Sciences, UK). However, the initial immobilized metal affinity chromatography (IMAC) purification step (*see* Subheading 3.3.2) can be disconnected from the size-exclusion chromatography step and automated using other ÄKTA purification systems.

3.3.1 Automated Protocol Using the ÄKTAxpress

1. Equilibrate either a HiLoad 16/600 Superdex S75 or S200 column with Gel Filtration Buffer (*see* **Note 7**).

2. Insert buffer lines A1 and A2 into Nickel Wash Buffer and manually wash the pumps to fill the lines with buffer.

3. Insert buffer line A3 into Nickel Elution Buffer. Place a large empty bottle or flask (this needs to be larger than the sample volume) on outlet line F3 and a 96 deep-well plate in the fraction collector.

4. Insert a pre-charged 5 mL HisTrap FF column into column position 1 of the ÄKTAxpress.

5. Carefully remove line A2 from the Nickel Wash Buffer and insert into the flask containing the filtered protein-containing medium.

6. Run the glycoprotein purification program transcribed in Nettleship et al. [27] (*see* **Note 8**).

7. This program will complete an automated IMAC purification followed by further purification by size-exclusion chromatography giving protein with over 95 % purity (Fig. 2).

3.3.2 Description of IMAC Purification for Use with Other Systems

1. 50 mL of medium is loaded through a 5 mL HisTrap FF column at 8 mL/min followed by 10 mL of Nickel Wash Buffer.

2. **Step 1** is then repeated until all the medium has been loaded through the column. This load/wash loop reduces the impact
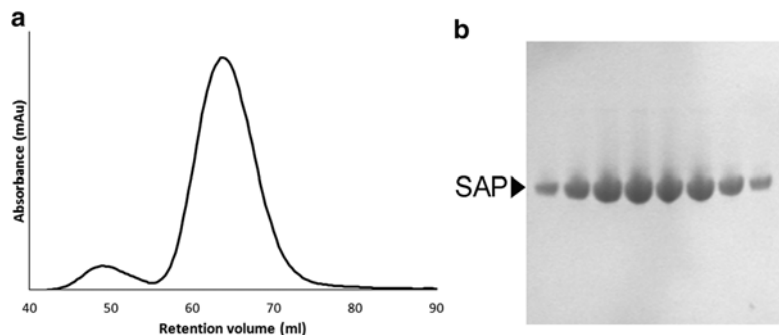
Fig. 2 Example showing the purification of SAP which gave 6 mg from 1 L of medium produced via the roller bottle protocol. (**a**) Size-exclusion chromatography trace and (**b**) SDS-PAGE analysis of the size-exclusion fractions

of IMAC incompatible components in the media as well as addressing pressure problems due to the viscosity of the mammalian culture medium particularly if it contains FCS.

3. The column is then washed with 50 mL (10× column volume) of Nickel Wash Buffer before elution of the protein with 25 mL (5× column volume) Nickel Elution Buffer collecting 2 mL fractions.

4. The product can then be further purified using size-exclusion chromatography.

3.4 Purification of Intracellular Protein Using a 3×FLAG® Tag

The protocol below describes a method of purification using the 3×FLAG® tag from a 1.2 L transfection. This can be scaled appropriately for larger-scale expression. This protocol, including the buffers stated, was developed for the purification of 3×FLAG®-HDAC3–SMRT-DAD complex (*see* **Note 9**). The method given is manual; however, various stages of the process may be automated using ÄKTA purification systems. After the initial Anti-FLAG® purification, a size-exclusion column is used to further purify the protein (including removing the TEV protease).

3.4.1 Initial Anti-FLAG® Purification

1. For a 1.2 L scale-up, defrost the cell pellet into ~30 mL FLAG Lysis Buffer.

2. Lyse the cells by sonication using five cycles of 30 s on/30 s off.

3. Remove the cell debris by centrifugation at $30,000 \times g$ for 30 min at 4 °C.

4. Meanwhile, equilibrate 1 mL of packed Anti-FLAG® M2 resin by washing three times with FLAG® Equilibration Buffer.

5. Incubate the supernatant from **step 3** with the Anti-FLAG® M2 resin in a 50 mL tube at 4 °C for 1 h with gentle mixing using a roller.

6. Centrifuge at $1,000 \times g$ for 5 min at 4 °C. Discard the supernatant and transfer the resin to a 15 mL tube. Wash the resin three times with FLAG Wash Buffer 1, then three times with FLAG Wash Buffer 2, and then three times with FLAG Cleavage Buffer.

7. After the last wash add 10 mL FLAG Cleavage Buffer to the resin along with 20 μg of His-TEV protease (*see* **Note 10**). Incubate overnight (ON) at 4 °C with gentle mixing using a roller.

8. Analyze the samples by SDS-PAGE. At this stage this fraction will contain the His-TEV protease as well as the purified protein of interest with the 3×FLAG® tag cleaved. The protein of interest is over 95 % pure discounting the protease (Figs. 3 and 4).

9. Before further purification, concentrate the protein to 0.5 mL using an appropriately sized Amicon Ultra Centrifugal filter (*see* **Note 11**) (Millipore, UK).
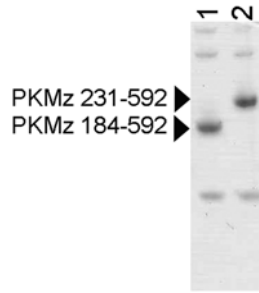
Fig. 3 SDS-PAGE showing the purification of PKMζ from a 30 mL HEK 293F experiment using Anti-FLAG® chromatography. Two constructs of PKMζ were purified with *lane 1* showing TEV-cleaved protein from a construct using amino acids 184–592 and *lane 2*, amino acids 231–592
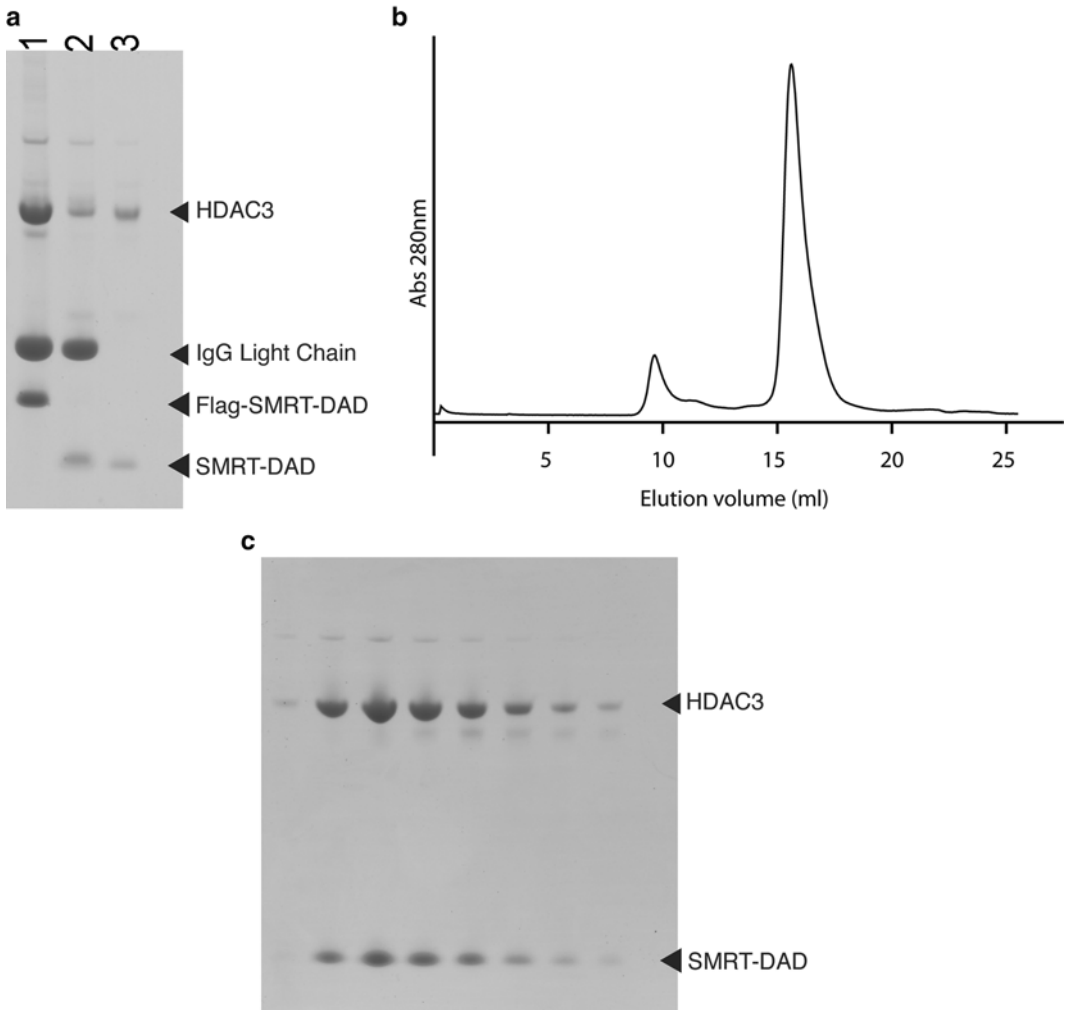
Fig. 4 (**a**) SDS-PAGE showing the purification of HDAC3–SMRT-DAD complex using Anti-FLAG® resin. *Lane 1* shows proteins bound to the FLAG® resin. *Lane 2* shows the Anti-FLAG® resin after elution of the protein with TEV protease, and *Lane 3* shows soluble proteins in the supernatant post-elution. (**b**) Size-exclusion chromatography trace (Superdex 200 column) and (**c**) SDS-PAGE analysis of the size-exclusion fractions

3.4.2 Further Purification by Size-Exclusion Chromatography

1. Based on the molecular weight of the protein of interest and the His-TEV protease, select either a HiLoad Superdex 10/300 S75 or S200 column (*see* **Note** 7).

2. Equilibrate the size-exclusion column in Gel Filtration Buffer.

3. Inject the fractions containing the protein of interest onto the column using a volume lower than 0.5 mL. Larger injection volumes can lead to a loss in resolution.

4. Analyze fractions collected by SDS-PAGE to assess separation of the protein of interest from the TEV protease. The protein of interest is now at sufficient purity (>99 %) for crystallization and structure determination [23] (Fig. 4).

4 Notes

1. Many vectors are available commercially for expression in mammalian cells. In the case of SAP, this was cloned into pOP-INTTG which is based on pTT [5] and uses the signal sequence from RTPTμ [28] and adds a C-terminal His$_6$-tag to the protein. The vector used for the expression of PKMζ and HDAC3 is based on pcDNA3 (Invitrogen, UK) and attaches an N-terminal His$_{10}$-3×FLAG®-TEV cleavage site tag onto the protein.

2. DNA for transfection needs to have an Abs$_{260nm}$/Abs$_{280nm}$ ratio of greater than 1.8. This can be obtained using standard commercial kits such as the PureLink HiPure Plasmid Megaprep kit from Life Technologies, UK.

3. Incubating for longer than 10 min can result in loss of transfection efficiency.

4. Kifunensine is an α-mannosidase I inhibitor which results in the secreted product containing only glycans of the form Man$_9$GlcNAc$_2$ which can be trimmed to one GlcNAc residue using endoglycosidase H (Man = mannose, GlcNAc = *N*-acetyl glucosamine). This is used to create homogeneous glycans in order to aid crystallogenesis [29].

5. One fully confluent T175 flask of attached cells is used per roller bottle.

6. Suitable roller incubators can be purchased from Wheaton Science Products, NJ, USA.

7. The Superdex S75 column resolves proteins in the 3–70 kDa molecular weight range and the Superdex S200 column in the 10–600 kDa range.

8. The full method for the glycoprotein purification program is written out in Nettleship et al. [27] and can be copied into the Method Editor section of the UNICORN™ software.

9. Depending on the 3×FLAG®-tagged protein being purified, the buffer system can be altered for optimal protein stability. For example, the manufacturer recommends 50 mM Tris–HCl, 150 mM NaCl, pH 7.4 (Sigma-Aldrich, UK).

10. Depending on the format of the vector, other proteases such as rhinovirus 3C protease or enterokinase can be used. In addition, the protein may be eluted from the column with its tag intact using the 3×FLAG® peptide (Sigma-Aldrich, UK) or a low-pH buffer such as glycine-HCl, pH 3.5.

11. When selecting the Amicon Ultra Centrifugal filter to be used, one should select the molecular weight cutoff (MWCO) based on half the molecular weight of the protein of interest. This is because the MWCO is calculated using a globular protein model, whereas the protein of interest may not be globular.

Acknowledgment

References

1. Hirose S, Kawamura Y, Yokota K et al (2011) Statistical analysis of features associated with protein expression/solubility in an in vivo *Escherichia coli* expression system and a wheat germ cell-free expression system. J Biochem 150:73–81

2. Geisse S, Henke M (2005) Large-scale transient transfection of mammalian cells: a newly emerging attractive option for recombinant protein production. J Struct Funct Genomics 6:165–170

3. Van Craenenbroeck K, Vanhoenacker P, Haegeman G (2000) Episomal vectors for gene expression in mammalian cells. Eur J Biochem 267:5665–5678

4. Boussif O, Lezoualc'h F, Zanta MA et al (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A 92:7297–7301

5. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res 30(2):E9

6. Aricescu AR, Owens RJ (2013) Expression of recombinant glycoproteins in mammalian cells: towards an integrative approach to structural biology. Curr Opin Struct Biol 23:345–356

7. Einhauer A, Jungbauer A (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J Biochem Biophys Methods 49:455–465

8. Ohana RF, Hurst R, Vidugiriene J et al (2011) HaloTag-based purification of functional human kinases from mammalian cells. Protein Expr Purif 76:154–164

9. Tennent GA, Dziadzio M, Triantafillidou E et al (2007) Normal circulating serum amyloid P component concentration in systemic sclerosis. Arthritis Rheum 56:2013–2017

10. Pepys MB, Booth DR, Hutchinson WL et al (1997) Amyloid P component. A critical review. Amyloid 4:274–295

11. Hawkins PN, Lavender JP, Pepys MB (1990) Evaluation of systemic amyloidosis by scintigraphy with 123I-labeled serum amyloid P component. N Engl J Med 323:508–513

12. Pepys MB, Herbert J, Hutchinson WL et al (2002) Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. Nature 417:254–259

13. Duffield JS, Lupher ML Jr (2010) PRM-151 (recombinant human serum amyloid P/pentraxin 2) for the treatment of fibrosis. Drug News Perspect 23:305–315

14. Hernandez AI, Blace N, Crary JF et al (2003) Protein kinase M zeta synthesis from a brain mRNA encoding an independent protein kinase C zeta catalytic domain. Implications for the molecular mechanism of memory. J Biol Chem 278:40305–40316

15. Kelly MT, Crary JF, Sacktor TC (2007) Regulation of protein kinase Mzeta synthesis by multiple kinases in long-term potentiation. J Neurosci 27:3439–3444

16. Glanzman DL (2013) PKM and the maintenance of memory. F1000 Biol Rep 5:4

17. Price TJ, Ghosh S (2013) ZIPping to pain relief: the role (or not) of PKMzeta in chronic pain. Mol Pain 9:6

18. Yang WM, Yao YL, Sun JM et al (1997) Isolation and characterization of cDNAs corresponding to an additional member of the human histone deacetylase gene family. J Biol Chem 272:28001–28007

19. Guenther MG, Barak O, Lazar MA (2001) The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol Cell Biol 21:6091–6101

20. Wagner JM, Hackanson B, Lubbert M et al (2010) Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy. Clin Epigenetics 1:117–136

21. Graff J, Rei D, Guan JS et al (2012) An epigenetic blockade of cognitive functions in the neurodegenerating brain. Nature 483: 222–226

22. Shirakawa K, Chavez L, Hakre S et al (2013) Reactivation of latent HIV by histone deacetylase inhibitors. Trends Microbiol 21:277–285

23. Watson PJ, Fairall L, Santos GM et al (2012) Structure of HDAC3 bound to co-repressor and inositol tetraphosphate. Nature 481: 335–340

24. Berrow NS, Alderton D, Sainsbury S et al (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res 35:e45

25. Berrow NS, Alderton D, Owens RJ (2009) The precise engineering of expression vectors using high-throughput in-fusion PCR cloning. Methods Mol Biol 498:75–90

26. Bird LE (2011) High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. Methods 55:29–37

27. Nettleship JE, Rahman-Huq N, Owens RJ (2009) The production of glycoproteins by transient expression in mammalian cells. Methods Mol Biol 498:245–263

28. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62: 1243–1250

29. Chang VT, Crispin M, Aricescu AR et al (2007) Glycoprotein structural genomics: solving the glycosylation problem. Structure 15:267–273

Chapter 12

Recombinant Glycoprotein Production in Human Cell Lines

Kamilla Swiech, Marcela Cristina Corrêa de Freitas, Dimas Tadeu Covas, and Virgínia Picanço-Castro

Abstract

The most important properties of a protein are determined by its primary structure, its amino acid sequence. However, protein features can be also modified by a large number of posttranslational modifications. These modifications can occur during or after the synthesis process, and glycosylation appears as the most common posttranslational modification. It is estimated that 50 % of human proteins have some kind of glycosylation, which has a key role in maintaining the structure, stability, and function of the protein. Besides, glycostructures can also influence the pharmacokinetics and immunogenicity of the protein. Although the glycosylation process is a conserved mechanism that occurs in yeast, plants, and animals, several studies have demonstrated significant differences in the glycosylation pattern in recombinant proteins expressed in mammalian, yeast, and insect cells. Thus, currently, important efforts are being done to improve the systems for the expression of recombinant glycosylated proteins. Among the different mammalian cell lines used for the production of recombinant proteins, a significant difference in the glycosylation pattern that can alter the production and/or activity of the protein exists. In this context, human cell lines have emerged as a new alternative for the production of human therapeutic proteins, since they are able to produce recombinant proteins with posttranslational modifications similar to its natural counterpart and reduce potential immunogenic reactions against nonhuman epitopes. This chapter describes the steps necessary to produce a recombinant glycoprotein in a human cell line in small scale and also in bioreactors.

Key words Glycosylated proteins, Recombinant proteins, Human cell lines, Lentiviral vectors, Virus production, Transient transfection, Cell transduction, Bioreactor culture

1 Introduction

Since the development of recombinant DNA technology in the late 1970s, the development of new strategies to produce recombinant proteins is in continuous expansion. These proteins can be produced in different expression systems including bacteria, fungi, yeasts, insect cells, and mammalian cells and have a variety of applications ranging from the use in diagnostic kits to veterinary and human therapeutic use. In the last two decades, recombinant proteins have been used in many human therapeutic approaches.

The number of approved proteins and the clinical trials using this kind of proteins is continuously growing.

The expression system most commonly used in recombinant protein production is the bacterial system. However, many human proteins expressed in *Escherichia coli* are in the insoluble form (in inclusion bodies—IBs). The formation of IBs is a major problem in the production of recombinant proteins in bacteria. Purification of the proteins from IBs usually requires the extraction of bacterial recombinant insoluble protein followed by solubilization. This process is very laborious, time consuming, and in many cases not effective. To overcome this problem, other expression systems can be used such as eukaryotic systems.

The expression of recombinant proteins in insoluble form is not the only problem encountered in bacterial expression system; several human proteins require posttranslational modifications that many species are not able to do. One posttranslational modification commonly found in human proteins is glycosylation. About 50 % of human proteins have some type of glycosylation. Glycosylation is the addition of the saccharide chain protein. This process is essential for the formation of secreted and membrane proteins. There are two types of glycosylation: a nitrogen-linked glycosylation (*N*-glycosylation), which occurs at the amide nitrogen of asparagine side chains, and oxygen-glycosylation, which occurs in the hydroxy oxygen of serine and threonine.

These polysaccharide chains added to the protein have several functions, maintaining the structure, stability, activity, and function of the protein [1–3]. Glycosylation has also a central role in cell-cell adhesion. Glycosylation also affects the half-life, pharmacokinetics, and immunogenicity of the protein. Recombinant proteins expressed in mammalian, yeast, and insect have shown significant differences in the glycosylation pattern [4]. The differences in the glycosylation patterns exist between intra- and also interspecies. Among the different mammalian cell lines used for the production of recombinant proteins, there is a significant difference in glycosylation pattern. This difference in some cases alters the production and/or activity of the protein. Human cell lines are an alternative host system for the production of human therapeutic proteins. These cell lines are the best choice since they are capable of producing recombinant proteins with posttranslational modifications most similar to the physiologic protein and the right patterns of glycosylation will reduce the potential for immunogenic responses against nonhuman epitopes.

1.1 Expression Systems for Recombinant Protein Production

The vast majority of human recombinant therapeutic proteins require posttranslational modifications to be biologically functional. As an example, it is important to consider that the expression of human glycosylated proteins in *E. coli* will result in the production of non-glycosylated and, therefore, nonfunctional molecules.

Among yeast expression systems, *Saccharomyces cerevisiae* and *Pichia pastoris* are the most widely used. *P. pastoris* expression system, which is licensed (http://www.rctech.com/licensing/gxt-pichia.php), allows the stable and lasting production of proteins, and it has the potential of performing many of the posttranslational modifications, including glycosylations (*O*- and *N*-linked). However, yeast glycosylation pattern differs from that of higher eukaryotes, since they add *O*-oligosaccharides composed just of mannose (Man) residues which negatively affect the half-life of the protein when injected in humans [5]. Recombinant protein production using plants will generate hyperglycosylated proteins containing xylose and fucose molecules that are immunogenic in humans [6].

Baculovirus expression vector system (BEVS) in insect cells represents a robust method for producing recombinant glycoproteins. However, the glycosylation pattern produced in insect cells differs from human cells.

Transgenic animals represent another option of expression system to produce recombinant proteins. Recombinant proteins can be produced in blood, egg white, urine, and milk, but all have many disadvantages. Many recombinant proteins when expressed in high levels may alter the health of the animals, and the cost of this type of production is very high. The advantages and the disadvantages of the different expression systems are shown in Fig. 1.

Currently, approximately 60 % of all the recombinant proteins used for therapeutic purposes are produced in mammalian cells.
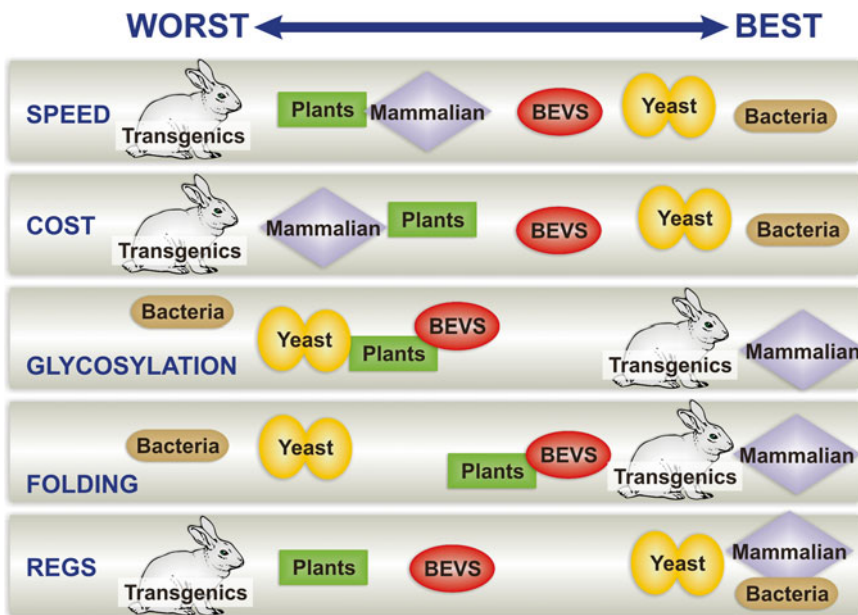
Fig. 1 Advantages and disadvantages of each expression system

The main advantage of using mammalian cells for the production of human recombinant proteins resides in the fact that these cells are capable of performing complex posttranslational modifications (glycosylation, carboxylation, hydroxylation, among others), thus generating proteins with characteristics similar to proteins present in the body.

1.2 Vector Integration Sites and Copy Number

For the production of a recombinant protein, it is important to know the protein structure to be produced in order to choose the best expression system. Another important point that should be considered is the vector that will be used for the genetic modification of these cells. Currently there are several types of vectors used to modify mammalian cells, and they can be integrated or not in the genome of the host cell. The non-integrative vectors such as plasmids and adeno-associated vectors are widely used for transient expression of recombinant proteins. The vectors that integrate into the cell genome such as those derived from retrovirus (e.g., *Gammaretrovirus* and *Lentivirus*) are capable of providing a long-lived and more stable expression of the protein of interest.

An important point that should be considered when opting for a viral vector is the number of copies that will integrate into the cell genome and the integration site of these vectors. Several studies have shown that viral vector integration is not random [7, 8], and it has been possible to elucidate some of the preferred primary retroviruses integration sites. Among these preferred insertion sites, it is possible to highlight the repetitive regions of the genome (SINE, LINE, LTR, etc.), CpG islands and fragile sites, gene regions, and active transcription start sites [9, 10].

Currently, there are several techniques that can be used for location tracking of integrating viral vectors such as LM-PCR (ligation-mediated PCR) and last-generation sequencing. There are also bioinformatics tools such as DAVID bioinformatics resources and QuickMap [11] that allow analyzing the data generated.

Regardless of the technique chosen, the study of integration sites of the viral vectors is important since the insertion site may influence the cell growth and recombinant protein expression.

1.3 Glycosylation Pattern in Human Cell Lines

Despite the ability to produce a product with acceptable therapeutic properties, the expression of recombinant proteins in mammalian cells has still some limitations. While other expression systems allow the production of the target protein at high levels in low-cost culture media, mammalian cell cultures require complex and costly media to maintain their growth and protein expression. Therefore, in order to have an economically viable manufacturing process, the genetic manipulation of the cell should result in an efficient protein expression to obtain the highest protein levels. Expression levels of the order from 5 to 100 mg of protein/10^6 cells/day are considered good. To achieve these high levels of expression, an efficient

method of selection and an appropriate expression vector are necessary. A suitable genetic manipulation and the method of transfection/transduction and gene amplification are key factors that can lead to a significant increase in the specific productivity of a specific production process.

Most of the proteins used in human therapeutic approaches are glycol-proteins, such as coagulation factors and monoclonal antibodies. The oligosaccharides added to the primary chain of the protein directly affect the efficacy and safety of the biopharmaceutical product. Glycosylation is the most complex posttranslational protein modification. This complex process comprises several families of proteins, including glycotransferases and glycosidases, which control the level and the type of glycol-structures in the protein [12].

Mammalian cells are able to produce fully glycosylated proteins; however, there are differences in the glycosylation patterns intra- and interspecifically. Thus, human cell lines are more appropriated to be used in the production of recombinant human therapeutic proteins.

There are two types of glycosylation: asparagine *N*-linked and serine/threonine *O*-linked, which are found in proteins. The presence of *N*- and *O*-glycan structures can alter the function and immunogenicity of the recombinant protein [13, 14]. When a presynthesized glycan structure is added to the amide nitrogen on the side chain of the asparagine in the polypeptide chain, this process is called *N*-linked glycosylation. *N*-glycosylation occurs in the endoplasmic reticulum (ER). If monosaccharides are attached to the hydroxyl oxygen on the side chain of serine or threonine, this type of glycosylation is called *O*-linked glycosylation. This different process from *N*-glycosylation occurs in the Golgi complex (GC) [15]. In addition to the two *N*- and *O*-glycosylation pathways, glycans can also be attached to arginine, tyrosine, hydroxylysine, hydroxyproline, and tryptophan residues [16].

Protein glycosylation is a complex process in which the insertion of glycol-structures can result in a large number of protein isoforms [17]. Potential glycosylation sites can be either occupied or not, and a different glycan structure can be incorporated in each site in different protein molecules. Glycosylation was long considered a process of little importance. However, although protein activity is determined by the primary amino acid sequence, there are several examples in which it is described that glycosylation may affect protein activity [18]. Altered glycosylation patterns are associated with various pathologic states, including various cancers, rheumatoid arthritis, Leroy, and a type of leukocyte adhesion deficiency II [19], as well as a number of infectious diseases [20–22].

CHO (Chinese hamster ovary) and BHK (baby hamster kidney) are the cell lines most used by the industry to produce recombinant therapeutic proteins. These murine cell lines are able

to perform glycosylation; however, murine cells present two important differences in glycosylation pattern compared to the human cells. Glycosylation process in murine cells presents differences in neutral oligosaccharides; these cells express NeuGc whereas human cells express NeuAc. N-Glycolylneuraminic acid (Neu5Gc) is a sialic acid molecule found in most mammals. However, in humans, Neu5Gc is not present because the human gene CMAH is inactivated by a mutation. The gene CMAH encodes for the enzyme CMP-N-acetylneuraminic acid hydroxylase, which is responsible to convert CMP-Neu5Gc into CMP-N-acetylneuraminic (CMP-Neu5Ac) acid [23]. Another important difference between murine and human cells is that human cells cannot synthesize a terminal Galα1-3Gal motif on N-glycans. As a consequence, they express antibodies against this structure inserted in recombinant proteins produced by murine cell lines [24].

Therefore, human cell lines are more suitable to produce human recombinant proteins, and the use of this expression system will diminish the presence of immunogenic glycan structures in the final recombinant proteins. However the production in human cell lines do not totally avoid the presence of antigenic structures, since Neu5Gc can be taken up from animal products present in the culture medium and incorporated into secreted glycoproteins [25]. Thus, culture media should be suitable for this purpose.

1.4 Human Cell Lines Human cell lines with human glycosylation and other posttranslational modification machineries have become an attractive alternative for biopharmaceutical industry. Currently, only few cell lines are used to produce recombinant proteins at large scale. Here, we will describe the characteristics of human cell lines, which are in use by the biopharmaceutical industry.

The HEK 293 cell line, derived from human embryonic kidney cells, was established almost 35 years ago. This cell line has been used for a long time in the production of pseudotyped viral vectors (adenovirus, lentivirus, retrovirus, and adeno-associated virus) and in the production of recombinant proteins to be used in research studies. Recently, this cell line has been the first human cell used to produce a commercial recombinant therapeutic product, Xigris® (activated protein C). Xigris is a gamma-carboxylated protein that requires a propeptide cleavage. The posttranslational modifications are essential for maintaining its biological activity. In contrast, CHO cells do not properly express gamma-carboxylated proteins [26]. Unfortunately, Xigris expressed in CHO fails to show a survival benefit for patients with severe sepsis and septic shock, and it was withdrawn from market by the FDA.

Recombinant HEK 293 cells constitutively express the Epstein-Barr virus nuclear antigen (EBNA), being an antigen responsible for high levels of plasmid amplification, and it results in an enhanced productivity. The HEK 293E cell line is the most widely used cell line for large-scale transient gene expression.

HEK 293 cell line is easily adapted to grow in serum-free suspension cultures and presents fast growth, and it is highly transfectable and transducible, so these cells are great candidate for use in high-throughput recombinant gene expression facilities [27] (*see* Chapter 10).

On the other hand, CEVEC Pharmaceuticals developed a new expression system based on human amniocytes derived from amniotic fluid cells obtained by amniocentesis. Primary human amniocytes were immortalized by adenovirus type 5 (including the E1 genes and the entire pIX sequence). CAP (CEVEC's amniocyte production) cells are optimized to grow in a variety of flask/wave formats up to large-scale processing in bioreactors and can be very efficiently transfected with commercially available transfection reagents. CAP cells, which show humanlike posttranslational modifications and authentic human glycosylation patterns, also grow in serum-free suspension cultures, allowing stable protein production.

HKB11 is a human hybrid cell line developed by Bayer Corporation. This cell line was made by the fusion of human embryonic kidney cells (293S) and modified Burkitt's lymphoma cells (2B8). The hybrid cell line is capable of secreting high levels of recombinant proteins with human glycosylation profiles and non-aggregating properties. HKB11 clone was selected for non-aggregating properties and possesses typical human glycosylation enzymes such as $\alpha(2,3)$ and $\alpha(2,6)$ sialyltransferases. The proteins produced by these cells were found to be capped with sialic acid of $\alpha(2,3)$ and $\alpha(2,6)$ linkages. The main advantages of using these hybrid cells include: (1) the cells are negative for immunoglobulin expression; (2) they grow easily in plasma protein-free medium (with or without the addition of recombinant insulin) as suspension cultures in a shake flask or in a bioreactor; (3) they are very susceptible for DNA transfection; (4) and they secrete high levels of recombinant proteins (such as recombinant monoclonal antibodies, soluble ICAM-1, rIL-4, and rFVIII).

The PER.C6 cell line was developed by Crucell and DSM Biologics. This cell line was established from human embryonic retinoblasts (PER.C6) transformed with adenovirus type 5 [28]. This cell line was initially developed for the safe production of pharmaceutical grade recombinant human adenoviral vectors used in vaccine and gene therapy purposes. PER.C6 was also used for producing classical vaccines including influenza and West Nile Virus. Recently, PER.C6 cells were also evaluated for the production of therapeutic proteins. These cells can grow in suspension or in adherent culture at high cell densities ($>10^7$ cells/mL) in serum-free and animal-component-free culture media. Therefore, more biological products might be harvested from smaller bioreactors [29]. The advantages of this expression system are that no gene amplification is required for high protein production levels, since low gene copy numbers are sufficient for very efficient IgG production.

PER.C6 cell line is considered one of the most advanced of the various human cell line alternatives to substitute CHO cells for recombinant therapeutic protein and antibody production.

2   Materials

2.1 Cell Culture

1. Adherent and nonadherent HEK 293 cells.

2. Vectors: pCMVR8.91 (viral proteins gag, pol, rev e tat from HIV-1) and pMD2 VSV-G (envelop).

3. Adherent HEK 293 cell medium: Dulbecco's modified Eagle's medium supplemented with 10 % (v/v) fetal bovine serum (complete medium). The medium base (i.e., without FBS) (Table 1) should be filter sterilized (0.2 μm) into sterile containers and stored at 4 °C.

4. Nonadherent HEK293 cell medium: FreeStyle 293 expression medium (Gibco). This medium is an animal origin-free,

Table 1
Base medium composition

Component	Concentration (mg/L)
Amino acids	
Glycine	30.0
L-Arginine hydrochloride	84.0
L-Cystine 2HCl	63.0
L-Glutamine	584.0
L-Histidine hydrochloride-H_2O	42.0
L-Isoleucine	105.0
L-Leucine	105.0
L-Lysine hydrochloride	146.0
L-Methionine	30.0
L-Phenylalanine	66.0
L-Serine	42.0
L-Threonine	95.0
L-Tryptophan	16.0
L-Tyrosine disodium salt dehydrate	104.0
L-Valine	94.0
Vitamins	
Choline chloride	4.0
D-Calcium pantothenate	4.0
Folic acid	4.0
Niacinamide	4.0
Pyridoxine hydrochloride	4.0
Riboflavin	0.4
Thiamine hydrochloride	4.0
i-Inositol	7.2

(continued)

Table 1
(continued)

Component	Concentration (mg/L)
Inorganic salts	
Calcium chloride ($CaCl_2$) (anhyd.)	200.0
Ferric nitrate ($Fe(NO_3)_3 \cdot 9H_2O$)	0.1
Magnesium sulfate ($MgSO_4$) (anhyd.)	97.67
Potassium chloride (KCl)	400.0
Sodium bicarbonate ($NaHCO_3$)	3,700.0
Sodium chloride (NaCl)	6,400.0
Sodium phosphate monobasic ($NaH_2PO_4–H_2O$)	125.0
Other components	
D-Glucose (dextrose)	4,500.0
Phenol red	15.0

chemically defined, and protein-free medium. It is also complete and ready-to-use, with no supplementation required.

5. Phosphate-buffered saline (PBS).

6. 10× trypsin solution: 2.5 % (w/v) solution.

7. T25 and T75 tissue culture flasks.

8. Sterile tissue culture dishes (60 and 100 mm).

9. Disposable sterile pipets.

10. Polypropylene centrifuge tubes (15 and 50 mL).

11. Freezing vials.

12. Sterile Eppendorf Tubes.

13. Trypan blue vital stain: Mix 0.9 % (w/v) NaCl and 1 % (w/v) trypan blue in 4:1 ratio. Use within 24 h.

14. Hematocytometer.

15. 100× penicillin-streptomycin solution.

16. Tissue culture grade dimethyl sulfoxide (DMSO).

17. Cryo 1 °C freezing container.

18. Liquid nitrogen storage system.

19. Cytodex 1 microcarrier (GE Healthcare).

20. Spinner flasks.

21. Stirred tank bioreactor with pitched-blade impeller.

22. Water bath.

23. Centrifuge.

24. Inverted microscope.

25. Ultrafiltration falcons with a 100,000-kDa cutoff membrane of regenerated cellulose.

2.2 Production ***of Virus by Transient*** ***Expression***	1. 2.5 M $CaCl_2$, filter sterilized (0.2 μm).

2. 2× HEPES-buffered saline (HBS) medium: 100 mM HEPES, 281 mM NaCl, 1.5 mM NaH_2PO_4, pH 7.12. Filter under sterile conditions. Accurate pH of this solution is critical. This solution should be aliquoted and stored at –20 °C to ensure its stability.

3. Sterile disposable syringes (10–50 mL).

4. Disposable syringe filters (0.22 μm).

5. Sterile polypropylene tubes (15–50 mL).

3 Methods

3.1 General Maintenance of Cell Lines

The most used cell line to produce pseudotyped lentiviral vectors is HEK 293 (human embryonic kidney) cell line. The cell line should be expanded as rapidly as possible and stocks (at least ten vials) frozen at a low passage number. It is advisable to grow cells in the absence of antibiotics because their use might mask persistent low-grade infections. The details of the cell expansion are in Subheading 3.1.1.

3.1.1 Recovery of a Cell Line from a Frozen Stock

1. Thaw the frozen cell vials in a 37 °C water bath with slowly manual agitation as rapidly as possible. When thawed, make sure the cells are evenly resuspended.

2. Transfer the ressuspended cells into 10 mL of the appropriate complete medium (i.e., basal medium/FBS) in a centrifuge tube and mix it by gently pipetting.

3. Centrifuge the cells at $1,000 \times g$ for 10 min at room temperature (RT).

4. Remove the supernatant and resuspend the cells in 7 mL of complete medium (25-cm² culture flask).

5. Incubate under appropriate conditions. For most cell lines it will be at 37 °C in 5 % CO_2 in a humidified incubator.

6. Replace culture medium every 2–3 days (*see* **Note 1**).

7. When the cell monolayer becomes confluent (90 %), subculture the cells (*see* **Note 2**).

3.1.2 Subculturing of Adherent Cell Lines

Adherent cell cultures should be subcultured as they become confluent (*see* **Note 3**):

1. Remove the culture medium from the flask and discard it.

2. Add 5 mL (for a 25-cm² flask) or 10 mL (for a 75-cm² flask) of PBS, and wash over the cell monolayer by gently rocking the flask.

3. Remove the PBS and add 3–5 mL (for a 25-cm² flask) or 5–10 mL (for a 75-cm² flask) of 0.25 % (w/v) trypsin (in PBS).

4. Incubate at RT for 3–5 min (the time is dependent on the cell type).

5. Examine the cell monolayer using an inverted microscope. The cells should become rounded and must be lifted in the solution (*see* **Note 3**). Pipet the solution several times to obtain a single cell suspension.

6. When the cells are in the supernatant, the trypsin should be inactivated by the addition of the same volume of complete medium.

7. Centrifuge the cells at $1,000 \times g$ for 10 min at RT, discard the supernatant, resuspend the cells in complete medium (5–7 mL for a 25-cm² flask, 10–15 mL for a 75-cm² flask), and incubate under the appropriate conditions.

8. Replace old medium with fresh medium every 2–3 days.

3.1.3 Seeding Cells at a Specific Density

1. Trypsinize and resuspend cells as described above and place into a sterile tube.

2. Add an equal volume of complete medium and mix.

3. Dilute a sample of the cell suspension into trypan blue vital stain and count live (cells that exclude trypan blue) and dead (cells that stain with trypan blue) cells using a hematocytometer.

4. Calculate live cell density in the original cell suspension and adjust if necessary.

5. Mix the cell suspension well and aliquot as desired.

3.1.4 Freezing Cell Stocks

1. Trypsinize cells as described above from a confluent 75-cm² flask and transfer to a centrifuge tube.

2. Add 2 mL of FBS to neutralize the trypsin.

3. Recover the cells by centrifugation and remove the supernatant.

4. Resuspend the cells in 1 mL of 90 % complete medium, 10 % DMSO (v/v), and transfer to a freezing vial.

5. Transfer the vial to a Cryo 1 °C freezing container and place in a –70 °C freezer (*see* **Note 4**).

6. After 24 h transfer the vial to liquid nitrogen for long-term storage.

3.2 Production of Lentiviral Vectors by Transient Expression

For safety reasons lentiviral vectors never carry the genes required for their replication. To produce a lentivirus, several plasmids are transfected into a so-called packaging cell line, commonly HEK 293. One or more plasmids, generally referred to as packaging plasmids, encode the viral proteins, such as the capsid and the reverse transcriptase. Another plasmid contains the genetic material to be delivered by the vector. It is transcribed to produce the single-stranded RNA viral genome and is marked by the presence of the

ψ (psi) sequence. This sequence is used to package the genome into the virion.

Calcium phosphate coprecipitation is adequate for this purpose; however, a wide range of equally suitable transfection reagents is available.

3.2.1 Production of Titer Virus by Transient Expression

The following protocol is for transfection of a 60-mm-diameter dish or 25-cm² flask of cells. It can be scaled up or down relative to the surface area of the culture dish/flask to be used. All solutions and procedures should be sterile. DNA solutions can be effectively sterilized by ethanol precipitation in a sterile tube followed by resuspension in sterile H_2O. The solutions to be used for preparing the calcium phosphate precipitate should be at RT before use. Generally, the highest titer of virus will be found 2–3 days after transfection. To harvest the virus, the conditioned medium is simply collected and passed through a 0.2-μm filter into a sterile container. To maximize virus collection, the medium can be collected at 48 h after transfection, and the cells refeed with fresh medium (prewarmed at 37 °C) with subsequent collections made in the same manner after a further 24 and 48 h. The virus can be stored at 4 °C for up to a few days or frozen at –70 °C for long-term storage. An approximate twofold decrease in titer generally results from the freeze/thaw cycle:

1. Seed 6×10^6 cells of a stable packaging cell line such as 293T in a 100-mm tissue culture dish and incubate it for 16–24 h. This should result in an even monolayer that is about 60–80 % confluent.

2. In an Eppendorf Tube, make up the DNAs. 10 μg of transgene vector, 13 μg of pCMVR8.91 (viral proteins gag, pol, rev e tat from HIV-1), and 7 μg of pMD2 VSV-G (envelop), a total of 20 μg of DNA. The DNA to be transfected is added to a final volume of 450 μL of H_2O and 50 μL of 2.5 mM $CaCl_2$.

3. Aliquot 500 μL of 2× HeBS into a second Eppendorf Tube, and then add the DNA/$CaCl_2$ mix dropwise while vortexing at high speed (*see* **Note 5**).

4. Continue vortexing for 5–10 s after all the solution has been added.

5. Allow the mixture to stand for 5 min.

6. Add the mixture dropwise to the cells and swirl to mix.

7. Incubate under normal culture conditions for 6–8 h, and then remove the medium from the cells and complete with fresh complete medium.

8. Incubate again for 48 h.

9. Collect the medium using a sterile disposable syringe, and then pass through a 0.2-μm filter unit into a suitable storage container such as a 10-mL centrifuge tube.

3.2.2 Preparation of High-Titer Viral Stocks	Viral stocks are simply prepared by collecting conditioned medium from the producer cell line and removing contaminating cells by filtration as described. Virus pseudotyped with VSV-G envelope can be concentrated to higher titers by ultrafiltration or ultracentrifugation. The titer can be increased 10- to 50-fold.

3.2.3 Concentration of Virus by Ultrafiltration

1. Set up a stirred cell ultrafiltration apparatus with a 100,000-kDa cutoff membrane in a 4 °C cold room or cold cabinet and rinse thoroughly with H_2O.

2. 20 mL 70 % ethanol is added to the membrane filter for sterilization, and it is centrifuged at $3,060 \times g$ for 10 min at 20 °C. Then, 15 mL PBS are added, and it is centrifuged at $3,060 \times g$ for 10 min at 20 °C to remove the ethanol from membrane.

3. Add viral supernatant (20 mL) to each membrane, and concentrate the sample at $3,060 \times g$ for 10 min at 20 °C. The volume can be concentrated up to tenfold after centrifugation for 30–60 min, getting a yield of approximately 60–100 % (*see* **Note 6**).

4. Collect the concentrated viral supernatant and rinse the apparatus with a small volume of medium and pool.

5. Rinse the membrane well with H_2O, wash in 1 M NaCl, and store it in 10 % ethanol.

3.2.4 Concentration of Virus by Ultracentrifugation

Ultracentrifugation is used to obtain a higher concentration factor. Before collecting the supernatant, turn on the vacuum of the ultracentrifuge; this helps it to cool at 4 °C quickly:

1. The culture medium containing the viral vectors is centrifugated at $50,000 \times g$ for 140 min at 10 °C (*see* **Note 7**).

2. The supernatant is discarded and the precipitated (viruses) are resuspended in 1:100 of the initial volume in PBS with 1 % HSA.

3. The aliquots are frozen at –80 °C (*see* **Notes 8** and **9**).

3.3 Transduction of Adherent Cells

The lentiviral vectors pseudotyped with VSV-G capsid can infect division cells and cells that are not in growing state. This is one of the advantages of using lentiviral vectors, which provide a greater number of modified cells. The exposure of the culture to several cycles of transduction will enhance the overall transduction efficiency. If the growth medium for the target cells is very different from the culture medium in which the viral producer cell line is grown, it should be considered to collect viruses in the target cell medium to ensure optimal growth of the target cells during transduction. Alternatively, supernatant virus can be mixed in a 1:1 ratio with target cell growth medium:

1. Plate a culture of the cells to be transduced and grow them up to 80 % confluence.

2. Remove the growth medium, replace with the viral supernatant, and add 8 µg/mL of polybrene.

3. Culture the cells for 8–16 h, and then remove the medium. After that, repeat another transduction cycle or add normal growth medium.

4. Expand the culture and use a suitable assay or select for transduced cells, for example: G418 or flow cytometer when the vector expresses GFP.

3.4 Transduction of Nonadherent Cells

Transduction of nonadherent cells is less efficient than the transduction of adherent cells, but the reason is not entirely clear. There are two basic approaches for the transduction of nonadherent cells: supernatant transduction and supernatant transduction plus centrifugation.

3.4.1 Supernatant Transduction

1. Recover target cells to be transduced by centrifuging at $1,000 \times g$ for 5 min.

2. Resuspend target cells directly in viral supernatant containing µg/mL of polybrene keeping the cells at the optimal density for logarithmic growth. If necessary, add some fresh growth medium to ensure optimal conditions for cell growth during the transduction procedure.

3. Culture for 8–24 h and repeat or grow out the cells for analysis.

4. Analyze the cells for transduction or select with antibiotic as described for adherent cells.

3.4.2 Supernatant Transduction in Orbital Shaker for 125-mL Erlenmeyer Flasks

1. Recover target cells to be transduced by centrifuging at $1,000 \times g$ for 5 min.

2. Resuspend target cells directly in 2 mL of viral supernatant and complete with 13 mL of fresh growth medium containing 8 µg/mL of polybrene.

3. Put the resuspended cells in a 125-mL Erlenmeyer flask under agitation (80 rpm) for 5 h, 37 °C in 5 % CO_2 in a humidified incubator.

4. After 5 h centrifuge the cells at $1,000 \times g$ for 5 min, remove the supernatant, and wash the cell pellet with PBS.

5. Repeat the transduction cycle or grow out the cells for analysis.

6. Analyze the cells for transduction or select with antibiotic as described for adherent cells.

3.5   Quality Control of the Cell Cultures: Cell Viability

This protocol describes how to perform a trypan blue staining which can be used to discriminate between viable and nonviable cells:

1. Dilute your cell sample in trypan blue dye of an acid azo exclusion medium by preparing a 1:1 dilution of the cell suspension using a 0.4 % trypan blue solution. Nonviable cells will be blue; viable cells will be unstained. Trypan blue should be sterile filtered before using it in order to get rid of particles in the solution that would disturb the counting process.

2. Carefully and continuously fill the hemocytometer chamber.

3. Incubate the hemocytometer and cells for 1–2 min at RT.
   For longer incubations, please use a humid chamber. Incubations exceeding 30 min may cause decreased cell viability due to trypan toxicity.

4. Count cells under the microscope in four 1-mm^2 squares of one chamber and determine the average number of cells per square (all hemocytometers consist of two chambers; each is divided into nine 1-mm^2 squares). For an accurate determination, the total number of cells overlying one 1 mm^2 should be between 20 and 50 cells/square. If the cell density is higher than 200 cells/square, you should dilute your cell suspension.

3.6   Bioreactor Cell Culture

3.6.1   Culture of Adherent Cells on Microcarriers in Spinner Flasks

1. Hydrate the desired amount of Cytodex 1 microcarrier in Ca^{2+}- and Mg^{2+}-free PBS (50–100 mL/g Cytodex) for at least 3 h at RT. Microcarrier cultures normally contain 1–5 g Cytodex/L of medium.

2. Wait for the microcarriers to decant and discard the supernatant. Add fresh Ca^{2+}- and Mg^{2+}-free PBS (30–50 mL/g Cytodex), and wash the microcarriers for a few min under agitation. Repeat the procedure.

3. Sterilize the PBS-microcarriers solution in spinner flasks at 115 °C and 15 psi for 15 min (*see* **Note 10**).

4. Prior to use, rinse the microcarriers in warm culture medium. When the microcarriers have settled, discard the supernatant and add 1/3 of the final culture volume. Leave equilibrate in CO_2 incubator until pH stabilized at 7.4.

5. Inoculate the cells in the range of 5×10^4–2×10^5 cells/mL. Stir the culture intermittently at 20–30 rpm (e.g., for 2 min every 30 min) at least for 4 h to ensure an even distribution of cells and microcarriers.

6. After this attachment period, add culture medium to the final culture volume and stir the culture continuously at a speed sufficient to prevent sedimentation of the microcarriers (usually 30–60 rpm).

7. Replace 50 % of the working volume with fresh medium as soon as the pH of the medium decreased to approximately 7.1.

8. Samples for quantification of cell density and viability, as well as substrate (glucose and glutamine) and by-products (ammonia and lactate) concentration, should be taken daily.

3.6.2 Culture of Nonadherent Cells in Bioreactor

1. Clean the vessel and adjacent tubes of the bioreactor with MQ H$_2$O.

2. Calibrate the pH probe outside the vessel using a two-point calibration method (buffer 7.0 for the zero and buffer 4.0 for the span) prior to autoclaving.

3. Make the appropriate connections, install the probes (pH and DO), and add PBS (volume equal to working culture volume) to sterilize the bioreactor (115 °C and 15 psi for 15 min) (*see* **Note 11**).

4. After autoclaving wait for the system to reach RT.

5. Allow for at least 12 h of polarization of the DO probe. Sparge the PBS solution with 100 % N$_2$ until all oxygen in the PBS solution is evacuated and the reading of the DO probe is stable to set the Zero. Sparge the PBS solution with 100 % of air and set the agitation to 100 rpm. Wait until the entire solution is saturated with air and the reading of the DO probe is stable to set 100 %.

6. Discard the PBS from vessel, add the culture medium, turn on agitation (50 rpm) and aeration, and wait for the system to reach the desired set points (e.g., pH 7.4 and 50 % of DO).

7. Inoculate the cells at 2×0–3×10^5 cells/mL (cells with viability greater than 90 %).

8. Take daily samples to cell concentration and viability determination and measure of glucose, glutamine, lactate, and ammonia concentration using commercial kits or an off-line analyzer.

4   Notes

1. It is very important to test cell viability. A health culture should contain 90–100 %.

2. The human cell lines are usually subcultured at 1:10 dilution. Dilutions such as 1:20 or 1:30 can be also used. However, over dilutions can affect cell viability.

3. Do not let the cell confluence reach 100 %. This will affect cell growth.

4. If the cells do not disattach, incubate for a further 5 min and/or replace trypsin solution with a fresh aliquot.

5. For freezing only FBS can be used with 10 % DMSO. The cost of this freezing is higher, but the cells are more feasible.

6. While vortexing the solution, slowly drip the DNA/CaCl$_2$ solution; this step is very important to form the DNA/CaCl$_2$ complex.

7. It is not possible to concentrate more than ten times the initial volume due to the FBS contained in the culture media, which makes the solution viscous, preventing the passage through the membrane.

8. Before collecting the supernatant, turn on the vacuum of the ultracentrifuge; this helps cool it to 4 °C quickly. Tubes may be balanced with serum-free media. Fill up the tubes and it is important for the level of the liquid to be 3–5 mm from the top.

9. The frozen viruses are stable for 6 months at –80 °C.

10. In order to prevent the microcarrier adhesion to the glass surface, the spinner flasks should be siliconized with Sigmacote (Sigma Aldrich).

11. Prior to autoclaving, confirm that all penetrations are plugged or have an appropriate process insert. Make sure that all lines connecting to submerged dip tubes are clamped off (sparger, harvest tube, sample line, etc.). All lines should be sealed, clamped, or protected by a filter and/or should be wrapped. All tubing connections should be secured with tie wraps. The exhaust line should be clear and protected by a filter. The jacket should be filled with water. DO and pH probes should be secured with autoclavable caps.

Acknowledgment

References

1. Leavitt R, Schlesinger S, Kornfeld S (1977) Impaired intracellular migration and altered solubility of nonglycosylated glycoproteins of vesicular stomatitis virus and Sindbis virus. J Biol Chem 252:9018–9023

2. Wallick SC, Kabat EA, Morrison SL (1988) Glycosylation of a VH residue of a monoclonal antibody against alpha (1–6) dextran increases its affinity for antigen. J Exp Med 168: 1099–1109

3. Walsh MT, Watzlawick H, Putnam FW et al (1990) Effect of the carbohydrate moiety on the secondary structure of beta 2-glycoprotein. I. Implications for the biosynthesis and folding of glycoproteins. Biochemistry 29: 6250–6257

4. Croset A, Delafosse L, Gaudry JP et al (2012) Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. J Biotechnol 161:336–348

5. Gemmill TR, Trimble RB (1999) Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim Biophys Acta 1426:227–237

6. Gomord V, Chamberlain P, Jefferis R et al (2005) Biopharmaceutical production in

plants: problems, solutions and opportunities. Trends Biotechnol 23:559–565

7. Russo-Carbolante EM, Picanco-Castro V, Alves DC et al (2011) Integration pattern of HIV-1 based lentiviral vector carrying recombinant coagulation factor VIII in Sk-Hep and 293T cells. Biotechnol Lett 33:23–31

8. Schroder AR, Shinn P, Chen H et al (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110:521–529

9. Ambrosi A, Cattoglio C, Di Serio C (2008) Retroviral integration process in the human genome: is it really non-random? A new statistical approach. PLoS Comput Biol 4:e1000144. doi:10.1371/journal.pcbi.1000144

10. Wu X, Li Y, Crise B et al (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 2300:1749–1751

11. Appelt JU, Giordano FA, Ecker M et al (2009) QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. Gene Ther 16:885–893

12. Varki A (1998) Factors controlling the glycosylation potential of the Golgi apparatus. Trends Cell Biol 8:34–40

13. Gross V, Heinrich PC, vom Berg D et al (1988) Involvement of various organs in the initial plasma clearance of differently glycosylated rat liver secretory proteins. Eur J Biochem 173:653–659

14. Rudd PM, Wormald MR, Wing DR et al (2001) Prion glycoprotein: structure, dynamics, and roles for the sugars. Biochemistry 40:3759–3766

15. Peter-Katalinic J (2005) Methods in enzymology: O-glycosylation of proteins. Methods Enzymol 405:139–171

16. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12:43R–56R

17. Hua S, Nwosu CC, Strum JS et al (2012) Site-specific protein glycosylation analysis with glycan isomer differentiation. Anal Bioanal Chem 403:1291–1302

18. Rajagopalan L, Organ-Darling LE, Liu H et al (2010) Glycosylation regulates prestin cellular activity. J Assoc Res Otolaryngol 11:39–51

19. Gornik O, Lauc G (2008) Glycosylation of serum proteins in inflammatory diseases. Dis Markers 25:267–278

20. Cooke CL, An HJ, Kim J et al (2009) Modification of gastric mucin oligosaccharide expression in rhesus macaques after infection with Helicobacter pylori. Gastroenterology 137:1061–1071

21. Dennis JW, Granovsky M, Warren CE (1999) Glycoprotein glycosylation and cancer progression. Biochim Biophys Acta 1473:21–34

22. Ohtsubo K, Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. Cell 126:855–867

23. Ghaderi D, Taylor RE, Padler-Karavani V et al (2010) Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. Nat Biotechnol 28:863–867

24. Galili U (2004) Immune response, accommodation, and tolerance to transplantation carbohydrate antigens. Transplantation 78:1093–1098

25. Bardor M, Nguyen DH, Diaz S et al (2005) Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. J Biol Chem 280:4228–4237

26. Suttie JW, Preusch PC (1986) Studies of the vitamin K-dependent carboxylase and vitamin K epoxide reductase in rat liver. Haemostasis 16:193–215

27. Chapple SD, Crofts AM, Shadbolt SP et al (2006) Multiplexed expression and screening for recombinant protein production in mammalian cells. BMC Biotechnol 6:49

28. Fallaux FJ, Bout A, van der Velde I et al (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. Hum Gene Ther 9:1909–1917

29. Jones D, Kroos N, Anema R et al (2003) High-level expression of recombinant IgG in the human cell line per.c6. Biotechnol Prog 19:163–168, PubMed PMID: 12573020

Part VII

Recombinant Protein Production in Other Systems

Chapter 13

Soluble Recombinant Protein Production in *Pseudoalteromonas haloplanktis* TAC125

Maria Giuliani, Ermenegilda Parrilli, Filomena Sannino, Gennaro Apuzzo, Gennaro Marino, and Maria Luisa Tutino

Abstract

Solubility/activity issues are often experienced when immunoglobulin fragments are produced in conventional microbial cell factories. Although several experimental approaches have been followed to solve, or at least minimize, the accumulation of the recombinant proteins into insoluble aggregates, sometimes the only alternative strategy is changing the protein production platform.

In this chapter we describe the use of Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 as host of choice for the production of the heavy-chain antibody fragment VHHD6.1. Combining the use of a regulated psychrophilic gene expression system with an optimized fermentation process in defined growth medium, we obtained the recombinant VHHD6.1 in fully soluble form and correctly translocated into host periplasmic space.

Key words *Pseudoalteromonas haloplanktis* TAC125, VHHD6.1, Psychrophilic gene expression system, LIV medium, Batch fermentation

1 Introduction

Till their first description [1], IgG antibodies from Camelidae (camels, dromedaries, and llamas) attracted attention of either basic or applied scientists due to their feature of "heavy-chain antibody" or HCAb. Indeed these antibodies are naturally devoid of light chains and it makes their size significantly lower than conventional IgG antibodies. Consequently, their binding domains consist only of the heavy-chain variable domains, referred to as VHHs [2], to distinguish them from conventional VHs. VHH is the smallest available intact antigen-binding fragment (~15 kDa) and it has a great potential in therapeutic and diagnostic application as multispecific fusion product [3].

Due to their reduced structural complexity, VHHs are often successfully produced in conventional microbial cell factories, such as *Escherichia coli* [4]. However, there may be still a fraction of

VHHs which escapes the binding evaluation tests due to poor stability in soluble form in the recombinant production host. These potentially valuable molecules are committed to be fatally overlooked, if the microbial production platform does not display optimized features for antibodies production.

Over the last years, our research group has been focused on the exploitation and implementation of the unconventional marine bacterium *Pseudoalteromonas haloplanktis* TAC125 as recombinant protein production host [5–7]. This psychrophilic Gram-negative bacterium, isolated from Antarctic sea water [8], displays several metabolic and physiological traits that justify a moderate interest as alternative protein production platform to be used when the other conventional microbial systems fail [9, 10].

Indeed, the combination of its optimal growth performances at reduced temperature—where hydrophobic interactions are significantly minimized—and a rich arsenal of folding factors and catalysts—supporting high-quality protein folding at low temperatures—allowed us to produce in soluble and active forms several difficult-to-express proteins [9–11]. Amongst them, a Fab antibody fragment [12] and a single-chain antibody fragment (unpublished results from this laboratory) highlighted an interesting proficiency of this bug in producing immunoglobulin-derived molecules.

To prove the ability of *P. haloplanktis* TAC125 to successfully produce soluble antibody fragments, an anti-human fibroblast growth factor receptor 1 (FGFR1) VHHD6.1 was chosen as model protein. It was selected by phage display from a pre-immune llama library [13] but its large-scale production in conventional *E. coli* expression systems was unsatisfactory due to inclusion bodies formation (De Marco A, personal communication). A new production process leading to improve soluble production of VHHD6.1 is therefore required for its further characterization.

In the present chapter, we describe the procedure for the cloning of *vhhD6.1* gene into a modified pUCRP psychrophilic gene expression system and its mobilization into *P. haloplanktis* TAC125 cells. Recombinant Antarctic strain was then grown at 15 °C in optimized culture conditions (LIV medium, batch cultivation in a 3 L STR automatic fermenter) and protein production followed by monitoring VHHD6.1 production and cellular localization over 60 h cultivation process, leading to the definition of optimal process conditions for the production of VHHD6.1 in soluble and fully periplasmic form.

2   Materials

2.1   Bacterial Strains   1. *P. haloplanktis* TAC125. This strain was kindly provided by C. Gerday, University of Liege, Belgium. The strain was isolated from the sea water in the surrounding of the Dumont

d'Urville Antarctic station (66°40′S, 40°01′E) during the 1988 summer campaign of the "Expeditions Polaires Française" in Terre Adélie [14].

2. *E. coli* DH5α [*supE*44, *ΔlacU*169 (ϕ80 *lacZΔM*15) *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1]. This strain was used as host for the gene cloning.

3. *E. coli* strain S17-1(λ*pir*) [*thi*, *pro*, *hsd* (r⁻ m⁺) recA::RP4:2-TCʳ::Mu Kmʳ::Tn7 Tpʳ Smʳ λ*pir*]. This strain was used as donor in intergeneric conjugation experiments [7].

| *2.2 Solutions* | 1. 100 mg/mL ampicillin stock solution: dissolve 1 g of ampicillin powder in 8 mL of deionized H_2O. Adjust the volume of the solution to 10 mL with deionized H_2O and sterilize by filtration through a 0.22 μm sterile filter. Split the obtained stock solution in 10 aliquots of 1 mL each in sterile polypropylene tubes and store them at –20 °C. |

2. 1× TAE buffer for agarose gel electrophoresis: 40 mM Tris–acetate, 1 mM EDTA, pH 8. Make a 50× TAE stock solution by mixing 242 g of Tris base, 57.1 mL of glacial acetic acid, 100 mL of 0.5 M EDTA, pH 8, and adjust the volume of the solution to 1 L with deionized H_2O. Store at room temperature (RT) up to 1 year.

3. 0.5 M EDTA, pH 8: dissolve 186.1 g of EDTA in 800 mL of deionized H_2O. Adjust the pH to 8 with NaOH (about 20 g of NaOH pellets) and adjust the volume of the solution to 1 L with deionized H_2O.

4. 3 M NaCl stock solution: dissolve 87.6 g of NaCl in 500 mL of dH_2O.

5. 6× agarose gel-loading buffer: 0.25 % bromophenol blue, 0.25 % xylene cyanol FF, and 30 % glycerol (Fermentas).

6. 10 mg/mL ethidium bromide: add 1 g of ethidium bromide to 100 mL of deionized H_2O. Stir on a magnetic stirrer for several hours to ensure that the dye has dissolved. Wrap the container in aluminum foil and store it at 4 °C.

7. 20 % (w/v) L-malate stock solution: Dissolve 10 g of L-malic acid in 40 mL of deionized H_2O. Adjust the pH to 7.0 with 5 M NaOH. Adjust the volume of the solution to 50 mL with dH_2O and sterilize by filtration through a 0.22 μm sterile filter. Store at 4 °C up to 2 months.

8. 1 M DTT stock solution: Dissolve 3.09 g of DTT in 20 mL of deionized H_2O. Sterilize by filtration. Dispense into 1 mL aliquots and store them at –20 °C.

9. 1× SDS-PAGE loading buffer: 62.5 mM Tris–HCl, pH 6.8, 10 % glycerol, 2 % SDS, 100 mM DTT, and 0.1 % bromophenol blue. This buffer lacking DTT can be stored at RT.

DTT should then be added just before that the buffer is used from a 1 M stock.

10. 0.5 M Tris–HCl, pH 6.8: Dissolve 60.55 g of Tris base in 800 mL of deionized H_2O. Adjust the pH to 6.8 with HCl and add dH_2O to make up a final volume of 1 L.

11. 0.5 % (w/v) bromophenol blue: Dissolve 0.25 g bromophenol blue powder in 45 mL of dH_2O. Shake well to dissolve the dye and then adjust the volume of the solution to 50 mL with dH_2O. Store at RT.

12. 5× Running buffer: Dissolve 15.1 g of Tris base, 94 g of glycine, and 5 g of SDS in 900 mL of dH_2O. Adjust the volume of the solution to 1 L with dH_2O.

13. 0.5 M Phosphate buffer: Dissolve 68.9 g of NaH_2PO_4 in 900 mL of dH_2O. Adjust the pH to 7.3 with NaOH and add dH_2O to make up a final volume of 1 L.

14. Borate buffer: Dissolve 7.63 g of $Na_2B_4O_7$, 0.76 g of NaCl in 90 mL of dH_2O. Add 1 mL EDTA 0.5 M, pH 8, shake and adjust the volume of the solution to 100 mL with dH_2O.

15. Western blot 1× Transfer buffer: Dissolve 3.03 g Tris base, 14.41 g glycine in 800 mL of deionized H_2O. Add 200 mL methanol. Adjust the volume of the solution to 1 L with dH_2O.

16. Western blot blocking buffer: Dissolve 50 g Skimmed Milk in 1 L of PBS buffer (5 % w/v). Add 1 mL Triton X-100 (0.1 % v/v) and mix.

17. Western blot washing buffer: Add 1 mL of Triton X-100 in 1 L of PBS buffer (0.1 % v/v) and mix.

18. 1× PBS: Dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 in 800 mL of distilled H_2O. Adjust the pH to 7.4 with HCl. Add H_2O to 1 L. Sterilize by autoclaving.

19. 20× SCHATZ Salts: Dissolve 20 g of KH_2PO_4, 20 g of NH_4NO_3, 4 g of $MgSO_4·7H_2O$, 0.2 g of $FeSO_4$, and 0.2 g of $CaCl_2·2H_2O$ in 1 L dH_2O. Adjust the pH to 7.0 by HCl addition. Sterilize by filtration through a 0.22 μm sterile filter. Store at 4 °C up to 2 months.

2.3 Media

1. LB medium (1 L) [15]: 10 g Bacto-Tryptone, 5 g Bacto-Yeast Extract, 10 g NaCl. Adjust 950 mL with deionized H_2O. Shake until the solutes have dissolved. Adjust the volume to 1 L with dH2O. Sterilize by autoclaving for 20 min at 1 atm on liquid cycle. Let it cool down and store at RT. When required, add 1 mL of sterile ampicillin stock solution. After antibiotic addition store the medium at 4 °C up to 2 weeks. To prepare solid medium, add 15 g/L Bacto-Agar just before autoclaving.

2. TYP medium (1 L) [7]: 16 g Bacto-Tryptone, 16 g Bacto-Yeast Extract, 10 g Marine mix, add 950 mL of deionized H_2O.

Shake until the solutes have dissolved. Adjust the pH to 7.5 with 5 N NaOH. Adjust the volume of the solution to 1 L with deionized H_2O. Sterilize by autoclaving for 20 min at 1 atm on liquid cycle. When required, add 1 mL of sterile ampicillin stock solution. After antibiotic addition store the medium at 4 °C up to 2 weeks. To prepare solid medium, add 15 g/L Bacto-agar just before autoclaving.

3. LIV medium (1 L) [12]: 1 g KH_2PO_4, 1 g NH_4NO_3, 10g NaCl, 0.2 g $MgSO_4 \cdot 7H_2O$, 10 mg $FeSO_4$, 10 mg $CaCl_2 \cdot 2H_2O$, 5 g L-leucine, 5 g L-isoleucine, 10 g L-valine, add 900 mL of deionized H_2O. Shake until the solutes have dissolved , adjust the volume to 1 L with dH_2O, and sterilize by filtration through a 0.22 μm sterile filter. Store at 4 °C. When required, add 1 mL of sterile ampicillin stock solution. After antibiotic addition store the medium at 4 °C up to 2 weeks. For protein induction add 2 mL of sterile 20 % w/v L-malate stock solution.

2.4 Reagents for Molecular Biology

– Phusion™ DNA Polymerase.
– Taq DNA Polymerase.
– Restriction enzymes.
– Calf Intestinal Phosphatase, CIP.
– T4 DNA ligase.
– PCR Product Purification Kit.
– Miniprep Kit.
– Nucleotide Removal Kit.
– pGem®-T Easy Vector System I.
– Anti c-Myc mAb produced in mouse.
– Peroxidase conjugated anti-mouse IgG.
– SuperSignal West Femto Chemiluminescent Substrate.

2.5 Storage Medium

Bacteria can be stored indefinitely in cultures containing 39 % of sterile glycerol (sterilize by autoclaving for 20 min at 1 atm on liquid cycle). At low temperature (from –20 to –70 °C).

3 Methods

3.1 VHHD6.1 Expression Vector Construction

For L-malate-inducible VHHD6.1 production in *P. haloplanktis* TAC125 cells, *vhhD6*.1 gene was cloned in pUCRP cold expression vector [16] previously modified by the addition of the sequences encoding the N-terminal *Ph*DsbA leader peptide for periplasmic secretion and C-terminal c-*myc* tag and 6xHis-tag (Fig. 1).
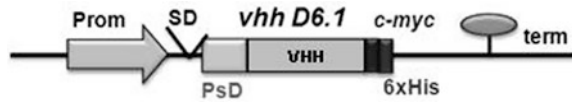
Fig. 1 Schematic representation of VHHD6.1 expression cassettes. *SD* Shine Dalgarno sequence, *Prom* psychrophilic L-malate inducible promoter, *term* psychrophilic terminator, *PsD Ph*DsbA signal peptide

The pUCRP-inducible expression vector [5] is a pUCLT/ Rterm derivative [17] containing the transcription promoter region of the *PSHAb0363* gene, which responds to the presence of L-malate into the culture medium. The pUCLT/Rterm plasmid, deriving from the pUC18 plasmid, is characterized by the presence of (1) the pJB3-derived OriT [18], a DNA fragment responsible for the initiation of the conjugative transfer between *E. coli* S17-1 λ*pir* strain (donor) and the psychrophilic cells (acceptor); (2) a pUC18-derived polylinker wherein the target gene can be cloned; (3) *E. coli bla*M gene, encoding a mesophilic β-lactamase which is used for the selection of the recombinant clones; (4) OriC, the origin of replication allowing the plasmid to replicate in *E. coli*; (5) the T/R box, a DNA fragment containing the cold-adapted origin of replication (OriR) [19]; (6) the *TaspC*, the transcription termination signal of the aspartate aminotransferase gene (*aspC*) isolated from *P. haloplanktis* TAC125 [14].

The addition of the molecular signal for periplasmic addressing is necessary in order to facilitate the correct folding of the recombinant product. The VHHs, like the other antibodies and antibody fragments, contain disulfide bonds in their immunoglobulin domains required for the binding activity. The oxidizing environment and the dedicated chaperones present in bacterial periplasm can prevent cysteine reduction and aid to the correct disulfide bond isomerization. In addition, the use of the leader peptide isolated from *P. haloplanktis* DsbA [20] addresses the recombinant secretion through a co-translational SRP-like secretion system [21], limiting the fast protein aggregation often observed in the cytoplasm of bacteria expressing recombinant antibody fragment secreted by posttranslational Sec-dependent mechanism.

The addition of C-terminal tandem tag will allow the easy protein detection by Western blot using monoclonal anti-c-Myc tag antibodies and IMAC purification trough the 6xHis-tag (Fig. 1).

1. The *vhhD6.1* gene is amplified on pHEN-D6.1 source vector (kindly provided by Dr. A. De Marco, IFOM-IEO campus Milan), in order to insert a 5′ *SalI* and a 3′ *NotI* restriction site by using primers VH-S-fw (5′-ATCGTGTCGACATGGC TGAGGTGC-3′) and VH-N-rv (5′-ATATATGCGGCCGC AATGGAGACGGTG-3′), respectively. PCR reaction is carried

out by Phusion® High-Fidelity DNA Polymerase following the manufacturer's protocols.

2. The amplified product is purified using a commercial purification kit and then digested with *SalI* and *NotI*. Restriction hydrolyses are performed by using five enzyme units/μg of DNA, in the reaction conditions defined by the manufacturer.

3. The pUCRP vector is digested with *SalI* and *NotI*.

4. The 5′ phosphate groups of the cleaved vector are dephosphorylated by treatment with calf intestinal alkaline phosphatase (0.5 U/pmol of 5′ phosphate ends) for 15 min at 37 °C and 45 min at 55 °C by using the appropriate buffer delivered with the enzyme. The CIP is heat-inactivated at 75 °C for 10 min.

5. The dephosphorylated DNA is then loaded on a 1 % agarose gel (containing ethidium bromide as fluorescent marker for the migrating DNA). The DNA is cut out of the gel and purified using a commercial gel-purification kit following the manufacturer's instructions.

6. The cleaved dephosphorylated vector is then ligated to the digested amplification products by using two consecutive ligation reactions, by the means of T4 DNA ligase according to the supplier's instructions.

7. The ligation reaction mixture is used directly for transformation of the chemically competent bacteria (DH5α *E. coli* strain) according to the procedure described by Hanahan [22].

8. Recombinant clones are selected on LB agar plates containing 100 μg/mL ampicillin as selection agent.

9. Plasmids are isolated from ampR clones and the presence of the appropriate insert is verified by restriction digestion analysis (*see* **Note 1**).

10. Finally the nucleotide sequences of the inserts are checked by DNA sequencing to rule out the occurrence of any mutation during synthesis.

The resulting expression vector, pUCRP-*vhhD6.1*, contains the *vhhD6.1* gene in-frame to N-terminal PsD and C-terminal c-Myc and 6xHis tag coding sequences.

3.2 Construction of *P. haloplanktis* TAC125 pUCRP-vhhD6.1 Recombinant Strain

1. The resulting vector pUCRP-vhhD6.1 is mobilized into *P. haloplanktis* TAC125 by intergeneric conjugation [7]. Cells are plated on TYP solid medium containing 50 μg/mL ampicillin and incubated at 4 °C to select recombinant *P. haloplanktis* TAC125 (the low temperature avoid *E. coli* growth as colony).

2. Three colonies are picked and inoculated in 3 mL of TYP liquid medium containing 100 μg/mL ampicillin and incubated at 15 °C under shaking (250 rpm) for 24 h.

3. Plasmidic DNA is extracted from each clone by using a commercial kit and recombinant plasmid clones were screened by PCR amplification of *vhhD6.1* gene.

**3.3 VHHD6.1
Production**

VHHD6.1 production is carried out according to the optimized protocol described in [12]. In detail, recombinant *P. haloplanktis* (pUCRP-*vhhD6.1*) batch cultivation was performed in a STR 3 L fermenter connected to a Bio-controller with a working volume of 1 L, in SCHATZ mineral medium supplemented with 0.5 % w/v L-leucine, 0.5 % w/v L-isoleucine, and 1.0 % w/v L-valine (LIV medium), 100 μg/mL ampicillin with additional 0.4 % w/v L-malate as inducer. The culture was carried out at 15 °C in aerobic conditions (Dissolved Oxygen Tension (DOT) ≥ 30 %), airflow of 20 L/h, and a stirring rate of 500 rpm. The culture pH was maintained at 7.00 by automatic addition of H_2SO_4 5 % v/v. The cell biomass from a pre-inoculum, performed in shaken flask with the same medium and temperature used for the successive experiment in batch, was used to inoculate batch cultures.

The controller automatically registers DOT, pH, and temperature values every minute during the whole process and acts on acid pump or the water bath connected to the water jacket to keep pH and temperature within the set point range. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600nm}) using a spectrophotometer.

3.3.1 Process Setup

1. Fill the vessel with the media carbon sources and NaCl dissolved in 1 L of dH_2O.

2. Insert the pH and DOT probes (the pH electrode must be previously calibrated according to manufacturer's instructions) and the stirrer.

3. Connect the tubes for sampling and for the inoculum and seal it with a clamp. Seal the open connections with aluminum foil. Connect a sterile 0.22 μM filter to the air inlet tube.

4. Fill in the water jacket.

5. Prepare a 250 mL Pyrex bottle with two ports cover containing H_2SO_4 5 % (v/v), connect a tube and seal it with a clump.

6. Sterilize both the vessel and the acid bottle at high temperature (120 °C for 50 min at 1 atm) in autoclave.

7. After autoclaving, discard the water from the jacket and let the vessel cool down to a comfortable handling temperature then connect the pH and DOT electrodes to the bio-controller and the temperature probe to the vessel. Connect the water jacket to a thermostated water bath set at 15 °C. Connect the tube from acid bottle to the peristaltic pump and to the vessel.

8. Complement the medium by adding 50 mL of SCHATZ Salts stock solution (20×) and 1 mL of ampicillin stock solution (1,000×) through the inoculum tube using a 50 mL syringe.

9. Turn on the Bio-controller and set the following parameters: $(DOT) \geq 30$ %.

pH 7.00 ± 0.2.

Temperature $15\ °C \pm 0.5$.

10. When the system reaches the desired temperature and pH, calibrate the DOT electrode by connecting the airflow inlet first to a nitrogen tank and setting the 0 % DOT when nitrogen saturation is obtained; then let the air in at maximum stirring rate to set the 100 % DOT.

11. After DOT calibration set the airflow at 20 L/h and stirring rate at 500 rpm.

3.3.2 Preculture

The viability of the precultured cells is crucial for a satisfying process outcome. The growth phase must be in middle exponential phase in the same medium that will be used for the fermentation.

1. From a glycerol stock streak the *P. haloplanktis* TAC125 pUCRP-*vhhD6.1* strain on a TYP agar plate containing 100 mg/L of ampicillin. Incubate it at 15 °C for about 36 h. The plate can be stored up to 3 days at 4 °C, carefully sealed with Parafilm to avoid oxygen availability to the cells (*see* **Note 2**).

2. Pick a single colony and inoculate it in 2 mL of liquid TYP medium supplemented with ampicillin 100 mg/L in a 14 mL snap-cap inoculation tube and incubate at 15 °C under vigorous shaking (250 rpm) for 36 h (*see* **Note 3**).

3. Dilute the inoculum in 50 mL of LIV medium supplemented with ampicillin 100 mg/L in a 250 mL flask and incubate for 16–18 h at 15 °C under vigorous shaking (250 rpm). The final biomass concentration should be 5–7 OD_{600nm}.

3.3.3 Fermentation Process

The fermentation process will follow the general procedure described below. However, each clone can behave differently needing further optimization depending on the properties of the protein to be produced. In the following procedure, the parameters set up at the beginning of the process can be changed according to specific requirements of individual processes. For instance, the airflow is set to 20 L/h at the beginning of the process but to guarantee the sufficient oxygen supply to the growing cells $(DOT > 30$ %) it can be increased manually during the process up to 40 L/h. The stirring rate indeed cannot be increased over 500 rpm due to system limitation and therefore the optimal DOT can be achieved only by playing with the airflow inlet.

1. Inoculate the amount of preculture required in order to obtain a starting concentration of $OD_{600nm} = 0.2$. To calculate it, register the optical density of the preculture at 600 nm using a spectrophotometer. Since the culture working volume in

bioreactor is 1 L, calculate the volume of inoculum by using the following formula:

$$\text{mL inoculum} = \left(\text{Culture OD}_{600\text{nm}} \times \text{mL culture volume}\right) / \text{Preculture OD}_{600\text{nm}}.$$

Use a 50 mL syringe connected to the inoculum tube. Carefully remove the aluminum seal and rapidly insert the syringe (to keep sterility, it can be useful to operate near the flame of a Bunsen burner). Remove the clamp before inserting the inoculum in the tube. Pipet up and down with the syringe several times to be sure that nothing lasts in the tube dead volume.

2. Wait a couple of minutes until the suspension becomes homogeneous, letting the stirrer on, then take a sample from the bioreactor to register the actual starting optical density at 600 nm. Use a 20 mL syringe connected to the sampling tube and pipet up and down several times to be sure to sample from the inside of the culture and not the dead volume of the tube.

3. Monitor the cell growth by measuring the $OD_{600\text{nm}}$ as described above. Register the data of at least two measurements to avoid the technical error. When the cell density reaches an $OD_{600\text{nm}}$ of 0.6–0.8 which corresponds to early exponential phase induce the recombinant gene expression by L-malate addition. Add 20 mL of 20 % w/v L-malate sterile stock solution to obtain the optimal inducer concentration of 0.4 % w/v using a syringe connected to the inoculum tube.

4. At different times after induction collect a sample corresponding to an $OD_{600\text{nm}}$ of 25 in triplicate. Calculate the volume of each sample using the following formula:

$$\text{mL sample} = 25 / \text{Culture optical density OD}_{600\text{nm}}.$$

Collect samples after about 24, 30, 42, and 60 h after induction by centrifuging the calculated volume for 15 min at $200 \times g$ at 4 °C. Discard the supernatant and store the biomass indefinitely at –80 °C.

5. Plot the optical density values versus the time of cultivation in graph. In a typical process (Fig. 2) the growth profile shows a diauxic growth due to the differential consumption of carbon sources during growth. During the first exponential the highest specific growth rate is reached ($\mu_{max} = 0.13$ h^{-1}); the L-valine is rapidly consumed within the first 24 h (data not shown). The second exponential growth phase starts at about 24 h of cultivation and lasts for the next 24 h with a very low specific growth rate. The total time of the process is about 60 h; afterwards the DOT increases and cell lysis starts (data not shown). Sampling point corresponds to the different phases of the fermentation process.
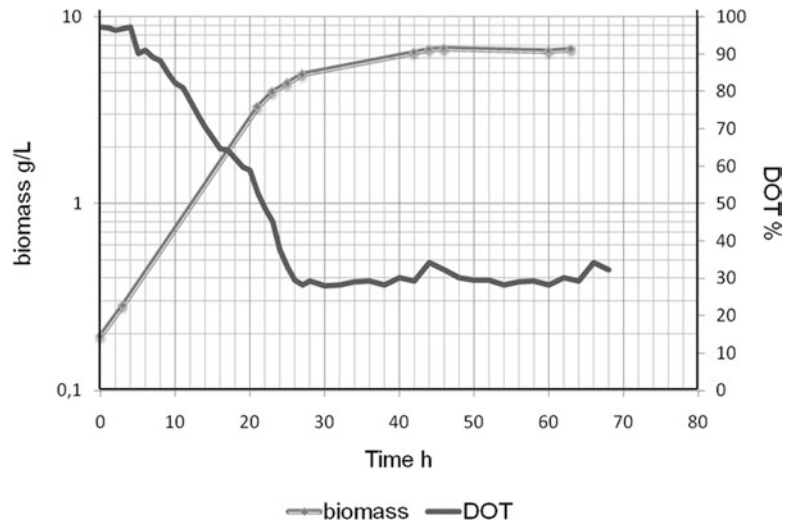
Fig. 2 *P. haloplanktis* TAC125 (pUCRP-*vhhD6.1*) fermentation profiles. The biomass concentration is reported as g/L of dry cell weight calculated according to [12]

3.4   VHHD6.1 Production Analysis

To test the process efficiency, the protein detection itself is not sufficient. In order to obtain a correctly folded and, consequently, a biologically active antibody fragment product, the protein should be localized in periplasmic compartment where, due to the proper chemical-physical properties and chaperones, the correct formation of disulfide bonds contained in VHH immunoglobulin domains can be achieved.

To analyze the VHHD6.1 production and periplasmic secretion, a cellular fractionation followed by Western blotting analysis is required.

3.4.1   Cell Lysis

1. Prepare 100 mL of Lysis buffer by diluting 10 mL of 0.5 M phosphate buffer stock solution and 10 mL of NaCl stock solution in 80 mL of dH_2O.

2. Resuspend the bacterial pellet collected ($OD_{600nm} = 25$) at different fermentation time point in 1 mL of lysis buffer by pipetting or vortexing.

3. Add 10 µL of 100 mM PMSF stock solution and 1 µL of 0.5 M EDTA stock solution and mix (*see* **Note 4**).

4. Apply five cycles of a benchtop French Press at 1.8 kbar to each sample.

5. Centrifuge the suspension at $8,200 \times g$ for 20 min at 4 °C. Recover the resulting supernatant containing the total soluble protein extract for further analysis. Keep the protein extract on ice or store it a 4 °C for no longer than 2 h.

3.4.2 Periplasm
Extraction

1. Resuspend an aliquot of bacterial pellet collected $(OD_{600nm} = 25)$ at different fermentation time point in 0.2 mL of borate buffer by gently pipetting.

2. Incubate for 16–18 h at 4 °C.

3. Centrifuge at $8,200 \times g$ for 15 min at 4 °C and store the supernatant containing the periplasmic protein extract for further analysis. Keep the extract on ice or store it at 4 °C for no longer than 2 h.

3.4.3 Western Blot
Protein Detection

1. Mix 12 μL of each sample prepared at Subheadings 3.4.1 and 3.4.2 with 4 μL of 4× SDS-PAGE loading buffer. Boil the samples for 5 min at 95 °C and load on 15 % SDS-PAGE gel. Load 16 μL of the samples coming from increasing time points alternating total protein extracts with periplasmic protein extracts. Run the gel for 45 min at constant 50 mA.

2. Wash the gel three times with transfer buffer for 10 min.

3. Transfer the protein on a 0.2 μm PVDF membrane previously activated in methanol according to the manufacturer's instruction.

4. Block the membrane for 1 h in blocking buffer under shaking at RT.

5. Dilute the anti-c-Myc monoclonal antibody (*see* Subheading 2) 1:5,000 in blocking buffer by diluting 2 μL of antibody in 10 mL. Incubate the membrane with the primary antibody solution for 1 h at RT under shaking.

6. Discard the primary antibody solution and wash the membrane three times with Western blot washing buffer for 10 min.

7. Dilute HRP anti-mouse antibody (*see* Subheading 2) 1:10,000 in blocking buffer by diluting 1 μL of antibody in 10 mL. Incubate the membrane with the primary antibody solution for 1 h at RT under shaking.

8. Discard the secondary antibody solution and wash the membrane five times with Western blot washing buffer for 10 min.

9. Develop the Western blot using chemiluminescence.

The analysis (Fig. 3) reveals VHHD6.1 production in soluble form during all fermentation and its correct periplasmic localization during early (22 h) and middle (29 h) exponential growth phase. In contrast, no recombinant VHHD6.1 was found in periplasmic fraction extracted from samples collected at late exponential growth phase (42 h) and stationary phase (60 h) while production titers seem to increase during exponential growth reaching the highest yield at late exponential phase (42 h). Furthermore, another specific signal showing an apparent molecular weight of about 30 kDa was detected in total soluble protein extracts probably corresponding to VHHD6.1 dimers. It is not
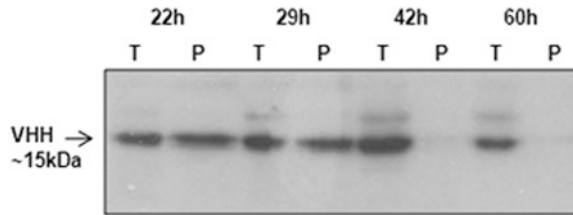
Fig. 3 Western blotting analysis. VHHD6.1 soluble production and cellular localization on total protein extracts (T) and periplasmic fraction (P) of recombinant *P. haloplanktis* TAC125 (pUCRP-*vhhD6.1*) cells collected at different times of fermentation. Expected recombinant VHHD6.1 molecular weight: 15 kDa

surprising since a strong tendency of multimerization has been reported for this and other formats of antibody fragments in vivo when their local concentration in recombinant host cells reaches a critical value [23]. It is worth noticing that as far as the high molecular weight signal relative intensity increases the secretion efficiency of recombinant product into periplasmic space seems to decrease. One explanation can be found in VHHs dimerization kinetics that could be faster than the product recruitment by the periplasmic secretion system. On the other hand, VHH dimers formation could be a consequence of its cytoplasmic localization. If protein secretion does not occur, its correct folding cannot be achieved and hydrophobic interactions can take place among partly folded intermediates thus causing protein molecules aggregation in soluble dimeric complexes. Although the co-translational SRP-mediated secretion system was successfully employed for Fab [12] and ScFv (unpublished results) formats model proteins, VHHD6.1 translocation across the inner membrane results to be somehow inhibited at high cell densities. Further investigation has to be carried out in order to find out the reason of this phenomenon and the best strategy to overcome it.

4   Notes

1. Alternatively, screen for recombinant plasmid clones by using PCR to directly amplify the insert from each bacterial recombinant colony.

2. The psychrophilic bacteria are able to grow at temperature as low as 4 °C. Limiting oxygen availability can reduce the growth but not avoid it.

3. This passage is optional but helps to overcome a long lag phase due to the adaptation of bacteria coming from a complex rich culture medium (TYP) to the defined LIV medium.

4. Add PMSF and EDTA to the lysis buffer to prevent proteolytic degradation of recombinant product.

Acknowledgement

References

1. Hamers-Casterman C, Atarhouch T, Muyldermans S et al (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446–448

2. Muyldermans S, Lauwereys M (1999) Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. J Mol Recognit 12:131–140

3. Joosten V, Lokman C, van den Hondel C et al (2003) The production of antibody fragments and antibody fusion proteins by yeasts and filamentous fungi. Microb Cell Fact 2:1

4. Harmsen MM, De Haard HJ (2007) Properties, production, and applications of camelid single-domain antibody fragments. Appl Microbiol Biotechnol 77:13–22

5. Rippa V, Papa R, Giuliani M et al (2012) Regulated recombinant protein production in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. Methods Mol Biol 824:203–218

6. Giuliani M, Parrilli E, Pezzella C et al (2012) A novel strategy for the construction of genomic mutants of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. Methods Mol Biol 824:219–233

7. Duilio A, Tutino ML, Marino G (2004) Recombinant protein production in Antarctic Gram negative bacteria. Methods Mol Biol 267:225–237

8. Médigue C, Krin E, Pascal G et al (2005) Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. Genome Res 15:1325–1335

9. Corchero JL, Gasser B, Resina D et al (2013) Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. Biotechnol Adv 31:140–153

10. Gasser B, Saloheimo M, Rinas U et al (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11

11. Vigentini I, Merico A, Tutino ML et al (2006) Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. J Biotechnol 127:141–150

12. Giuliani M, Parrilli E, Ferrer P et al (2012) Process optimization for recombinant protein production in the psychrophilic bacterium *Pseudoalteromonas haloplanktis*. Process Biochem 46:953–959

13. Monegal A, Ami D, Martinelli C et al (2009) Immunological applications of single-domain llama recombinant antibodies isolated from a naïve library. Protein Eng Des Sel 22:273–280

14. Birolo L, Tutino ML, Fontanella B et al (2000) Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. Cloning, expression, properties, and molecular modelling. Eur J Biochem 267:2790–2802

15. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor

16. Papa R, Rippa V, Sannia G et al (2007) An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. J Biotechnol 127:199–210

17. Tutino ML, Parrilli E, Giaquinto L et al (2002) Secretion of α-amylase from *Pseudoalteromonas haloplanktis* TAB23: two different pathways in different hosts. J Bacteriol 184:5814–5817

18. Blatny JM, Brautaset T, Winther-Larsen HC et al (1997) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. Appl Environ Microbiol 63:370–379

19. Tutino ML, Duilio A, Parrilli E et al (2001) A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. Extremophiles 5:257–264

20. Madonna S, Papa R, Birolo L et al (2006) The thiol-disulphide oxidoreductase system in the cold-adapted bacterium *Pseudoalteromonas haloplanktis* TAC 125: discovery of a novel disulfide oxidoreductase enzyme. Extremophiles 10:41–51

21. Schierle CF, Berkmen M, Huber D et al (2003) The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. J Bacteriol 185: 5706–5713

22. Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

23. Hollinger P, Hudson PJ (2005) Engineered antibody fragments and the rise of single domains. Nat Biotechnol 23:1126–1136

Part VIII

Insoluble Protein Purification

A Screening Methodology for Purifying Proteins with Aggregation Problems

Mario Lebendiker, Michal Maes, and Assaf Friedler

Abstract

Many proteins are prone to aggregate or insoluble for different reasons. This poses an extraordinary challenge at the expression level, but even more during downstream purification processes. Here we describe a strategy that we developed for purifying prone-to-aggregate proteins. Our methodology can be easily implemented in small laboratories without the need for automated, expensive platforms. This procedure is especially suitable for intrinsically disordered proteins (IDPs) and for proteins with intrinsically disordered regions (IDRs). Such proteins are likely to aggregate due to their lack of tertiary structure and their extended and flexible conformations. Similar methodologies can be applied to other proteins with comparable tendency to aggregate during the expression or purification steps.

In this chapter, we will mainly focus on protein solubility and stability issues during purification and storage, on factors that can prevent aggregation or maintain solubility, and on the importance of the early elimination of aggregates during protein purification.

Key words Protein aggregation, Insoluble proteins, Intrinsically disordered proteins, Protein storage, Protein concentration, Stabilizers, Aggregation suppressors, Chaotropes, Kosmotropes, Buffer conditions, Aggregation analysis

1 Introduction

1.1 Insoluble Proteins, Instability, and Aggregation

Stability is an extremely important issue in protein production, due to the fact that once destabilized, proteins are susceptible to chemical and physical alteration that lead to loss of activity. Chemical alteration as protein cleavage or related to covalent bond modifications like oxidation and disulfide bond shuffling. Physical changes include protein unfolding, undesirable binding to surfaces, and aggregation [1]. These undesirable changes can be reversible or irreversible. They can produce aggregates that range in size from soluble aggregates, only detectable by size exclusion chromatography (SEC), to particles that may contain trillions (or more) of monomer units visible by the eye [2]. There is a great concern about the presence of aggregates in therapeutic proteins because of
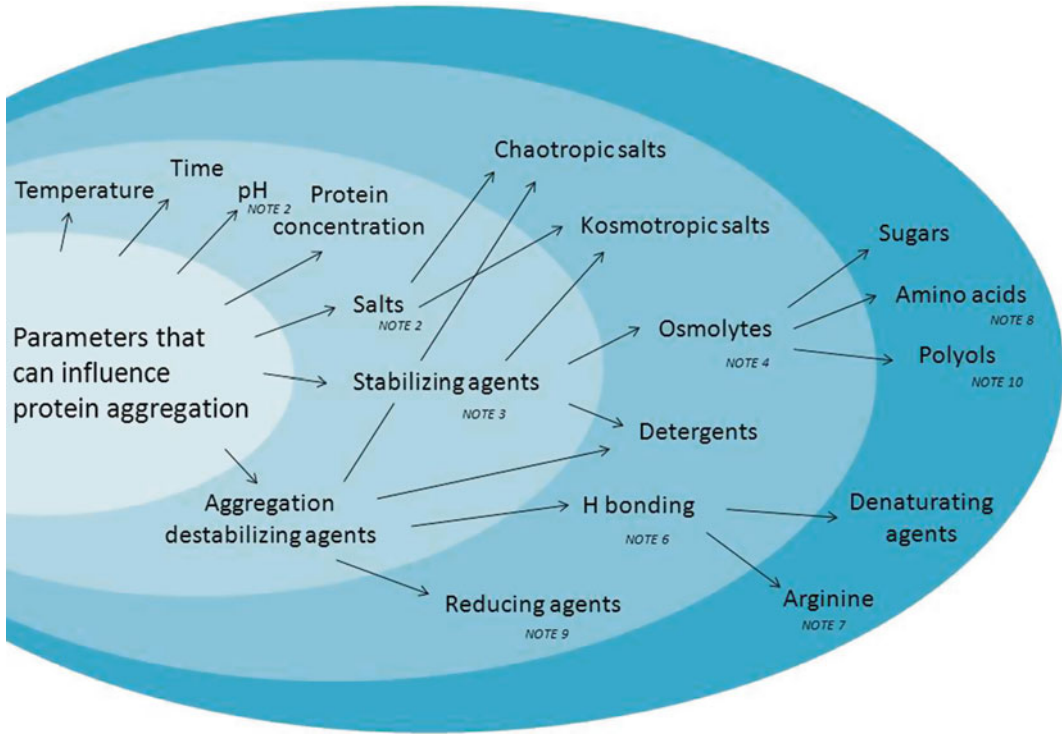
Fig. 1 Critical issues to be considered in order to prevent aggregation during protein purification

their unpredictable ability to give rise to adverse toxicological and immunological responses, which in extreme cases can be life-threatening [2]. As the number of therapeutic proteins increases, finding ways to understand and prevent this problem continues to gain importance. The same issues of protein instability and aggregation cause many problems in basic as well as in applied research: protein production yields are decreased, aggregated proteins are unable to crystallize, their specific activity is highly affected, and the credibility of the results using aggregated proteins in all kinds of experiments is questionable.

Aggregation is an undesired interaction between protein monomers. This process can be influenced by temperature, protein concentration, buffer conditions, etc. (Fig. 1). There is an extended lag phase before large aggregates appear and accumulate in an abrupt way [3].

Protein aggregates may be classified in numerous ways, including soluble/insoluble, covalent/non-covalent, reversible/irreversible, native/denatured, or by size, conformation, and morphology [4, 5]. Some efforts are made for nomenclature standardization and classification [4]. Five major mechanisms of aggregation have been proposed: concentration-induced aggregation, aggregation induced by conformational changes, aggregation induced by

chemical reactions, nucleation-dependent aggregation, and surface-induced aggregation [3, 6]. A fundamental understanding of the mechanism of aggregation is not only valuable for identifying the cause of the problem but is also helpful for developing methods to suppress aggregation [3, 6].

Insolubility of recombinant proteins may be encountered already at the expression level. Several solutions can overcome this problem. These include screening of different bacterial strains, decreasing culture temperatures, different culture mediums, different fusion protein constructs such as maltose binding protein (*see* Chapter 2), alternative expression systems such as cell-free expression (*see* Chapter 6) or baculovirus (*see* Chapter 9), using constructs with either amino or carboxyl-terminal deletions, expression of homologs of a protein of interest, removing flexible loops or residues that affect solubility, and refolding of denatured proteins [7]. As detailed, many of these approaches will be extensively discussed in other chapters.

1.2   Factors that Influence Protein Aggregation and Insolubility During Expression and Purification

Although protein solubility during expression is an essential prerequisite before purification, this does not prevent aggregation problems from arising at later stages of the protein production process (*see* **Note 1**). In this chapter, we will focus mainly on protein stability issues that must be considered from the very early purification steps until storage. Some general issues that can lead to denaturation and aggregation and should be considered are purification time and temperature, protein concentration at each step, and prevention of mechanical or nonmechanical stresses (freezing, exposure to air, interactions with metal surfaces, etc.). Other factors that can influence aggregation are pH or ionic strength. In addition, the protein environment can be affected by cosolutes such as chaotropes and kosmotropes (*see* **Note 2**), osmolytes and ligands, protein–protein interaction inhibitors, reducing agents, surfactants, and non-denaturative detergents. All of these can be divided into two main categories: factors that stabilize proteins and factors that inhibit aggregation or inhibit protein–protein interactions (Fig. 1, *see* **Note 3**).

A change in solution conditions such as a decrease in protein concentration or changes in pH or salt concentration can dissociate the aggregates in some cases. This is especially true for aggregates where the molecules are held together by relatively weak, non-covalent interactions. However, such changes rarely affect other types of aggregation. Such pH- or salt-dependent reversibility is indicative of equilibrium between the monomer and high-order forms [5].

Protein stabilizers are additives that inhibit aggregation by stabilizing the native structure of the protein [8]. There is correlation between additives that stabilize proteins against thermal stress in cells and additives that stabilize proteins during isolation and

storage. These stabilizing cosolutes are also termed osmolytes, since they are utilized in nature to increase the osmotic pressure of the cellular environment and are compatible with the macromolecular function and cell viability (*see* **Note 4**) [8]. Examples for such osmolytes are trehalose and trimethylamine *N*-oxide (TMAO), both used for protein refolding [9]. Other examples include sucrose, glycerol, sorbitol, mannitol, glycine betaine (betaine) [10], and proline [11]. Polyethylene glycol (PEG) (*see* **Note 5**) and kosmotropic salts as magnesium or ammonium sulfate [8, 12, 13] and potassium citrate [7] also act as protein stabilizers. Alcohols such as ethanol can be used to stabilize folding intermediates by weakening hydrophobic interactions that facilitate aggregation [12] (unpublished data).

Aggregation suppressors can work in several ways. The H-bonding agents, like urea or guanidine HCl (GdnHCl), work as chaotropic agents at low concentration (0.5–2 M). They decrease the net hydrophobic effect of prone-to-aggregate hydrophobic regions in proteins by disordering the water molecules adjacent to the protein surface (*see* **Note 6**). The way L-arginine hydrochloride (L-ArgHCl) protects proteins from aggregation is more complicated. It can act as an H-bonding agent like urea or GdnHCl, but it has certain kosmotropic properties, allowing it to interact with aromatic side chains of the protein (*see* **Note 7**) [14]. Other amino acids such as proline, histidine, and beta-alanine, as well as the naturally occurring polyamines putrescine, spermidine, and spermine, were also reported as aggregation suppressors [15].

Aggregation can be induced by chemical modifications such as incorrect disulfide bond or arrangement or the formation of bityrosine (*see* **Note 8**). The presence of weak reducing agents and oxidants can reverse this problem or lead to changes in protein conformation that may alter the function of the protein. Reducing agents can break disulfide bonds and lead to dissociation of parts of the protein chain(s) that are normally associated. Oxidants can cause the formation of disulfide bonds and consequent association of parts of the protein chain that are normally not associated (*see* **Note 9**).

Surfactants are used in biotechnology to stabilize therapeutic proteins, suppress aggregation, and assist in protein refolding. They can prevent protein adsorption on surfaces, which would result in loss of activity and/or surface-induced aggregation. Surfactants can also bind hydrophobic regions in proteins and thus prevent aggregation [6]. Some widely used surfactants are polysorbate, poloxamers, and non-detergent sulfobetaines (NDSBs) (*see* **Note 10**) [6].

Although recommended additive concentrations are found in the literature [12, 13, 16, 17], the optimal range for each protein is highly specific, and the buffer conditions must be fine-tuned for each project (*see* **Note 11**). Moreover, there could be a synergistic

effect between some of these agents. This could prevent different aggregation mechanisms, for example, osmolytes, as cosolutes will favor protein structures with minimal surface area, while addition of surfactants can mask exposed hydrophobic regions [13]. Each family of additives will improve solubility of some proteins while decreasing the solubility of others. The same kosmotrope environment that stabilizes folded proteins can enhance protein–protein interactions and subsequent aggregation in partially unfolded proteins. On the other hand, chaotropic agents that destabilize aggregation of proteins in the native state can induce or enhance aggregation of partially unfolded proteins [13].

Finding the optimal buffer conditions can be performed using functional biological assays, but this is not applicable to all proteins. There could be cases where no assay is available, the assay is not reliable, or alternatively time, effort, and cost make the assay unfruitful. In any case, such assays do not provide information regarding yield, oligomeric homogeneity, and protein purity.

Several experimental methods are routinely used to determine aggregation: visual observation of turbidity, size exclusion chromatography (SEC), circular dichroism (CD), light scattering (LS), fluorescence-based thermal shift (ThermoFluor) assay, and more (*see* **Note 12**). No single method is optimal for all aggregates. Since there is a large number of variables to determine (different buffers, pH, additives, salt, etc.), there is a need for a progressive and rational experimental methodology that can be used to identify the optimal buffer conditions and additive concentrations to maintain protein solubility. High throughput screening (HTS) assays are not always available, so alternatives must be found (*see* **Note 13**). In a recent publication, Leibly et al. used a screening methodology with 144 additives, but only the classical ones gave the best results (*see* **Note 11**) [7]. Their findings confirm our assumption that for nonautomated laboratories, using a shorter list of additives covering most of the aggregation mechanisms can considerably reduce cost and efforts.

1.3 Our Approach for Minimizing Aggregation

We present a new approach for minimizing aggregation. Our approach is based on a hierarchical buffer selection using a small group of additives, covering different mechanism of aggregation inhibition. A similar approach has been previously reported by the Bondos' lab for pure or almost pure proteins (*see* **Notes 12** and **13**) [12, 13]. To maximize yield and information, we prefer to tackle the solubility issue early, starting from the cell lysate, and then continue analyzing the oligomeric state of the partially pure protein during the different purification steps until the final pure product.

Our strategy (Fig. 2) begins with a screening of solubility-promoting buffers during cell lysis, followed by a quick capture step by parallel small-scale immobilized metal chelate
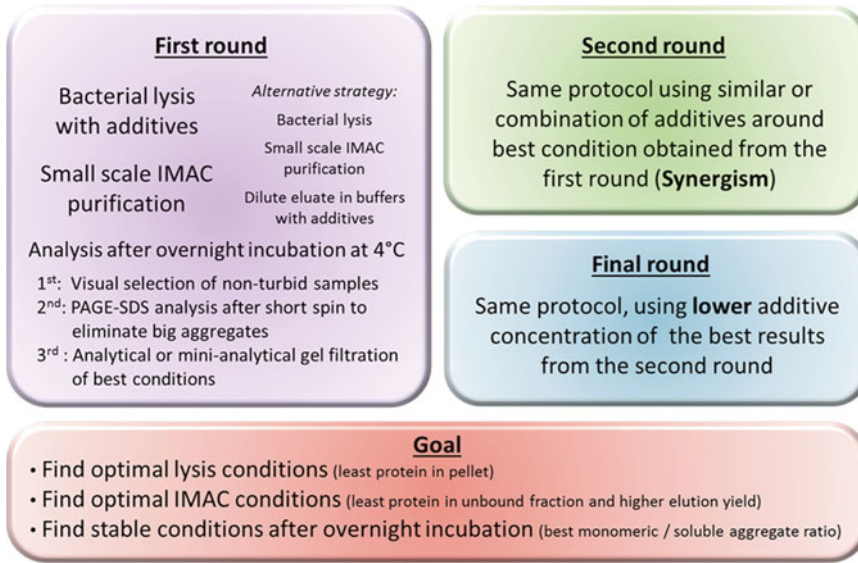
Fig. 2 Screening methodology for purifying proteins with aggregation problems

chromatography/Ni column (IMAC) purification (or any other capture method) in the presence of selected additives. Analysis is performed by SDS-PAGE of the insoluble lysis extract, unbound fraction, and eluted protein. Only the best elution conditions are further analyzed by analytical SEC, immediately after elution and after 24 h at 4 °C (time dependent aggregation), searching for the best monomer/soluble aggregate ratio (Fig. 3b). From this first screening, it is possible to estimate the influence of different additives groups on insoluble aggregated proteins, allowing better binding to the capture resin and as a result obtaining the best yield of native oligomeric conformation. This strategy not only provides maximum information on solubility issues but also improves the final output, since it rescues the protein fraction that was initially soluble in the bacteria but was then secluded as insoluble protein [7]. The following optimization rounds (Fig. 2) check if other additives from the same category may give better results, together with a combination of agents that can synergize protein solubility. In the final step, the additives concentration is optimized together with stability over time (Fig. 3b). For some projects, different buffers, pH, and additive type and concentration must be matched for each purification step. In these cases, before scaling up, the best additive(s) must be found for all intermediate steps and for storage (*see* **Note 14**). In some cases, the beneficial effect of the chosen additives may be maximal during cell lysis and early purification steps. This will allow drastic reduction of their concentrations at later stages [7].

Since aggregation is a nucleation-growth process, the presence of soluble aggregates during bacterial lysis can accelerate the
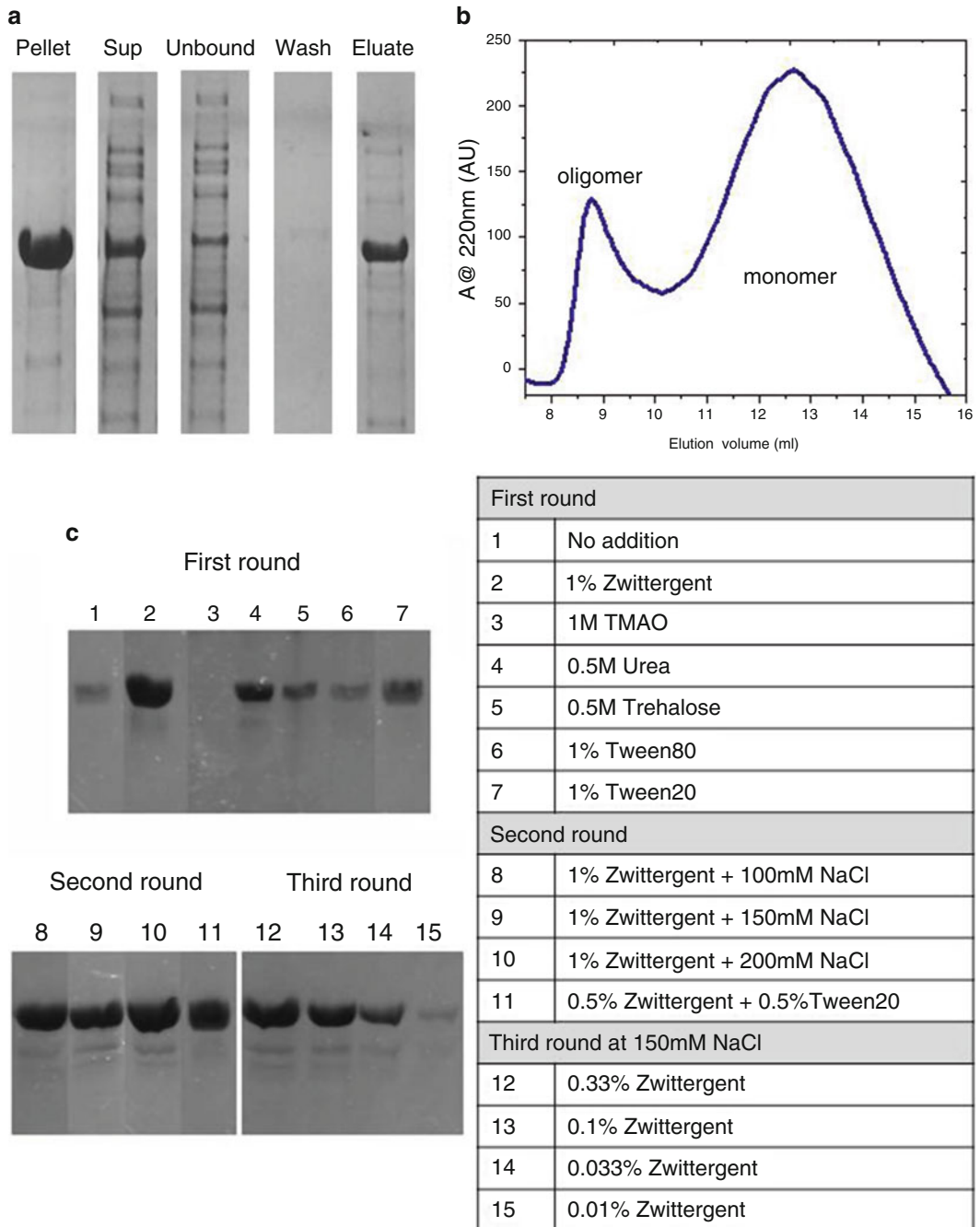
Fig. 3 Example of screening methodology. (**a**) small-scale IMAC purification; (**b**) spectrum of analytical gel filtration – the protein appears in two distinct peaks, oligomer and monomer; (**c**) three rounds of optimization for additives, according to the table (Data contributed by Dr. Ronen Gabizon)

insolubilization process. Thus, the classical strategy used by many of the protein-producing laboratories (first capture on IMAC column, followed by protease cleavage under dialysis and negative IMAC, and final polishing by size exclusion chromatography) [18, 19] can be harmful while processing of prone-to-aggregate proteins. The risk is that the presence of soluble aggregates after the first IMAC column will trigger the insolubility of more protein molecules and decrease total yield (unpublished data, *see* **Note 15**). A better strategy for such proteins is to try and remove the soluble aggregates as soon as possible by performing SEC (or other chromatographic procedures, *see* **Note 16**) immediately after the IMAC purification, followed by tag cleavage. In extremely problematic projects, we observed that high protein concentration during cell lysis or at the top of the column during chromatographic loading could sometimes speed up the aggregation process. These problems can be overcome by higher lysis volume or batch purification (*see* **Note 17**) or by immediate dilution of the concentrated protein after elution (*see* **Note 18**).

Once the conditions that give the optimal ratio between active protein and unusable aggregates are found, they must be checked for suitability with long-term storage or certain particular experimental requirements (NMR, crystallography, etc.). In some cases, an additional screening will be required to determine the buffer conditions appropriate for storage and specific usage (unpublished data, *see* **Note 19**).

Finally, the importance of designing a "quick strategy of purification" must be emphasized, since process time is one of the most critical points to consider. Pure protein should be produced and stored as quickly as possible. For this reason, maximum efforts must be made to optimize and fine-tune each purification step before scale-up, guaranteeing that the whole process can be performed quickly and smoothly.

The experience accumulated in our laboratory using these approaches with many IDPs and IDRs is useful for project-oriented protein production of prone-to-aggregated proteins in academic and nonautomated laboratories (without standard HTS).

2   Materials

2.1   First Round of Buffer Additives

1. Basic lysis/wash buffer: 50 mM Tris–HCl, 500 mM NaCl with 10 % glycerol, pH 8.0, with/without β-mercaptoethanol (BME) (*see* **Note 9**), and different *additives*.

2. Lysis buffer: wash buffer, 1 mM PMSF, lysozyme 0.2 mg/mL, DNase 50 μg/mL, protease inhibitor cocktail.

3. Elution buffer: wash buffer, 300 mM imidazole, and *additives*.

4. *Additives*: (a) 1 M guanidine HCl, (b) 1 M urea, (c) 0.5 % Tween 20, (d) 0.5 % *n*-tetradecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate (Zwittergent 3–14), (e) 0.5 M trehalose, (f) 500 mM L-ArgHCl (only in the elution buffer).

2.2 Cell Lysis	1. Microfluidizer (LV1, Microfluidics Corp., Newton, MA) or Sonicator (Sonics Vibra Cell VCX 750) for small scale (less than 10 mL).
	2. Microfluidizer (M-110 EHIS, Microfluidics Corp., Newton, MA) for large volumes (more than 10 mL).
2.3 Small-Scale IMAC Purification	1. IMAC beads: Ni-NTA or similar beads for small-scale batch purification.
	2. IMAC beads: Ni Sepharose High Performance or similar beads for large-scale column purification.
2.4 Analytical and Mini-Analytical Size Exclusion Chromatography (SEC)	1. ÄKTA explorer system (GE Healthcare).
	2. Analytical SEC Superdex™ 200 or 75 HR 10/30 or Superose 12 30×1 (GE Healthcare). Use according to molecular weight of the protein. Flow: 0.7 mL/min.
	3. Mini-Analytical SEC: homemade columns using Superdex™ 200, Superdex™ 75, Superose 12 resin, and Tricorn 5/200 column (~4 mL) (GE Healthcare). Flow: 0.3 mL/min.

2.5 Second Round: Buffer Optimization

pH optimization: prepare several buffers changing two variables: pH and conductivity

1. 50 mM MES, pH 6.0.
2. 50 mM phosphate buffer, pH 7.0.
3. 50 mM Tris–HCl, pH 8.5.
4. 50, 300, and 500 mM NaCl.
5. Different *additives* to each buffer (*see* **Note 9**).

2.6 Set Up Concentration Limit

Disposable 0.5 mL ultrafiltration devices or protein concentrators with molecular weight cutoff lower than that of the native protein.

3 Methods

3.1 First Buffer Selection: Different Types of Additives (Fig. 2)

Prepare wash, lysis, and elution buffers with the different *additives* and with or without BME. Each *additive* represents a different mechanism of protein stabilization or suppression of aggregation (*see* Subheading 2.1 and **Note 9**). Add L-ArgHCl only in the elution buffer.

3.2 Cell Lysis and Clarification	1. Grow bacterial cells and induce protein expression according to the best overexpression conditions found (temperature, time, [IPTG], induction time, etc.).

2. Harvest cells and keep aliquots of 15, 100, and 500 mL pellet cells at –80 °C until further processing (100 and 500 mL aliquots will be used for future scale-up).

3. Resuspend different pellets from 15 mL cell culture in 1.5 mL lysis buffer with different additives (*see* Subheading 2.1) and lyze mechanically using a Microfluidizer at 21,000 psi at 4 °C or sonication on ice for 3×10 s or more (*see* Subheading 2.2) if the cells are not completely disrupted (lysis is complete when the cloudy cell suspension becomes translucent; avoid protein denaturation by frothing and extensive sonication). Remove insoluble cell debris from the cell lysate by centrifugation at 4 °C for 20 min $18,000 \times g$. Separate clear supernatant (lysate) from the pellet. Keep sample of supernatant for further analysis by SDS-PAGE or Western Blot: supernatant. Continue with supernatant (*see* Subheading 3.3).

4. Resuspend pellet (insoluble cell debris) in 1.5 mL buffer and keep sample for further analysis by SDS-PAGE or Western Blot: pellet.

3.3 Small-Scale IMAC Purification

1. Equilibration of IMAC beads: place 200 μL IMAC beads in a 2 mL plastic centrifuge tube for each condition. Wash once with 1.5 mL H_2O and twice with 1.5 mL lysis buffer (washing: mix, spin 3 min at $1,200 \times g$ discard supernatant).

2. Mix supernatant of each condition with its equivalent equilibrated resin at 4 °C for 60 min.

3. Spin for 3 min at $1,200 \times g$ at 4 °C. Discard supernatant and keep sample of 40 μL (unbound proteins) for PAGE-SDS or Western Blot.

4. Wash beads with 1.5 mL buffer (of each condition) at least three times: mix, spin 3 min $1,200 \times g$, keep supernatant (wash). Be careful not to remove the resin.

5. Elute recombinant protein twice with 300 μL buffer with 300 mM imidazole (incubate 5 min each time before spinning 3 min, $1,200 \times g$ at 4 °C). Elution sample is obtained. Keep sample for PAGE-SDS or Western Blot.

6. Keep elution pools at 4 °C for further use.

3.4 Analysis of First Round: Different Types of Additives. Alternative and Less Comprehensive Screen

The emphasis in this first screening is on checking additives that act by different mechanisms to suppress or avoid aggregation. For certain projects, this first run could be enough to determine the best conditions. An alternative screening use partially purified protein after the first capture step (*see* **Note 20**). This alternative screening, although faster and simpler, is less comprehensive. The best results

of these screens can be later applied to all the steps during medium- and large-scale purifications.

1. For each condition, run samples on SDS-PAGE: pellet, supernatant, unbound to IMAC, and eluted proteins (*see* Subheading 3.3). Analyze them by Coomassie staining.

2. Profile for the best additive: less target protein in the pellet and in the unbound fraction and higher protein concentration in the elution.

3. Keep best elution samples overnight (ON) at 4 °C.

4. Visual selection of non-turbid samples. Spin best samples 15 min 18,000×*g* at 4 °C and discard pellet. Only the best elution conditions are analyzed immediately by analytical or mini-analytical SEC, searching for the best monomeric/soluble aggregate ratio along time.

5. A simple but less informative option is to run SDS-PAGE after ON incubation at 4 °C, and spin: higher soluble protein along time, without any indication about the oligomeric conformation.

6. Check Western Blot only in case of low protein concentration or to verify the presence of the target and absence of cleavage products.

7. Final evaluation for this round: estimation of the influence of different additives groups on lowering insoluble aggregated proteins, allowing better binding to the IMAC resin, with the healthiest oligomeric conformation along time (Fig. 2).

3.5 Second Round of Optimization: Similar Additives or Combination of Additives from the First Round (Fig. 2)

In this optimization round, the emphasis is on finding alternative additives from the same group of the best additives from the first round and testing possible synergism of different additives with different modes of action.

1. Repeat small-scale IMAC purification using other additives of the same group as the best results from the first round (similar to [13]; for more information, *see* [12, 17]).

2. If trehalose gives the best results, try other osmolytes: 1 M TMAO, sorbitol or sucrose, 0.05 % polyethylene glycol 3,350 or 6,000.

3. If a detergent such us Tween 20 gives the best results, try 0.5–1 % of other surfactants like Nonidet P40, Tween 80, or Brij 35, or detergents used for crystallization of membrane proteins, octyl glucoside (*n*-octyl-β-D-glucoside) (OG) or *n*-dodecyl-β-D-maltoside (DDM).

4. If 0.5 % Zwitergent 3–14 gives the best results, try 1 M non-detergent sulfobetaines (NDSBs), 0.5 % 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), or Lauryl-dimethylamine *N*-oxide (LDAO).

5. If L-ArgHCl in the elution buffer is the best, try other amino acids as proline or a combination of 50 mM L-Arg with 50 mM L-Glu [20, 21].

6. Mix additives such as osmolytes and surfactants if both work or try other possible synergistic combinations.

3.6 Third Round: Lower Concentration of Additives (Fig. 2)

This optimization step is used for projects in which maximal decrease in additive concentration is important (e.g., detergents that can affect downstream applications, expensive additives, or undesirable chemicals such as urea or GdnHCl). Repeat low-scale IMAC purification using sequential dilutions of the target additive(s) (Fig. 3c). Alternatively, the additive concentration can be drastically reduced during elution or during later purification steps [7] (unpublished data).

3.7 Buffer Optimization Designed for Subsequent Purification Steps

This optimization step is performed when additives used for the first capture step are incompatible or undesirable in the following purification steps. Some other parameters not checked in the first purification step can be checked here: different pH, different salt concentrations (very important for ion or hydrophobic exchange columns), and other types of reducing agents or surfactants.

1. Dilute protein samples after first IMAC column 1:4 with a matrix of different buffers (*see* Subheading 2.5). The two main variables should be pH and conductivity. Additional additives and reducing agents can be added according to previous results.

2. Keep ON at 4 °C (alternative: experimental stresses; *see* **Note 21**).

3. Spin 20 min at $18,000 \times g$ and 4 °C. Run SDS-PAGE or perform Western Blot on supernatants.

4. Profile of best conditions: most protein in the supernatant after long incubation at 4 °C.

5. Only the best conditions are analyzed immediately by analytical or mini-analytical SEC, searching for the best monomer/soluble aggregate ratio.

3.8 Set Up Concentration Limit and Best Stability/ Storage Conditions

For many biochemical and structural studies, there is a need for highly concentrated protein. Reaching such concentrations is a difficult task for prone-to-aggregate proteins. This screening is applied to purified protein in order to find the best buffer conditions for maintaining maximal protein concentration and long-term stability during storage. In this round, like in the previous round, different parameters should be tested including pH, salt concentration, and other types of reducing agents or surfactants.

1. Select concentrator (*see* Subheading 2.6). As a general rule, the pore size of the concentrator membrane should be two times smaller than the molecular weight of the protein. Select the concentrator volume size according to your needs.

2. Add some buffer to the concentrator and rinse the membrane. Use the concentrator immediately after washing and avoid drying the membrane. Always start with a small sample to determine the upper limit before concentrating the total amount of protein.

3. Spin according to the manufactures' instructions for a few minutes and check the protein concentration. If losses are higher than 20–30 %, check for protein concentration in the flow through. If protein is detected in the flow through, it may be that the unit is damaged or a smaller MWCO should be used.

4. Continue protein concentration by incremental steps. Take samples after each step. Aliquot the sample and keep part of the samples at –80 °C (*see* **Note 19**) and the rest of the sample ON at 4 °C.

5. Spin aliquots that were at 4 °C (20 min 18,000 × g, 4 °C). Run SDS-PAGE or check the protein concentration.

6. Profile evaluation of best conditions: highest protein concentration in the supernatant after long incubation at 4 °C.

7. Only the best conditions are then analyzed by analytical or mini-analytical SEC, looking for the best monomeric/soluble aggregate ratio.

8. Repeat same evaluation with aliquots keeps at –80 °C (*see* **Note 19**).

9. Use this information to concentrate and store your protein during scale-up.

4   Notes

1. Many laboratories use a simple protocol based on a small screening by SDS-PAGE to check the presence of soluble and insoluble proteins after cell lysis and centrifugation. We emphasize the importance of minimal presence of soluble aggregates as well as insoluble aggregates. The quality of the overexpressed product must be evaluated in order to minimize undesired aggregates during purification down the line. To reach this goal, we coupled small-scale expression and analysis by SDS-PAGE and analytical gel filtration for the optimal ratio of monomer/soluble aggregate in the bacterial lysates (similar to [22]). During expression, conditions must be found that give minimal presence of aggregates (both soluble and insoluble) and maximal yield of the native overexpressed protein.

2. Heat increases the kinetic energy of the protein chain, and this increase can break relatively weak H-bonds, electrostatic

interactions, and hydrophobic interactions, speeding up the aggregation process. pH change can affect the charge of acidic or basic functional groups in the protein and thus disrupt or create electrostatic repulsion that will alter the protein structure. Ionic strength can affect protein aggregation in different ways by reducing desired electrostatic interactions at high salt or increasing undesired electrostatic interactions at low salt. This can result in either stabilization or destabilization of proteins, or even denaturation [15]. This effect can differ for chaotropic or kosmotropic ions (mainly anions). Kosmotropic salts such as ammonium or magnesium sulfate stabilize the native protein state favoring protein–water interactions (so-called water-structure makers) [12, 13]. They are usually small ions with low polarizability and a bigger "salting-out" effect according to the Hofmeister series. Chaotropic salts, like magnesium chloride (with higher "salting-in" effect according to the Hofmeister series), are water-structure breakers and protein destabilizers. They can also inhibit protein–protein interactions by shielding charges and preventing stabilization by salt bridges [12, 13].

3. Factors that enhance protein stability interact mainly with the solvent. On the other hand, factors that suppress protein aggregation operate mainly by binding to the protein surface or by competitive binding to the interface that has the potential to destabilize the protein structure or cause aggregation [8].

4. Through the interaction of water molecules with osmolytes, water molecules are excluded from protein surface, thus stabilizing the native state of the protein with the smallest surface area [12, 23]. The addition of such cosolutes not only stabilizes many proteins but also deters ice formation, thus inhibiting the harmful effects of freezing on protein structure [12].

5. The amphiphilic polymer polyethylene glycol (PEG) is intensively used for protein refolding [24] and for protein stabilization by chemical modification (i.e., PEGylation) [25]. PEG interacts with the hydrophobic side chains that become exposed upon unfolding. Because of their high water solubility, low toxicity, and low antigenicity, PEGs are used in protein engineering to enhance refolding, assist in crystallization, increase water solubility, and prolong the blood circulation time of proteins [26]. Polyvinylpyrrolidone (PVP), a similar amphiphilic polymer, is applied in pharmaceutical products due to its low toxicity [15].

6. H-bonding agents, such as urea or GdnHCl, interfere with intramolecular interactions mediated by non-covalent forces such as hydrogen bonds, van der Waals forces, and hydrophobic effects. High concentration of these additives can lead to

protein unfolding by either a direct interaction with the protein [15] or an indirect effect on the surrounding water structure. Most likely, these two mechanisms are not mutually exclusive [21]. At low concentration (0.5–2 M), they act as chaotropic agents.

7. The mode of interaction between L-ArgHCl and proteins is still under extensive investigation [8, 14]. L-ArgHCl as an aggregation suppressor during refolding was first reported in a patent application [27]. Stepwise decrease of denaturant concentration in combination with the addition of L-ArgHCl is a conventional method for protein refolding [28]. It is also a versatile additive for protein formulation and affinity column chromatography [15]. It was shown to reduce nonspecific protein binding in SEC, to facilitate elution of antibodies from protein A columns, to enhance elution of resin-bound proteins, and as a solvent for elution in hydrophobic interaction chromatography (HIC) and ion exchange chromatography (IEC) [14].

   A well-known synergistic enhancement of protein solubility is achieved by the combination of L-ArgHCl and L-glutamic acid (L-Glu). They interact with oppositely charged residues on the protein surface and mask the surrounding exposed hydrophobic patches [20, 21]. Only 50 mM of each compound are necessary, instead of the high concentrations (around 0.5–1 M) of L-ArgHCl alone. The mixture can be added to eluted protein after the first IMAC column and to all subsequent buffers.

8. Bi-tyrosine formation as a consequence of tyrosine oxidation is a chemical modification that can stimulate aggregation [5]. Oxygen scavengers such as methionine or sodium thiosulfate can avoid this aggregation [6].

9. Reducing agents must be used during extraction and purification if cysteines in the target protein are predicted or known to be free. This would prevent protein aggregation by inhibiting the formation of nonnative disulfide bonds. The most common reducing agents are dithiothreitol (DTT), β-mercaptoethanol (BME), or tris-(2-carboxyethyl) phosphine hydrochloride (TCEP). TCEP is a non-thiol and odorless compound, stable in aqueous solutions, and resistant to air oxidation. Unlike DTT, TCEP retains its reducing ability at acidic pH and at pH above 7.5 [29].

   It is best to use BME during IMAC purification, since DTT or TCEP are incompatible with many of the IMAC resins. Using 5–15 mM, BME can avoid the formation of nonnative disulfide bonds. In other chromatographic procedures, BME can be replaced by other reducing agents.

No reducing agents must be used if only disulfide bonds are predicted. A problematic crossroad is a mixture of free cysteines and disulfide bonds in the same protein target. Our approach in this case is not to use reducing agents at all, or a very low BME concentration (2 mM), as a compromise solution.

There are several websites that can predict the bonding state of cysteines on proteins, such as Cyspred (http://gpcr.biocomp.unibo.it/cgi/predictors/cyspred/pred_cyspredcgi.cgi), DiANNA (http://clavius.bc.edu/~clotelab/DiANNA/), and DISULFIND (http://disulfind.dsi.unifi.it/).

10. Polysorbate 80 (polyoxyethylene sorbitan monooleate) and polysorbate 20 (polyoxyethylene sorbitan monolaurate) are surfactants that are widely incorporated in marketed protein pharmaceuticals. Used in the 0.0003–0.3 % range [8], they are reported to suppress aggregation upon agitation, shaking, freeze-drying, and freeze-thawing processes and can prevent protein adsorption at solid surfaces [6].

    Poloxamers like the triblock copolymers of polyethylene oxide–polypropylene oxide–polyethylene oxide (PEO–PPO–PEO) or commercially available poloxamers such as Pluronics® or Synperonics™ are used in pharmaceutical formulations [30]. Poloxamer 188 (BASF Pluronic® F68) is widely used for the large-scale production of mammalian cell culture, especially when bioreactors are used to amplify a cell population [6].

    Non-detergent sulfobetaines (NDSBs) are very good aggregation suppressors. They have a short hydrophobic group and a hydrophilic sulfobetaine head group, which is a zwitterion over a wide pH range. NDSBs do not behave like detergents, since their hydrophobic group is too short to form micelles even at concentrations as high as 1 M. This property allows them to be easily removed by dialysis. Moreover, they weakly bind proteins. All these reasons make them sometimes more useful than detergents [15].

11. Buffer conditions can potentially alter protein conformation or activity. These effects can vary at different cosolvent concentrations, using different cosolvents from the same family, changing protein concentration, or depending on the protein purification stage.

    Screening of 144 additive conditions for increasing the solubility of recombinant proteins expressed in *E. coli* was recently described [7]. The classical additives gave the best results: trehalose, glycine betaine, mannitol, L-ArgHCl, potassium citrate, $CuCl_2$, proline, xylitol, NDSB 201, cetyltrimethylammonium bromide (CTAB), and K_2PO_4.

12. An easy alternative aggregation test is the visual observation of turbidity as a result of precipitation. This can be performed by

observing aggregates under a microscope (*Giladi, O., 2012 Rational optimization of protein stability. P4EU Workshop on Protein Purification*) or by optical observation at different wavelengths (340, 490, or 600 nm). These approaches, although fast and easy to perform, require large volumes of concentrated protein (to allow screening by buffer dilution) and can only detect highly insoluble and very large protein aggregates, while soluble aggregates remain undetected.

There are more laborious analytical methods to check for soluble aggregates. The most popular of these methods is SEC [22]. Other methods like CD or light scattering (LS) are not always available in all laboratories, and their results are more difficult to interpret. Analytical ultracentrifugation is the most accurate, but it is very expensive. Native gels are much cheaper, but need to be optimized for each protein and do not give an analytical result.

A filter-based aggregation assay used on crude cell extract or partially purified proteins was described [12]. After incubation in different buffers and under different conditions, the soluble, non-aggregated protein was separated from the big aggregates. This was performed using little ultracentrifugation devices where the MWCO was selected such that soluble protein was allowed to pass through the filter, while aggregate forms were retained. Analysis was done by SDS-PAGE or Western blotting [12].

Another method, used mainly for protein characterization for crystallography or NMR, is the fluorescence-based thermal shift (ThermoFluor) assay. An environmentally sensitive dye, Sypro Orange, is used to monitor the thermal stability of a protein. This assay can be used to investigate the effect of factors (buffers, additives, or ligands) on protein stability [31, 32]. RT-PCR machines with fluorescent detectors are used to compare shifts of T_m (midpoint of the unfolding transition on the melting curve).

There are several commercial assays with similar approach. All these can be employed to streamline protein processing and optimize formulation procedures:

- OptiSol™ Protein Solubility Screening Kit (Dilyx Biotechnologies) based on a filtration assay.
- ProteoStat™ protein aggregation assay (Enzo-Life Sciences or BioTek), using Thioflavin T as a fluorescence dye and a multi-mode microplate reader.
- Optim1000 (Avacta) combines fluorescence and static light scattering technologies.

As is described in the methods section, we prefer to use standard SDS-PAGE to select the best buffer and additives (electrophoresis of supernatant after ON incubation with

different additives), followed by SEC or mini-SEC for a more precise analysis. SEC completes the information about protein purity with information about the oligomerization state and allows a rapid estimation of the presence and amount of soluble aggregates, although larger aggregates seem to be lost in the pre-column filters [22]. Since aggregation is time dependent, we use ON incubation as a relative compromise. In addition, SEC can be coupled in-line to a light scattering device to measure the absolute molar mass, size, and shape of macromolecules in solution. Although not using a high amount of protein, the main disadvantage of SEC is that it is time consuming for nonautomated laboratories.

13. In a recent report, the first buffer selection was performed by the type of chemical that best improves solubility, followed by identifying the optimal chemical and its most effective concentration [13]. The report describes a filter-based aggregation assay used on crude cell extract to rapidly identify buffers that maintain protein solubility for purification and subsequent assays (*see* **Note 12**). A similar work was published some years ago with a very good table of agents that may promote protein solubility [12]. In spite of its simplicity, this approach yields less information regarding optimal purification conditions.

14. The isolation and purification of a tagged protein can be achieved by using a cheap and convenient affinity column that can yield tagged protein with 70–90 % purity following a single-capture step. Further purification is done by ion exchange, hydrophobic exchange, size exclusion chromatography, and the new mixed-mode chromatography columns (*see* **Note 16**) in order to achieve a higher degree of purification, which is often required for downstream applications.

Ion exchange chromatography is essential as an intermediate step for separating target proteins from protein contaminants such as chaperones and other host cell proteins. It also allows separating the target protein from heterogeneously folded forms that are a consequence of the expression and purification conditions used and from heterogeneous post-translational modifications. Sometimes ion exchange chromatography does not sufficiently separate the impurities, and additional chromatographic methods are required. These should be based on different principles, such as hydrophobic exchange, mixed mode, or hydroxyapatite. SEC is often recommended as a final purification step in order to eliminate protein contaminants and low molecular weight molecules and to obtain a homogeneous oligomeric form [33].

15. For some projects, we found that changing the order of the purification steps gave better results. This way the soluble aggregates were eliminated after the first capture step by SEC

before tag cleavage by specific proteases. SEC increased the purity of the protein and adjusted the initial buffer conditions for next columns. Long cleavage incubation times can be circumvented by increasing the protease concentration.

16. SEC is the method of choice to separate different oligomers. Symmetric elution profiles are characteristic of homogeneous proteins, whereas asymmetric profiles reflect nonhomogeneous, partially aggregated samples or large aggregates if eluting in the void volume of the chromatogram (or when the column is in poor condition) [34]. Recently, a great effort has been done to produce resins with high capacity and high flow rates, to be used for separating recombinant proteins from aggregates. Since these operate on a "mixed-mode" mechanism, based on a combination of electrostatic and hydrophobic properties of the proteins and ligands, they are called "multimodal" or "mixed-mode" resins. Examples of them are Capto adhere or Capto MMC (GE Healthcare); HEA, PPA, and MEP HyperCel (PALL); MX-Trp-650 M (Tosoh); Eshmuno HCX (Merck); and Hydroxyapatite (BioRad).

    We have observed several times that high selective ion exchange columns can also separate different oligomeric states (unpublished data).

17. For prone-to-aggregate proteins, the ratio of lysis buffer to cell mass is extremely important and can lead to aggregation before the first purification step. We suggest to use at least twice or more lysis buffer for the same cell mass (1:5 to 1:10 of initial culture).

    For some difficult projects, we preferred to use a batch binding of the crude lysate to the resin, in order to avoid the aggregation of the protein in the upper part of the column during loading. An alternative option is to use an excess of resin to avoid molecular crowding, although this approach can compromise the purity of the final product. A similar approach is used for purification of membrane proteins.

18. Since proteins are concentrated in the upper side of the columns during all chromatographic procedures except SEC, it happens that proteins with an extreme tendency toward aggregation start to precipitate immediately after elution. A small volume of buffer can be added to the collection tubes in order to obtain an immediate dilution of the protein and avoid or inhibit aggregation.

19. It is prudent to use a small sample to examine the stability of the protein for both protein concentration and freeze-thaw cycles before processing the entire batch. Be aware that during ultrafiltration (centrifuge-driven filter devices with adequate MWCO) a local over-concentration and irreversible precipitation

or aggregation of the protein on the filtration membrane can take place [34]. Small aliquots should be frozen in liquid nitrogen and then stored at –80 °C to avoid damaging freeze-thaw cycles. Moreover, aliquots should always be thawed on ice [34].

20. Alternative screening evaluation: low-scale IMAC purification is performed without *additives*. The eluted protein is diluted 1:4 in buffers with different *additives* (concentrate protein with disposable ultrafiltration devices if the eluate is not concentrated enough). Then proceed to **step 3**, Subheading 3.4.

21. Experimental stresses: *OptiSol™ Protein Solubility Screening Kit Application Manual* http://www.dilyx.com/protein_solubility_screen_home2.

Acknowledgment

References

1. Chi EY, Krishnan S, Randolph TW et al (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm Res 20:1325–1336

2. Berkowitz SA, Engen JR, Mazzeo JR et al (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. Nat Rev Drug Discov 11:527–540

3. Philo JS, Arakawa T (2009) Mechanisms of protein aggregation. Curr Pharm Biotechnol 10:348–351

4. Narhi LO, Schmit J, Bechtold-Peters K et al (2012) Classification of protein aggregates. J Pharm Sci 101:493–498

5. Cromwell ME, Hilario E, Jacobson F (2006) Protein aggregation and bioprocessing. AAPS J 8:E572–E579

6. Lee HJ, Mcauley A, Schilke KF et al (2011) Molecular origins of surfactant-mediated stabilization of protein drugs. Adv Drug Deliv Rev 63:1160–1171

7. Leibly DJ, Nguyen TN, Kao LT et al (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. PLoS One 7:e52482

8. Ohtake S, Kita Y, Arakawa T (2011) Interactions of formulation excipients with proteins in solution and in the dried state. Adv Drug Deliv Rev 63:1053–1073

9. Bandyopadhyay A, Saxena K, Kasturia N et al (2012) Chemical chaperones assist intracellular folding to buffer mutational variations. Nat Chem Biol 8:238–245

10. Singh LR, Dar TA, Rahman S et al (2009) Glycine betaine may have opposite effects on protein stability at high and low ph values. Biochim Biophys Acta 1794:929–935

11. Street TO, Bolen DW, Rose GD (2006) A molecular mechanism for osmolyte-induced protein stability. Proc Natl Acad Sci U S A 103:13997–14002

12. Bondos SE, Bicknell A (2003) Detection and prevention of protein aggregation before, during, and after purification. Anal Biochem 316:223–231

13. Churion KA, Bondos SE (2012) Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification. Methods Mol Biol 896:415–427

14. Arakawa T, Ejima D, Tsumoto K et al (2007) Suppression of protein interactions by arginine: a

proposed mechanism of the arginine effects. Biophys Chem 127:1–8

15. Hamada H, Arakawa T, Shiraki K (2009) Effect of additives on protein aggregation. Curr Pharm Biotechnol 10:400–407

16. Kissmann J, Joshi SB, Haynes JR et al (2011) H1n1 influenza virus-like particles: physical degradation pathways and identification of stabilizers. J Pharm Sci 100:634–645

17. Lebendiker M. Protein purification facility. http://wolfson.huji.ac.il/purification/Protocols/Additives_Folding.html

18. Arnau J, Lauritzen C, Petersen GE et al (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 48:1–13

19. Nallamsetty S, Waugh DS (2007) A generic protocol for the expression and purification of recombinant proteins in escherichia coli using a combinatorial his6-maltose binding protein fusion tag. Nat Protoc 2:383–391

20. Golovanov AP, Hautbergue GM, Wilson SA et al (2004) A simple method for improving protein solubility and long-term stability. J Am Chem Soc 126:8933–8939

21. Shukla D, Schneider CP, Trout BL (2011) Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. Adv Drug Deliv Rev 63:1074–1085

22. Sala E, De Marco A (2010) Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography. Protein Expr Purif 74:231–235

23. Timasheff SN (1998) Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. Adv Protein Chem 51:355–432

24. Lee LL, Lee JC (1987) Thermal stability of proteins in the presence of poly(ethylene glycols). Biochemistry 26:7813–7819

25. Roberts MJ, Bentley MD, Harris JM (2002) Chemistry for peptide and protein pegylation. Adv Drug Deliv Rev 54:459–476

26. Muraoka T, Adachi K, Ui M et al (2013) A structured monodisperse peg for the effective suppression of protein aggregation. Angew Chem Int Engl 52:2430–2434

27. Rudolph R, Fisher S (1990) Process for obtaining renatured proteins. US Patent 4,933,434

28. Umetsu M, Tsumoto K, Hara M et al (2003) How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system. Spectroscopic evidence for highly efficient refolding of a single-chain fv fragment. J Biol Chem 278:8979–8987

29. Willis MS, Hogan JK, Prabhakar P et al (2005) Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions. Protein Sci 14:1818–1826

30. Kabanov AV, Lemieux P, Vinogradov S et al (2002) Pluronic block copolymers: novel functional molecules for gene therapy. Adv Drug Deliv Rev 54:223–233

31. Nettleship JE, Brown J, Groves MR et al (2008) Methods for protein characterization by mass spectrometry, thermal shift (thermofluor) assay, and multiangle or static light scattering. Methods Mol Biol 426:299–318

32. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2:2212–2221

33. Lebendiker M, Danieli T (2011) Purification of proteins fused to maltose-binding protein. Methods Mol Biol 681:281–293

34. Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium et al (2008) Protein production and purification. Nat Methods 5:135–146

Chapter 15

Solubilization and Refolding of Inclusion Body Proteins

Anupam Singh, Vaibhav Upadhyay, and Amulya K. Panda

Abstract

High-level expression of recombinant proteins in *Escherichia coli* often results in accumulation of protein molecules into aggregates known as inclusion bodies (IBs). Isolation of properly folded, bioactive protein from IBs is a cumbersome task and most of the times results in poor recovery. The process of recovering bioactive proteins from IBs consists of solubilization of IB aggregates using denaturants, followed by refolding of the solubilized protein. Here, we describe a simple protocol for screening of buffers for solubilization of IB proteins. Various IB aggregate solubilization methods including organic solvents have been described.

Key words Inclusion bodies, Solubilization, Pulsatile renaturation, Aggregation, Protein refolding

1 Introduction

During expression of recombinant protein in heterologous hosts, high concentration of partially folded intermediates often results in aggregation of protein into inclusion bodies (IBs). Apart from high concentration of partially folded protein molecules, reducing environment of bacterial cytoplasm, lack of chaperones, and posttranslational modifications also contribute toward IB formation [1, 2]. Protein aggregation leading to IB formation has been reported to be highly specific; thus, by optimal washing procedures, IBs having more than 90 % purity can be prepared [3]. The strategy to recover bioactive protein from IB involves four general steps: isolation and purification of IBs from *E. coli* cells, solubilization of the IB aggregates, refolding of solubilized IB protein into native conformation, and purification of the refolded protein employing various chromatographic techniques [4, 5]. Among these steps, solubilization of IBs and refolding of the solubilized protein are the most crucial steps and it is necessary to pay attention on them to finally get a high protein recovery. This may help in reducing the number of steps as well as requirement of tags for protein purification.

In general, proteins expressed as IBs are solubilized by the use of high concentration (6–8 M) of chaotropes like urea and guanidine hydrochloride (GdnHCl). A major issue concerning these conventional solubilization agents is that they completely denature the solubilized protein molecules, which often aggregate again during refolding step. Chaotropic agents such as urea and GdnHCl, in the presence of low concentration of detergents such as sodium dodecyl sulfate (SDS) [6], sodium deoxycholate, and sodium N-lauroyl sarcosine [7] along with reducing agents like β-mercaptoethanol and dithiothreitol, have been extensively used for solubilizing the IB proteins. During the last years, there has been a great amount of research aiming to develop new strategies for solubilization of IBs. The use of extreme pH with a combination of low concentration of denaturing agent has been used for the solubilization of IB proteins [8]. Solubilization of IBs by applying high hydrostatic pressure has also been reported [9]. Another novel solubilization method has been developed in our laboratory, which employs the use of organic solvents like *n*-propanol and β-mercaptoethanol with low concentration of urea [10, 11]. This method helps in high-throughput recovery of bioactive proteins from IBs.

Solubilized proteins, in general, are refolded into their native state by removal of chaotropic agents and other salts by dialysis [12] or dilution of the solubilized protein directly into the renaturation buffer. The biggest hurdle often faced while using these methods is the aggregation of protein molecules. Pulse dilution method, which involves the addition of small amounts of solubilized protein to the refolding buffer at successive time intervals, improves the overall performance of the refolding process [13]. Dropwise addition ensures low protein concentration and at the same time provides enough time for the protein molecules to refold properly. Moreover, once a protein molecule is properly folded, it does not interact with the unfolded or partially unfolded protein from the subsequently added drop. This method decreases protein aggregation during refolding and leads to high recovery of bioactive protein.

The choice of solubilization agent varies from protein to protein and no single, universal solubilization method works in every case. Here we describe a general protocol for purification of IBs, their solubilization employing different methods and refolding of the solubilized protein into bioactive form. In this context, the selection of a suitable solubilization process is crucial to get a high recovery of bioactive protein. The schematic of inclusion body solubilization with subsequent refolding process is described in Fig. 1. This chapter aims to give the readers a simple strategy to screen different solubilization buffers for IB aggregates and optimize a protocol best suited for their protein of interest.
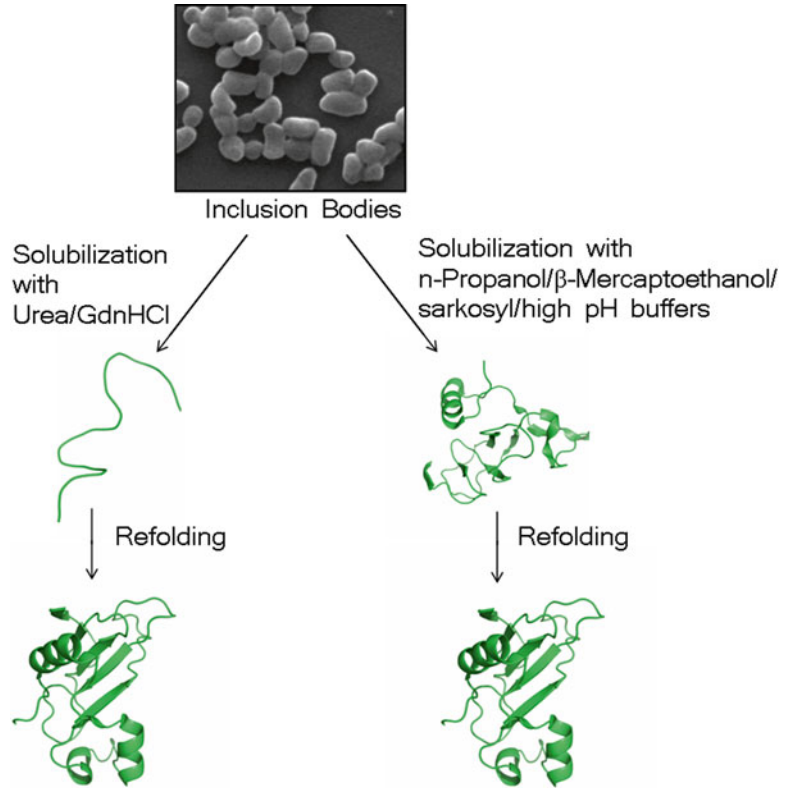
Fig. 1 Inclusion body refolding. IBs have been shown to have native-like secondary structures. Solubilization of IBs in high concentration of urea or GdnHCl usually results in complete loss of protein structure. On the other hand, mild solubilization agents (organic solvents, detergents, or high pH buffers) preserve some of the secondary and tertiary structures. In both the cases, refolding process can be optimized to improve the recovery of correctly folded native protein. Mild solubilization of IBs results in lower protein aggregation during refolding and thus helps in improving the recovery of bioactive protein from IBs

2 Materials

2.1 Cell Culture

1. Transformed *E. coli* cells carrying the desired construct with isopropyl-β-D-thio-galactopyranoside (IPTG) inducible promoter.

2. Modified Luria Bertani (LB) medium: 10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl, and 5 g glucose per L of MQ H_2O.

3. Antibiotics.

4. IPTG (filter-sterilized).

2.2 Isolation of Inclusion Bodies from E. coli Cells

1. Lysis buffer: 50 mM Tris–HCl, 5 mM ethylenediaminetetraacetic acid (EDTA), and 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.5.

2. Wash buffer A: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, and 0.5–2 % deoxycholic acid (DOC), pH 8.5.

3. Wash buffer B: 50 mM Tris–HCl, pH 8.5.

2.3 Solubilization of Inclusion Bodies

1. Solubilization buffer A: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, and 8 M urea, pH 8.5.

2. Solubilization buffer B: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, and 6 M GdnHCl, pH 8.5.

3. Solubilization buffer C: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, and 2 M urea, pH 12.

4. Solubilization buffer D: 50 mM Tris–HCl, 5 % glycerol, 0.1 mM EDTA, 50 mM NaCl, and 0.4 % sarkosyl, pH 7.9 [5].

5. Solubilization buffer E: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, 6 M *n*-propanol, and 2 M urea, pH 8.5.

6. Solubilization buffer F: 50 mM Tris–HCl, 5 mM EDTA, 1 mM PMSF, 6 M β-mercaptoethanol, and 2 M urea, pH 8.5.

2.4 Refolding of Solubilized Protein

1. Refolding buffer: 50 mM Tris–HCl, 1 mM EDTA, 10 % (w/v) sucrose, and 1 mM PMSF, pH 7.5.

3   Methods

3.1 Cell Culture

1. Inoculate *E. coli* cells carrying the desired construct in 10 mL of modified LB medium (in a 50 mL Falcon tube). Add suitable quantity of appropriate antibiotic for selection before inoculation.

2. Let the cells grow overnight (ON) in an orbital shaker set at 37 °C and 200 rpm.

3. Add appropriate quantity of antibiotic in 1 L of modified LB medium in shaker flasks. Inoculate medium with 10 mL of primary culture.

4. Transfer flasks in an orbital shaker set at 37 °C and 200 rpm.

5. Allow the cells to grow while monitoring OD_{600nm} at regular time intervals.

6. Induce cells with IPTG (final concentration 1 mM) when OD_{600nm} reaches 0.6–0.8.

7. Allow cells to grow for another 3 h in orbital shaker for expression of protein.

8. Transfer the culture to Oakridge tubes and centrifuge at $10,000 \times g$, for 20 min at 4 °C.

9. Discard the supernatant and store pellet at −20 °C till further use.

**3.2 Purification
of Inclusion Bodies**

Purification of IBs from *E. coli* cells is an important step in the recovery of bioactive proteins. It has been reported that the quality of IBs affects the solubilization process and controls the refolding yield of proteins [14]. Also, purified IBs can yield pure protein without multiple protein purification steps. General protocol for IB purification from *E. coli* cells is given below. Check notes for possible modifications in this protocol:

1. Resuspend cell pellet from 1 L culture in 40 mL of lysis buffer in a 50 mL beaker. Homogenize the suspension at 5,000 rpm using a homogenizer for 1 min on ice.

2. Sonicate cell suspension for 10 cycles of 1 min each (short pulses of 1 s followed by a gap of 1 s) with 1 min gap between the cycles. Operate the sonicator at 40 % amplitude. Use 13 mm probe for sonication. Maintain low temperature while sonication using ice bath (*see* **Note 1**).

3. Centrifuge the cell suspension at $20,000 \times g$ for 20 min at 4 °C. Discard supernatant.

4. Resuspend the pellet in 40 mL of wash buffer A in a 50 mL beaker. Homogenize the suspension at 5,000 rpm using a homogenizer for 1 min on ice.

5. Repeat **steps 2** and **3**.

6. Wash the resulting pellet by resuspending it in 25 mL of wash buffer B. Homogenize at 5,000 rpm using a homogenizer for 1 min on ice. Centrifuge the suspension at $20,000 \times g$ for 20 min at 4 °C. Discard supernatant (*see* **Note 2**).

7. Repeat **step 6**.

8. Resuspend the final pellet (purified IBs) in 2 mL MQ H_2O. Use fresh IBs for solubilization and refolding (*see* **Notes 3** and **4**).

**3.3 Solubilization
of Inclusion Body
Aggregates**

In this step, purified IBs are subjected to chemical treatment so as to solubilize the protein aggregates. There is no single solubilization method that works perfectly with every protein. The best strategy is to screen different solubilization agents and then choose one on the basis of solubilization efficiency and refolding yield. The choice also depends upon the nature of IBs. Some nonclassical IBs are so soft that a mild, non-denaturing solubilization agent such as low concentration of dimethyl sulfoxide (DMSO) or sodium N-lauroyl sarcosine results in efficient extraction of properly folded molecules [15, 16]. However, most of the IBs will not be solubilized in these agents making the use of high concentration of denaturants such as urea and GdnHCl inevitable. Protocol to screen a variety of solubilization agents is described below:

1. Determine the total protein concentration of IBs by solubilizing a 2 μL of IBs in 98 μL of 2 % SDS solution followed by protein quantification by bicinchoninic acid-based Micro BCA kit.

2. Suspend IBs (5 mg) in 1 mL of different solubilization buffers (A–F). Mix well by vortexing (*see* **Notes 5–8**).

3. Allow the mixture to incubate at room temperature for at least 1 h. Vortex the mixture 3–6 times during incubation.

4. Centrifuge solubilized protein sample at $20,000 \times g$ for 30 min at 4 °C.

5. Calculate the percentage solubilization by estimating protein quantity in supernatant by Micro BCA method.

3.4 Refolding of Solubilized Protein

Refolding of solubilized protein molecules into their native form is a crucial step. Protein aggregation during this step is a challenging problem. Refolding of solubilized protein depends upon the nature of protein, solubilization agent used, as well as the composition of refolding buffer (*see* **Notes 9–12**). Pulsatile renaturation protocol for refolding of solubilized proteins described below works efficiently for the majority of proteins (*see* **Notes 13** and **14**):

1. Take 9 mL of refolding buffer and cool it by keeping it in an ice bath under stirring conditions.

2. Add 1 mL solubilized protein to the refolding buffer either by using a peristaltic pump operating at 0.1–0.5 mL/min or by using a micropipette adding small amount of solubilized protein at regular intervals (protein is diluted 10 times during refolding).

3. Once all the solubilized protein is added, keep the refolded sample at 4 °C for 3–6 h.

4. Centrifuge the refolded protein sample at $20,000 \times g$ for 30 min at 4 °C to remove protein aggregates (if any). Alternatively, the aggregates can also be removed by filtration using 0.45 μM low protein-binding PVDF membrane filter.

5. Estimate the protein concentration of the refolded sample by Micro BCA method and homogeneity by SDS-PAGE.

Protein refolding should be followed by further purification of recombinant protein using appropriate chromatographic techniques. To estimate the quality of refolded protein, its tertiary and secondary structure should be determined by fluorescence and circular dichroism (CD) spectroscopy, respectively. Also, if the protein is an enzyme, it should be checked for the presence of activity. Sometimes during refolding process, protein molecules tend to aggregate into invisible, soluble aggregates which cannot be separated by centrifugation at low speeds. To determine the presence of such aggregates, analytical gel filtration chromatography should be used.

4 Notes

1. Other than sonication, French press can also be used for mechanical disruption of cells [17, 18]. Many protocols also use lysozyme in the lysis buffer to aid cell lysis.

2. Deoxycholate or other detergents are commonly employed during IB preparation to remove membrane fragments and membrane-bound proteins (mainly proteases) that get adsorbed to the hydrophobic surface of IBs. But this step may also result in the loss of the protein of interest. So, the concentration of the detergent should be optimized.

3. Homogeneous and pure IBs can also be obtained by density gradient ultra centrifugation. This method is very useful for obtaining highly pure IBs [17, 19].

4. Analyze the pellet and supernatant obtained after every wash by SDS-PAGE to determine the loss of protein during washes.

5. It is generally seen that IB preparation in a buffer of high pH results in clean and pure IBs. But in case of IBs of some proteins, high pH can result in the loss of protein of interest. In such cases, the pH of the buffer used for preparing IBs should be optimized.

6. The solubilization methods listed above don't contain a reducing agent. But, if the protein to be solubilized contains cysteines, a reducing agent like β-mercaptoethanol (10–20 mM) or dithiothreitol (0.3–1 mM) should be added to the solubilization buffer to keep sulfhydryl groups in reduced state.

7. Chaotropic agents, like urea and GdnHCl, are hygroscopic in nature. For accurate concentration measurements of these agents in solutions, refractometer should be used.

8. Urea preparations can contain impurities like cyanates formed due to degradation of urea. Cyanates can carbamylate proteins, mostly by reacting to the free amino groups and can alter protein's stability, function, and efficiency [20]. In order to decrease the buildup of cyanates, it is advisable to make fresh urea solutions and use them immediately after preparation.

9. For efficient solubilization, high concentration of denaturing agent is normally required. But certain IBs can be effectively solubilized in low concentrations of denaturants. As IBs are considered to be containing partially folded protein molecules, solubilization in low denaturant concentration would result in preservation of that partial structure, which can in turn lead to efficient refolding [14, 21]. Thus, before deciding the denaturant concentration, one should determine the lowest possible concentration of denaturant that can be used for solubilization.

10. If the protein has a pI above 7, acidic pH buffer (pH 4–5) should be used instead of high pH in solubilization buffer 3.

11. During refolding it is advisable to add certain additives (sugars, polyhydric alcohols, amino acids, chaotropes, kosmotropes) so as to decrease aggregation and/or to stabilize native fold. A variety of additives (cosolvents) can be screened for this purpose [22].

12. In cysteine containing proteins, there are chances of incorrect disulfide bond formation during refolding. Reduced and oxidized glutathione in different ratios (5:1 or 10:1) should be used to aid proper disulfide bond formation in such cases. A combination of oxidized and reduced glutathione helps in achieving disulfide shuffling, which allows the disulfide bonds to be oxidized and reduced repeatedly. Any incorrect disulfide bond formed during the refolding process will be reduced and the refolding process can proceed further toward achieving the proper fold [5].

13. The use of pulsatile refolding is the method of choice as it presents certain advantages over the other methods. The intermolecular protein interaction can be kept to a minimum by using this method. This method also uses lower volume of buffer for protein refolding. But, it is not necessary that pulsatile refolding always provide superior results over the other refolding processes. So, the other methods of refolding like flash dilution, reverse dilution, or step dialysis can also be tried.

14. Certain proteins require the presence of metal ions (cofactors) to be active. These cofactors are also known to be stabilizing the native fold of the protein. It is advisable to add appropriate metal ions in the refolding buffer as and when required.

Acknowledgment

References

1. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24:179–185

2. García-Fruitós E, Sabate R, de Groot NS et al (2011) Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J 278:2419–2427

3. Khan RH, Rao KB, Eshwari AN et al (1998) Solubilization of recombinant ovine growth hormone with retention of native-like secondary structure and its refolding from the inclusion bodies of *Escherichia coli*. Biotechnol Prog 14:722–728

4. Clark ED (2001) Protein refolding for industrial processes. Curr Opin Biotechnol 12:202–207

5. Burgess RR (2009) Refolding solubilized inclusion body proteins. Methods Enzymol 463:259–282

6. Stockel J, Doring K, Malotka J et al (1997) Pathway of detergent-mediated and peptide ligand-mediated refolding of heterodimeric class II major histocompatibility complex (MHC) molecules. Eur J Biochem 248:684–691

7. Burgess RR (1996) Purification of overproduced *Escherichia coli* RNA polymerase sigma factors by solubilizing inclusion bodies and refolding from Sarkosyl. Methods Enzymol 273:145–149

8. Panda AK (2003) Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*. Adv Biochem Eng Biotechnol 85:43–93

9. St John RJ, Carpenter JF, Balny C et al (2001) High pressure refolding of recombinant human growth hormone from insoluble aggregates. Structural transformations, kinetic barriers, and energetics. J Biol Chem 276:46856–46863

10. Panda AK, Singh SM, Upadhyay AK (2008) Process for obtaining bioactive recombinant protein from inclusion bodies. Patent no. US20100273234 (PCT/IN2008/000297)

11. Singh SM, Sharma A, Upadhyay AK et al (2012) Solubilization of inclusion body proteins using *n*-propanol and its refolding into bioactive form. Protein Expr Purif 81:75–82

12. Hutchinson MH, Morreale G, Middelberg AP et al (2006) Production of enzymatically active ketosteroid isomerase following insoluble expression in *Escherichia coli*. Biotechnol Bioeng 95:724–733

13. De Bernardez CE, Schwarz E, Rudolph R (1999) Inhibition of aggregation side reactions during *in vitro* protein folding. Methods Enzymol 309:217–236

14. Upadhyay AK, Murmu A, Singh A et al (2012) Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. PLoS One 7:e33951

15. Peternel S, Grdadolnik J, Gaberc-Porekar V et al (2008) Engineering inclusion bodies for non denaturing extraction of functional proteins. Microb Cell Fact 7:34

16. Jevsevar S, Gaberc-Porekar V, Fonda I et al (2005) Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnol Prog 21:632–639

17. Singh SM, Eshwari AN, Garg LC et al (2005) Isolation, solubilization, refolding, and chromatographic purification of human growth hormone from inclusion bodies of *Escherichia coli* cells: a case study. Methods Mol Biol 308:163–176

18. Georgiou G, Valax P (1999) Isolating inclusion bodies from bacteria. Methods Enzymol 309:48–58

19. Bowden GA, Paredes AM, Georgiou G (1991) Structure and morphology of protein inclusion bodies in *Escherichia coli*. Biotechnology (N Y) 9:725–730

20. Street TO, Courtemanche N, Barrick D (2008) Protein folding and stability using denaturants. Methods Cell Biol 84:295–325

21. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99:303–310

22. Bondos SE, Bicknell A (2003) Detection and prevention of protein aggregation before, during, and after purification. Anal Biochem 316:223–231

Chapter 16

Bacterial Inclusion Body Purification

Joaquin Seras-Franzoso, Spela Peternel, Olivia Cano-Garrido, Antonio Villaverde, and Elena García-Fruitós

Abstract

Purification of bacterial inclusion bodies (IBs) is gaining importance due to the raising of novel applications for this type of submicron particulate protein clusters, with potential uses in the biomedical field among others. Here, we present two optimized methods to purify IBs adapting classical procedures to the material nature as well as the requirements of its final application.

Key words Bacterial inclusion body, Purification, Nanoparticles, Bacterial cell free, Cell disruption

1 Introduction

Bacterial IBs have been regarded for many years as inert waste by-products of the recombinant protein production process and therefore either straightforward discarded or isolated in order to resolubilize and refold the aggregated protein. However, recent studies have shown high levels of molecular organization as well as significant extents of biologically active polypeptides within these protein nanoparticles [1, 2] (Fig. 1). This change in the perception of IB structure prompted the appearance of new applications, becoming bacterial IBs a final product itself with potential uses in industry, biomedicine, or diagnostics [3]. In this regard, bacterial IBs have shown their ability to act as naturally immobilized biocatalysts [4–7] or stimulate mammalian cell response in terms of adhesion, proliferation, or differentiation when these protein nanoparticles are used as topographical modifiers of cell culture interfaces [8–10].

So far, IB purification protocols consisted basically of mechanical, chemical, or enzymatic cell disruption methods followed by series of washing steps exhibiting high recovery yields (up to 95 %). However, disregarding removing other impurities such as viable bacteria or bacterial debris, that typically co-sediment with
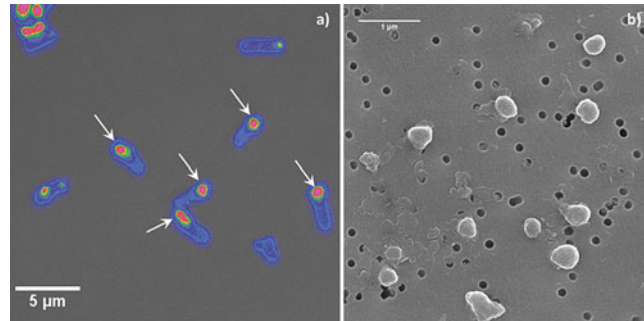
Fig. 1 Inclusion bodies. (**a**) Confocal image of *E. coli* cells producing mGFP IBs, pointed by *white arrows*. (**b**) Scanning Electron Microscopy micrograph of purified mGFP

insoluble bacterial products, make the use of IBs in biological interfaces incompatible [11]. Here, we present two general protocols focused on obtaining cell-free IBs ready to be used in mammalian cell cultures or as biocatalysts. One of the described procedures is a general protocol aiming to disrupt bacterial cells forming IBs without compromising the protein conformational quality of the nanoparticle, by means of an enzymatic attack directly performed in the bacterial cell culture, to increase the efficiency of the reaction. This treatment, in turn, is followed by series of freeze/thaw cycles. The combination of both enzymatic digestion and soft mechanical disruption allows also to completely remove viable bacteria from the samples, being the presence of bacterial contaminants monitored along the process by seeding small sample volumes on LB plates. This monitoring is crucial to tailor the number of freeze/thaw cycles required in each case to obtain bacterial cell-free biomaterial. Besides, extensive washing steps with mild detergents and further enzymatic treatments are performed to improve the purity of IBs. As a result of the whole procedure, active, bacterial cell-free submicron protein particles are obtained.

On the other hand, a second procedure aiming to specifically recover IBs composed from proteins that are prepared at low cultivating temperatures is detailed. Considering that these IBs are extremely soluble in mild detergents [2, 12–14], chemical cell disruption is not appropriate for the isolation of such nanoparticles. Nonetheless, mechanical disruption alone can also compromise the structure or the quality of protein trapped inside these IBs. Thus, the process for isolation of whole IBs with functional proteins produced at low cultivating temperatures, which merges together two mechanical processes, was designed. Firstly bacterial cells are exposed to several freeze/thaw cycles where soft mechanical forces facilitate bacterial cell walls to crack. This is then followed by homogenization, where high pressure enables total cell disruption. Extensive washing in buffers that does not compromise

IB structure facilitate IB purification. During this IB isolation process, viable bacteria are removed from the sample, and IBs can be used as whole nanoparticles as well as for further downstream protein isolation [2, 13].

2 Materials

Prepare all solutions and buffers using ultrapure H_2O at RT and store the stocks at the indicated temperature. Magnetic stirrers and beakers are required. All buffers are filtrated (0.22 μm) before storage.

2.1 IB Purification Protocol

1. EDTA-free Protease Inhibitor Cocktail Tablet.

2. 100 mM PMSF: Dissolve 0.0174 g phenylmethylsulfonyl in 1 mL 2-Propanol. Leave in 500 μL aliquots in Screw-Cap microcentrifuge tubes and store them at –20 °C (*see* **Note 1**).

3. 0.5 mg/mL lysozyme: Dissolve 0.5 mg lysozyme in 1 mL water. Store aliquots at –20 °C (*see* **Note 2**).

4. Triton X-100.

5. Nonidet P-40.

6. 1 M $MgSO_4$: Dissolve 123.24 g $MgSO_4$ in 100 mL water. Autoclave the solution and store at room temperature (RT).

7. 1 mg/mL DNAse I: Dissolve 1 mg in 1 mL water. Store aliquots at –20 °C.

8. Lysis buffer + 0.5 % Triton X-100: 20 mM Tris–HCl, 320 mM NaCl, 2 mM $MgCl_2$, 5 mM dithiothreitol (DTT), EDTA-free Protease Inhibitor Cocktail tablet, pH 7.5. Dissolve 0.1 g $MgCl_2$, 0.385 g DTT in 25 mL 1 M Tris–HCl, pH 8, and 6.25 mL 4 M NaCl. Add an EDTA-free Protease Inhibitor Cocktail Tablet dissolved in 50 mL water. Adjust volume until 500 mL and pH at 7.5. Add 0.5 mL Triton X-100. Finally, filter the solution and store it at RT.

9. PBS buffer 1×: Prepared from PBS buffer 10×: 25 mM $Na_2HPO_4 \cdot 2H_2O$, 1.5 M NaCl, 75 mM $NaH_2PO_4 \cdot H_2O$. Dissolve 13.35 g $Na_2HPO_4 \cdot 2H_2O$, 81.18 g NaCl, and 3.45 g $NaH_2PO_4 \cdot H_2O$ in 1 L water. Adjust at pH 7.4. Store at RT.

10. LB plates: Dissolve 10 g NaCl, 10 g tryptone, and 5 g yeast extract in 1 L water. Add 15 g agar at the bottle and autoclave it. Plate the solution. Store at 4 °C.

2.2 Purification Protocol of IBs Produced at Low Temperatures

1. PBS buffer 1×: Prepared from PBS buffer 10×: 25 mM $Na_2HPO_4 \cdot 2H_2O$, 1.5 M NaCl, 75 mM $NaH_2PO_4 \cdot H_2O$. Adjust at pH 7.4. Store at RT.

2. Buffer B50/30; 50 mM Tris–HCl buffer with 30 mM NaCl, prepared from 1 M Tris–HCl, pH 8.0. Adjust pH at RT with

the addition of concentrated (32 %) HCl. Add water to 1 L. For 50 mM Tris–HCl buffer with 30 mM NaCl, dissolve 50 mL of 1 M Tris–HCl, pH 8.0, and 1.75 g NaCl in 1 L water. Store at 4 °C.

3. LB plates: Dissolve 10 g NaCl, 10 g tryptone, and 5 g yeast extract in 1 L water. Add 15 g agar at the bottle and autoclave it. Plate the solution. Store at 4 °C.

2.3 IBs Quantification by Western Blot

1. Solution B: Dissolve 0.4 g sodium dodecyl sulfate (SDS) and 18.2 g Tris base in 100 mL water. Adjust at pH 8.8. Store at 4 °C.

2. 10 % ammonium persulfate: Dissolve 1 g ammonium sulfate in 10 mL water. Store 1 mL aliquots at –20 °C.

3. Solution C: Dissolve 0.4 g sodium dodecyl sulfate (SDS) and 6 g Tris base in 100 mL water. Adjust at pH 6.8. Store at 4 °C.

4. Denaturing buffer (Laemmli 4×): 1.28 g of Tris base, 8 mL of glycerol, 1.6 g of sodium dodecyl sulfate (SDS), 4 mL of β-mercaptoethanol, and 9.6 g of urea. Store at RT.

5. Electrophoresis buffer SDS-free (10×): Dissolve 144 g glycine and 30.3 g Tris-HCl in 1 L water. Store at RT.

6. 10 % SDS: Dissolve 50 g in 500 mL water. Store at RT.

7. Electrophoresis buffer: Dilute 100 mL electrophoresis buffer SDS-free (10×) and 10 mL 10 % SDS in water up to 1 L. Store at 4 °C (*see* **Note 3**).

8. Transference buffer: Dilute 100 mL electrophoresis buffer SDS-free (10×) and 200 mL methanol in water up to 1 L. Store at –20 °C (*see* **Note 4**).

9. Blocking solution: Dissolve 2 g skimmed milk in 40 mL PBS 1×.

10. 40 % Acrylamide/Bis Solution, 37. 5:1.

3 Methods

3.1 General IB Purification Protocol (Fig. 2)

Carry out all procedures in a laminar flow hood;

1. Add into the bacterial cell culture EDTA-free Protease Inhibitor Cocktail (1 tablet/500 mL culture) (*see* **Note 5**), PMSF (0.4 mM), and lysozyme (1 μg/mL) (*see* **Note 2**). Incubate 2 h at 250 rpm and 37 °C. That step can be performed in the same shake flask in which the bacterial cell culture has been performed.

2. Fill a beaker with the bacterial cell culture. Freeze at –80 °C O/N (*see* **Note 6**).

3. Thaw the culture at RT (*see* **Note 7**). Add Triton X-100 (0.4 mL/100 mL sample). Incubate 1 h under gentle agitation at RT (*see* **Notes 8** and **9**).
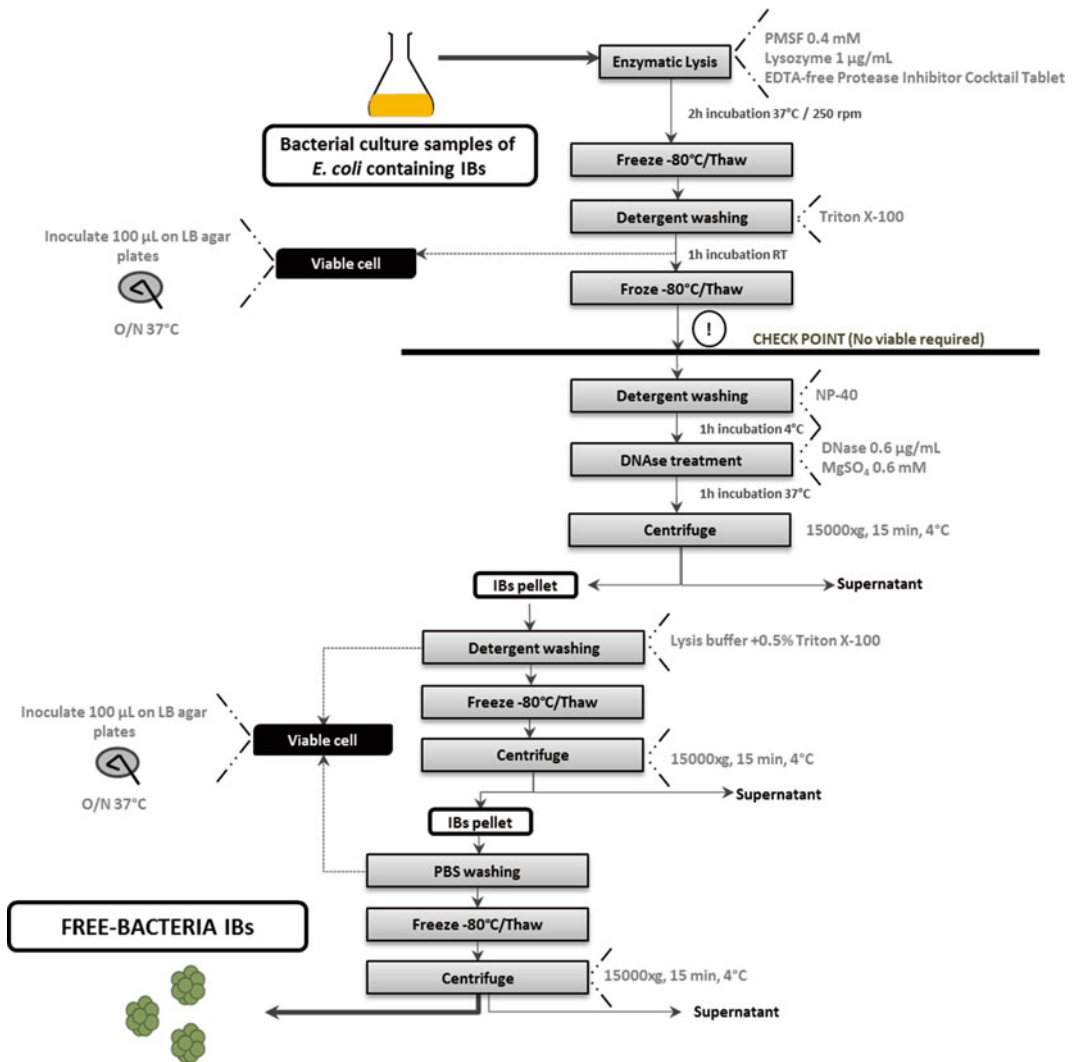
Fig. 2 IB purification protocol: enzymatic lysis, detergent washing treatment, and repeated freeze and thaw rounds

4. Test bacterial contamination by inoculating 100 μL on LB plates. Incubate the plates O/N at 37 °C. Freeze the IB sample at –80 °C O/N.

5. Thaw the IB sample at RT (*see* **Note 7**). Check the contamination level by counting bacterial colonies on the LB plates. Repeat the freeze/thaw process until obtaining no viable on LB plate.

6. Once obtained a free viable bacteria IB sample, add Nonidet P-40 (25 μL/100 mL sample) and incubate 1 h under gentle agitation at 4 °C (*see* **Notes 8** and **9**).

7. Add $MgSO_4$ (60 μL/100 mL sample) and DNAse I (60 μL/100 mL sample). Incubate 1 h under gentle agitation at 37 °C.

8. Harvest the IBs at $15,000 \times g$ for 15 min at 4 °C. Remove the supernatant and resuspend the pellet with lysis buffer + Triton X-100 (5 mL/100 mL initial sample).

9. Inoculate 100 μL on LB plates and leave O/N at 37 °C. Freeze the sample at –80 °C O/N.

10. Thaw the sample at RT (*see* **Note 7**). Harvest at $15,000 \times g$ for 15 min at 4 °C. Remove the supernatant and resuspend the pellet with PBS buffer (5 mL/100 mL initial sample).

11. Inoculate 100 μL on LB plates and leave O/N at 37 °C. Freeze the sample at –80 °C O/N.

12. Thaw the sample at RT. Harvest at $15,000 \times g$ for 15 min at 4 °C. Remove the supernatant and do aliquots. Store the aliquots at –80 °C.

13. In order to consider the pellets free of bacterial cells, plates from **steps 9** and **11** must be without any colony. Moreover, if IBs pellets will be used in mammalian cell cultures, bacterial contamination must be further tested in mammalian cell medium (**step 14**).

14. Resuspend an IB pellet (**step 12**) in 1 mL of mammalian cell medium. Add 200 μL of this solution, as well as its dilutions 1:10, 1:100, and 1:1,000, to a 96-well plate in triplicate. Leave at least 2–3 days at the mammalian cell culture conditions. If there is no contamination, the IB stock can be validated as sterile.

3.2 Purification Protocol of IBs Produced at Low Temperatures (Fig. 3)

The process can be performed in the single centrifuge tube (*see* **Note 10**):

1. After the protein production, harvest the bacterial cells at $5,000 \times g$ for 5 min at 4 °C. Remove the supernatant (medium) and wash the pellet twice in PBS or B50/30 buffer (*see* **Note 11**). Discharge supernatant. The pellet can be stored at –80 °C (*see* **Note 12**).

2. Resuspend the bacterial pellet in the selected buffer (PBS or B50/30); the volume of the suspension can be from ¼ to ½ of the original culture volume. Freeze the suspension for 2 h at –80 °C (*see* **Note 13**).

3. Thaw the culture on ice. This freeze/thaw cycle should be repeated at least three times.

4. Harvest the bacterial cells at $5,000 \times g$ for 5 min at 4 °C. Remove the supernatant.

5. Resuspend the cells in the selected buffer (PBS or B50/30) and keep the suspension on ice (*see* **Note 10**).

6. Bacterial cells are disrupted in high-pressure homogenizer (e.g., EmulsiFlex-C5, Avestin) at operating pressure 75–100 MPa (*see* **Note 14**).
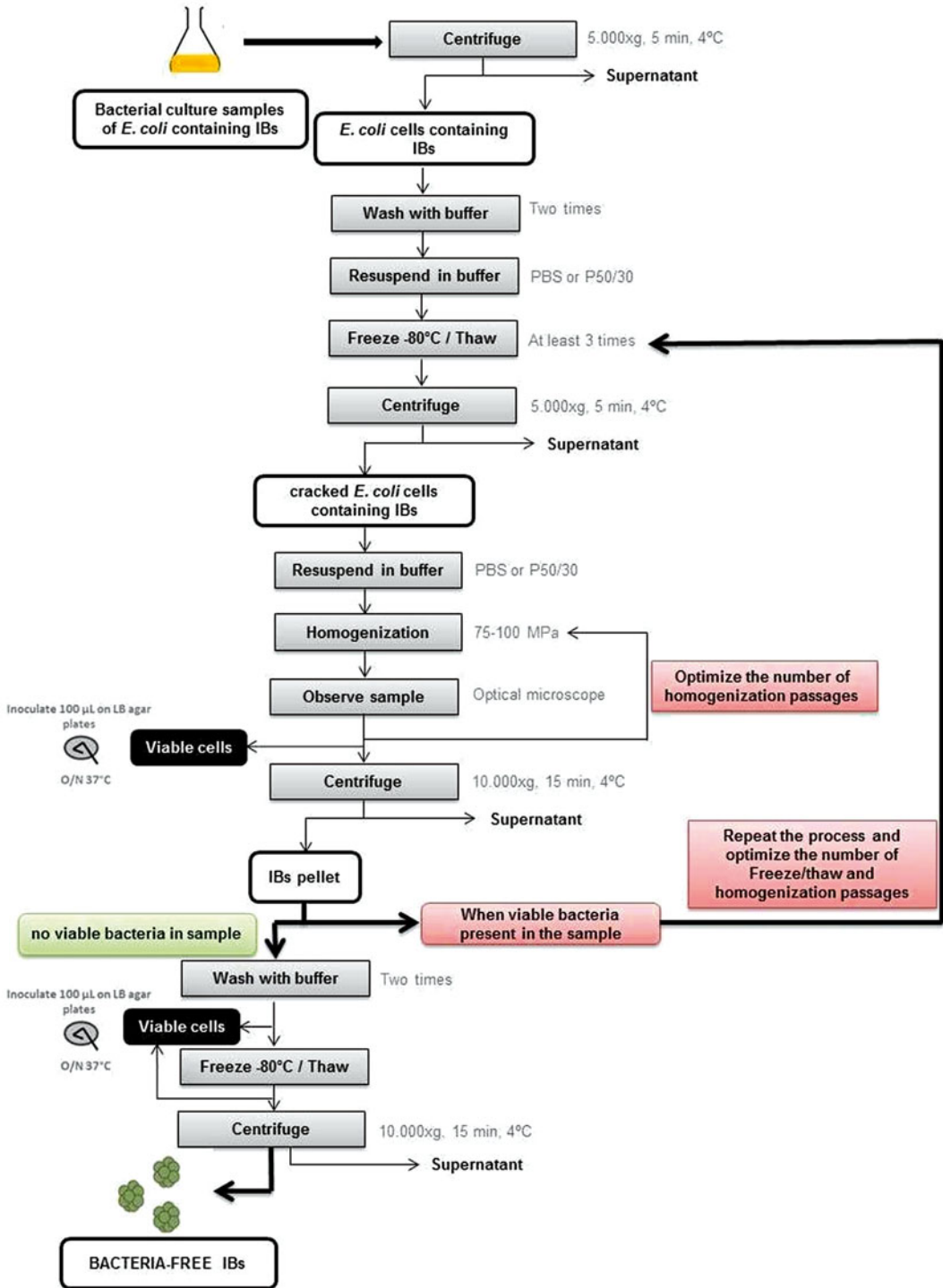
Fig. 3 IB purification protocol at low temperatures: homogenization step and freeze and thaw steps

7. Observe the sample under the optical microscope to assess the effectiveness of cell disruption. Depending on the host strain and the nature of the produced protein, together with the efficiency of previous freeze/thaw cycles, the number of passages has to be optimized (*see* **Note 15**).

8. Test bacterial contamination by inoculating 100 μL on LB plates. Incubate the plates O/N at 37 °C.

9. Harvest IBs at $10,000 \times g$ for 5 min at 4 °C. Discharge supernatant.

10. Freeze the IB sample at −80 °C O/N.

11. Check the contamination level by counting bacterial colonies on the LB plates. Repeat the freeze/thaw process in combination of homogenization until no viable on LB plate is obtained .

12. Once obtained a viable bacteria-free IB sample, wash the IB pellet twice with the selected buffer. Thoroughly resuspend the IB pellet in the buffer with gentle agitation at RT (10 min) for better washing efficacy.

13. Test bacterial contamination by inoculating 100 μL on LB plates. Incubate the plates O/N at 37 °C.

14. Harvest at $10,000 \times g$ for 15 min at 4 °C. Remove the supernatant and resuspend the pellet with PBS or P50/30 buffer (5 mL/100 mL initial sample).

15. Aliquotate the pellet. Harvest the IBs at $10,000 \times g$ for 15 min at 4 °C. Remove the supernatant.

16. Store the aliquots at −80 °C.

17. In order to consider the pellets free of bacterial cells, plates from **steps 11** and **13** must be without any colony. Moreover, if IB pellets will be used in mammalian cell cultures, bacterial contamination must be further tested in mammalian cell medium.

18. Resuspend an IB pellet (**step 12**) in 1 mL of mammalian cell medium. Add 200 μL of this solution, as well as its dilutions 1:10, 1:100, and 1:1,000, to a 96-well plate in triplicate (Fig. 4). Leave at least 2–3 days at the mammalian cell culture conditions. If there is no contamination, the IB stock can be validated as sterile.

3.3 IB Quantification by Western Blot

3.3.1 Acrylamide Gels Preparation (1 Gel)

1. First of all prepare the running gel (10 % acrylamide) (*see* **Note 16**). Mix 4.93 mL MQ H$_2$O, 2.5 mL solution B, 2.5 mL acrylamide, 60 μL ammonium persulfate (APS) 10 %, and 8 μL TEMED. Immediately, pour 7 mL solution between both glass plates. To make the top of the separating gel be horizontal, fill in 2-Propanol into the gap until the side.

2. Once gelated, remove the 2-Propanol by decantation. Dry remnants of 2-Propanol with filter paper.
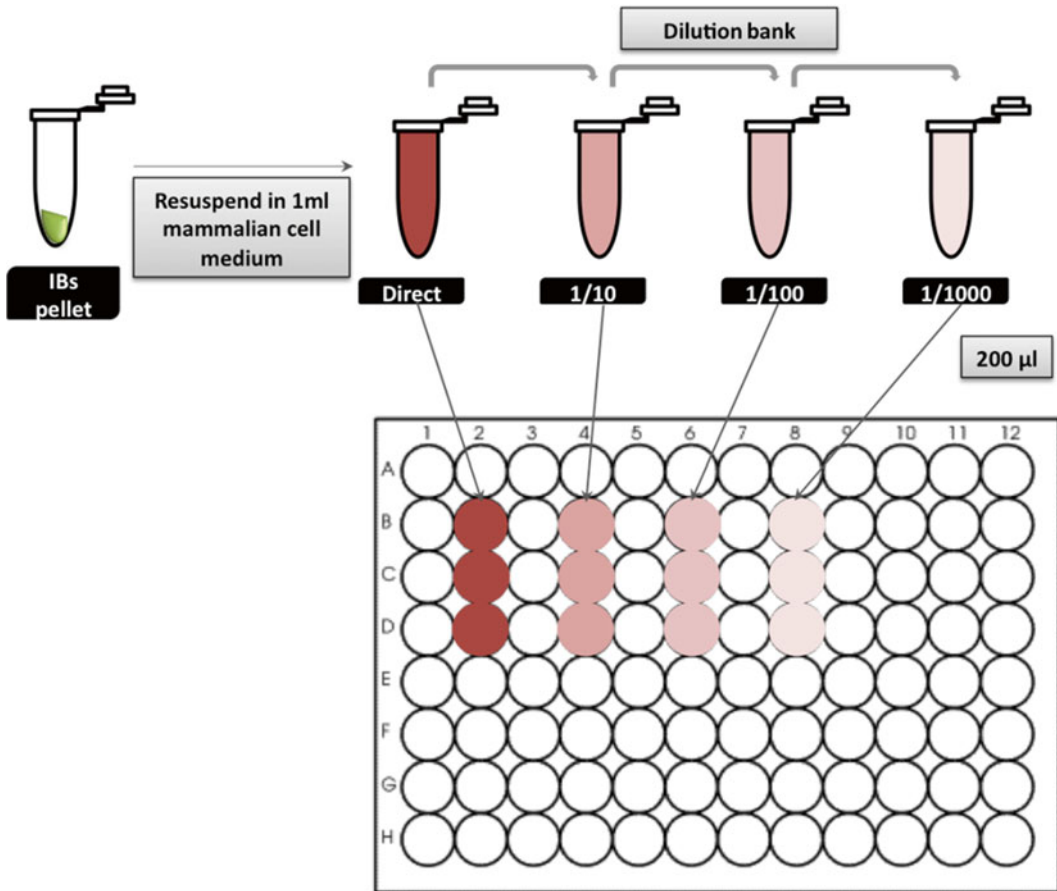
Fig. 4 IB sterility protocol performed in mammalian cell medium

3. Prepare the stacking gel (3.5 % acrylamide). Mix 4.124 mL MQ H_2O, 1.575 mL solution C, 0.7 mL acrylamide, 70 μL 10 % ammonium persulfate (APS), and 7 μL TEMED. Immediately add the stacking gel solution until the side and insert the well-forming comb.

3.3.2 Sample Preparation and Blotting

1. Resuspend IBs in denaturing buffer (Laemmli 4×) at appropriate ratios to obtain a 1× denaturing buffer concentration and boil the samples for 45 min (*see* **Note 17**). Concurrently, prepare a calibration curve using diluted series of protein samples of known concentration (*see* **Note 18**).

2. Place the acrylamide gels in the tank and fill in the cold phoresis buffer (*see* **Note 3**).

3. Take out the comb and load the sample onto, the volume of sample to load will depend on the well size. Load also a protein marker.
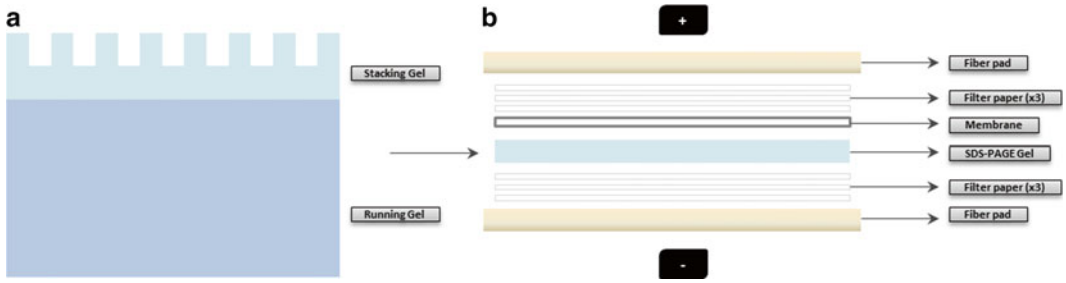
Fig. 5 SDS-PAGE analysis. (**a**) Acrylamide gel sketch: stacking gel at the top (*light blue*) and running gel at the bottom (*dark blue*). (**b**) Western blot transference preparation of the cassette (Color figure online)
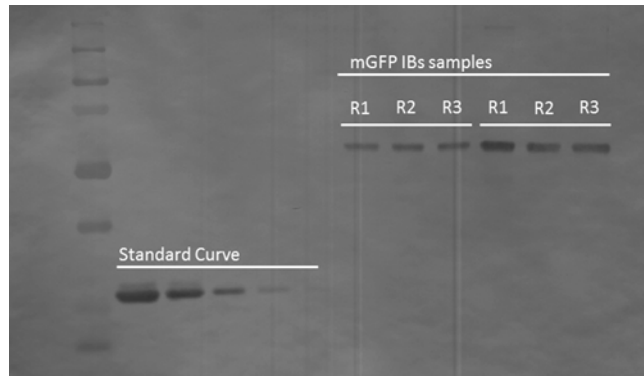
Fig. 6 Western Blot Quantification of mGFP IBs. Soluble GFP was used as standard curve. Lanes from 1 to 5 correspond to 500, 250, 125, 75, and 37.5 ng of protein, respectively. Bands from lanes 6 to 9 correspond to 10 μL of IB sample in triplicate. Bands from 10 to 12 correspond to 20 μL of IB sample in triplicate

4. Set an appropriate voltage and run the electrophoresis (*see* **Note 19**).

5. Once the electrophoresis is finished, prepare the western blot transference. Cut the nitrocellulose membrane and filter papers (six per gel). Equilibrate gels and membranes in transfer buffer.

6. Place the membrane and the gel in the cassette forming a sandwich as it is represented in the image (Fig. 5).

7. Place the cassette in the transfer cell filled with transfer buffer and the cooler.

8. Run the transference for 1 h at 100 V.

3.3.3 Immunodetection (Fig. 6)

1. Incubate the membrane O/N in blocking solution.

2. Dilute the primary antibody in the blocking solution at the suitable concentration. Incubate the membrane for 2 h at RT.

3. Wash the membrane with PBS 1× and twice with PBS-Tween for 15 min.

4. Dilute the secondary antibody in PBS 1× at the suitable concentration. Incubate the membrane for 1 h at RT.

5. Wash the membrane with PBS 1× and twice with PBS-Tween for 15 min.

6. Finally perform the detection with chosen method (colorimetry, chemiluminescence, etc.).

4   Notes

1. PMSF aliquots stored at –20 °C tend to precipitate, but once at RT PMSF becomes soluble again. Make sure that PMSF is completely solubilized.

2. A high lysozyme concentration not always helps to improve *E. coli* lysis, since at high concentrations this enzyme tends to aggregate and remain a contaminant present at the final IB purification step.

3. It should be prepared before the sample preparation and stored at 4 °C to be cold enough during the protein electrophoresis.

4. It should be prepared before running the SDS-PAGE and stored at –20 °C to be cold enough during the protein transference.

5. Add the suitable volume of EDTA-free Protease Inhibitor Cocktail Tablet suspended in sterile H_2O.

6. Make sure the beaker can be frozen at –80 °C.

7. So as to thaw faster the sample, it can be left in warm water.

8. Gentle agitation can be performed with a magnetic stirrer.

9. Warm up at 37 °C Triton X-100 and NP40 detergents prior to use in order to reduce their viscosity and facilitate their pipetting.

10. Keep the samples on cold (ice or a suitable substitute) through the isolation process.

11. The buffer should be chosen regarding the stability of the protein inside IBs.

12. Make sure the centrifugation tube can be frozen at –80 °C.

13. When proteins inside IBs are sensitive to proteolysis, the EDTA-free Protease Inhibitor Cocktail Tablet (1 tablet/500 mL) can be added to the buffer.

14. During homogenization the sample chamber as well as the sample-collecting tube should be kept on cold (ice or a suitable substitute) through the isolation process when the homogenizer is not cooled or located in the cold room.

15. The number of freeze/thaw cycles and number of homogenization passages has to be optimized for every single case, while the level of stress for the host organism varies depending on the protein produced, on expression vectors used, as well as on the different genetic backgrounds of different *E. coli* strains. Therefore, even with the same bacterial strain and very similar proteins produced in similar vectors, the number of homogenization passages needed to totally disrupt bacterial cell was found to be different [13].

16. Depending on the protein weight, a different acrylamide percentage should be used.

17. Aggregated protein need longer boiling times than soluble protein.

18. A quantified protein which can be detected with the same antibodies as the sample is necessary. A soluble protein can be used as calibration curve.

19. Generally 60 V for the stacking gel and then change to 100 V for the running gel.

Acknowledgment

References

1. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27

2. Jevsevar S, Gaberc-Porekar V, Fonda I et al (2005) Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnol Prog 21:632–639

3. Garcia-Fruitos E, Vazquez E, Diez-Gil C et al (2012) Bacterial inclusion bodies: making gold from waste. Trends Biotechnol 30:65–70

4. Nahalka J, Nidetzky B (2007) Fusion to a pull-down domain: a novel approach of producing

Trigonopsis variabilis D-amino acid oxidase as insoluble enzyme aggregates. Biotechnol Bioeng 97:454–461

5. Nahalka J, Gemeiner P, Bucko M et al (2006) Bioenergy beads: a tool for regeneration of ATP/NTP in biocatalytic synthesis. Artif Cells Blood Substit Immobil Biotechnol 34:515–521

6. Nahalka J (2008) Physiological aggregation of maltodextrin phosphorylase from Pyrococcus furiosus and its application in a process of batch starch degradation to alpha-D-glucose-1-phosphate. J Ind Microbiol Biotechnol 35:219–223

7. Garcia-Fruitos E, Villaverde A (2010) Friendly production of bacterial inclusion bodies. Korean J Chem Eng 27:385–389

8. Seras-Franzoso J, Diez-Gil C, Vazquez E et al (2012) Bioadhesiveness and efficient mechanotransduction stimuli synergistically provided by bacterial inclusion bodies as scaffolds for tissue engineering. Nanomedicine (Lond) 7:79–93

9. Garcia-Fruitos E, Seras-Franzoso J, Vazquez E et al (2010) Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. Nanotechnology 21: 205101

10. Diez-Gil C, Krabbenborg S, Garcia-Fruitos E et al (2010) The nanoscale properties of bacterial inclusion bodies and their effect on mammalian cell proliferation. Biomaterials 31: 5805–5812

11. Rodriguez-Carmona E, Cano-Garrido O, Seras-Franzoso J et al (2010) Isolation of cell-free bacterial inclusion bodies. Microb Cell Fact 9:71

12. Peternel S, Jevsevar S, Bele M et al (2008) New properties of inclusion bodies with implications for biotechnology. Biotechnol Appl Biochem 49:239–246

13. Peternel S, Grdadolnik J, Gaberc-Porekar V et al (2008) Engineering inclusion bodies for non denaturing extraction of functional proteins. Microb Cell Fact 7:34

14. Peternel S, Komel R (2010) Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. Microb Cell Fact 9:66

Chapter 17

Characterization of Intracellular Aggresomes by Fluorescent Microscopy

Lianwu Fu and Elizabeth Sztul

Abstract

Correct folding of newly synthesized proteins is essential to cellular homeostasis and cells have evolved sophisticated means to fold and modify proteins. When misfolding occurs, the misfolded proteins often expose normally buried hydrophobic domains, causing localized aggregation. Individual small aggregates appear to be transported towards the microtubule-organizing center and there coalesce to form larger aggregates called aggresomes. Both cytoplasmic and nuclear proteins can form aggresomes. The study of aggresomes has progressed rapidly because numerous human diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, various myopathies, and prion disease are characterized by the formation of aggresomes. Importantly, aggresomes sequester many cellular proteins and the pathology of aggresomal disease is at least partially caused by the deregulation of cellular components. Thus, it is essential to identify and characterize the composition of aggresomes formed by different proteins. However, most protein aggregates are insoluble even in buffers with high concentration of detergent, which makes them very difficult to analyze by biochemical approaches. An alternative approach that has been used successfully is the in situ characterization of protein components within aggresomes by immunofluorescent microscopy. Here, we provide detailed protocols to study the characteristic features of aggresomes by fluorescent microscopy.

Key words Inclusion bodies, Aggresomes, Protein aggregates, Immunofluorescent microscopy, Protein misfolding, Protein degradation, Molecular chaperones, Intermediate filament, Proteasome, Autophagy

1 Introduction

Newly synthesized proteins must be folded and modified properly to function correctly. The cells employ precise machineries that facilitate correct folding. Nevertheless, misfolding can occur due to mutations within a protein, outside stresses, or overexpression of proteins. About 1/3 of newly synthesized peptides is estimated to be incorrectly folded and is degraded [1]. Misfolded proteins often expose their hydrophobic domains, which leads to nonproductive protein association and aggregation. Aggregated proteins tend to coalesce and form large deposits termed inclusion bodies,

Russell bodies, or aggresomes. Various human diseases have been linked to the formation of protein aggresomes. These so-called protein aggregation diseases include amyloid-β and Tau inclusions in Alzheimer's disease, α-synuclein inclusions (Lewy bodies) in Parkinson's disease, polyglutamine-containing protein aggregates in Huntington's disease, TDP-43 inclusions in amyotrophic lateral sclerosis, skeletal muscle fibers in patients with myopathies, aggregates of mutated α1-antitrypsin in patients with α1-antitrypsin deficiency, Mallory bodies in patients with alcoholic steatohepatitis, and PrP protein aggregates in prion disease. Because of their relevance to a large variety of diseases, the analysis and characterization of specific protein aggresomes will provide obvious diagnostic and therapeutic implications.

The biochemical processes of protein aggregation have been actively investigated (for review: [2–4]). Aggregation of proteins most likely occurs cotranslationally, while nascent peptide chains are synthesized on a polyribosome and folded by protein chaperones. If the newly synthesized peptides cannot be folded on time, they will expose the hydrophobic β-sheet structures that promote oligomerization of the protein to form small aggresomal particles. These aggresomal particles are quickly transported towards the microtubule (MT)-organizing center (MTOC), where they coalesce to form large protein aggresomes [5, 6]. Immuno-histochemical studies have demonstrated that cytoplasmic protein aggresomes are often enriched in molecular chaperones including Hsc70, Hdj1, and Hdj2. Increasing evidence suggests that the aggresomes may sequester the cellular components involved in ubiquitin proteasomal degradation and autophagy [7, 8]. This indicates that the formation of protein aggresomes may play a role to attenuate the quality control systems in the cells. Another characteristic feature of the cytosolic aggresomes is that they are often enclosed within a cage formed by vimentin or other filament proteins [6, 9]. The function and consequence of this vimentin cage remain unclear. Aggresomes appear to be formed by the misfolded proteins that escaped from the proteasomal degradation but can be eventually cleared from the cells by autophagy [10]. The removal of aggresomes by autophagy requires p62 that functions as a bridge connecting ubiquitinated protein aggregates and autophagosomes and HDAC6 (the tubulin deacetylase histone deacetylase 6) that mediates the fusion of autophagosomes with lysosomes [11].

In addition to cytoplasmic aggresomes formed around the MTOC, nuclear inclusions are often found in polyglutamine diseases such as Huntington's disease (HD) or spinocerebellar ataxia (SCA) [12]. HD and SCA are neurodegenerative diseases caused by expanded polyglutamine repeats in huntingtin and ataxins, respectively. The mutant proteins aggregate to form both cytoplasmic and intranuclear inclusions. The formation of nuclear

inclusions depends on the length of the polyglutamine repeats and factors in the host cells. However, our studies using a green fluorescent protein (GFP)-tagged fusion protein (GFP170*) suggest that even proteins without polyglutamine tracts can form aggregates in the nucleus [9, 13, 14]. Others also demonstrated that some viral proteins without polyglutamines such as a viral AP-1 homolog also deposit as protein aggresomes in the nucleus [15].

Studies from cell culture experiments and transgenic mice show that the nuclear inclusions also contain a subset of molecular chaperones, ubiquitin, and proteasomal subunits [16]. These findings suggest that the nuclear inclusions might be analogous to cytoplasmic aggresomes, and that the nucleus might also contain specific sites to compartmentalize misfolded proteins. Little is known about how the nuclear aggresomes are formed except that they often recruit nuclear factors such as PML (promyelocytic leukemia) protein and transcriptional factors such as p53.

Most of the detection methods for protein–protein interaction require the proteins are soluble in solutions. This is true for protein electrophoresis, chromatographic separation, mass spectrometry, and other diffraction-based methods. However, most of the proteins associated with aggresomes are insoluble even in solutions with high concentrations of detergent, which makes the biochemical analysis of protein aggresomes difficult [17]. Other methods such as electron microscopy (EM) and immuno-EM were used to visualize the morphology of the protein aggregates but require sophisticated chemical staining and expensive equipment. Fluorescent microscopy provides a relatively easy and rapid method to visualize and characterize the formation and composition of intracellular aggresomes. In most experiments the aggresome-forming proteins are tagged with fluorescent proteins such as GFP and this provides a convenient means of monitoring aggresome formation. In addition, protein components present within the aggresomes can be detected by labeling with specific primary and secondary antibodies with different fluorochrome conjugates. By monitoring the localization of the aggregating proteins and various cellular components by immunofluorescent microscopy, the localization and the compositions of the aggregates can be analyzed.

2 Materials

2.1 Cell Culture and Introduction of Aggresome-Forming Plasmids

1. General materials for cell culture: COS7 cells were chosen as an example here to host the cellular aggresomes for their convenience of transfection and appearance under the microscope.

   (a) Cell culture media (DMEM medium), fetal bovine serum (FBS), 0.25 % trypsin, 10,000 I.U./mL penicillin, and 10,000 μg/mL streptomycin.

(b) Cell culture plates/flasks, sterile pipets and tips, adjustable pipettors.

(c) Humidified CO_2 incubators.

2. Cover glasses: Circular microscope cover glasses with 12 mm diameter and ethanol bath were used for flame sterilization (*see* **Note 1**).

3. Plasmid constructs that form protein aggresomes: We have previously described two model proteins, GFP250 and GFP170*, that when expressed in cells form cytosolic and nuclear aggresomes, respectively. The GFP250* construct contains the GFP fused at its COOH terminus to the NH_2-terminal first 252 amino acids of p115, a protein transport factor essential for endoplasmic reticulum (ER) to Golgi traffic [6]. The GFP170* consists of GFP fused to an internal fragment (amino acids 566–1,375) of the Golgi complex protein 170 (GCP170) [9]. Both GFP250 and GFP170* form protein aggregates 24–48 h after standard transfection without proteasome inhibition.

2.2 Fixation and Permeabilization for Immunofluorescent Staining

1. Phosphate-buffered saline (PBS) solution: Dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 in 800 mL of distilled H_2O. Adjust the pH to 7.4 with HCl. Add H_2O to 1 L. Dispense the solution into 50 mL aliquots by filter sterilization. If necessary, PBS can be supplemented with 1 mM $MgCl_2$ and 0.1 mM $CaCl_2$.

2. Fixation solution: 3 % paraformaldehyde dissolved in PBS or 100 % methanol. Add 3 g paraformaldehyde to 100 mL PBS. Filter before use (*see* **Note 2**).

3. Quenching solution: 10 mM NH_4Cl. Dissolve 53.49 mg NH_4Cl in 100 mL PBS. Filter before use.

4. Permeabilization solution: 0.1 % (V/V) Triton X-100 diluted in PBS. Pipette 100 μL Triton X-100 and mix with 100 mL PBS. Filter before use (*see* **Note 3**).

2.3 Immunofluorescent Staining with Primary and Secondary Antibodies

1. Blocking solutions: 2.5 % goat serum in PBS, 0.2 % Tween 20 (PBST). Add 0.25 mL of 100 % serum to 10 mL of PBST, mix by vortexing, filter, and keep in –20 °C freezer before use. 0.4 % fish-skin gelatin diluted in PBST. Add 40 μL of 100 % fish-skin gelatin to 10 mL of PBST and mix. The blocking solutions were filtered and stored at –20 °C freezer (*see* **Note 3**).

2. Primary antibodies: Primary antibodies from different manufacture companies were used in 1:1,000 to 1:100 dilutions with PBST containing 0.4 % fish-skin gelatin (*see* **Note 4**).

3. Secondary antibodies: Goat anti-mouse or anti-rabbit IgG (H + L) conjugated with Alexa Fluor® was used in 1:200 dilution with PBST containing 2.5 % goat serum.

4. Hoechst 33258 for nuclear staining: Measure 10 mg Hoechst 33258 and dissolve in 1 mL of H_2O to make 10 mg/mL stock. Use the stock solution in 1:10,000 dilutions.

2.4   Immuno-fluorescent Microscopy

1. Mounting solution: 1 mg/mL *para*-phenyl diamine in PBS containing 75 % glycerol. Adjust the pH of PBS to 8.5 using 5 M NaOH. Add 10 mg *para*-phenyl diamine to 2.5 mL PBS, pH 8.5 and 7.5 mL glycerol. Mix thoroughly by vortexing. Filter and dispense the solution to 100 μL aliquots. Store the solution in –80 °C freezer. Keep the stock in dark at room temperature.

2. Fluorescent microscopy: Any fluorescent microscope with sufficient objective and filters can be used. We used a Leitz epifluorescence microscope equipped with a step motor, filter-wheel assembly (Ludl Electronics Products), and an 83,000-filter set (Chroma Technology). Images were obtained with a SenSys-Cooled, charge-coupled high-resolution camera (Photometrics). IpLab Spectrum software (Signal Analytics) was used for image acquisition and analysis.

3   Methods

All the experimental procedures were conducted at room temperature (RT, about 25 °C) unless otherwise specified.

3.1   Preparation of Cover Slips

1. Pick up the number of cover slips glasses (12 mm) needed with a fine-tip tweezers and put them into a Petri dish. Add 100 % ethanol to soak the cover slips for 10 min.

2. Inside the tissue culture hood, carefully pick up the cover glass one at a time and dry it above gas flame. Put the dried cover glass onto a 35mm tissue culture plate and spread them evenly in the plate (*see* **Note 5**).

3. Add cell culture medium to the plate containing the cover slips. Press the cover slips to the bottom of the culture dish (*see* **Note 6**).

3.2   Cell Culture and Transfection

1. Subdivide proper amount of COS7 cells to the culture dish containing the cover slips to make sure that the cell density reaches about 40–50 % confluence (*see* **Note 7**). Grow cells overnight (ON). Check density to have an optimal density of ~60–70 %.

2. Next day, set up transfection reactions: Add 250 μL serum-free RPMI medium to two 1.5 mL sterilized microcentrifuge tubes.

In one tube, add 1 μg plasmid expressing either the GFP250 or GFP170*. In the other one, add 3 μL of TransIT® L1 transfection reagent. Incubate at RT for 5 min. Mix the plasmid with the transfection reagent and incubate for another 25 min.

3. Remove medium from cells growing in a dish on cover slips. Add 1.5 mL of DMEM medium to the dish containing the cells and add the 500 μL DNA mixture dropwise to the dish. Incubate the cells in the CO_2 incubator at 37 °C ON. Next morning remove the transfection mixture, add 2 mL of growth medium, and incubate for 24–48 h at 37 °C in an incubator (*see* **Note 8**).

3.3 Fixation

1. Wash the cells in the 35 mm dishes three times with ice cold PBS.

2. Add 2 mL of 3 % paraformaldehyde solution or cold 100 % methanol to fix the cells for 10 min.

3. Quench with 10 mM NH_4Cl in PBS for 10 min. Wash the cells three times with PBS (*see* **Note 9**).

3.4 Permeabilization

1. Pick up each individual cover glass from the 35 mm dish with a fine-tip tweezers and put them in a homemade humidified chamber (*see* **Note 10** and Fig. 1).

2. Add 100 μL of 0.1 % Triton X-100 in PBS on top of each cover glass and incubate for 7 min (*see* **Note 11**).

3. Wash the cover glasses with 200 μL PBS three times for 2 min each (s*ee* **Note 12**).

3.5 Antibody Incubation

1. Add 200 μL blocking solution (2.5 % goat serum in PBST) on the top of each cover glass and incubate for 5 min.

2. Switch to 200 μL of blocking solution (0.4 % fish-skin gelatin in PBST) and incubate for another 5 min.
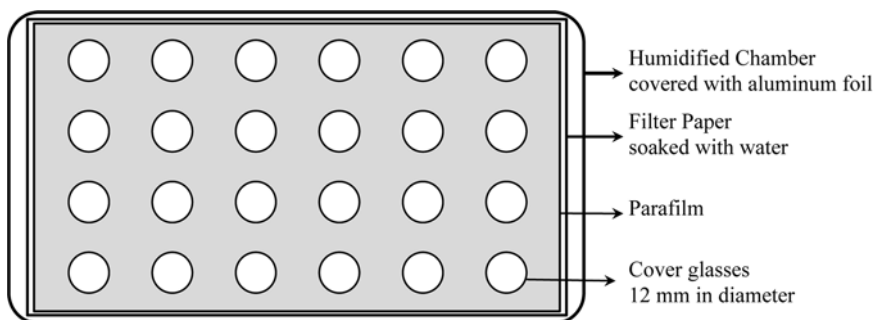
Fig. 1 Assembly of humidified chamber for incubation with antibodies. Take covers from two 6-well plates and wrap each with aluminum foil, making sure to mold it to the cover. Tape the covers together along the long side to make a hinged chamber. Cut thick (3 mm) filter paper to the size of the chamber, place it in the bottom of the chamber, and soak the filter paper with water. Cut a piece of parafilm to the size of the chamber and place it on the top of the wetted filter paper. Put the cover glasses on top top the parafilm The parafilm keeps the solutions on the cover slips from spreading out

3. During the blocking, dilute the primary antibodies in 0.4 % fish-skin gelatin blocking solution in a 1.5 mL microcentrifuge tube. Mix by vortexing and spin for 10 min in microcentrifuge at $15,000 \times g$ to clear possible precipitates. Add 200 μL of diluted antibody on top of cover glass and incubate for 60 min (see **Note 13**).

4. In the end of the incubation, remove the primary antibody by aspiration. Wash the cells with 200 μL PBST five times for 5 min each.

5. Block with 200 μL of 0.4 % fish skin gelatin in PBST for 5 min and then block again with 200 μL of 2.5 % goat serum in PBST for 5 min.

6. During the blocking, dilute the secondary antibodies in 2.5 % goat serum blocking solution in a microcentrifuge tube. Spin for 10 min in microcentrifuge at $15,000 \times g$. Add 200 μL of diluted antibody on the top of cover glass and incubate for 45 min (see **Note 13**).

7. Wash the cells with 200 μL PBST three times for 5 min each.

8. Dilute Hoechst 33258 from a 10 mg/mL stock 10,000 times and add 200 μL solution onto the cover glass to stain the nuclei for 5 min if it is desired.

9. Wash the cells with 200 μL of PBS two times for 5 min each.

3.6 Mounting

1. Clean some microscope slides (size: $25 \times 75 \times 1$ mm) with Kimwipes tissue and add 3 μL of mounting solution for each cover slip (see **Note 14**).

2. Aspirate the washing solution from the cover glass and carefully put them on the top of the mounting solution cells side down. Draw off any excess mounting solution with filter paper (see **Note 15**).

3. Seal the edge of the cover glass with nail polish and let them air-dry for 30 min. Examine immediately or keep the slides in a slide-holder in 4 °C refrigerator before examination under microscope (see **Note 16**).

3.7 Fluorescent Microscopy

1. Take out the slides to RT 30 min before imaging.

2. Examine the cells under fluorescent microscope and take pictures using IpLab Spectrum Software. Merge the pictures taken in different wavelength and save them as .tif files.

3. Analyze the colocalization of different proteins with the aggresomal markers using IpLab software. The colocalization indicates that the protein might play a role in the formation of the cellular aggregates or in the clearance of the aggregated proteins. Figure 2 shows some examples of cellular proteins colocalized with the cytosolic GFP-250 aggregates. Figure 3 shows that nuclear GFP-170* aggregates often localize adjacent to PML bodies and are enriched in the p53 oncoprotein.
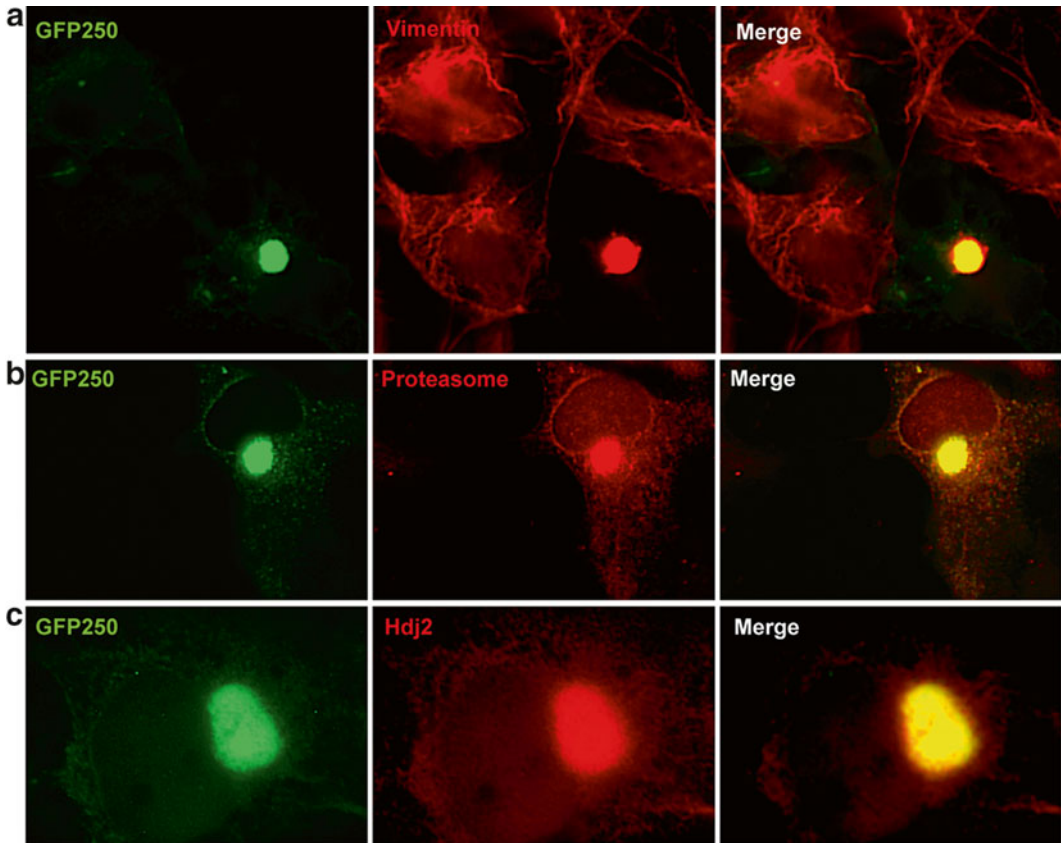
Fig. 2 Recruitment of cellular proteins to cytosolic aggresomes. The construct encoding the aggregation-prone GFP250 protein was transfected into COS7 cells for 48 h. The cells were then fixed, permeabilized, and incubated with primary antibodies as follows: (**a**) mouse monoclonal antibody against vimentin (Sigma-Aldrich, Cat. #V6630); (**b**) rabbit polyclonal antibody against the α-subunit of the 20S proteasome (EMD Millipore Chemicals, Cat. #539145); or (**c**) rabbit polyclonal antibody against the co-chaperone Hdj2 (Abcam, Cat. #AB107066). The cells were subsequently incubated with secondary antibodies: goat anti-mouse or anti-rabbit conjugated with Alexa Fluor 594 (Invitrogen). The aggresomes formed by GFP250 (*green*) are found only within the cytoplasm and also contain vimentin (*red*), components of the proteasomal degradative machinery (*red* 20S), and components of the refolding chaperone machinery (*red* Hdj2)

4   Notes

1. The cover slips can be sterilized by autoclave in a petri dish. If primary neuronal cells are used, the cover glasses need to be coated with 0.1 % poly-L-lysine.

2. The choice of fixation depends on the primary antibodies. Some antibodies give high background staining after paraformaldehyde fixation, and vice versa when 100 % methanol fixation is used. Chill the methanol in –20 °C freezer ON before use.
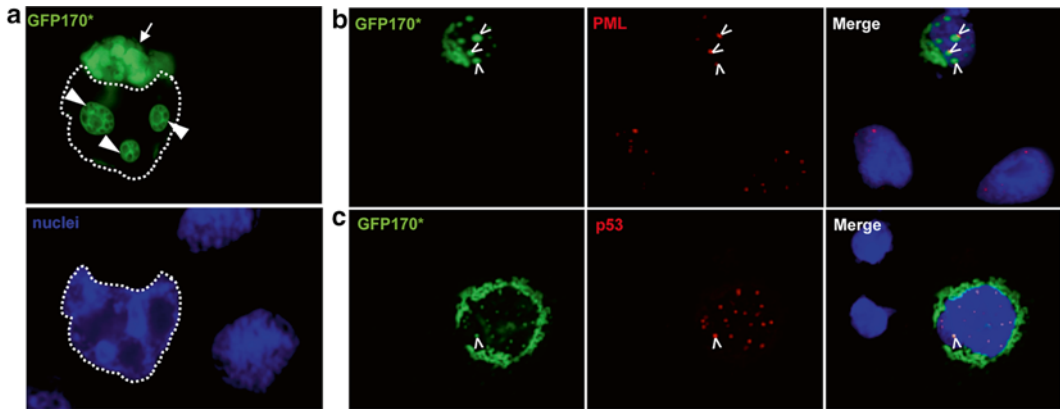
Fig. 3 Recruitment of cellular proteins to nuclear aggresomes. The construct encoding the aggregation-prone GFP170* protein was transfected into COS7 cells for 48 h. The cells were then fixed and permeabilized and the nuclei were stained with Hoechst (**a**) or incubated with primary antibodies as follows: (**b**) mouse monoclonal antibody against PML (Santa Cruz Biotechnology, Cat. #sc-966) or (**c**) mouse monoclonal antibody against p53 (Santa Cruz Biotechnology, Cat. #sc-98). The cells in (**b**) and (**c**) were subsequently incubated with goat anti-mouse secondary antibodies conjugated with Alexa Fluor 594 (Invitrogen, Cat. #A11032) and the nuclei were stained with Hoechst 33258. (**a**) The aggresomes formed by GFP170* (*green*) are found both within the cytoplasm (*arrow*) and within the nucleus (*arrowheads*). (**b**) The nuclear aggresomes are always adjacent to the PML bodies (*open arrowheads*). (**c**) The formation of the GFP170*aggresomes causes the accumulation of the p53 oncogene in association with the aggresomes (*open arrowheads*)

3. Some solutions such as Triton X-100, Tween-20, and fish-skin gelatin are very sticky. Cut the tip to make the opening wider and pipette slowly and mix with PBS by rotating for 30 min.

4. The optimal concentration of primary antibody needs to be determined by titration to achieve a maximal signal-to-noise ratio. A typical start point for antibody dilution is 1:200.

5. The number of cover slips in each culture plate is determined by the number of antigens to be tested along with the aggresomal marker. One 35 mm dish can hold up to five cover slips (12 mm in diameter). If more antigens need to be tested, use bigger dish and more cover slips to culture the cells.

6. The dried cover slips tend to float on the surface of the medium. Press them to the bottom of the culture plate and make sure that they are not overlapping with each other before adding the cells.

7. To make sure that the cell confluence reaches about 40–50 % before transfection, dilute the cells 1:4- to 1:5-fold from a confluent culture and then seed the plate containing the cover slips.

8. The efficacy of transfection needs to be tested by titration of the amount of plasmids and transfection reagent according to the manufacturer's recommendation.

9. No quenching is needed when methanol is used for fixation. If the procedure needs to be paused temporarily, the cells can be stored in PBS at 4 °C after fixation for up to a couple of days.

10. Add 200 μL PBS to each cover slip after placing it in the humidified chamber to avoid the cells being dried out. Be careful not to let the cells dry at any time during the process.

11. Methanol can make the cells permeable to the antibodies. This step can be skipped if methanol was used for fixation.

12. The 2-min incubation between washes with PBS is necessary to eliminate the possible effect of Triton X-100 on the following steps.

13. Dilute the amount of primary and secondary antibodies right before the usage. Spin the diluted antibody to get rid of any possible precipitates. When adding the antibodies to the cover glasses, try to avoid the precipitates in the bottom in the microcentrifuge tube because they will show as high background on the slides.

14. One standard microscope slide (size: $25 \times 75 \times 1$ mm) can hold 3–4 cover slips of 12 mm in diameter. When placing the cover slips onto the microscope slides, be careful not to trap air bubbles between the slides and cover slips. Hold the cover slip on one side, put the other side down against the drop of mounting solution, and slowly place the cover slip in position without moving it around or dropping it.

15. Draw out any excess amount of mounting solution around the edge of each cover slip using filter paper cut to small pieces. Be careful not to remove the mounting solution between the slide and the cover slip. Otherwise, air bubbles will be generated between them, which will affect the microscopic examination of the slides.

16. After sealing with finger nail polish, put the slides in a holder and store in the dark to avoid exposure to light and photodamage. The nail polish will dry in about 30 min to 1 h at RT. Store the slides in the 4 °C refrigerator or –20 °C freezer till ready for microscopy.

Acknowledgment

References

1. Schubert U, Anton LC, Gibbs J et al (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404:770–774

2. Chen B, Retzlaff M, Roos T et al (2011) Cellular strategies of protein quality control. Cold Spring Harb Perspect Biol 3(8):a004374. doi:10.1101/cshperspect.a004374, cshperspect.a004374 [pii]

3. Takalo M, Salminen A, Soininen H et al (2013) Protein aggregation and degradation mechanisms in neurodegenerative diseases. Am J Neurodegener Dis 2:1–14

4. Garcia-Mata R, Gao YS, Sztul E (2002) Hassles with taking out the garbage: aggravating aggresomes. Traffic 3:388–396

5. Johnston JA, Ward CL, Kopito RR (1998) Aggresomes: a cellular response to misfolded proteins. J Cell Biol 143:1883–1898

6. Garcia-Mata R, Bebok Z, Sorscher EJ et al (1999) Characterization and dynamics of aggresome formation by a cytosolic GFP-chimera. J Cell Biol 146:1239–1254

7. Keller JN, Hanni KB, Markesbery WR (2000) Impaired proteasome function in Alzheimer's disease. J Neurochem 75:436–439

8. Bence NF, Sampat RM, Kopito RR (2001) Impairment of the ubiquitin-proteasome system by protein aggregation. Science 292:1552–1555

9. Fu L, Gao YS, Tousson A et al (2005) Nuclear aggresomes form by fusion of PML-associated aggregates. Mol Biol Cell 16:4905–4917

10. Fu L, Sztul E (2009) ER-associated complexes (ERACs) containing aggregated cystic fibrosis transmembrane conductance regulator (CFTR) are degraded by autophagy. Eur J Cell Biol 88:215–226

11. Richter-Landsberg C, Leyk J (2013) Inclusion body formation, macroautophagy, and the role of HDAC6 in neurodegeneration. Acta Neuropathol 126:793–807

12. Perez MK, Paulson HL, Pendse SJ et al (1998) Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. J Cell Biol 143:1457–1470

13. Fu L, Gao YS, Sztul E (2005) Transcriptional repression and cell death induced by nuclear aggregates of non-polyglutamine protein. Neurobiol Dis 20:656–665

14. Tower C, Fu L, Gill R et al (2010) Human cytomegalovirus UL97 kinase prevents the deposition of mutant protein aggregates in cellular models of Huntington's disease and ataxia. Neurobiol Dis 41:11–22

15. Park R, Wang'ondu R, Heston L et al (2011) Efficient induction of nuclear aggresomes by specific single missense mutations in the DNA-binding domain of a viral AP-1 homolog. J Biol Chem 286:9748–9762

16. Cummings CJ, Mancini MA, Antalffy B et al (1998) Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. Nat Genet 19:148–154

17. Bagriantsev SN, Kushnirov VV, Liebman SW (2006) Analysis of amyloid aggregates using agarose gel electrophoresis. Methods Enzymol 412:33–48

Part IX

Insoluble Protein Characterization

Chapter 18

Dialysis: A Characterization Method of Aggregation Tendency

Mireia Pesarrodona, Ugutz Unzueta, and Esther Vázquez

Abstract

All researchers immersed in the world of recombinant protein production are in agreement that often the production and purification process of a protein can become a nightmare due to an unexpected behavior of the protein at different protocol stages. Once the protein is purified, scientists know that they still cannot relax. There is a decisive last step missing: performing a protein dialysis in a suitable buffer for subsequent experimental trials. Here is when we can find proteins that precipitate during dialysis by buffer-related factors (ionic strength, pH, etc.), which are intrinsic to each protein and are difficult to predict. How can we find the buffer in which a protein is more stable and with less tendency to precipitate? In this chapter we go over possible factors affecting the protein precipitation tendency during the dialysis process and describe a general dialysis protocol with tricks to reduce protein aggregation. Furthermore, we propose a fast method to detect the most appropriate buffer for the stability of a particular protein, performing microdialysis on a battery of different buffers to measure afterwards precipitation by a colorimetric method, and thus being able to choose the most suitable buffer for the dialysis of a given protein.

Key words Protein aggregation, Dialysis, Microdialysis, Protein precipitation, Protein stability

1 Introduction

Dialysis is a common laboratory process in which small solute molecules diffuse from a highly concentrated to a lower concentrated solution, until the equilibrium is reached, through a semipermeable membrane of defined pore size that selectively allows smaller molecules to pass while retaining larger species [1]. It usually corresponds to the last step of protein purification process where the composition of protein containing elution buffer is replaced by a more suitable one, regarding protein stability and experimental applicability [2]. This phase generally represents a critical step where proteins can precipitate if the appropriate buffer is not used, which concerning to proteins is very difficult to predict.

1.1   Types of Dialysis   Different types of dialysis systems have been developed since the first dialysis procedures were performed at the end of the nineteenth century [3, 4]. All of them can generally be classified into two major categories including "static" and "dynamic" dialysis.

1.1.1   Static Dialysis   Static dialysis corresponds to classical dialysis systems in which the diffusion process is performed in a batch where the sample containing semipermeable membrane is suspended in a fixed volume of the dialysate (buffer) for several hours until the equilibrium is reached. Different dialysis devices have been developed within this category, each of them offering different advantages and disadvantages [5].

Membrane dialysis is one of the most simple and widely known systems, in which the sample is loaded into a semipermeable membrane sac closed by threads (first-generation devices) (Fig. 1a) or clamps (second-generation devices) (Fig. 1b) on each edge. Although they are simple and cheap devices, sample loading and removal require considerable dexterity and the whole procedure results in a quite slow process.

Dialysis cassettes (third generation of membrane devices) are prefabricated plastic devices with already fitted membranes that make the sample loading and removing process much easier (Fig. 1c). However, the process is quite expensive and time consuming and additional accessories such as syringes and needles are required. Moreover, the cassettes do not show volume range flexibility and different size cassettes have to be used for different sample volume ranges.
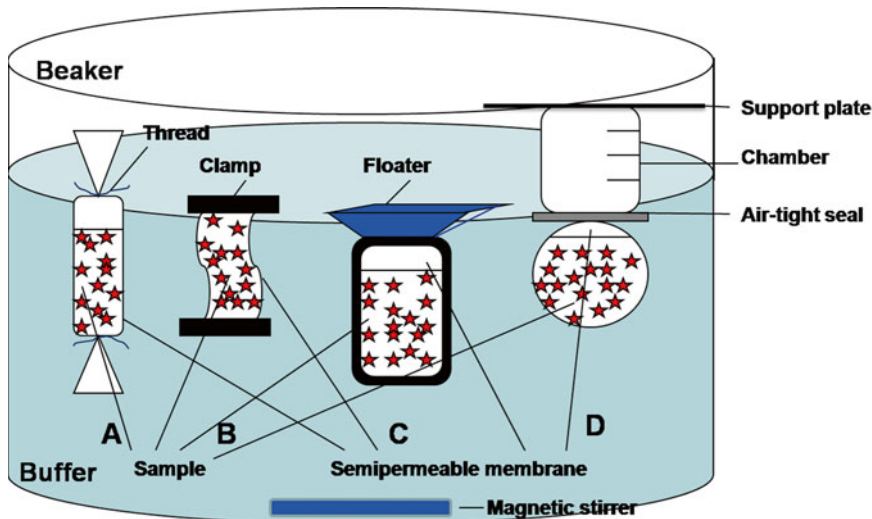
Fig. 1 Four generations of static membrane dialysis devices. A (first generation), B (second generation), C (third generation), and D (fourth generation). More information above (*see* Subheading 1.1.1). Modified from [5]

Bandal's device or "small wonder lyzer" (fourth generation of membrane devices) is a recently invented dialysis device that does not employ any accessory and combines the easy sample loading and removing capacity of third-generation devices and the sample volume versatility of the first- and second-generation devices. It is a non-floating bubble shape membrane which dips into the dialyzing buffer while all other attachments remain above the buffer (Fig. 1d). Moreover, it is the only system that allows the online monitoring of all the sample volume changes without sample removal from device.

Microdialysis is a method based on the general aspects and characteristics of common dialysis but being indicated for small-volume samples (<80 μL) in which the sample is deposited onto a membrane floating on a petri dish filled with the selection buffer. Microdialysis has been proposed as a rapid and simple screening method of protein solubility and conformational quality under different experimental conditions and as a routine method to identify suitable buffers for new and non-characterized proteins [6].

1.1.2   Dynamic Dialysis   In dynamic dialysis systems, the buffer or dialysate is usually circulated to create higher concentration gradient and thus, significantly reduce the dialysis time. The maximum dialysis efficiency is achieved in the countercurrent system in which both the sample and the buffer are circulated on different sides of the membrane in opposite directions creating the largest possible concentration gradient between them. This system results to be very appropriate when the sample is very delicate and a very fast dialysis process is required. Moreover, since it is a very fast and efficient system, a lot of time can be saved when very-large-volume samples have to be dialyzed [7].

Tubular membrane devices have been developed for this type of approaches among others. The sample is introduced in a fixed volume tubular shape membrane and the dialysate flow is circulated through the membrane surrounding chamber using a simple peristaltic pump (Fig. 2a). Hollow fiber membrane devices have been developed for countercurrent dynamic dialysis applications in which the sample is circulated through the inside part of semipermeable hollow fiber membranes while the dialysate is circulated through the fibers surrounding chamber in the opposite direction (Fig. 2b) [7].

1.2   Buffer Characteristics Affecting Protein Stability   One of the most challenging tasks within the whole protein purification process is the selection of the most appropriate protein storage buffer which has to assure both chemical and physical protein stability. Protein aggregation, which is the most common protein physical instability process, can be induced by many different factors. Thus, careful buffer formulation is required to avoid this kind of events to occur.
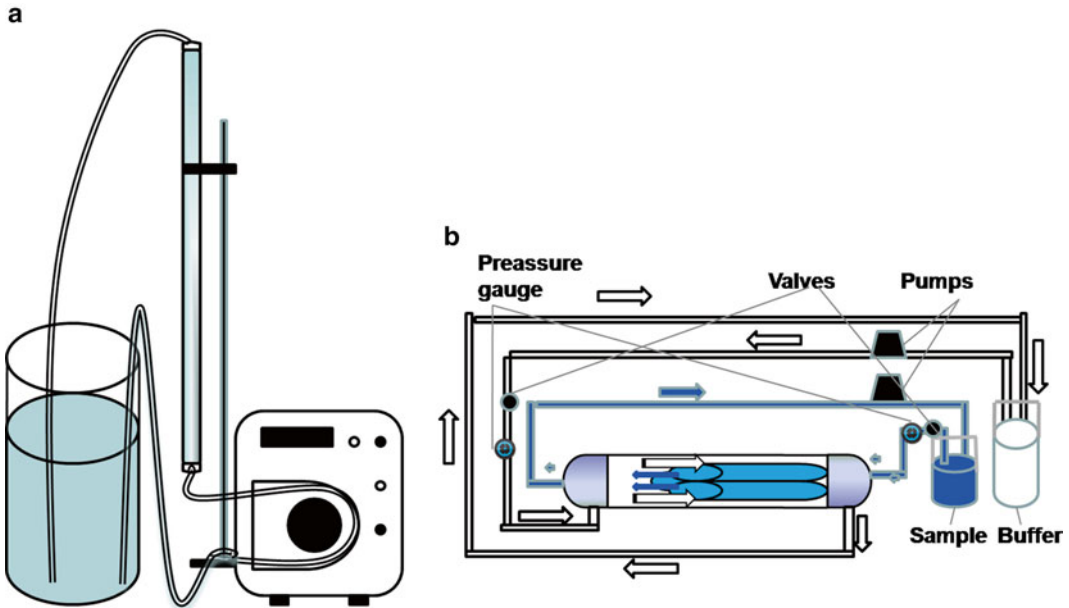
Fig. 2 Dynamic dialysis. (**a**) Tubular membrane device. (**b**) Hollow fiber membrane device. Adapted from [7]

Main buffer characteristics affecting protein stability include temperature, pH, ionic strength, and buffer composition among which many protein-stabilizing excipients have been described such as sugars, polyols, or surfactants. Unfortunately there is no a simple way to predict which is the perfect composition for each protein's stability, so usually proteins have to be empirically evaluated on a case-by-case basis.

1.2.1 Temperature

Temperature is an important parameter in protein stability. Usually proteins are stable in a certain range of temperature and higher temperatures generally decrease their stability [8].

1.2.2 pH

Proteins also are generally stable only at a narrow pH range and as the pH becomes closer to p*I* the protein's net charge tends to neutralize being zero at pH=p*I*. At that pH, electrostatic repulsion between individual proteins disappears and electrostatic attractions may occur resulting in a precipitate generation [1]. At extreme pHs far away from p*I*, the electrostatic repulsion within the protein is so high that it can show a tendency to unfold [8–10].

1.2.3 Ionic Strength

The effect of salts on protein stability is difficult to predict. They can stabilize and destabilize or have no effect depending on the nature of ionic interactions within the protein, their charged residues, and used salt types [8, 11]. At low salt concentration range, electrostatic shielding weakens ionic attractions or repulsions positively or negatively affecting protein stability depending on the

protein properties. However, at high salt concentrations, since the electrostatic shielding is already saturated, the effect of salt changes solvent properties, affecting water-protein interface surface tension and consequently altering protein's hydrophobic interactions [8].

1.2.4 Buffer Composition   No general rules have been reported for the selection of the appropriate buffer species. They are important since they not only stabilize the protein but also help maintaining the solution's pH in the adequate range for the protein. Many different buffers have been widely used for protein storage such as Tris–HCl, PBS, sodium phosphate, sodium citrate, sodium borate, sodium acetate, or sodium carbonate among others [8].

Sugars and polyols such as sucrose, dextrose, trehalose, and glycerol are commonly used excipients since their stabilizing capacity by preferential exclusion mechanism has been widely demonstrated, their effect being concentration dependent [8, 12, 13].

Surfactants decrease protein solution surface tension and reduce the forces driving protein aggregation at hydrophobic surfaces. Non-ionic surfactants such as Tween 20 or Triton X-100 are the most used ones since very low concentrations of those compounds are enough to reduce protein aggregation. Ionic surfactants such as SDS are usually avoided since they can bind both to polar and non-polar groups of proteins and cause their denaturation [8, 14].

2   Materials

Prepare all solutions using distilled H_2O.

2.1 Common Components
1. Dialysis buffer: 1,000 times sample volume.
2. Appropriate beaker for buffer volume.
3. Magnetic stirrer device.
4. Magnetic stir bar.
5. Dialysis device.

2.2 Membrane Dialysis Components
1. Dialysis membrane: big enough to contain sample volume.
2. Two clamps.

2.3 Cassette Dialysis Components
1. Appropriate size dialysis cassette to contain sample volume.
2. Syringe.
3. Needle.
4. Floater.

2.4 Microdialysis
1. VSWP02500 Millipore (Billerica, MA).
2. Petri plates.
3. Coomassie staining.

3 Methods

3.1 Dialysis

Carry out all procedures at room temperature (RT) unless otherwise specified.

1. Choose the desired type of dialysis (*see* Subheading 1).

2. Choose an adequate membrane pore size (*see* **Note 1**).

3. Prepare the appropriate dialysis buffer and pour it in a stir bar containing beaker (*see* Subheading 1 and **Note 2**).

4. Hydrate the semipermeable membrane: Dip membrane or membrane containing cassette into the dialysis buffer for approximately 5 min.

5. Load the sample into dialysis membrane:

   (a) Membrane dialysis:

   - Close the low end of the membrane with a clamp.
   - Introduce the sample through the upper side of the membrane (*see* **Note 3**).
   - Close upper side of the membrane with another clamp (*see* **Note 4**).

   (b) Cassette dialysis:

   - Fill a syringe with the sample (*see* **Note 5**).
   - Insert the needle of the syringe through one of the ports of the cassette until the membrane cavity is reached (*see* **Note 6**).
   - Inject the sample.
   - Withdraw the air from the membrane cavity using the syringe (*see* **Note 7**).
   - Remove the syringe while retaining the air.

6. Float sample containing dialysis device into the dialysis buffer (*see* **Note 8**).

7. Place the beaker on a magnetic stirring device.

8. Leave the sample dialyzing in agitation at 4 °C overnight (ON) (*see* **Note 9**).

9. Remove the sample from dialysis membrane:

   (a) Membrane dialysis:

   - Extract the membrane from the buffer.
   - Open the upper side clamp.
   - Carefully remove the sample from the membrane.

   (b) Cassette dialysis:

   - Extract the cassette from the buffer.
   - Fill a syringe with air (*see* **Note 10**).

- Insert the needle of the syringe through another port of the cassette taking care of not touching the membrane.
- Slowly introduce the air into the membrane cavity to separate membranes.
- Turn the cassette until the needle is in the bottom and the sample is near the needle.
- Remove the sample with the syringe.
- Centrifuge the sample at $15,000 \times g$ for 15 min to separate proteins aggregated during the dialysis process from soluble protein.

3.2 Microdialysis (Fig. 3)

Carry out all procedures at RT unless otherwise specified.

1. Prepare 50 mL of the screening buffers ensuring a correct concentration and pH equilibration.
2. Add 25 mL of each buffer into different Petri plates.
3. Float a VSWP02500 Millipore membrane filter disc (Billerica, MA) on the buffer with the bright side facing up. Avoid shiny membrane side contact with the buffer and the presence of air bubbles trapped under the filter.
4. Let the membrane filter equilibrate for 5 min.
5. Deposit several drops of 20 μL each (up to 6 to avoid overlapping) of model protein (previously prepared or purified) with known concentration on top of the membrane filter.
6. Leave the drops dialyzed for 30 min.
7. Collect the drops without perturbing the membrane.
8. Centrifuge the samples at $14,841 \times g$ for 15 min at 4 °C to separate the soluble fraction from the insoluble one and quantify it.
9. Let the membrane filters dry in new Petri dishes for 30 min.
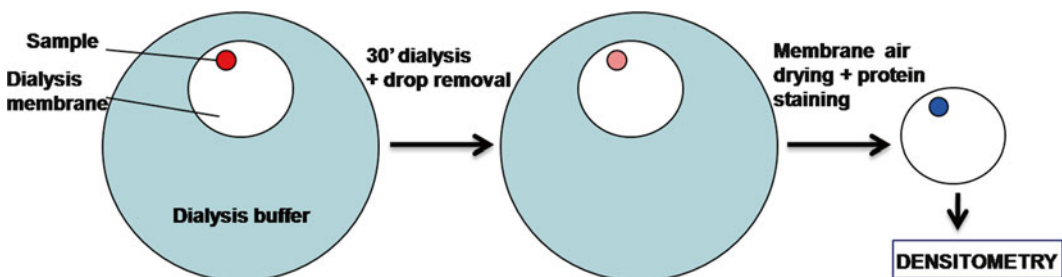
Fig. 3 Schematic representation of the microdialysis procedure, shown as a sequential pattern. The insoluble protein present in the dialyzed samples can be determined by densitometry of the filters after soluble protein removal and Coomassie staining. Modified from [6]

10. Stain the membrane filters with Coomassie for 15 min (*see* **Note 11**).

11. Destain the membrane filters until the surfaces of the drops are defined.

12. Dry the membrane filters for 30 min.

13. Scan by using a densitometer and quantify.

4 Notes

Tricks to reduce the aggregation tendency: (a) Perform the dialysis at RT for 1 h. This condition is generally enough to obtain an efficient sample equilibration against the buffer [15] and usually proteins show lower aggregation tendency. (b) Dilute the sample: Protein aggregation is generally concentration dependent [8, 16, 17] and accelerated aggregation of proteins at high concentrations has been reported in many cases [8, 18–20].

1. In dialysis, the membrane pore size is usually defined as molecular weight cutoff (MWCO), which is the solute size that is retained by at least 90 %. It is recommended to select an MWCO of 50 % the size of the molecular weight of the species to be retained (protein) and 100 times larger than the species that has to pass through (buffer).

2. Buffer volume: Prepare at least 1,000 times the sample volume to assure an efficient buffer interchange.

3. After hydration process, carefully remove any buffer traces remaining inside the membrane before applying the sample.

4. Avoid as much as possible any air bubble to be retained inside the membrane in order to maximize the sample's contact surface during the dialysis and avoid air/buffer interfaces which are detrimental for protein stability.

5. Fill the syringe with small amount of air before taking the sample to avoid any sample loss by the syringe's death volume.

6. Do not allow the needle to contact with the membrane to avoid piercing it.

7. Removing the air from membrane cavity will maximize sample's contact surface with the buffer.

8. In membrane dialysis, plastic clamps will confer floating capacity to the membrane. For cassette dialysis, couple the cassette to a floater.

9. Adjust agitation intensity to allow the membrane to be completely sunken into the buffer but avoiding to be too strong.

10. Fill the syringe with enough air volume to fill the membrane cavity with air and completely separate both side membranes.

11. Protein precipitated can be calculated by comparing soluble protein concentration calculation through absorbance A280 and Lambert-Beer law (A280 = $\varepsilon \cdot l \cdot$CM) (after **step 8**) from initial and after dialysis sample. However, it has been demonstrated [1] that aggregated fraction can also be measured by densitometry assay using Coomassie blue staining or by fluorometric assay (for fluorescent proteins) of protein deposited on the filter surfaces [6].

Acknowledgment

References

1. Wood EJ (1989) Protein purification methods, a practical approach. Oxford University Press, Oxford. ISBN 0-19-963003-8

2. Bollag DM, Edelstein SJ (1991) Protein methods. Wiley-Liss, New York. ISBN 0471568716

3. Graham T (1854) The Bakerian lecture: on osmotic force. Phil Trans R Soc Lond 144: 177–228

4. Graham T (1866) On the absorption and dialytic separation of gases by colloid septa. Phil Trans R Soc Lond 156:399–439

5. Bansal P, Ajay D (2012) Laboratory dialysis—past, present and future. Recent Pat Biotechnol 6:32–44

6. Toledo-Rubio V, Vazquez E, Platas G, Domingo-Espin J, Unzueta U, Steinkamp E, Garcia-Fruitos E, Ferrer-Miralles N, Villaverde A (2010) Protein aggregation and soluble aggregate formation screened by a fast microdialysis assay. J Biomol Screen 15: 453–457

7. Spectrumlabs. http://eu.spectrumlabs.com/dialysis/Fund.html. web page commercial. 1-2-0013

8. Wang W (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 185:129–188

9. Goto Y, Fink AL (1989) Conformational states of beta-lactamase: molten-globule states at acidic and alkaline pH with high salt. Biochemistry 28:945–952

10. Volkin DB, Klibanov AM (1989) Minimizing protein inactivation. In: T. E. Creighton (ed), Protein function: a practical approach. Information Press, Oxford, UK, pp 1–24

11. Kohn WD, Kay CM, Hodges RS (1997) Salt effects on protein stability: two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs. J Mol Biol 267:1039–1052

12. Xie G, Timasheff SN (1997) Mechanism of the stabilization of ribonuclease A by sorbitol: preferential hydration is greater for the denatured then for the native protein. Protein Sci 6:211–221

13. Timasheff SN (1998) Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. Adv Protein Chem 51:355–432

14. Giancola C, De SC, Fessas D et al (1997) DSC studies on bovine serum albumin denaturation. Effects of ionic strength and SDS concentration. Int J Biol Macromol 20: 193–204

15. Reiner M, Fenichel RL (1948) Dialysis of protein solutions for electrophoresis. Science 108: 164–166

16. Fields G, Alonso D, Stiger D et al (1992) Theory for the aggregation of proteins and copolymers. J Phys Chem 96:3974–3981

17. Ruddon RW, Bedows E (1997) Assisted protein folding. J Biol Chem 272:3125–3128

18. Won CM, Molnar TE, McKean RE, Spenlehauer GA (1998) Stabilizers against heat-induced aggregation of RPR114849, an acidic fibroblast growth factor (AFGF). Int J Pharm 167:25–36

19. Roefs SP, De Kruif KG (1994) A model for the denaturation and aggregation of b-lactoglobulin. Eur J Biochem 226:883–889

20. Gu LC, Erdos EA, Chiang HS et al (1991) Stability of interleukin 1 beta (IL-1 beta) in aqueous solution: analytical methods, kinetics, products, and solution formulation implications. Pharm Res 8:485–490

Chapter 19

Applications of Mass Spectrometry to the Study of Protein Aggregation

Sílvia Bronsoms and Sebastián A. Trejo

Abstract

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio of charged particles. Nowadays mass spectrometry-based approaches play a pivotal role in both detection and characterization of proteins. Here we describe two applications to study insoluble proteins: (a) hydrogen/deuterium exchange combined with mass spectrometry to analyze structural properties of amyloid fibrils and (b) the screening for inhibitors of the aggregation process by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Key words Mass spectrometry, Hydrogen/deuterium exchange, Aggregation, Inhibitor screening, Aβ42 amyloid peptide, Liquid chromatography, MALDI-TOF

1   Introduction

Mass spectrometry (MS) is a powerful technique for the study of protein structure, folding, and interaction in solution. Other techniques as X-ray crystallography and nuclear magnetic resonance spectroscopy can also fulfill these purposes with excellent results. However, they have important limitations, including the use of large amounts of material or the availability of high-quality crystals. MS analyses are fast and very sensitive and, more importantly, do not require large amounts of protein.

Nowadays MS represents a key technique to investigate insoluble proteins. The different applications of the methodology together with the coupling with other experimental approaches have forged MS as a very successful technique for the study of disordered proteins [1]. It has been applied to the detection of ordered/disordered regions in insoluble proteins, the analysis of conformational changes, the interaction of aggregated proteins with metals, the screening of inhibitors of aggregation, or the detection and quantification of proteins in vitro and in vivo [2].

Some novel MS-based methodologies have also achieved promising results in the study of insoluble proteins. MS imaging gives direct information on peptide/protein localization on tissue sections and it has been applied to the detection of Aβ peptides in mouse brain sections [3]. Ion mobility MS separates complex mixtures based on their shape and charge, giving conformational or structural information of the components of the mixture, and it has revealed new insights into protein oligomerization during amyloid assembly [4].

In this chapter we describe two MS applications for the study of insoluble proteins. In the first one the use of hydrogen/deuterium exchange (HDX) combined with electrospray ionization (ESI) MS is applied to the characterization of structural properties of amyloid fibrils. In the second one matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS is employed to screen for inhibitors of the aggregation process in a high-throughput format.

2   Materials

2.1   Structural Characterization of Protein Aggregates by HDX_LC-MS

Prepare all solutions using ultrapure H_2O (18 mΩ cm at 25 °C) and use MS-grade solvents and reagents.

2.1.1   Amyloid Fibril Preparation

1. PBS buffer (phosphate-buffered saline): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH 7.4. Prepare a stock solution (10×), autoclave, and store at 4 °C. Prepare working solution by dilution of one part of stock solution with nine parts ultrapure H_2O.

2.1.2   Hydrogen Exchange Labeling and Quenching of the Samples

1. Labeling buffer: 2 mM Tris–HCl pH 7.5 in 95 % D_2O. Weigh 1.21 mg Trizma base, add 9 mL of D_2O, adjust pH with DCl, add 0.5 mL of MS-grade deionized H_2O, and make up to 10 mL with D_2O.

2. Quenching buffer: 50 μL of formic acid (FA) MS grade is diluted to 50 mL with MS-grade deionized H_2O. Store at 4 °C and keep it on ice a few minutes before use.

2.1.3   Protein Digestion

1. Enzymes: Pepsin, protease type XIII from *Aspergillus saitoi* and protease type XVIII from Rhizopus species can be used. Prepare 0.1 % FA solution to use as digestion buffer. Store at 4 °C and keep it on ice a few minutes before use.

2. Peptide separation is performed by reverse-phase chromatography using a C18 column (e.g., 3 μm C_{18}, 75 μm id × 10 cm) in a high-pressure liquid chromatography (HPLC) system. The use of a trap column (e.g., 1 mm id × 8 mm microtrap column) is also recommended. LC mobile phases: Phase A is

0.1 % FA in MS-grade water and phase B is 0.1 % FA in MS-grade acetonitrile. Prepare the mobile phases in advance and cool them prior to the chromatographic separation. Keep them on ice during the chromatography.

2.2 Screening of Inhibitors of Aβ42 Aggregation

Aβ42 lyophilized powder, 1,1,1,3,3,3-hexafluoro-propan-2-ol (HFIP), acetonitrile (MS grade), sinapinic acid (MS grade), trifluoroacetic acid (MS grade), and water (MS grade). Any other reagents used in the assays are of normal use in the laboratory.

2.2.1 Aβ42 Sample Preparation

1. 300 μM Na_2CO_3 solution: Dissolve 31.8 mg of Na_2CO_3 in 1 L H_2O and store at room temperature (RT).

2. 250 μM NaOH solution: Dissolve 10.0 mg of NaOH in 1 L H_2O and store at RT.

3. Starting solution: Mix 480 μL of ACN, 480 μL of 300 μM Na_2CO_3, and 34 μL of 250 μM NaOH. This solution should be prepared at the moment of use.

2.2.2 Detection of Inhibitory Activity of Aβ42 Aggregation

1. Solution A: Dissolve 27.6 g of $NaH_2PO_4 \cdot H_2O$ in 1 L H_2O (200 mM) and store at RT.

2. Solution B: Dissolve 53.62 g of $NaH_2PO_4 \cdot H_2O$ in 1 L H_2O (200 mM) and store at RT.

3. Aggregation solution (10 mM phosphate buffer–11 mM NaCl, pH 7.7): Mix 10.5 mL of solution A with 89.5 mL of solution B, add 127.6 mg of NaCl, and make up to 200 mL with H_2O to obtain 10× aggregation solution. It is strongly recommended to check the final pH with a pH meter and adjust the final pH to 7.7 with HCl. Finally, at the moment of use dilute 1:10 with H_2O to obtain the working solution.

2.2.3 MALDI-TOF MS Data Acquisition

1. MALDI matrix solution: Dissolve 30 mg of sinapic acid with 499 μL of H_2O, 500 μL of ACN, and 1 μL of trifluoroacetic acid.

3 Methods

3.1 Structural Characterization of Protein Aggregates by HDX_LC-MS

HDX combined with ESI MS can be applied to the study of structural properties of amyloid fibrils. This approach allows us to identify the core of the amyloid fibril, as defined by systematically H-bonded structures [5]. Backbone amide hydrogens located in protein unstructured regions exchange very fast; on the other hand amide hydrogens participating in H-bonded structures or buried in the core of a protein show a slow exchange rate. Since amide protons located in the β-sheet core of amyloid fibrils are extremely resistant to exchange, these proteins are very good models to be studied by HDX.

HDX can be used to probe the core structure of amyloid fibrils and protofibrils, as developed in the following protocol, or it can be applied to monitor protein conformational changes as a function of time [6]. In both cases, to obtain detailed information about HDX rates along the molecule, the protein can be proteolyzed. In this way the spatial resolution is improved and we are able to define differences in HDX rates in peptides ranging from 5 to 15 amino acids.

3.1.1 Amyloid Fibril Preparation

1. To obtain amyloid fibrils a concentrated sample of the protein, around 100 μM, is dissolved on ice in PBS buffer, and is further incubated for several days at 37 °C (*see* **Note 1**).

2. Fibrils are pelleted via centrifugation at $20,000 \times g$ for 20 min at 4 °C and the supernatant is removed (*see* **Note 2**).

3.1.2 Hydrogen Exchange Labeling and Quenching of the Samples

For these experiments a minimum of triplicate runs should be performed.

1. The fibril pellet is resuspended at 15–20 μM in 2 mM Tris–HCl pH 7.5, 95 % D_2O, and 5 % H_2O (*see* **Note 3**).

2. The labeling proceeds at RT and aliquots are taken at different time points, usually from 1 min to 24 h, but the final time point depends on the protein (*see* **Note 4**).

3. Each labeled aliquot is quenched by rapid acidification to pH 2.5 on ice with fully non-deuterated buffer. The sample is diluted 1:9–1:19 in final 0.1 % FA and the solution may be supplemented with a denaturant (0.5–1 M urea, guanidinium thiocyanate, or guanidine hydrochloride) (*see* **Note 5**). The HDX rate reaches a minimum at pH 2.5, and together with the low temperature it minimizes the so-called exchange-in and also the back-exchange. Finally, the acidic pH and the presence of denaturant help to the rapid dissociation of the amyloid fibrils to monomers (*see* **Note 6**). A scheme of the whole procedure is displayed in Fig. 1.

3.1.3 Protein Digestion

1. Samples are digested with an acidic nonspecific protease at 0 °C. There are three proteases that work at acidic pH and give good proteolytic fragmentation: pepsin, protease type XIII from *Aspergillus saitoi*, and protease type XVIII from *Rhizopus* species [7]. Digestion can be performed in solution or in column (*see* **Note 7**) and many overlapping peptides are obtained.

   For in-solution digestion, a 2-min incubation using a ratio around 1:1 (w/w) enzyme/protein is usually convenient (*see* **Note 8**).

   For immobilized proteases, samples are passed through the column at acidic conditions (i.e., 0.1 % FA).
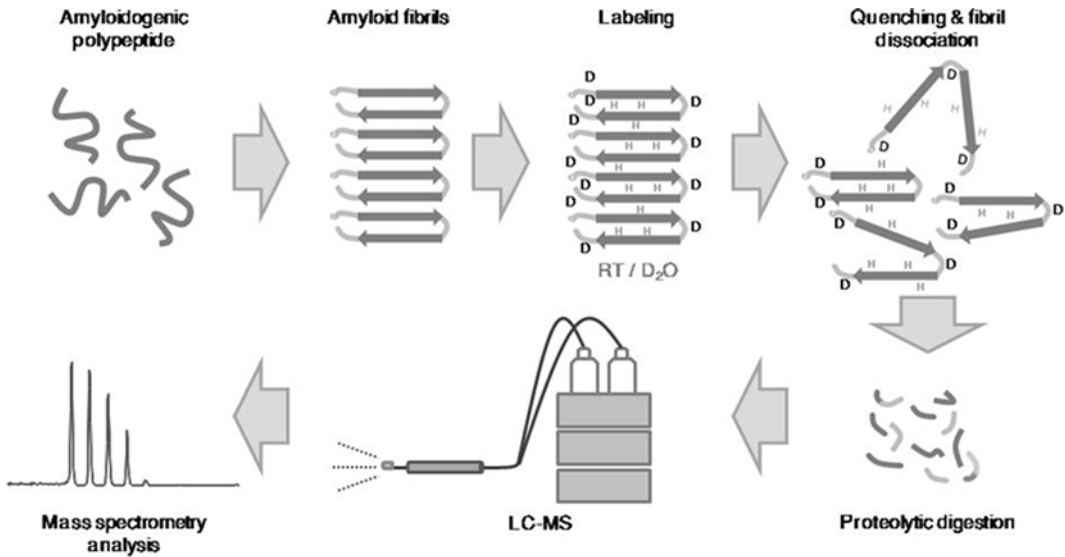
Fig. 1 Workflow of the hydrogen/deuterium exchange protocol followed by liquid chromatography mass spectrometry

2. Digested samples are purified through a C18 reverse-phase column working at pH 2.5 and 0 °C (*see* **Note 9**). Phase A is 0.1 % FA in water and phase B is 0.1 % FA in acetonitrile. A typical 5–50 % phase B gradient in 15–30 min may give a good separation of the digested peptides (*see* **Note 10**). This step allows both desalting and removal of undigested protein. It is important to cool buffers, valves, injector, and column in the chromatographic step to minimize back-exchange effects.

3.1.4   In- and Back-Exchange Controls

These controls are necessary to correct the HDX results for the presence of deuterium atoms in the quenched state that can be incorporated by the peptides (in-exchange) and for deuterium back-exchange due to the presence of H_2O in the sample during protein digestion, separation, and MS analysis.

1. For back-exchange control, a small sample of the protein is digested using the same acidic protease (*see* Subheading 3.1.3), diluted 1:19 in deuterated buffer and incubated for 10–20 h. Afterwards the sample is quenched with 0.1 % FA at 0 °C, separated by a reverse-phase column, and analyzed by MS.

2. For in-exchange control, the sample used for the HDX experiment is diluted 1:19 with a deuterated quenching solution at 0 °C. The sample is digested as previously stated, separated by a reverse-phase column, and analyzed by MS.

3.1.5 Mass Spectrometry and Data Analysis

1. The eluent of the reverse-phase column is injected directly into a mass spectrometer. Any spectrometer with a good resolving power may be used to analyze HDX experiments (e.g., an ion trap, a quadrupole-time of flight, an Orbitrap).

2. Non-deuterated samples are used to characterize the peptide map obtained after the digestion with the acidic protease. Each protease generates a certain amount of peptides, most of them from 5 to 15 amino acids, which cover the protein sequence. This proteolytic pattern is reproducible when working with the same protein and under the same digestion conditions. Thus, we will obtain roughly the same pattern of peptides for the deuterated and non-deuterated samples.

3. Deuterated samples are analyzed in the same conditions and their peptidic maps are characterized.

   Each peptide may acquire one, two, or more positives charges upon ionization; this fact depends largely on the type of mass spectrometer that we use, and therefore the instrument of choice cannot be changed along a whole study. Besides, a peptide does not contain a single peak, but a group of signals that correspond to the isotopic distribution. The isotopic pattern starts with the monoisotopic mass (it is the lowest mass of the peptide and it accounts for those molecules of the peptide for which all atoms are the lightest naturally occurring isotopes) and also includes the masses of the molecules containing heavier naturally occurring isotopes (e.g., ^{13}C, ^{34}S, ^{15}N), or in the case of deuterated samples, also the ^{2}H introduced by the HDX experiments (Fig. 2 shows the typical isotopic pattern of a non-deuterated and a deuterated peptide). Although high-resolution mass spectrometers may assign a peptide by its exact mass-to-charge ratio, for most of the equipment (ion trap, quadrupole-time of flight) peptides must be identified by tandem mass spectrometry before being assigned to a particular fragment of the protein sequence (*see* **Note 11**).

4. For each peptide its average molecular weight (the centroid mass of the peak envelope) is calculated in deuterated and non-deuterated samples (in-exchange control); in deuterium-labeled samples there is a mass shift of the peptides to higher mass values. The difference between the centroid mass of the labeled and unlabeled peptides accounts for the deuterium content of peptides. Since the exchange rates of deuterons located on side chains of peptides are several orders of magnitude faster than the exchange rates of backbone amide hydrogens, they are completely back-exchanged during the analysis [8]. Accordingly the observed mass shift on the deuterated peptides is a direct measure of the exchanged amide protons during the experiment.
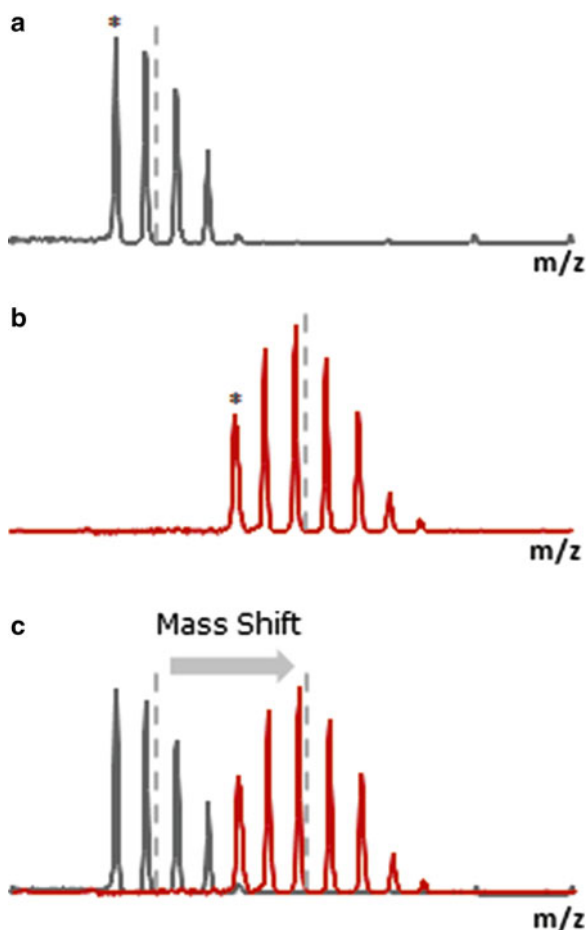
Fig. 2 Scheme of the isotopic distribution of a non-deuterated (**a**) and deuterated (**b**) peptide. The monoisotopic peak is indicated (*asterisk*); the average mass of the distribution is marked with a *dashed line*. The superimposition of the two distributions (**c**) shows the mass shift produced by the peptide deuteration

5. Peptide deuteration levels are finally corrected for in- and back-exchange effects during quenching, proteolysis, and chromatography using the following equation [9]:

$$D(t) = \frac{m(t) - m(0)}{m(100) - m(0)} \times N$$

$$\text{Deuterium incorporated}\,(\%) = \frac{D(t)}{N}$$

where $m(t)$ is the centroid mass of a given peptide at the incubation time t, $m(0)$ is the centroid mass of the same unlabeled peptide (in-exchange control), $m(100)$ is the centroid mass for the fully exchanged peptide (back-exchange control), N is the total number of exchangeable amide protons in the peptide, and finally $D(t)$ is the corrected number of incorporated deuteriums for a given peptide at the incubation time t.
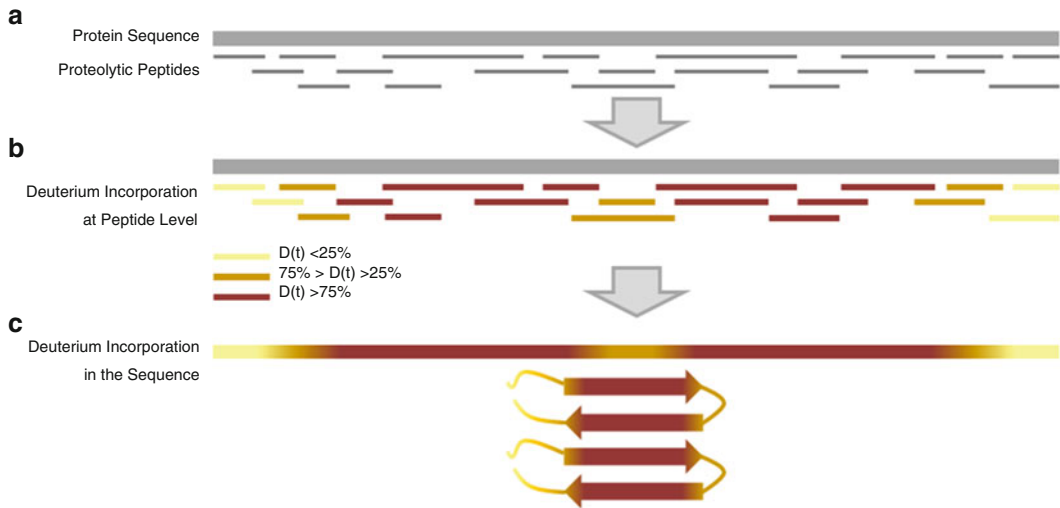
Fig. 3 Exchange pattern of a protein at peptide level. The reporter peptides of the protein (**a**) are analyzed by LC-MS to determine their level of deuterium incorporation (**b**) and the information is visualized on the protein structure (**c**). Each *horizontal block* represents one reporter peptide and its *color* shows the degree of labeling as indicated in the *inset*

6. Classify the peptides depending on the amount of incorporated deuterons, i.e.:

   – Peptides belonging to highly protected regions: Deuterium incorporation (%) < 25 %.

   – Peptides belonging to moderately protected regions: Deuterium incorporation (%) between 25 and 75 %.

   – Peptides belonging to low protected regions: Deuterium incorporation (%) > 75 %.

   In Fig. 3 the workflow of the deuterium incorporation analysis is described.

7. Deuterium labeling kinetics: The labeling kinetics of each peptide may also be calculated by representing the number of incorporated deuterons as a function of time (Fig. 4).

3.2 Screening of Inhibitors of Aβ42 Aggregation

Many efforts have been done to develop efficient screening methods to detect inhibitors of Aβ42 aggregation using high-throughput MS approaches. Some of them have been carried out by ESI MS and others by MALDI-TOF MS, both with their own advantages and disadvantages [2, 10–12]. Nowadays, we have a deeper understanding about the myeloid fibril formation grateful to the efforts of many scientists [2, 13], and we can summarize the Aβ42 self-assembly with a simple scheme (Fig. 5).

MALDI-TOF MS detection was selected for this chapter of screening of inhibitors of Aβ42 aggregation due to is speed, high
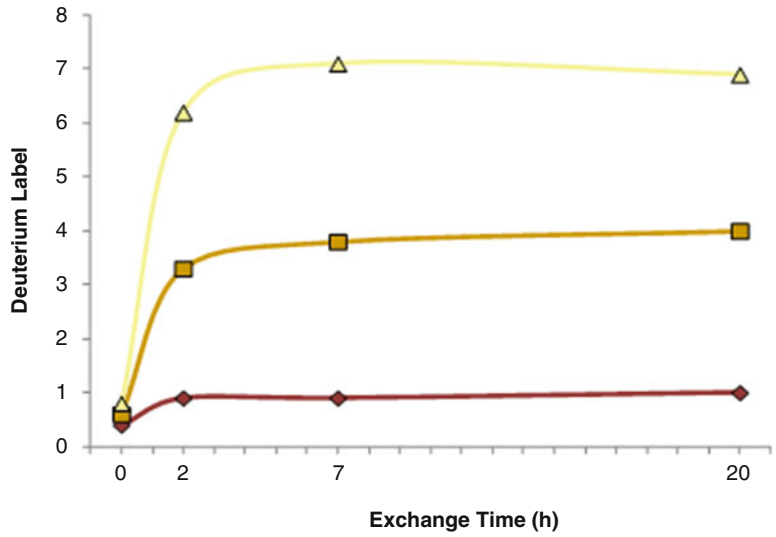
Fig. 4 Deuterium labeling kinetics of three proteolytic peptides measured by mass spectrometry. The *color* of the curve shows the degree of labeling as indicated in the *inset* of Fig. 3
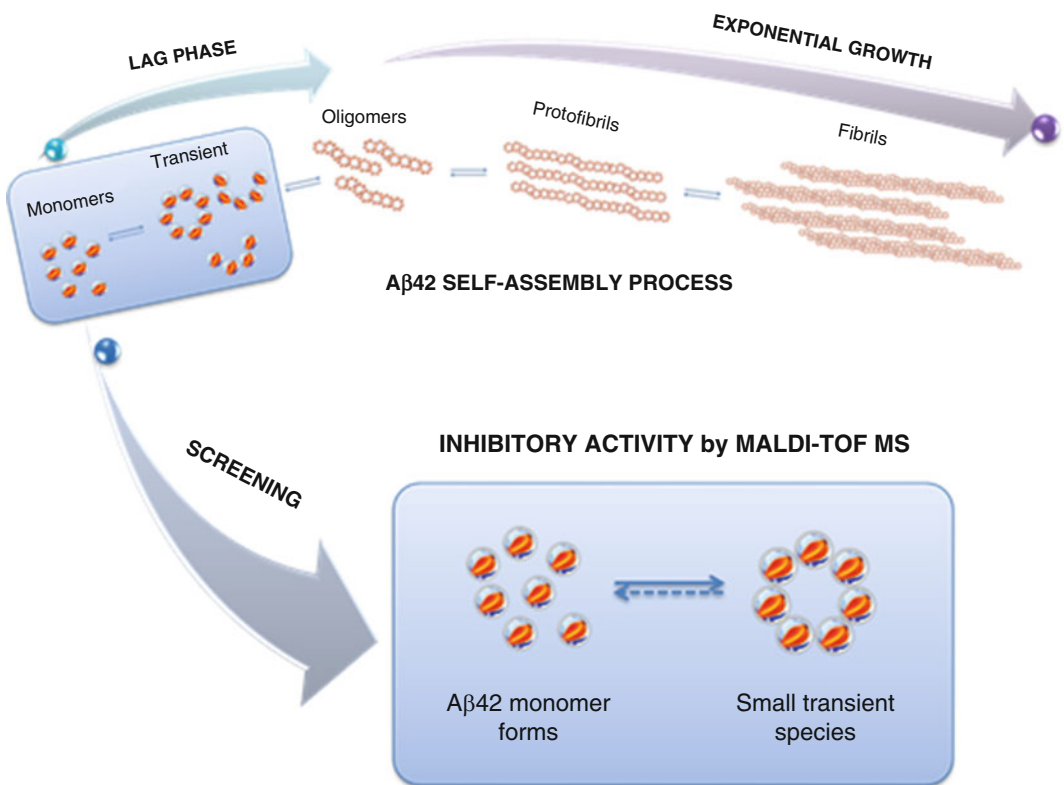
Fig. 5 Scheme of Aβ42 self-assembly process. Aβ42 monomers form low-molecular-weight conformers that may lead into oligomers, protofibrils, and fibrils [2, 19]

sensitivity, and automation possibilities. Transient species appearing at early phases of the Aβ42 oligomerization process have been successfully studied by MALDI-TOF MS [14]. Moreover, the study of the inhibition of the oligomerization on this phase has also been analyzed with this approach [15].

The following methodology is specifically applied to study the inhibition of the formation of low-molecular-weight Aβ42 oligomers.

*3.2.1 Aβ42 Sample Preparation (See **Note 12***)

1. Aβ42 lyophilized powder is dissolved in HFIP at 0.15 mM concentration by sonication and vortexing. This solution is kept overnight (ON) at RT.

2. The solution is split in several microcentrifuge tubes (prepare aliquots of 100 μL) and HFIP is evaporated on a vacuum centrifuge. In these conditions, the resulting Aβ42 film can be stored for months at –20 °C.

3. To start the assay, a vial of Aβ42 film is redissolved in 30 μL of aggregation solution (*see* Subheading 2) by brief sonication and vortexing. This will be the Aβ42 starting solution (alkaline solution of 500 μM Aβ42).

3.2.2 Aβ42 Aggregation Inhibition Studies

1. The starting solution is diluted 1:9 with 10 mM buffer phosphate–NaCl to obtain a 50 μM Aβ42 solution at final pH 8.0. For inhibition studies, the Aβ42 starting solution is diluted 1:9 with 10 mM buffer phosphate–NaCl containing 55 μM of the tested inhibitor (*see* **Note 13**) to obtain a molar ratio of inhibitor/Aβ42 = 1.

2. The Aβ42 solution is briefly sonicated and incubated at 30 °C (without stirring).

3. Aliquots are collected at different time points (*see* **Note 14**) for further analysis by MALDI-TOF MS to determine the effect of the tested compound over the Aβ42 aggregation process.

The final concentration of all components is described in Table 1.

Table 1
Composition of the screening assay

Component	Aβ42 (μM)	Inh. (μM)	Phosphate buffer (mM)	NaCl (mM)	Na_2CO_3 (μM)	NaOH (mM)	ACN (% v/v)	pH
Fc – inh	50	–	8.7	10	14.5	0.85	8.2	8.0
Fc + inh	50	50	8.7	10	14.5	0.85	8.2	8.0

Final concentration of the components in the assay without (Fc – inh) and with (Fc + inh) inhibitor
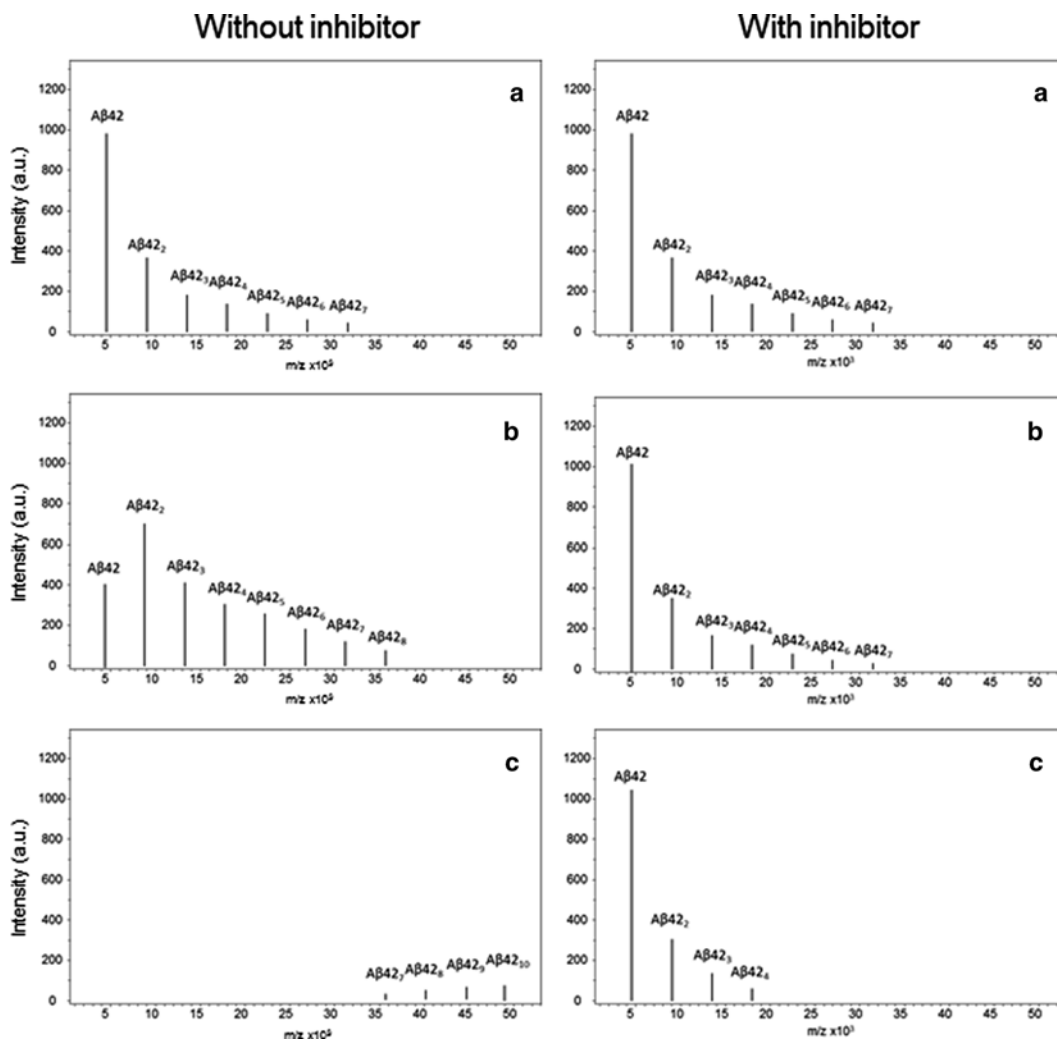
Fig. 6 Screening assay. Comparison of MALDI-TOF MS spectra of Aβ42 peptide at time zero (**a**), at an intermediate incubation time (**b**), and at the end of the assay (**c**) in the absence (*Left panel*) or presence (*Right panel*) of an inhibitor of the oligomerization process

3.2.3   MALDI-TOF MS Data Acquisition

1. MALDI-TOF MS analyses are performed using a MALDI-TOF mass spectrometer.

2. The analysis is carried out by spotting on the target plate 1.0 μL of the sample mixed with an equal volume of the matrix solution.

3. Spectra are acquired in linear mode at *m/z* range from 2,000 to 50,000 (*see* **Note 15**).

3.2.4   Data Analysis

1. A full-time course of MS spectra is obtained for each condition that was tested in the experiment (Fig. 6).

2. The time courses of the experiments with and without the tested compound are compared.

   The monomeric form and some low-molecular-weight molecules of Aβ42 oligomeric forms populate the time zero spectra of both experiments, with and without inhibitor. The presence of inhibitor prevents the disappearance of Aβ42 monomeric form and the accumulation of oligomers; therefore the MALDI-TOF MS spectra do not change substantially throughout the time course with inhibitor (Fig. 6 (Right panel)). On the other hand, in the absence of inhibitor Aβ42 oligomeric forms populate the first stages of the time course and result in a decrease of Aβ42 monomeric form, which almost disappears at long incubation times (Fig. 6 (Left panel).

3. It is strongly recommended to validate the detection of inhibitory activity in the screening assay using a kinetic method (*see* **Note 16**).

4   Notes

1. For each protein there are specific experimental conditions to obtain amyloid fibrils. For example, for Aβ (1–40) a 6-day incubation at 37 °C without shaking [8], for α-synuclein a 4-day incubation at 37 °C with shaking [16], for PrP (89–143) prion protein a 3-week incubation at 10 mg/mL in a 50 % acetonitrile solution at RT [17], or for PAPf39 a 2-day period at 37 °C with agitation [18].

2. One or two washes of the amyloid fibrils with PBS or 10 mM sodium phosphate pH 7.5 are optional but recommended. At this point the presence of fibrils in the sample should be confirmed, for example by thioflavin T fluorescence measurements, atomic force microscopy, transmission detection microscopy, or Congo red binding assays.

3. The fibril pellet can also be resuspended in a small volume of non-deuterated buffer, final concentration around 100–200 μM, and then it can be diluted 1:9–1:19 in deuterated buffer.

4. Usually a first experiment is performed to set up the time course. In this first experiment samples are taken at increasing times and at the end of the experiment its labeling kinetics is represented (Fig. 4). The labeling curve displays a plateau at its final part; the suitable final time point for the experiment should belong to this part of the curve.

5. Many different quenching buffers may be used, but in all cases it must be a cold buffer, with a pH value around 2.5 and with the convenient amount of detergent, salts, or other compounds necessary to ensure the complete dissociation of the fibrils.

6. At this point, the samples can be flash-frozen and stored at –80 °C or analyzed immediately. However, direct analysis should be the first choice whenever possible.

7. A combination of two enzymes can be used to increase the number of proteolytic peptides. In this way, the spatial resolution of the experiment is increased by shortening the length of the peptides.

8. Before we run the HDX experiment we may follow the protein digestion by electrophoresis to determine the optimum protein/enzyme ratio. To minimize autolysis we should choose the lowest protease concentration required to obtain a complete protein digestion at 0 °C in 2 min.

9. For samples with denaturant or high salt content, a previous desalting step with a trap column is recommended.

10. The duration of the gradient can be shorter or longer depending on the amount of peptides present in the sample. The gradient can be set up before the HDX experiments using a digested protein sample (obtained in the same experimental conditions as followed in the HDX experiment) and optimizing the conditions to minimize the gradient duration while maximizing the chromatographic peptide resolution. The buffers can also be acidified using 0.1 % trifluoroacetic acid instead of FA.

11. In both cases, sequential information of the reporter peptides is obtained by performing tandem MS by application of strong electric field or collision with inert gas.

12. Bartolini et al. [19, 20] reported the importance to work with a homogeneous system for studies of the early steps of Aβ42 self-aggregation. In these works, the authors described in detail the principles that lead to obtain homogeneous and stable starter solutions of Aβ42 monomers for this kind of assays.

13. The preparation of this solution depends on the solubility of the tested compounds. The recommended protocol is to dissolve the compound to obtain a 55 μM solution in 10 mM buffer phosphate–NaCl. More concentrated/diluted solutions of the compound can also tested; in this case the concentrations of the remaining components of the assay should be recalculated. If the solubilization of the compound requires the addition of a specific solvent, make sure to add the same amount of solvent in the control experiment (without inhibitor).

14. The time course of Aβ42 aggregation has already been reported by thioflavin T fluorescence and by MALDI-TOF MS [14]. The authors have reported that after 12 h of Aβ42 self-assembly only traces of monomer are detected. Thus, 12 h can be chosen as the final time point for the study of inhibitory

activities. This final time point depends on the experimental conditions; therefore it should be defined again whenever these conditions are modified.

15. The quality of the obtained spectra can be improved by collecting the data in two independent ranges of m/z, for example from 2,000 to 20,000 and from 20,000 to 50,000.

16. Zovo et al. [21] developed a powerful method by MALDI-TOF MS which could be used for this purpose. If preferred, a classical method based on colorimetric or fluorometric measures can also be used.

References

1. Beveridge R, Chappuis Q, Macphee C et al (2013) Mass spectrometry methods for intrinsically disordered proteins. Analyst 138:32–42

2. Grasso G (2011) The use of mass spectrometry to study amyloid-β peptides. Mass Spectrom Rev 30:347–365

3. Stoeckli M, Knochenmuss R, McCombie G et al (2006) MALDI MS imaging of amyloid. Methods Enzymol 412:94–106

4. Woods L, Radford SE, Ashcroft E (2013) Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly. Biochim Biophys Acta 1834: 1257–1268

5. Lu X, Wintrode PL, Surewicz WL (2007) Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. Proc Natl Acad Sci U S A 104:1510–1515

6. Carulla N, Zhou M, Arimon M et al (2009) Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation. Proc Natl Acad Sci U S A 106:7828–7833

7. Cravello L, Lascoux D, Forest E (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. Rapid Commun Mass Spectrom 17:2387–2393

8. Qi W, Zhang A, Patel D et al (2008) Simultaneous monitoring of peptide aggregate distributions, structure, and kinetics using amide hydrogen exchange: application to Abeta(1-40) fibrillogenesis. Biotechnol Bioeng 100:1214–1227

9. Zhang Z, Smith DL (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci 2:522–531

10. Nerelius C, Gustafsson M, Nordling K et al (2009) Anti-amyloid activity of the C-terminal domain of proSP-C against amyloid beta-peptide and medin. Biochemistry 48:3778–3786

11. Skribanek Z, Baláspiri L, Mák M (2001) Interaction between synthetic amyloid-β-peptide (1-40) and its aggregation inhibitors studied by electrospray ionization mass spectrometry. J Mass Spectrom 36:1226–1229

12. Cheng X, van Breemen RB (2005) Mass spectrometry-based screening for inhibitors of β-amyloid protein aggregation. Anal Chem 77: 7012–7015

13. Necula M, Kayed R, Milton S (2007) Small molecule inhibitors of aggregation indicate that amyloid beta oligomerization and fibrillization pathways are independent and distinct. J Biol Chem 282:10311–10324

14. Bartolini M, Naldi M, Fiori J et al (2011) Kinetic characterization of amyloid-beta 1-42 aggregation with a multimethodological approach. Anal Biochem 414:215–225

15. Fiori J, Naldi M, Bartolini M et al (2012) Disclosure of a fundamental clue for the elucidation of the myricetin mechanism of action as amyloid aggregation inhibitor by mass spectrometry. Electrophoresis 33:3380–3386

16. Del Mar C, Greenbaum E, Mayne L et al (2005) Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. Proc Natl Acad Sci U S A 102:15477–15482

17. Damo SM, Phillips AH, Young AL et al (2010) Probing the conformation of a prion protein fibril with hydrogen exchange. J Biol Chem 285:32303–32311

18. French KC, Makhatadze GI (2012) Core sequence of PAPf39 amyloid fibrils and mechanism of pH-dependent fibril formation: the role of monomer conformation. Biochemistry 51:10127–10136

19. Bartolini M, Bertucci C, Bolognesi ML et al (2007) Insight into the kinetic of amyloid beta

(1-42) peptide self-aggregation: elucidation of inhibitors' mechanism of action. Chembiochem 8:2152–2161

20. Bartolini M, Bertucci C, Cavrini V et al (2003) beta-Amyloid aggregation induced by human acetylcholinesterase: inhibition studies. Biochem Pharmacol 65:407–416

21. Zovo K, Helk E, Karafin A et al (2010) Label-free high-throughput screening assay for inhibitors of Alzheimer's amyloid-β peptide aggregation based on MALDI MS. Anal Chem 82:8558–8565

Chapter 20

Insoluble Protein Assemblies Characterized by Fourier Transform Infrared Spectroscopy

Antonino Natalello and Silvia M. Doglia

Abstract

Fourier transform infrared (FTIR) spectroscopy is a useful tool for the structural characterization of insoluble protein assemblies, as it allows to obtain information on the protein secondary structures and on their intermolecular interactions. The protocols for FTIR spectroscopy and microspectroscopy measurements in transmission and attenuated total reflection modes will be presented and illustrated in the following examples: bacterial inclusion bodies, self-assembling peptides, thermal aggregates, and amyloid fibrils.

Key words Amyloid, β-sheet, Fourier transform infrared spectroscopy, Hydrogel, Inclusion body, Microspectroscopy, Peptide scaffold, Protein aggregation, Secondary structures, Spectroscopy

1 Introduction

The formation of insoluble protein assemblies is a phenomenon widely observed in vitro and in vivo, occurring under several conditions and leading to final products characterized by very different structural properties. This process is of great interest in numerous fields (such as medicine, biotechnology, and food processing) and can be in different cases either a desirable or detrimental event [1–3]. In particular, the aggregation of a specific protein in the form of oligomers and fibrils is the hallmark of a number of neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases, and of non-neuropathic localized and systemic amyloidosis, such as type II diabetes and hemodialysis-related amyloidosis. However, several proteins have been also found to form amyloid-like fibrils unrelated to diseases but, on the contrary, associated with a normal biological activity and therefore called "functional amyloids" [2]. The current view is that every protein can adopt an amyloid structure under specific conditions, namely, that the capability to form such a structure is an inherent property of polypeptide chains. Indeed, the amyloid fibrils of different proteins share common structural and morphological properties, such as a

fibrillar morphology with core structures of intermolecular β-sheets with β-strands perpendicular to the long fibril axis (called cross-β-sheet) and the binding to specific amyloid dyes (i.e., Congo red, Thioflavin T and S) [2]. The ordered supramolecular structure of protein aggregates can be also exploited for the development of new biomaterials [4]. Indeed, very recently, peptide- and protein-based assemblies in the form of hydrogel have attracted the attention for their possible applications, for instance, in tissue engineering [5]. The chemical or physical cross-linking of the constituents makes hydrogels insoluble, even if their polymeric network and the 3D structure allow the adsorption of large amount of water. These supramolecular assemblies can be obtained by natural or designed polypeptides, which in the final materials can assume different conformations: the intermolecular β-sheet or the coiled-coil structures in particular [5]. The relevance of protein aggregation in food processing can be illustrated considering the generation of gelling and foaming agents [3, 6]. In biotechnology, aggregation can occur during the expression, purification, and storage of the protein of interest, eventually leading to a strong decrease of the final yield in the native and functional protein. In particular, the overexpression of a recombinant protein in bacteria is often associated to its deposition in cytoplasmic or periplasmic aggregates, called inclusion bodies (IBs). In recent years, the scientific and industrial interest for these aggregates is growing, and, instead of being regarded as waste by-products of protein production, IBs are now considered as new promising nanomaterials [7] and model systems in biotechnology and protein science [8]. Indeed, it has been found that the functional and structural features of the polypeptides embedded in IBs can be tuned by the expression conditions. The constituent proteins can retain their function and can assume different conformations from the native-like to the amyloid structures. In this way, IBs can be exploited as a source of functional proteins for the controlled delivery of therapeutic polypeptides and drugs, as reusable biocatalysts, as model systems for amyloid studies, and as biocompatible materials that can stimulate mammalian cell proliferation, with possible uses in tissue engineering [7, 9–11].

The examples reported above emphasize the high interest for these insoluble protein assemblies and thus also for the methods able to characterize their aggregation as well as the structural properties of the final assemblies. Among the available methods, Fourier transform infrared (FTIR) spectroscopy has been extensively applied to the study of protein aggregation both in the test tube and in situ on intact biological samples [12–16]. In particular, as it will be discussed, this technique allows to study the protein secondary structures and the formation of intermolecular β-sheet structures of samples under very different physical states

[12, 15]: protein solutions [15, 17, 18], hydrated thin films [19–21], solid films, powder of lyophilized proteins [22], protein crystals [23], as well as suspensions of protein precipitates and hydrogels [12]. Moreover, the FTIR spectra of proteins can be measured by different sampling techniques using a transmission cell, a device for measurements in attenuated total reflection (ATR) or an IR microscope, allowing to collect spectra in transmission, ATR, or reflection modes. The FTIR absorption spectrum of a protein is dominated by the absorption bands due to the vibrational modes of the peptide bond, called amide bands, including the amide I (1,700–1,600 cm^{-1}), amide II (1,600–1,500 cm^{-1}), and amide III (1,400–1,200 cm^{-1}). The most used for protein conformational characterizations is the amide I band, mainly due to the stretching vibration of the C=O peptide group, which is very sensitive to the C=O environment and therefore to the secondary structure of the protein. The amide I band of a protein is therefore the result of the overlapping of the C=O bond absorption in the different secondary structures, including intermolecular β-sheets [14, 15, 24–26].

The FTIR characterization of protein-insoluble assemblies can be performed by employing different sampling techniques, each requiring peculiar conditions and procedures for sample preparation and spectral collection. Therefore, in this chapter, we will present protocols for FTIR spectroscopy and microspectroscopy measurements in transmission and ATR modes of different types of protein aggregates, namely, peptide scaffolds, amyloid fibrils, thermal aggregates, and IBs—both in solution/suspension and in the form of thin hydrated film.

2   Materials

2.1   Reagents, Peptides, and Proteins

Solutions were prepared using ultrapure water (Milli-Q) and analytical grade reagents.

1. Deuterated phosphate buffered saline (PBS): 50 mM sodium phosphate, 100 mM NaCl, and pH = 7.4. It was obtained starting from the non-deuterated buffer that was then lyophilized and resuspended in D$_2$O. The final pH was adjusted by deuterated hydrochloric acid and sodium hydroxide (*see* **Note 1**).

2. The B24 peptide [27, 28], ataxin-3 protein with 55 glutamines in the polyglutamine (poly-Q) region (AT3-Q55) [19], β-2 microglobulin [23], and AT3-Q55 IBs [29] were obtained as previously described.

   See **Note 2** for a general remark on the typical protein and peptide concentrations required for the FTIR studies.

2.2 FTIR Spectrometer	1. Varian 670-IR spectrometer (Varian Australia Pty Ltd, Mulgrave VIC, AU) equipped with a nitrogen-cooled mercury cadmium telluride detector and carefully purged by dry air to avoid the interference of water vapor.
	2. Varian 610-IR microscope coupled with the Varian 670-IR spectrometer for micro-FTIR studies.
2.3 ATR Devices	Two ATR accessories, both with a diamond crystal, were employed:
	1. Single-reflection Golden Gate device (Specac Ltd, UK).
	2. Nine-reflection DuraSamplIR II device (Smith Detection, USA) equipped with a temperature controller.
2.4 Transmission Cell	For FTIR measurements in the transmission mode of protein solutions, we employed a temperature-controlled transmission cell (Wilmad, Buena, NJ, USA) with two BaF_2 windows (Ø 13 mm×2 mm dimensions; from Korth kristalle, GmbH, Kiel, Germany) and an optical path whose length was determined by the thickness of the Teflon spacer.
2.5 Software	Spectra collection and Fourier self-deconvolution were performed with the Varian Resolutions Pro software (Varian Australia Pty Ltd, Mulgrave VIC, AU). All the other spectral analyses were done using the GRAMS/AI 8 software (Thermo Fisher Scientific Inc., USA).

3 Methods

3.1 FTIR Measurements of Peptide Scaffolds and Amyloid Fibrils in the ATR Mode	The attenuated total reflection approach has been extensively applied for the FTIR characterization of protein aggregates in different states, namely, in solution/suspension and in the form of protein hydrated film [12]. As illustrated in Fig. 1a, the sample is deposed on the crystal element of the ATR device having a refractive index higher than that of the sample. The infrared beam reaches the ATR element at the interface with the sample with an angle allowing the total reflection of the light. After one or more total internal reflections, the beam reaches the detector. At each reflection, an evanescent beam wave penetrates into the sample where it can be absorbed, producing an attenuation of the beam itself. The penetration of the evanescent wave depends on the beam wavelength, on the refractive indices of the ATR element and of the sample, as well as on the beam incident angle [20] (*see* **Note 3**). In the following, the procedures required to measure the ATR spectra of protein aggregates both in solution/suspension and as hydrated film will be reported. We will take as an example a biotinylated self-assembling peptide developed from the bone marrow homing peptide 1 (BMHP1) and designed as B24 peptide
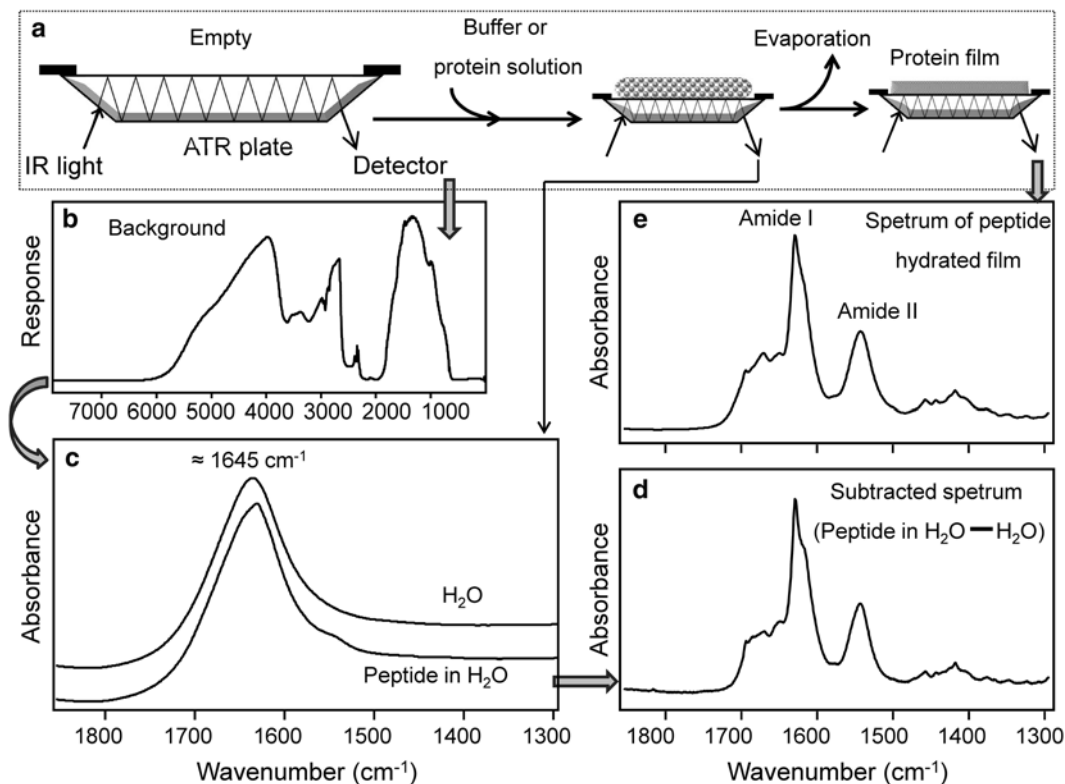
Fig. 1 Attenuated total reflection ATR/FTIR measurements on a peptide scaffold. (**a**) Scheme of a multireflection ATR device and of the procedure to collect the ATR absorption spectra of a protein/peptide solution and of a protein/peptide hydrated film. (**b**) Background spectrum of the nine-reflection DuraSamplIR II diamond device (Smith Detection, USA) coupled to the Varian 670-IR spectrometer (Varian Australia Pty Ltd, Mulgrave VIC, AU). (**c**) The absorption spectra of water (H$_2$O) and of the B24 peptide-assembled scaffold in H$_2$O [27, 28] are presented. (**d**) The absorption spectrum of the peptide-assembled scaffold in H$_2$O is reported after the subtraction of water contribution. (**e**) Absorption spectrum of the peptide-assembled scaffold measured in the form of hydrated film

(sequence: biotin-GGG-AFASTKT-CONH2) [27, 28]. This peptide has been found to self-assemble into a hydrogel scaffold able to sustain neural stem cell adhesion and proliferation [27].

The second example is given by the amyloid fibrils of ataxin-3 [19], a protein containing a stretch of consecutive glutamines, which is associated to the neurodegenerative disorder spinocerebellar ataxia type 3 when the protein poly-Q region is expanded beyond a critical threshold [30].

3.1.1 Peptide Scaffold Measured in Water

1. Set the temperature of the ATR plate to that required for the sample characterizations (37 °C). In this case, we employed the nine-reflection DuraSamplIR II device (Smith Detection, USA) with a diamond crystal.

2. Collect the background spectrum of the empty ATR plate (*see* **Note 4**) under the very same conditions that will be employed for the samples (MCT detector, 2 cm^{-1} spectral resolution, 25 kHz scan speed, triangular apodization, and 1,000 scan coadditions—*see* **Note 5**). Pay attention to the accurate purging of the spectrometer before starting measurements to reduce the interference of the water vapor (*see* **step 4** in Subheading 3.1.3). The background spectrum of the nine-reflection diamond ATR plate used here is reported in Fig. 1b. The shape of this single beam spectrum depends on the instrumental configuration (i.e., detector, beam splitter, IR source) and on the crystal employed. For instance, the low energy observed around 2,000 cm^{-1} in Fig. 1b is due to the strong absorption of infrared light by the diamond crystal.

3. Depose few microliters of the buffer (20 μL of water in this case) on the ATR crystal, and cover the plate to avoid solvent evaporation. Wait a few minutes to reach the desired temperature, and collect the buffer spectrum (labeled as H_2O in Fig. 1c).

4. Repeat the previous steps to collect the spectrum of the peptide solution. In this case, we deposed on the ATR plate 20 μL of the B24 peptide at a concentration of 1 % in water. This approach can allow to monitor the self-assembling process by collecting several ATR spectra at different times of incubation on the ATR plate. The spectrum of the assembled B24 scaffold is reported in Fig. 1c in the 1,850–1,300 cm^{-1} spectral range.

5. Subtract the buffer. As can be seen in Fig. 1c, the spectrum of B24 in water solution and that of the water solvent is very similar, and it is not possible to easily recognize the peptide amide I and amide II bands directly in the measured spectrum. Therefore, it is necessary to subtract the spectrum of the buffer (only H_2O in this case) to that of the peptide aggregates in order to obtain the infrared response of the peptide without the interference of the solvent. The subtracted spectrum is reported in Fig. 1d. It displays a maximum in the amide I region at ~1,629 cm^{-1}, assigned to the intermolecular β-sheet structures of the peptide assemblies (*see* Subheading 3.4.3 for the amide I band assignment to the polypeptide secondary structures).

We should stress that the subtraction of the aqueous buffer is a crucial point in the FTIR characterization of proteins and peptides because of the very high absorption of H_2O around 1,645 cm^{-1}, due to the water bending vibrations (*see* **Notes 6** and **7**). The buffer spectrum is subtracted from that of the protein solution/suspension adjusting iteratively the subtraction factor until a straight baseline is obtained in the region 2,300–1,750 cm^{-1}. Indeed, in this region, no polypeptide

absorption bands are observed, while only the water band due to the combination of the bending and librational modes is present. The accuracy of water subtraction can be improved also by examining a second subtraction range at 4,000–3,600 cm^{-1}, as well as the ratio of the amide I and amide II bands. It is important to know that in the correction procedure, the formation of negative bands due to buffer over-subtraction should be avoided. For a successful subtraction, it is mandatory to use the very same buffer composition, pH, and temperature of the sample and to collect its spectrum under the same instrumental conditions [14, 31–33].

6. Another crucial point for a reliable FTIR characterization of polypeptide secondary structures is the control of the interference of the water vapor, as it will be discussed in the **step 4** of Subheading 3.1.3.

7. *See* Subheading 3.4 for the spectral analyses.

3.1.2   Peptide Scaffold Measured in the Form of Hydrated Film

To overcome the difficulties associated to the subtraction of H$_2$O-based buffers, polypeptides can be measured in D$_2$O buffers (*see* Subheading 3.2) or in the form of hydrated film, as reported in the following:

1. After the background collection on the empty ATR plate (Fig. 1b), depose few microliters (3 μL in our case) of the B24 assembled scaffold in water on the ATR element (Fig. 1a).

2. Allow the formation of a protein hydrated film by rapid evaporation of the solvent using a gentle flow of nitrogen (Fig. 1a, *see* **Note 8**).

3. Collect the spectrum of the polypeptide film. The measured ATR/FTIR spectrum of B24 assembled scaffold in the form of film is reported in Fig. 1e. It appears almost superimposable to that measured in water suspension, without water evaporation, and obtained after the subtraction of the solvent spectrum. Indeed, it is desirable to perform control experiments to compare the infrared spectra of polypeptide in solution and in the form of film, since dehydration can affect in some cases the protein secondary structures [34].

4. *See* **step 4** of Subheading 3.1.3 for the correction of the possible interference of the water vapor and Subheading 3.4 for the spectral analyses.

3.1.3   Fibrils Measured in the Form of Hydrated Film

Fibrils of an ataxin-3 variant containing a pathological number of glutamines (55 in our case) in its poly-Q region have been obtained as previously described [19]. In particular, the purified protein was incubated at 37 °C for 1 week at a concentration of 1 mg/mL in PBS (25 mM potassium phosphate, pH 7.2, 0.15 M NaCl). Under these conditions, the spectra of the AT3-Q55 fibrils can be

measured using few microliters of the aggregate suspension by means of the approach here reported for B24 (*see* Subheading 3.1.2). Interestingly, the same method can be applied to measure the infrared spectrum of protein aggregates obtained at lower protein concentrations and/or in the presence of buffers with high absorption in the amide I region. Indeed, protein aggregates can be easily collected, for instance, by centrifugation, thanks to their insolubility. The pellet can be then transferred in a suitable buffer (such as PBS or only H_2O) at the appropriate aggregate concentration in order to obtain spectra with high signal-to-noise ratio.

1. After the background collection on the empty ATR plate, depose 3 μL of the pelleted AT3-Q55 fibrils on the ATR element. Here, we employed the Golden Gate ATR system (Specac Ltd, UK) with a single-reflection diamond crystal.

2. Allow the formation of the protein hydrated film by rapid evaporation of the solvent using a gentle flow of nitrogen.

3. Collect the spectrum of the fibrils. The ATR/FTIR spectrum of the AT3-Q55 fibrils is reported in Fig. 2a. It displays an amide I maximum at ~1,624 cm^{-1}, due to the formation of intermolecular β-sheet structures, characteristic of protein aggregates (*see* Subheading 3.4 for further discussion on band component identification and assignment).

4. Vapor subtraction. The infrared spectrum of water vapor in the 1,800–1,500 cm^{-1} spectral region is reported in Fig. 2a. It is characterized by several very sharp peaks that can strongly affect the analyses of the amide I and amide II bands of protein spectra. Therefore, it is crucial to efficiently reduce the environmental vapor in the FTIR spectrometer by an accurate purging of the instrument using a flow of dry air or nitrogen. Moreover, residual vapor interference can be corrected by subtracting the vapor spectrum (collected as described in **Note 9**) to that of the protein. To illustrate possible artifacts in the vapor correction procedure, we reported in Fig. 2a an ATR spectrum of AT3-Q55 fibrils not affected by the vapor interference (spectrum "1") as well as the ATR spectra of the same sample with positive or negative vapor peaks (spectrum "2" and "3," respectively). After spectral smoothing (Subheading 3.4.1), the three spectra appeared almost identical (Fig. 2b). However, only the second derivative (*see* Subheading 3.4.1 and **Note 10**) of spectrum "1" allowed a reliable identification of the amide I components (Fig. 2c). Indeed, the second derivatives of the spectra "2" and "3" clearly displayed additional components due to artifacts (Fig. 2c). The examples reported above illustrate the procedure that can be followed for the correction of the vapor interferences on the protein spectra. In particular, subtract the
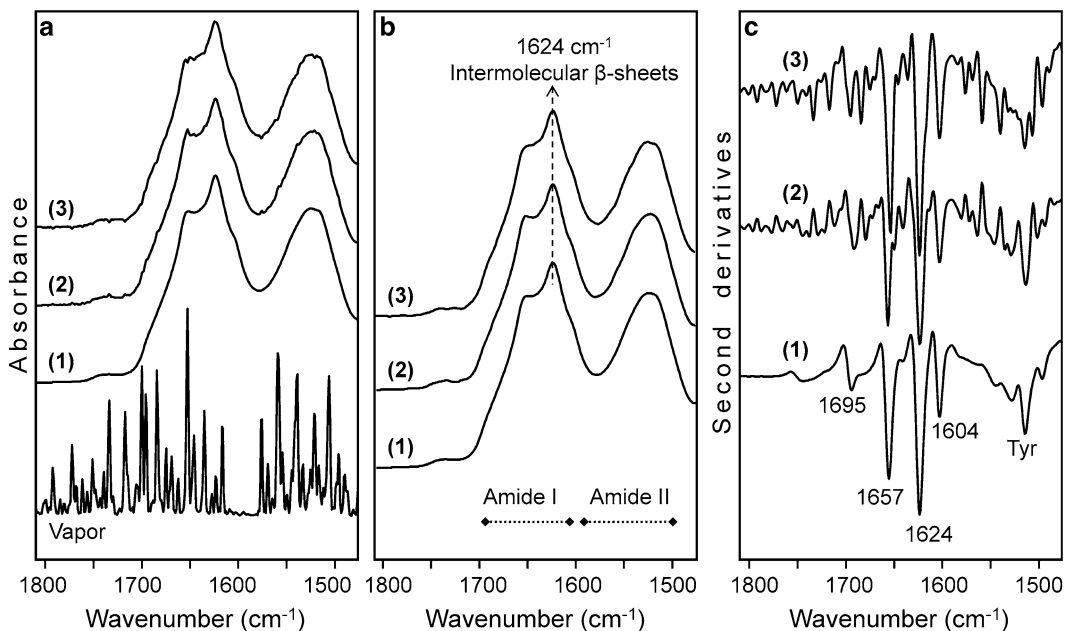
Fig. 2 ATR/FTIR measurements on ataxin-3 fibrils in the form of protein film. (**a**) The ATR/FTIR spectrum of AT3-Q55 fibrils not affected by the vapor interference (spectrum "1") and of the same sample with positive or negative vapor peaks (spectrum "2" and "3," respectively) are reported together with that of water vapor. (**b**) The spectra of (**a**) are given after binomial smoothing (11 points). (**c**) Second derivatives of the spectra reported in (**b**). The peak positions of the main amide I components are also indicated. The 1,624 cm^{-1} and 1,695 cm^{-1} peaks can be assigned to the intermolecular β-sheet structures of the AT3-Q55 fibrils. The 1,657 and 1,604 cm^{-1} peaks have been previously [19] assigned, respectively, to the C=O stretching and NH$_2$ bending modes of the glutamine side chains, as indicated by H/D exchange experiments. These Gln marker bands, not detectable in the freshly purified proteins and in the early aggregates, appeared and increased in intensity during the formations of SDS-insoluble fibrils. The reported results [19] demonstrated that these marker bands can be employed to study the rearrangement of glutamine side chains and the role of the resulting side chain hydrogen bonding in the formation of SDS-insoluble aggregated of poly-Q proteins and peptides

vapor spectrum to that of the protein adjusting the subtraction factor to fulfill the following conditions: (1) the absorption spectrum of the protein has to be smooth in the 1,800–1,500 cm^{-1} spectral region—pay particular attention on the vapor peak at ~1,653 cm^{-1}; (2) no vapor peaks should be present in the protein's second derivative spectrum within the region 1,820–1,720 cm^{-1}, where proteins do not absorb.

We should stress that ideally the IR measurements should be performed under very accurate purging in order to obtain a measured spectrum not requiring vapor corrections.

3.2 FTIR Measurements in the Transmission Mode: Thermal Aggregates in D₂O

Protein solution/suspension can be measured in the transmission mode using a transmission cell made by two windows that have to be transparent to the IR light and insoluble in water, such as BaF$_2$ and CaF$_2$ windows. The cell path-length is given by a spacer (usually in Teflon) between the two windows (Fig. 3a). The high
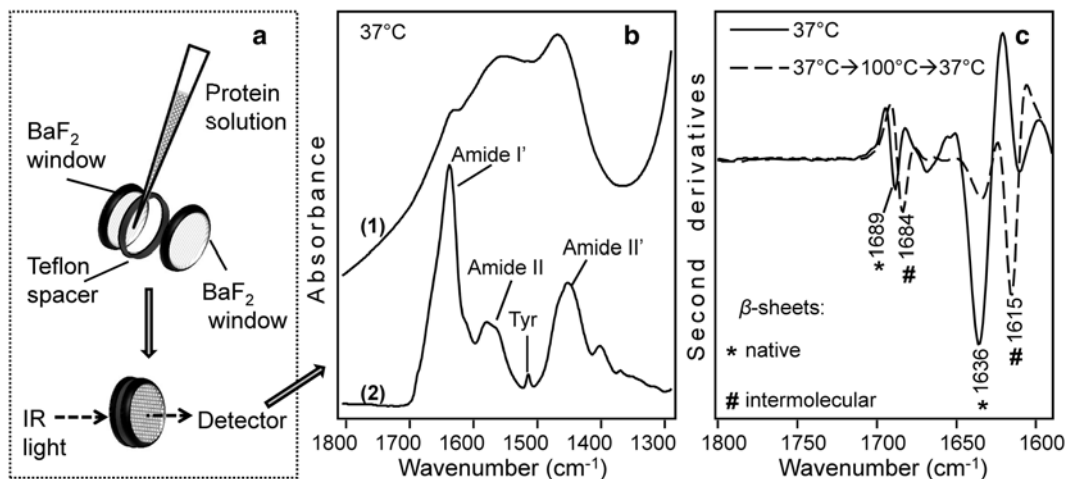
Fig. 3 FTIR measurements in the transmission mode of β-2 microglobulin in D_2O. (**a**) For FTIR measurements in the transmission mode, the sample solution is placed between two IR windows. The path-length is determined by the thickness of the Teflon spacer. (**b**) The absorption spectrum of β-2 microglobulin at 37 °C (in deuterated phosphate buffer), measured in the transmission mode, is reported before (1) and after (2) the subtraction of the buffer absorption. (**c**) Second derivative spectra of β-2 microglobulin, in deuterated phosphate buffer at 37 °C, are reported before and after the thermal treatment at 100 °C. The assignment of the main amide I components is also indicated [23]

absorption of H_2O limits the path-length of the cell to 6–15 μm and imposes a protein concentration typically higher than 5–10 mg/mL (*see* also Subheading 3.1.1 and **Notes 2**, **6**, and **7**). These limitations can be overcome by measurements on protein hydrated films or in D_2O solution. Indeed, D_2O displays a low absorption in the 1,700–1,600 cm^{-1} spectral range (*see* **Note 11**), allowing measurements at higher path-lengths (50–150 μm) and lower concentrations (down to 1–5 mg/mL) compared to those in H_2O.

In the example described below, we reported an FTIR study on the thermal aggregation of the protein β-2 microglobulin in a D_2O solution (protein at 2.5 mg/mL concentration in deuterated 50 mM phosphate buffer, 100 mM NaCl, pH 7.4) [23].

1. Set the temperature of the empty transmission cell to 37 °C, and collect the background spectrum under the same instrumental conditions that will be employed for the sample: nitrogen-cooled mercury cadmium telluride detector, 2 cm^{-1} spectral resolution, 25 kHz scan speed, 1,000 scan coadditions, and triangular apodization.

2. Load 20 μL of the wild-type β-2 microglobulin in deuterated phosphate buffer in the transmission cell made by two BaF$_2$ windows (*see* **Note 12**) separated by a 150 μm Teflon spacer (Fig. 3a).

3. Wait 10–20 min to reach the thermal equilibrium and to allow the purging system to remove the water vapor from the sample compartment.

4. Collect the infrared spectrum of the sample at 37 °C (Fig. 3b).

5. Heat the sample from 37 to 100 °C at a rate of 0.4 °C/min while collecting consecutive spectra. Under the above conditions, each transmission spectrum is collected every 3.6 °C. In this way, it is possible to monitor the structural changes induced by the thermal treatment and, therefore, to study the thermal stability and aggregation of the investigated proteins [17, 18, 23].

6. Cool down the sample temperature to 37 °C, and collect again the infrared spectrum to evaluate the possible reversibility of the process (Fig. 3c).

7. In an equivalent experiment, repeat the above steps to collect the infrared spectra of the buffer under the same conditions employed for the protein solution.

8. Subtract to each protein solution spectrum that of the corresponding buffer collected at the same temperature. To this aim, adjust iteratively the subtraction factor until a straight baseline in the region 2,300–1,750 cm^{-1} is obtained (*see* **Note 11**). The infrared spectrum of β-2 microglobulin at 37 °C is reported in Fig. 3b before and after buffer subtraction. In particular, the subtracted spectrum is characterized in the 1,700–1,400 cm^{-1} spectral region by the amide I′, the amide II, and the amide II′ bands (*see* **Note 13**). The second derivative spectrum of the β-2 microglobulin thermal aggregates is reported in Fig. 3c. It displays two negative peaks (*see* Subheading 3.4.1 and **Note 10**) at 1,615 and 1,684 cm^{-1} due to the formation of intermolecular β-sheets, typical of protein aggregates (*see* Subheading 3.4.3). On the contrary, the second derivative spectrum of native β-2 microglobulin displays two negative peaks at 1,636 and 1,689 cm^{-1} assigned to the intramolecular β-sheet structures [23] (*see* Subheading 3.4.3).

9. *See* **step 4** of Subheading 3.1.3 for the correction of the possible interference of the water vapor and Subheading 3.4 for the spectral analyses.

3.3 FTIR Microspectroscopy of Inclusion Bodies (IBs)

FTIR microspectroscopy, obtained by coupling an IR microscope to a spectrometer, allows to collect the IR spectra of selected sample areas with a spatial resolution of a few tens of microns. This technique has been widely applied for the study of intact complex biological systems and of protein aggregation in situ [13, 21, 35–37]. In this regard, a method for the micro-FTIR measurements of intact cells overexpressing a recombinant protein in the form of IBs has been proposed [38–40] and recently discussed by Ami,

Natalello, and Doglia [13]. In the following, we will describe a method for the micro-FTIR measurements in the transmission mode of extracted IBs [39–41]. In particular, an ataxin-3 variant containing a pathological numbers of glutamines (55 Glns) in its poly-Q region was overexpressed in *E. coli* in the form of IBs, and the protein aggregates were extracted from the cells at different times after induction [29]. We will report here in detail the procedures required to obtain reliable micro-FTIR spectra. Concerning the expression of the protein of interest and the extraction of its aggregates from the cells, it is suggested to follow the available protocols already optimized for the specific protein and employed host cells.

1. Collect the extracted protein aggregates by centrifugation and resuspend the pellet in water. This washing should be repeated at least three times to remove traces of the extraction buffers and detergents.

2. Resuspend the washed aggregates in water at few different concentrations, and depose drops of about 1–5 μL (*see* **Note 14**) on the infrared transparent support (a BaF$_2$ window in our case).

3. Allow the evaporation of the solvent in order to obtain protein films on the window (Fig. 4a). We generally let dry the sample at room temperature (RT), for at least 30 min, in a laminar flow hood.

4. Place the sample window on the microscope stage; pull down the objective plastic cylinder to reduce the environmental vapor interference; focus the beam on the sample, and set the microscope diaphragm aperture; select an empty area of the window, and optimize the IR beam focus and intensity by adjusting the Cassegrain condenser and the spectrometer parameters; collect the background spectrum under the same conditions that will be employed for the sample: in this case, nitrogen-cooled mercury cadmium telluride detector (narrow band, 250 μm), 2 cm^{-1} spectral resolution, 25 kHz scan speed, 512 scan coadditions, and triangular apodization.

5. Keeping constant the microscope diaphragm aperture, select the protein film area without cracks, obtained by the above procedure, and collect the FTIR spectra on different areas of the same film. Repeat the measurements on several protein films to evaluate possible sample heterogeneity. Exclude from the data analyses those spectra with very low or too high absorption to avoid data with low signal-to-noise ratios or artifacts (*see* **Note 14**). The infrared absorption spectra of the extracted aggregates obtained in vivo at different times after the induction of the AT3-Q55 protein are reported in Fig. 4.

6. *See* **step 4** of Subheading 3.1.3 for the correction of the possible interference of the water vapor and Subheading 3.4 for the spectral analyses.
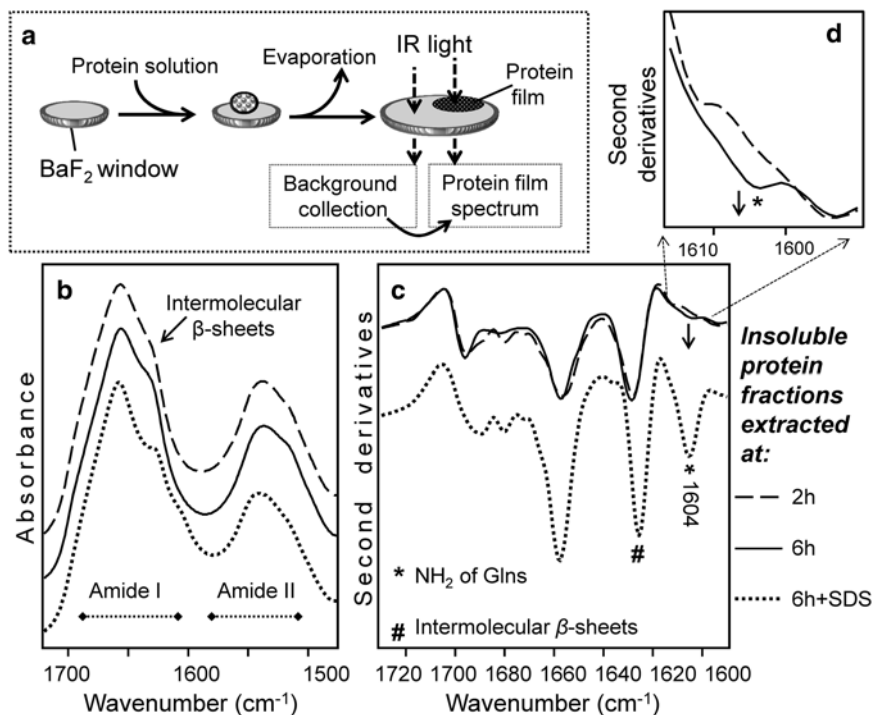
Fig. 4 Micro-FTIR characterization of extracted bacterial inclusion bodies. (**a**) Scheme of the procedure employed for micro-FTIR measurements of IBs (*see* text). (**b**) The FTIR absorption spectra of the insoluble protein fractions extracted at 2 and 6 h after induction of the AT3-Q55 expression strains are reported. The FTIR spectrum of the insoluble protein fraction extracted at 6 h is also given after SDS treatment. (**c**) Second derivatives of the absorption spectra reported in (**b**). (**d**) The second derivatives as in (**c**) are reported in an enlarged scale in the spectral region of NH_2 bending modes of glutamines. Second derivative spectra are normalized at the ~1,515 cm^{-1} peak of tyrosines (*see* **Note 15**). The appearance of the 1,604 cm^{-1} component at longer incubation time and the formation of SDS-insoluble aggregate indicate that the structural ordering of the glutamine side chains, involved in a network of hydrogen bonds, takes place also in vivo [29], in agreement with the results [19] obtained in vitro (Fig. 2)

3.4 Spectral Analyses and Structural Information

The amide I band in a protein FTIR spectrum consists of the different secondary structure components of the polypeptide backbone. However, these components usually cannot be directly resolved in the measured spectra because of their inherent widths and close spectral proximity, which lead to their overlapping. Therefore, after buffer and vapor subtraction, the crucial step for the FTIR structural characterization of proteins is the identification of the amide I band components and their assignment to the different protein secondary structures.

The band components that overlap in the amide I region can be resolved following different approaches: the most employed are the second derivative of the absorption spectrum [42] and the Fourier self-deconvolution (FSD) [43, 44].

3.4.1 Second Derivative of the Absorption Spectrum

The mathematical calculation of the second derivative of the protein absorption spectrum allows to resolve the overlapped components of the amide I band, since it reduces the band half-widths of about a factor 2.7. We should recall that the minima in the second derivative spectra correspond to the maxima in the absorption spectra. The peak intensities in the second derivative are directly proportional to those of the absorption components, but inversely proportional to the square of the absorption band half-widths [42]. It is therefore important to note that this inverse correlation leads to a strong enhancement of the relative contribution of sharp peaks in the second derivative spectrum, such as noise, interference fringes, and residual vapor (Fig. 2). This effect strongly impairs the second derivative analysis of the absorption spectra. Therefore, before computing the second derivative, the absorption spectra should be corrected for the vapor interference and smoothed. We usually apply an 11-point binomial smoothing of the absorption spectrum followed by second derivative calculation using the Savitzky-Golay algorithm (third grade polynomial, five smoothing points). When comparing spectra of the same sample obtained under different conditions, it is very important to apply the same number of smoothing points to the raw spectra and then to calculate the second derivatives. In this way, it is possible to correlate the spectral changes observed in the second derivatives to conformational changes of the protein (*see* **Note 15**).

The second derivative spectra of AT3-Q55 fibrils, β-2 microglobulin (native and thermal aggregates), and AT3-Q55 IBs are reported in Figs. 2c, 3c, and 4c, d, respectively.

It is possible to evaluate the significance of the protein second derivative analyses by inspecting the region between 1,750 and 1,850 cm^{-1}, where it is easy to check the noise level and the vapor interference in the spectrum (Figs. 2c and 3c).

3.4.2 Fourier Self-Deconvolution

FSD is a resolution enhancement approach in which a narrowing of the bands is obtained by decreasing the decay rate of the corresponding components in the Fourier time domain. Therefore, the measured spectrum is subjected to a Fourier reverse transformation, followed by a multiplication with an exponentially increasing weighting function and by an appropriate apodization. The Fourier transform of this new interferogram leads to the deconvoluted spectrum with an enhanced spectral resolution [43, 44]. To perform the FSD of an absorption spectrum, the following parameters are required: the peak width narrowing factor (K-factor, its value should be lower than 3 and depends on the signal-to-noise ratio of the measured spectrum), the bandwidth at half-height of the components (values between 6.5 and 18 cm^{-1} are typically employed for the deconvolution of the amide I band), and the apodization function. The spectral region selected for the deconvolution should include a portion of the baseline, such as the 1,750–1,850 cm^{-1}

region, where the noise level can be monitored. To evaluate the reliability of the deconvolution, the number and peak position of the components resolved by FSD should be the same as those obtained by the second derivative analysis.

3.4.3 Amide I Band Assignment

The different band components identified in the amide I spectral region by FSD and/or second derivative analyses can be assigned to the protein secondary structures, thanks to the experimental FTIR investigations and computational studies on model compounds, polypeptides, and proteins. Indeed, the absorption of the different protein secondary structures and of intermolecular β-sheets occurs at specific spectral regions (Fig. 5) [12, 14–16, 24–26, 43, 45]. However, their partial overlapping makes the band assignment sometimes difficult and arbitrary in the absence of an appropriate validation that can be obtained, for instance, by additional FTIR studies, such as hydrogen/deuterium exchange, thermal and chemical denaturation experiments, and complementary circular dichroism measurements.

The absorption of the main protein secondary structures will be briefly described below and schematized in Fig. 5.

Intramolecular and Intermolecular β-Sheet Structures

In the amide I band, intramolecular β-sheet secondary structures typically display two components of different intensities around $1,633$ cm^{-1} (the most intense; extremes $1,640$–$1,623$ cm^{-1}) and $\sim$$1,686$ cm^{-1} (extremes $1,695$–$1,674$ cm^{-1}), when measurements are performed in H$_2$O. In D$_2$O, both bands are downshifted to $\sim$$1,630$ cm^{-1} (extremes $1,638$–$1,615$ cm^{-1}) and to $\sim$$1,679$ cm^{-1} (extremes $1,694$–$1,672$ cm^{-1}), respectively.
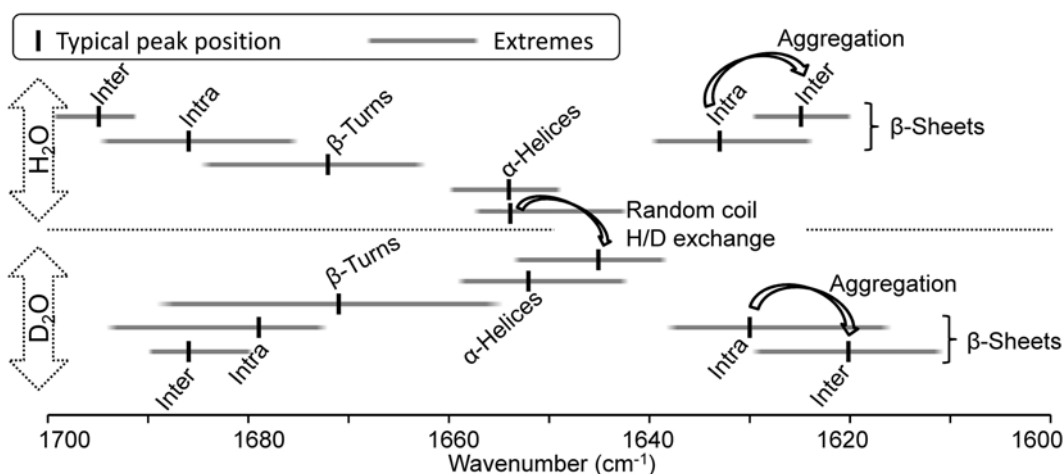
Fig. 5 Amide I band assignment. The typical band positions and spectral ranges of the principal protein secondary structures and of intermolecular β-sheets, characteristic of aggregates, are reported. The interval extremes and positions vary for measurements in H$_2$O- and in D$_2$O-based buffers

Protein aggregation is typically associated with the formation of intermolecular β-sheet structures, as, for instance, observed for amyloids, thermal aggregates, and IBs. These structures display similar absorption features of the intramolecular β-sheets but shifted in peak position (Fig. 5). In particular, intermolecular β-sheets are characterized by two components at ~1,625 cm⁻¹ (extremes 1,630–1,620 cm⁻¹) and at ~1,695 cm⁻¹ (extremes 1,698–1,692 cm⁻¹) in H₂O; at ~1,620 cm⁻¹ (extremes 1,630–1,611 cm⁻¹) and at ~1,686 cm⁻¹ (extremes 1,690–1,680 cm⁻¹) in D₂O.

The typical shift of the native β-sheet bands observed after protein aggregation can be appreciated in the example reported in Fig. 3c. Indeed, native β-2 microglobulin at 37 °C displays two components at ~1,689 and ~1,636 cm⁻¹ assigned to the intramolecular β-sheet structures of the protein, while its thermal aggregates are characterized by two peaks at ~1,684 and ~1,615 cm⁻¹, due to the formation of intermolecular β-sheets [23].

Usually, it is very difficult to discriminate between parallel and antiparallel β-sheets [24, 46, 47] since they are characterized by a similar infrared spectrum. However, recent works suggested that it is possible to assign the IR response to parallel intermolecular β-sheets when the low wavenumber component is present in the spectrum without the high wavenumber β-sheet band [34, 48–51]. In this way, it has been possible to differentiate oligomers from fibrils of different polypeptides by their IR response [51], such as in the case of the prion peptide PrP82–146 [34]. In the above examples, while the IR spectra of oligomers were characterized by two β-sheet components, fibrils displayed only the low-frequency band, suggesting that the spectral features of oligomers are the antiparallel β-sheets, while those of fibrils are the parallel β-sheets—*see* for a recent review [51].

A systematic comparison between the IR response of β-sheets in native proteins and in amyloid fibrils has been reported in ref. 16. This work showed a downshift and narrowing of the low-frequency β-sheet band when going from native to amyloid structures. Finally, we should note that the band positions and intensities of the IR β-sheet bands critically depend on the geometry of the structure, such as the strand twist angle, the number of the strands per sheet, and the H-bond strength [12, 16, 23, 52]. For instance, an upshift of the low-frequency band has been observed in folding intermediates and misfolded proteins, indicating the weakening of the H-bonds in looser β-sheet structures [53, 54]. On the other hand, a downshift of the band can indicate tighter β-sheet structures [12, 23, 24, 26], as observed, for instance, in the comparison of native β-2 microglobulin in solution and in the crystalline state [23].

α-Helices   The IR absorption of α-helices occurs at around 1,654 cm⁻¹ (extremes 1,660–1,648 cm⁻¹) in H₂O and downshifts to a few wavenumbers in D₂O. Their peak position is related to the helical

structural properties. For instance, long and rigid helices absorb at lower wavenumbers compared to short and flexible ones [43, 55]. Moreover, the formation of H-bonds with the solvent induces an important downshift (1,640–1,630 cm^{-1} in D$_2$O), as observed for hydrated α-helices in proteins [56] and peptides [57, 58]. α-Helical structures have been also observed by FTIR in protein aggregates formed in bacteria and related to the presence of native-like structures in the proteins embedded in IBs [21].

Random Coil and Other Secondary Structures

The absorption of random coil structures occurs around 1,654 cm^{-1} (extremes 1,657–1,642 cm^{-1}) in H$_2$O and, therefore, in the same spectral region of α-helices. However, it is possible to discriminate among the two structures performing the FTIR analyses both in H$_2$O- and in D$_2$O-based buffers, since random coil structures will undergo a much higher H/D exchange than α-helices, downshifting to 1,645 cm^{-1} in D$_2$O (extremes 1,654–1,639 cm^{-1}) [43, 45]. Moreover, random coils typically display a large bandwidth, which could hide their contributions in the second derivative spectra dominated instead by sharper bands due to the other protein secondary structures [14].

Among the protein secondary structures, the absorption of open loops occurs around 1,643 cm^{-1} both in H$_2$O and D$_2$O [59] and that of β-turns in the spectral regions 1,686–1,662 cm^{-1} in H$_2$O and 1,691–1,653 cm^{-1} in D$_2$O [45].

Amino Acid Side Chains

The absorption of the amino acid side chains occurs in the wide mid-IR range, as reported in the comprehensive review of Barth [60]. Here, we will mention only the tyrosine peak at 1,518–1,515 cm^{-1} in H$_2$O (at 1,517–1,513 in D$_2$O; Figs. 2 and 3) and the glutamine peaks occurring in the amide I spectral range. The C=O stretching of Gln side chains falls around 1,687–1,668 in H$_2$O and around 1,654–1,635 cm^{-1} in D$_2$O (*see* **Note 16**), while the Gln NH$_2$ bending absorption occurs at 1,610–1,586 cm^{-1} in H$_2$O (~1,163 cm^{-1} in D$_2$O). Moreover, the C=O stretching band is very sensitive to H-bonding (lower peak position corresponds to stronger H-bond). Interestingly, these spectral features have been exploited to study the role of glutamine side chain hydrogen bonding in the aggregation process of the poly-Q protein ataxin-3, in vitro [19] as well as during the formation of IBs in bacterial cells [29] (Figs. 2 and 4).

3.4.4 Structural Information

The procedure to obtain structural information on the investigated protein employing FTIR spectroscopy can be schematized as follows: (1) sample preparation and choice of the FTIR approach (i.e., ATR or transmission mode, solution or protein film, H$_2$O or D$_2$O); (2) collection of the FTIR spectra of the protein samples, buffers, and vapor; (3) buffer subtraction followed by vapor subtraction, when necessary; (4) identification of the amide I

components, through the second derivative and/or FSD analyses of the subtracted spectrum; and (5) band assignment to the protein secondary structures and to the intermolecular β-sheets, typical of protein aggregates.

After band assignment, it is possible to follow the relative variation of the protein secondary structures monitoring changes in peak intensities and peak positions in the second derivative spectra collected during the sample exposure to different conditions (*see* **Note 15**). This approach allows, therefore, to study protein folding, unfolding, misfolding, and aggregation induced by thermal and chemical treatments, interactions with partners, pH changes, and other environmental conditions. For instance, in this way it will be possible to compare the native protein structures with those of its aggregates obtained in vitro or in vivo by the direct comparisons of their second derivative FTIR spectra.

Moreover, the percentage of each protein secondary structure can be quantitatively evaluated by several methods [61–63], such as the curve fitting of the amide I band in the absorption spectrum. The curve fitting analysis can be also performed on the Fourier self-deconvoluted spectrum and on the inverted second derivative spectrum. Since the different amide I components are more resolved in these spectra, than in the measured absorption spectrum, a more efficient and stable fitting can be obtained. However, the distortion of the spectral shape induced by these resolution enhancement procedures raises the question on the reliability of the fitting results.

The critical point of this analysis is the selection of the input parameters. For the curve fitting of the protein absorption spectrum, the following procedure can be used [17, 63]:

1. Input data and parameters: (a) the absorption spectrum to be fitted (already corrected for the buffer and vapor absorptions), (b) baseline and spectral region (linear baseline and amide I band), (c) number and peak positions of the amide I components (detected by second derivative and/or FSD analyses) that will be approximated by a specific function (usually Gaussian, Lorentzian, or mixed Gaussian-Lorentzian), (d) band width of the components (they can be evaluated from the second derivative and/or FSD analyses), and (e) band heights of the components (they can be set at 90 % of that of the absorption spectrum for the main components and at 70 % for the other components).

2. Perform the iterative fitting of the spectrum by fixing the peak positions and letting free the other parameters.

3. Perform a second fitting using the results of the first one as input and letting free also the peak positions in order to find the final set of the best fitting functions.

4. The fractional area of each component of the best fitting functions over the total component area will represent the percentage of the corresponding secondary structures.

4 Notes

1. The pH values measured on the pH meter can be converted in pD values by adding +0.4 [64].

2. Here, we reported an indication on the protein concentrations required for FTIR characterizations. For measurement of protein solution in water: typically 5–10 mg/mL or higher (5–20 μL of volume, 6–15 μm of path-length). For protein solution in heavy water: typically 1–5 mg/mL or higher (10–20 μL of volume, 50–150 μm of path-length). For measurements of proteins in the form of hydrated films: typically 1 mg/mL (1–15 μL of volume, significant lower concentrations can be employed in absence of buffer salt interference).

3. The different penetration depth of the evanescent wave at varying wavenumbers through the spectrum will affect the relative intensities of the different band components of the spectrum. When necessary, this effect can be compensated using an ATR correction function [20].

4. The ATR element of the device should be very accurately cleaned before and after each measurement, since polypeptides and other sample components have a high propensity to be adsorbed on the surface. At the end of the ATR/FTIR characterizations, check the cleaning progression collecting the spectrum of the empty ATR plate using the initial background to get the absorption spectrum.

5. The number of scan coadditions of the background spectrum should be the same or higher than that of the sample spectrum.

6. Another very strong H_2O infrared band occurs around 3,400 cm^{-1}, and it has three components due to the symmetrical and asymmetrical stretching vibrations (at ~3,450 and ~3,600 cm^{-1}) and to the overtone of the bending vibrations at ~3,250 cm^{-1}. The water libration band occurs below 800 cm^{-1} [32].

7. The high H_2O absorption leads to a strong limitation in the path-length of the infrared cell when performing measurements in the transmission mode of protein water solution (typically path-lengths of 6–15 μm are employed). Under these conditions, protein concentrations higher than 5–10 mg/mL are required.

8. If salt crystals are formed on the ATR plate after solvent evaporation, they can interfere with the infrared response. In this case, rinse gently the protein on the ATR plate with deionized water [14, 50].

9. The absorption spectrum of water vapor can be obtained as follows: collect the background spectrum of the empty infrared support (namely, the ATR crystal, the BaF$_2$ windows, etc.) after an accurate purging of the spectrometer and of the sample compartment, reduce the flow rate of dry air or nitrogen of the purging system (in some cases, it is possible to open the sample comportment, such as in the IR microscope where a cylinder around the objective is usually pulled down to protect the sample from the environmental water vapor), and collect the spectrum when the water vapor bands reach a satisfactory intensity (Fig. 2a).

10. The absorption maxima of the absorption spectrum appear as negative peaks, at the same wavenumbers, in the second derivative spectrum.

11. The bending vibrations of D$_2$O occur around 1,209 cm^{-1}. Other D$_2$O infrared bands occur around 1,555 cm^{-1} (combination); a complex band at 2,500 cm^{-1} (it displays more components due to the symmetrical and asymmetrical stretching vibrations and bending overtone); ~3,404 and ~1,462 cm^{-1}, respectively, due to stretching and bending vibrations of HOD from traces of H$_2$O in D$_2$O [32]. These bands are useful to perform an accurate buffer subtraction from the spectra of proteins in D$_2$O solution.

12. We should note that the polishing of the IR windows is important for the success of the measurements, for instance, reducing bubble formation. A diamond (size 1–9 μm) or aluminum oxide (using methanol as solvent) polishing paste can be employed.

13. The amide II band (1,600–1,500 cm^{-1}) is mainly due to the backbone NH bending and CN stretching, with also the contribution of other vibrations. During H/D exchange, the amide II band strongly decreases, and a new deuterated band (called amide II′) appears around 1,490–1,460 cm^{-1}. The high sensitivity of this band to H/D exchange has been extensively exploited to monitor the kinetics of the process and to obtain new insight on the protein conformational dynamics [20, 65].

14. The protein film should be thin (to avoid excessive absorption) and without cracks that can lead to baseline and spectral feature alterations. The deposition on the IR window of the sample drops of different volumes and protein concentrations can help to obtain a uniform film of the proper thickness.

15. To compare absorption or second derivative spectra of the same protein measured at different concentrations or under different conditions, it is often required to normalize the spectra to compensate for possible differences in protein content. To this aim, it is advisable to normalize the absorption spectra at the amide I band area before performing the second derivative calculation. Alternatively, the second derivative spectra can be normalized at the well-resolved tyrosine peak around 1,515 cm^{-1}.

16. This downshift is larger than that expected for the protein secondary structures helping to discriminate between side chain and backbone absorptions [19, 60].

Acknowledgment

References

1. Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3:R9–R23

2. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366

3. Adamcik J, Mezzenga R (2012) Proteins fibrils from a polymer physics perspective. Macro molecules 45:1137–1150

4. Cinar G, Ceylan H, Urel M et al (2012) Amyloid inspired self-assembled peptide nanofibers. Biomacromolecules 13:3377–3387

5. Jonker AM, Lowik DWPM, van Hest JCM (2012) Peptide- and protein-based hydrogels. Chem Mater 24:759–773

6. Mezzenga R, Schurtenberger P, Burbidge A et al (2005) Understanding foods as soft materials. Nat Mater 4:729–740

7. Garcia-Fruitos E, Vazquez E, Diez-Gil C et al (2012) Bacterial inclusion bodies: making gold from waste. Trends Biotechnol 30:65–70

8. Gatti-Lafranconi P, Natalello A, Ami D et al (2011) Concepts and tools to exploit the potential of bacterial inclusion bodies in pro-

tein science and biotechnology. FEBS J 278:2408–2418

9. Garcia-Fruitos E, Seras-Franzoso J, Vazquez E et al (2010) Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. Nanotechnology 21:205101

10. de Groot NS, Sabate R, Ventura S (2009) Amyloids in bacterial inclusion bodies. Trends Biochem Sci 34:408–416

11. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27

12. Seshadri S, Khurana R, Fink AL (1999) Fourier transform infrared spectroscopy in analysis of protein deposits. Methods Enzymol 309:559–576

13. Ami D, Natalello A, Doglia SM (2012) Fourier transform infrared microspectroscopy of complex biological systems: from intact cells to whole organisms. In: Uversky VN, Dunker AK (eds) Intrinsically disordered protein analysis: volume 1, vol 895, Methods and experimental

tools. Methods in molecular biology. Humana, New York, pp 85–100. doi:10.1007/978-1-61779-927-3_7

14. Natalello A, Ami D, Doglia S (2012) Fourier transform infrared spectroscopy of intrinsically disordered proteins: measurement procedures and data analyses. In: Uversky VN, Dunker AK (eds) Intrinsically disordered protein analysis: volume 1, methods and experimental tools. Methods in molecular biology, vol 895. Humana, New York, pp 229–244. doi:10.1007/978-1-61779-927-3_16

15. Barth A (2007) Infrared spectroscopy of proteins. Biochim Biophys Acta 1767:1073–1101

16. Zandomeneghi G, Krebs MRH, McCammon MG et al (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. Protein Sci 13:3314–3321

17. Natalello A, Ami D, Brocca S et al (2005) Secondary structure, conformational stability and glycosylation of a recombinant Candida rugosa lipase studied by Fourier-transform infrared spectroscopy. Biochem J 385:511–517

18. Natalello A, Doglia SM, Carey J et al (2007) Role of flavin mononucleotide in the thermostability and oligomerization of Escherichia coli stress-defense protein WrbA. Biochemistry 46:543–553

19. Natalello A, Frana AM, Relini A et al (2011) A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation. PLoS One 6:e18789

20. Goormaghtigh E, Raussens V, Ruysschaert JM (1999) Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes. Biochim Biophys Acta 1422:105–185

21. Doglia SM, Ami D, Natalello A et al (2008) Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. Biotechnol J 3:193–201

22. Carpenter JF, Prestrelski SJ, Dong A (1998) Application of infrared spectroscopy to development of stable lyophilized protein formulations. Eur J Pharm Biopharm 45:231–238

23. Ami D, Ricagno S, Bolognesi M et al (2012) Structure, stability, and aggregation of β-2 microglobulin mutants: insights from a Fourier transform infrared study in solution and in the crystalline state. Biophys J 102:1676–1684

24. Barth A, Zscherp C (2002) What vibrations tell us about proteins. Q Rev Biophys 35:369–430

25. Arrondo JLR, Goni FM (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. Prog Biophys Mol Biol 72:367–405

26. Tamm LK, Tatulian SA (1997) Infrared spectroscopy of proteins and peptides in lipid bilayers. Q Rev Biophys 30:365–429

27. Gelain F, Silva D, Caprini A et al (2011) BMHP1-derived self-assembling peptides: hierarchically assembled structures with self-healing propensity and potential for tissue engineering applications. ACS Nano 5:1845–1859

28. Silva D, Natalello A, Sanii B et al (2013) Synthesis and characterization of designed BMHP1-derived self-assembling peptides for tissue engineering applications. Nanoscale 5:704–718

29. Invernizzi G, Aprile FA, Natalello A et al (2012) The relationship between aggregation and toxicity of polyglutamine-containing ataxin-3 in the intracellular environment of Escherichia coli. PLoS One 7:e51890

30. Costa MC, Paulson HL (2012) Toward understanding Machado–Joseph disease. Prog Neurobiol 97:239–257

31. Rahmelow K, Hubner W (1997) Infrared spectroscopy in aqueous solution: difficulties and accuracy of water subtraction. Appl Spectrosc 51:160–170

32. Venyaminov SY, Prendergast FG (1997) Water (H_2O and D_2O) molar absorptivity in the 1000–4000 cm^{-1} range and quantitative infrared spectroscopy of aqueous solutions. Anal Biochem 248:234–245

33. Dong A, Huang P, Caughey WS (1990) Protein secondary structures in water from 2nd-derivative amide-I infrared-spectra. Biochemistry 29:3303–3308

34. Natalello A, Prokorov VV, Tagliavini F et al (2008) Conformational plasticity of the Gerstmann-Straussler-Scheinker disease peptide as indicated by its multiple aggregation pathways. J Mol Biol 381:1349–1361

35. Ami D, Natalello A, Lotti M et al (2013) Why and how protein aggregation has to be studied in vivo. Microb Cell Fact 12:17

36. Choo LP, Wetzel DL, Halliday WC et al (1996) In situ characterization of beta-amyloid in Alzheimer's diseased tissue by synchrotron Fourier transform infrared microspectroscopy. Biophys J 71:1672–1679

37. Diomede L, Cassata G, Fiordaliso F et al (2010) Tetracycline and its analogues protect Caenorhabditis elegans from beta amyloid-induced toxicity by targeting oligomers. Neurobiol Dis 40:424–431

38. Ami D, Bonecchi L, Cali S et al (2003) FT-IR study of heterologous protein expression in recombinant Escherichia coli strains. Biochim Biophys Acta 1624:6–10

39. Ami D, Natalello A, Gatti-Lafranconi P et al (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett 579:3433–3436

40. Ami D, Natalello A, Taylor G et al (2006) Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. Biochim Biophys Acta 1764:793–799

41. Gonzalez-Montalban N, Natalello A, Garcia-Fruitos E et al (2008) In situ protein folding and activation in bacterial inclusion bodies. Biotechnol Bioeng 100:797–802

42. Susi H, Byler DM (1986) Resolution-enhanced Fourier-transform infrared-spectroscopy of enzymes. Methods Enzymol 130:290–311

43. Arrondo JLR, Muga A, Castresana J et al (1993) Quantitative studies of the structure of proteins in solution by Fourier-transform infrared-spectroscopy. Prog Biophys Mol Biol 59:23–25

44. Kauppinen JK, Moffatt DJ, Mantsch HH et al (1981) Fourier self-deconvolution – a method for resolving intrinsically overlapped bands. Appl Spectrosc 35:271–276

45. Goormaghtigh E, Cabiaux V, Ruysschaert JM (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. III. Secondary structures. Sub-cell Biochem 23:405–450

46. Khurana R, Fink AL (2000) Do parallel beta-helix proteins have a unique Fourier transform infrared spectrum? Biophys J 78:994–1000

47. Susi H, Byler DM (1987) Fourier-transform infrared study of proteins with parallel beta-chains. Arch Biochem Biophys 258:465–469

48. Chitnumsub P, Fiori WR, Lashuel HA et al (1999) The nucleation of monomeric parallel beta-sheet-like structures and their self-assembly in aqueous solution. Bioorg Med Chem 7:39–59

49. Yamada N, Ariga K, Naito M et al (1998) Regulation of b-sheet structures within amyloid-like b-sheet assemblage from tripeptide derivatives. J Am Chem Soc 120:12192–12199

50. Cerf E, Sarroukh R, Tamamizu-Kato S et al (2009) Antiparallel β-sheet: a signature structure of the oligomeric amyloid β-peptide. Biochem J 421:415–423

51. Sarroukh R, Goormaghtigh E, Ruysschaert JM et al (2013) ATR-FTIR: a "rejuvenated" tool to investigate amyloid proteins. Biochim Biophys Acta 1828:2328–2338

52. Torii H, Tatsumi T, Tasumi M (1998) Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and reso-nance raman intensities of N-methylacetamide. J Raman Spectrosc 29:537–546

53. Natalello A, Mattoo RUH, Priya S et al (2013) Biophysical characterization of two different stable misfolded monomeric polypeptides that are chaperone-amenable substrates. J Mol Biol 425:1158–1171

54. Kauffmann E, Darnton NC, Austin RH et al (2001) Lifetimes of intermediates in the beta-sheet to alpha-helix transition of beta-lactoglobulin by using a diffusional IR mixer. Proc Natl Acad Sci U S A 98:6646–6649

55. Tatulian SA, Biltonen RL, Tamm LK (1997) Structural changes in a secretory phospholipase A(2) induced by membrane binding: a clue to interfacial activation? J Mol Biol 268:809–815

56. Walsh STR, Cheng RP, Wright WW et al (2003) The hydration of amides in helices; a comprehensive picture from molecular dynamics, IR, and NMR. Protein Sci 12:520–531

57. Martinez G, Millhauser G (1995) Ftir spectroscopy of alanine-based peptides – assignment of the amide I' modes for random coil and helix. J Struct Biol 114:23–27

58. Javor S, Natalello A, Doglia SM et al (2008) Alpha-helix stabilization within a peptide dendrimer. J Am Chem Soc 130:17248–17249

59. Roque A, Iloro I, Ponte I et al (2005) DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. J Biol Chem 280:32141–32147

60. Barth A (2000) The infrared absorption of amino acid side chains. Prog Biophys Mol Biol 74:141–173

61. Haris PI, Severcan F (1999) FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media. J Mol Catal B: Enzym 7:207–221

62. Hering JA, Innocent PR, Haris PI (2002) Automatic amide I frequency selection for rapid quantification of protein secondary structure from Fourier transform infrared spectra of proteins. Proteomics 2:839–849

63. Vila R, Ponte I, Collado M et al (2001) Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA. J Biol Chem 276:30898–30903

64. Salomaa P, Schaleger LL, Long FA (1964) Solvent deuterium isotope effects on acid–base equilibria. J Am Chem Soc 86:1–7

65. Dong AC, Hyslop RM, Pringle DL (1996) Differences in conformational dynamics of ribonucleases A and S as observed by infrared spectroscopy and hydrogen-deuterium exchange. Arch Biochem Biophys 333:275–281

Chapter 21

Insoluble Protein Characterization by Circular Dichroism (CD) Spectroscopy and Nuclear Magnetic Resonance (NMR)

Shaveta Goyal, Haina Qin, Liangzhong Lim, and Jianxing Song

Abstract

Besides misfolded proteins, which still retain the capacity to fold into uniquely defined structures but are misled to "off-pathway" aggregation, there exists a group of proteins which are unrefoldable and insoluble in buffers. Previously no general method was available to solubilize them and consequently their solution conformations could not be characterized. Recently, we discovered that these insoluble proteins could in fact be solubilized in pure water. Circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) characterization led to their classification into three groups, all of which lack the tight tertiary packing and consequently anticipated to unavoidably aggregate in vivo with ~150 mM ions, thus designated as "intrinsically insoluble proteins (IIPs)." It appears that eukaryotic genomes contain many "IIP," which also have a potential to interact with membranes to trigger neurodegenerative diseases. In this chapter, we provide a detailed procedure to express and purify these proteins, followed by CD and NMR spectroscopy characterization of their conformation and interaction with dodecylphosphocholine (DPC).

Key words Insoluble protein, Unsalted water, Inclusion body, Dodecylphosphocholine (DPC), Circular dichroism (CD) spectroscopy, Nuclear magnetic resonance (NMR) spectroscopy, Two-dimensional ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) spectroscopy

1 Introduction

Proteins play the most important functional roles in living cells. A protein is a linear heteropolymer of 20 common amino acids which can be categorized into polar (P) and hydrophobic (H) groups based on their different interactions with water. Despite the huge number of possible amino acid sequences, it seems that only a small portion of sequences fold into uniquely defined three-dimensional structures [1]. For a well-folded protein, more than 80 % of hydrophobic side chains are buried in the internal core while most hydrophilic residues are exposed to polar water phase, thus being soluble in aqueous buffers. Previously, it has been believed that the unique, crystal-like structures are essential for the

biological functions of all proteins. However, now it has been realized that many proteins lacking well-defined structures are fully functional, thus called intrinsically unstructured proteins (IUPs) [2, 3], and there is a sharp increase in IUPs with the transition from prokaryotic to eukaryotic cells [3].

One commonly encountered problem associated with protein research and application is their insolubility/aggregation, which not only imposes a practical challenge to the biopharmaceutical industry, but also are responsible for a large spectrum of human diseases, in particular neurodegenerative diseases including spongiform encephalopathies, amyotrophic lateral sclerosis (ALS), and Alzheimer's and Parkinson's diseases [4, 5]. It has been estimated that ~35–50 % of the proteins expressed in *Escherichia coli* cells were in inclusion bodies (IBs) [6, 7]. Proteins have been extensively demonstrated to get aggregated through "misfolding." In other words, they still retain the capacity to fold into uniquely folded structures but are misled to "off-pathway" aggregation, as proteins are only marginally stable, with only ~5–20 kcal/mol in free energy more stable than unfolded states under physiologic conditions. As a result, even under native conditions, proteins undergo inherently dynamic fluctuations among different conformations and thus have access to partially unfolded states in which the hydrophobic side chains are more exposed to the bulk solvent [5].

Usually misfolded proteins can be refolded into their native structures by optimizing a variety of conditions in vitro. On the other hand, despite extensive efforts in fusing insoluble proteins with highly soluble tags, coexpressing folding catalysts and chaperones, reducing culture temperature, and modifying culture media, a portion of proteins were not refoldable and soluble in buffer systems with any available methods. Previously, no general method was available to solubilize these proteins without addition of detergents and/or denaturants such as sodium dodecyl sulfate (SDS), urea, or guanidine hydrochloride at high concentrations, and consequently their structural properties remained unknown. In 2005, we discovered that in fact, such unrefoldable and insoluble proteins could in fact be solubilized in unsalted water for detailed circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy characterization, which led to classification into three groups based on their solution conformations characterized by CD and NMR ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) spectroscopy (Fig. 1): group 1, with no stable secondary structure as evidenced by CD and no significant tertiary packing as judged by narrowly dispersed and sharp HSQC peaks; group 2, with secondary structure by CD but no significant tertiary packing as indicated by narrowly dispersed and sharp HSQC peaks; and group 3 with secondary structure by CD and rudimentary tertiary packing like molten globules as implied by broadened HSQC peaks [8, 9]. Intriguingly, we failed to find any insoluble protein with the tight tertiary packing. Therefore, unlike "misfolded
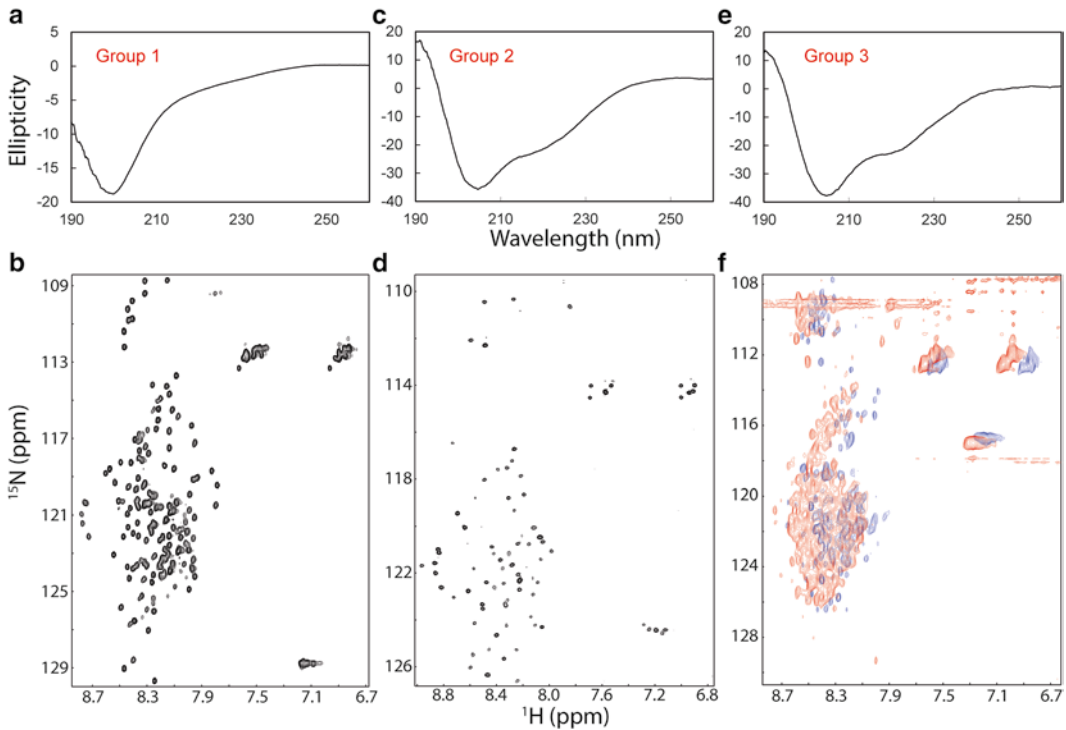
Fig. 1 Classification of unrefoldable and insoluble proteins based on their conformations as characterized by CD and NMR HSQC spectroscopy. Group 1 as represented by the hemagglutinin receptor-binding domain (152 residues) of avian influenza A virus solubilized in unsalted water at pH 6.2, which has no stable secondary structure as evidenced by the far-UV CD spectrum (**a**) with the maximal signal at ~198 nm and no positive signal at 190 nm, and also has no significant tertiary packing as judged by narrowly dispersed and sharp HSQC peaks (**b**). Group 2 as represented by the human Nogo-60 (60 residues) at pH 4.0, which has helical secondary structure as evidenced by the far-UV CD spectrum (**c**) with two maximal signals at ~206 nm and 222 nm, respectively, and a positive signal at 192 nm, but has no significant tertiary packing as judged by narrowly dispersed and sharp HSQC peaks (**d**). Group 3 as represented by of the entire extracellular domain of the human Nogo receptor (420 residues) at pH 6.2, which has helical secondary structure as evidenced by the far-UV CD spectrum (**e**) with two maximal signals at ~206 nm and 222 nm, respectively, and a positive signal at 192 nm and also has rudimentary tertiary packing as judged by broad HSQC peaks (**f**). Consequently, only a small set of peaks are detectable in unsalted water (*blue*) and a large amount of peaks appeared upon adding urea to 8 M (*red*)

proteins" which still retain the capacity to fold into uniquely defined structures, unrefoldable and insoluble proteins completely lack this ability and consequently exist in highly disordered or partially folded states with many hydrophobic side chains exposed to the bulk solvent. Marvellously enough, for insoluble proteins and even one of the most hydrophobic transmembrane sequences, in unsalted water with the pH value several units away from the pI value of the protein, their repulsive electrostatic interactions appear to be sufficient to balance out the hydrophobic clustering to prevent the occurrence of intermolecular aggregation [8–10]. By contrast, for such proteins, the presence of ions even at a very low concentration

is sufficient to screen out repulsive interactions and consequently allow the hydrophobic clustering/aggregation to occur [9, 10].

This discovery has offered us and other groups [11–13] a powerful tool to characterize insoluble proteins, and recently we have further addressed several fundamental and disease-relevant issues [14–17]. One novel finding is that even for well-folded proteins such as SH3 and MSP folds, improper truncation, splicing variation, and even a point mutation are sufficient to render them into unrefoldable and insoluble proteins by completely eliminating the native structures as characterized in unsalted water for the SH3 [14], viperin [15], P56S mutant [16], and splicing variant [17] of the VAPB-MSP domain which leads to amyotrophic lateral sclerosis (ALS). Based on our characterization, these proteins are only soluble in unsalted water but anticipated to unavoidably aggregate in vivo with ~150 mM ions, thus designated as "intrinsically insoluble proteins (IIPs)" [10]. It appears that eukaryotic genomes contain many IIPs due to at least three well-known characteristics, namely (1) increase of intrinsically unstructured proteins with low-complexity sequences; (2) emergence of the splicing variation; and (3) accumulation of random mutations [10].

As protein aggregation in buffer is mainly owing to the improper exposure of hydrophobic patches, it is thus tempted to speculate that such proteins may have strong tendency to interact with lipid membranes, which represent a hydrophobic phase in cells [10]. With our unique ability to solubilize the insoluble proteins, we have recently tested this speculation and found that indeed, insoluble proteins could interact with lipid molecules such as dodecylphosphocholine (DPC) to different degrees. In particular, as shown in Fig. 2, a gradual addition of DPC triggered the insertion of the ALS-causing P56S MSP domain into the membrane as well as formation of the helical conformation as deciphered by CD and NMR HSQC investigation [17]. Our discovery thus rationalizes the in vivo observation that the ALS-causing P56S mutation could remodel endoplasmic reticulum (ER) to form stacked cisternae [10, 18]. This implies that insoluble proteins have the potential to interact with biological membranes in vivo which may represent a critical mechanism for protein aggregation to trigger neurodegenerative diseases [10].

In this chapter, we describe the detailed procedure to express, purify, and characterize the unrefoldable and insoluble proteins in unsalted water by CD and NMR HSQC spectroscopy. Experimentally, Milli-Q water is sufficient for solubilizing all insoluble proteins (>60) we studied so far. On the other hand, however, to allow the solubilization, proteins have to be very pure, which request the final purification by the reversed-phase high-pressure liquid chromatography (RP-HPLC) to remove particularly nonprotein impurities such as attached lipids, small molecules, and ions; and the water pH value needs to be several units away from the pI of the protein to generate the repulsive electrostatic interaction.
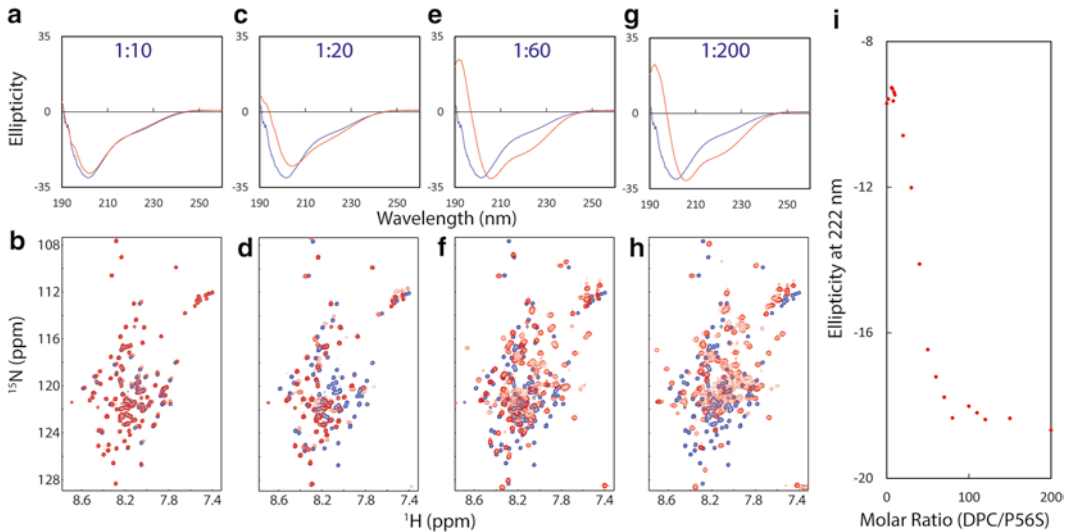
Fig. 2 Interaction of insoluble proteins with lipid mimetics DPC as characterized by CD and NMR HSQC spectroscopy. Superimposition of far-UV CD (**a**, **c**, **e**, **g**) and HSQC (**b**, **d**, **f**, **h**) spectra of the ALS-causing P56S MSP domain (125 residues) solubilized in unsalted water at pH 4.0, in the absence (*blue*) and in the presence (*red*) of DPC at molar ratios of 1:10, 1:20, 1:60, and 1:200. (**i**) Plot of the CD ellipticity values at 222 nm vs. the ratios (DPC/P56S)

2   Materials

2.1   Protein Expression

1. LB broth: 25 g of premix LB medium and dissolve it in dH_2O to final volume of 1 L or dissolve 10 g tryptone, 5 g yeast extract, and 10 g NaCl in dH_2O to the final volume of 1 L. Autoclave the medium for 20 min at 15 psi.

2. LB broth with ampicillin: Add ampicillin stock to make final concentration of 100 μg/mL to the autoclaved LB broth.

3. LB ampicillin agar plates: 7.5 g of premix LB agar powder and dissolve it in dH_2O to make the final volume to 500 mL or weigh out 5.0 g tryptone, 2.5 g yeast extract, 5.0 g NaCl, and 7.5 g agar and dissolve in 500 mL dH_2O in 1 L flask. Autoclave the medium and let it cool to 55 °C before adding ampicillin. Add 0.5 mL of ampicillin stock, mix well, and pour into petri dishes.

4. 100 mg/mL ampicillin stock concentration: Dissolve 5 g of ampicillin in 5 mL of dH_2O. Filter using 0.22 μm syringe filter, dispense into 1 mL aliquots, and store the stock solution at –20 °C (final concentration = 100 μg/mL).

5. 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) stock concentration: 2.383 g of IPTG into dH_2O to the final volume of 10 mL and filter sterilize using 0.22 μm filters. Store the 1 mL dispensed aliquots at –20 °C.

6. M9 minimal medium: 17.1 g/L $Na_2HPO_4 \cdot 12H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L $(^{15}NH_4)_2SO_4$ (Cambridge Isotope Laboratories Inc.), 4 g/L glucose (2 g/L ^{13}C glucose if preparing double-labeled protein sample for 3D/4D experiments), 1 mM $MgSO_4$, 2 mg/L thiamine, 75 mg/L ampicillin. Add all the components except labeled salt, glucose, $MgSO_4$, thiamine, and ampicillin to 1 L bottle; make the final volume to 1 L using Milli-Q water; and autoclave. Prepare the labeled salt in a new Falcon tube by mixing $(^{15}NH_4)_2SO_4$, glucose (^{13}C glucose), $MgSO_4$, thiamine, and ampicillin. Filter using 0.22 μm syringe filter and add this to the autoclaved M9 medium.

2.2 Protein Purification

1. Nickel beads.

2. 20 mL column.

3. Imidiazole: 1.7 g of imidiazole and dissolve in lysis buffer (*see* **step 8** in this section) to final volume of 100 mL. Adjust pH to 8.0 using HCl/NaOH and protect imidiazole from light. Prepare fresh imidiazole.

4. Dithiothreitol (DTT): 1.545 g DTT in dH_2O to final volume of 100 mL. Filter sterilize using 0.22 μm syringe filter and store 1 mL aliquots at –20 °C.

5. Urea: 48 g of urea into lysis buffer to the final volume of 100 mL and adjust the pH to 8.0 using NaOH/HCl.

6. 10× Phosphate-buffered saline (PBS) stock solution: 80 g NaCl, 2 g KCl, 11.5 g $Na_2HPO_4 \cdot H_2O$, 2 g KH_2PO_4, pH 7.4.

7. Pellet washing buffer: 1× PBS buffer with 0.1 % Triton X-100, pH 7.4.

8. Lysis buffer: 8 M Urea, 50 mM Tris, 150 mM NaCl, 10 mM beta-mercaptoethanol (β-ME), pH 8.0.

9. Wash buffer: Lysis buffer with 20 mM imidiazole, pH 8.0.

10. Elution buffer: Lysis buffer with 250 mM imidiazole, pH 8.0.

11. Reversed-phase high-pressure liquid chromatograph (RP-HPLC).

12. Analytical and semi-preparative columns (*see* **Note 1**).

13. Acetic acid.

14. Buffer A: Filtered and degassed Milli-Q water with 0.1 % (v/v) trifluroacetic acid (TFA) (*see* **Note 2**).

15. Buffer B: Acetonitrile with 0.1 % (v/v) TFA.

16. Buffer C: Acetonitrile, sample filters with 0.22 μM porosity.

17. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE): 50× Tris–acetate–EDTA (TAE): 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5 M ethylenediamine tetraacetic acid (EDTA), pH 8.0. Make the final volume to 1 L with deionized water. SDS-PAGE gel preparation: make up 30 mL of resolving gel using the composition of the ingredients from the table below. Add TEMED and APS at

last, when the separating gel solution is ready to be poured. Mix the solution well before pouring into the gel assembly. After pouring resolving gel solution into gel cassette, overlay it with 1–2 mm of isopropanol to exclude oxygen and ensure a flat interface between the resolving and stacking gels. Allow the gel to polymerize for 30 min.

Resolving gel composition table:

Percentage (%)	7	10	12	15
H$_2$O (mL)	15.3	12.3	10.2	7.2
1.5 M Tris–HCl, pH 8.8 (mL)	7.5	7.5	7.5	7.5
20 % (w/v) SDS (mL)	0.15	0.15	0.15	0.15
Acrylamide/Bis-acrylamide (30 %/0.8 % w/v) (mL)	6.9	9.9	12.0	15.0
10 % (w/v) ammonium persulfate (APS) (mL)	0.15	0.15	0.15	0.15
TEMED (mL)	0.02	0.02	0.02	0.02

Stacking gel: Remove the isopropanol from the top of the resolving gel and prepare stacking gel as per the composition shown in table below. Fill the top of the cassette with the solution and insert the comb. The comb should rest so that the tops of the well dividers are level with the top of the short plate. This excludes oxygen while ensuring that the dividers will fully separate the wells. Allow the stacking gel to polymerize for approx. 30 min.

Stacking gel composition table:

H$_2$O (mL)	3.075
0.5 M Tris–HCl, pH 6.8 (mL)	1.25
20 % (w/v) SDS (mL)	0.025
Acrylamide/Bis-acrylamide (30 %/0.8 % w/v) (mL)	0.67
10 % (w/v) ammonium persulfate (APS) (mL)	0.025
TEMED (mL)	0.005

18. SDS-PAGE running buffer (10×): 144 g glycine, 30.2 g Tris base, 10 g SDS, deionized water to make the final volume to 1 L.

19. SDS sample loading buffer (40 mL): 5 mL 0.5 M Tris, pH 6.8, 8 mL 50 % glycerol, 8 mL 10 % SDS, bromophenol blue, 2 mL 2-β-mercaptoethanol, 16 mL ddH$_2$O.

20. Coomassie staining solution: 1 g Coomassie R250, 100 mL glacial acetic acid, 400 mL methanol, deionized water for final volume of 1 L.

2.3 Circular Dichroism	1. Lyophilized protein samples.
	2. MQ H_2O.
	3. Quartz cuvette.
	4. NaOH solution.
	5. 8 M urea.
2.4 NMR	1. Lyophilized labeled protein samples.
	2. Milli-Q water.
	3. D_2O.
	4. NMR tubes.
2.5 Interaction with Lipid Mimetics	1. Dodecylphosphocholine (DPC) (Avanti Polar Lipids, Inc.).

3 Methods

3.1 Protein Expression and Purification

1. Pick a single colony from the freshly transformed plate and inoculate into the tube containing 10 mL LB medium along with ampicillin (100 mg/mL). Incubate at 37 °C at 200 rpm overnight (ON). For ^{15}N-labeled sample for NMR, inoculate a single colony in 1 mL LB broth supplemented with ampicillin and incubate at 37 °C at 200 rpm for 8 h, and later transfer this 1 mL culture to 100 mL M9 medium supplemented with 1 g/L ([15]NH$_4$)$_2$SO$_4$ (*see* Subheading 2.1) ^{15}N medium and incubate ON at 37 °C.

2. Inoculate the ON seed culture to 2.8 L flask containing 1 L LB medium and ampicillin (100 mg/mL).

3. Incubate at 37 °C at 200 rpm till the OD_{600nm} reaches 0.6.

4. Induce the culture with IPTG at the concentration of 1 mM per L of medium.

5. After 4 h of induction, harvest the cells using centrifuge at the speed of $7,903 \times g$ for 15 min at 4 °C.

6. The harvested cells are stored at –80 °C until used for protein purification.

3.2 Affinity Purification

1. The harvested cells are resuspended in the lysis buffer and set for sonication at the rate of 1 s pulse on and 2 s pulse off for 10 min. During the sonication, it is important to keep the sample at 4 °C.

2. After sonication the cells are centrifuged at the high speed of $38,465 \times g$ for 30 min, so as to remove the supernatant from the pellet. Insoluble proteins are in the pellet fraction.

3. Wash the pellet with the washing buffer three times, to remove the lipids of the *E. coli* membrane (*see* **Note 3**).

4. Redissolve the pellet into 25 mL lysis buffer and keep it for stirring at room temperature (RT) for 1 h (*see* **Note 4**).

5. Spin down the pellet and collect the supernatant.

6. Allow the supernatant to bind with His-tagged beads for approximately 1 h at RT (*see* **Note 5**).

7. Load the supernatant containing beads onto the 20 mL column, and collect the supernatant and beads will remain in the column. Take out small amount of supernatant sample and dissolve in SDS loading dye.

8. Wash the beads with the wash buffer; this will remove the unspecific bound proteins and also collect some sample to mix up with SDS loading buffer.

9. Elute the protein with elution buffer and collect some sample from each eluted fraction to mix up with SDS loading buffer.

10. Run the SDS-PAGE gel to check the protein expression level and purity.

3.3 Reversed-Phase High-Pressure Liquid Chromatography (RP-HPLC) Purification

RP-HPLC is a very powerful and widely used technique for separating biomolecules, and the separation mechanism depends on the hydrophobic binding interactions between the solute molecules in mobile phase and the immobilized hydrophobic stationary phase. RP-HPLC is able to resolve very similar polypeptides, with even a single amino acid difference.

1. Connect the analytical column to the solvent delivery system according to the HPLC system requirements and turn on the HPLC software along with pump, detector, and degasser.

2. Remove the organic solvent (e.g., acetonitrile) from the column by washing the column with 100 % buffer B followed by 100 % buffer A for 10 min, each with a flow rate of 1 mL/min.

3. To the elution add 10 mM DTT, so as to reduce the protein and incubate it at RT for 1 h. (DTT is only required if the protein contains free cysteine residues.)

4. Before injecting the sample for HPLC purification, samples are acidified with 10 % acetic acid and sample loading loop is washed thoroughly with water.

5. Filter the sample, inject 100 μL sample into the HPLC column, and use a linear gradient from 0 to 100 % buffer B over 60 min to elute out the protein sample.

6. The separation for proteins/peptides is monitored at 215 nm wavelength, which is specific for the peptide bond and 254 or 280 nm which is useful for monitoring the aromatic amino acid residues.

7. Collect the peaks and check the molecular mass by a Voyager STR matrix-assisted laser desorption ionization time-of-flight-mass spectrometer (Applied Biosystems), and also check the purity by SDS-PAGE gel.

8. Large-volume sample separation is further done by using semi-preparative column. For semi-preparative separation follow **steps 2–6** but with a flow rate of 6 mL/min (or depending on column specifications) and with a gradient program based on analytical separation.

9. Collect and lyophilize the fractions containing the target proteins.

10. Wash the columns with 100 % buffer B and store columns in 100 % buffer C.

11. The lyophilized powder is stored at –80 °C for biophysical characterizations with CD and NMR.

3.4   Circular Dichroism

CD spectroscopy is based on the differential absorption of left- and right-handed circularly polarized light by optically active molecules in the sample. CD is used for characterization of secondary structure content of protein and also tertiary packing, conformational stability with change in temperature, pH or solvent, and also protein-protein and protein-ligand interactions. To average the random noise, the CD spectrum is typically averaged from three to five scans. For detailed instruction and interpretation of the CD spectroscopy, please *see* ref. 19.

3.4.1   Determination of Protein Secondary Structures by Far-UV CD Experiment

The far-UV CD spectrum of proteins is utilized to characterize the secondary structure content. The obtained spectra can be attributed to the combination of three types of secondary structures [19].

1. The CD samples are prepared by dissolving the lyophilized powder in MQ H_2O to the final concentration of ~20–50 μM (*see* **Note 6**) and 200 μL of this sample is loaded into a 1 mm path length quartz cuvette (*see* **Note 7**).

2. First record the spectrum with the wavelength window from 190 to 250 nm for the blank, i.e., MQ H_2O.

3. Record the spectrum of the protein sample.

4. Subtract the blank spectrum and plot ellipticity vs. wavelength (Fig. 1), or further calculate mean residue ellipticity [culusing]:

$$[\theta] = \frac{\theta obs. \times \text{Mean residual weight}}{10 \text{Concentration} \left(\frac{g}{ml} \right) \times \text{pathlength} \, (cm)}$$

5. Estimate the secondary structure contents of the protein by deconvoluting its far-UV CD spectrum by using special programs such as provided on Dichro Web (dichroweb.cryst.bbk. ac.uk/html/references.shtml).

3.4.2   Assessment of Protein Tertiary Structure by Near-UV CD Experiment

For near-UV CD spectrum, data is collected from 250 to 350 nm wavelength and it depends on the orientations and environments of the side chains in proteins [19]. In a folded protein, side chains such as aromatic ones are relatively rigid within a chiral environment which will have detectable CD signal. However, if a protein is partially folded or highly disordered, these side chains become dynamic and consequently the chirality is lost and no or reduced CD signal is observed. Near-UV CD spectroscopy can thus be sensitive to the change of the protein tertiary structure including those upon ligand binding and protein–protein interactions [19].

1. Protein CD samples are dissolved in MQ H_2O in the absence of and in the presence of 8 M urea and the final protein concentration for near-UV CD is usually tenfold more concentrated than that for far-UV CD.

2. Record the CD spectrum with the wavelength window from 250 to 360 nm for the blanks (one, MQ H_2O water in which protein is dissolved, another, MQ H_2O with 8 M urea).

3. Record the near-UV spectrum for the protein in MQ H_2O.

4. Record the near-UV spectrum for the protein in MQ H_2O with 8 M urea.

5. Subtract the blank spectra from respective spectra of the protein.

6. Compare the near-UV spectra of the protein without and with 8 M urea. If two spectra are very different, this means that the protein has the tight tertiary packing in the native condition (without urea). By contrast, no significant difference between two near-UV spectra indicates the lacking of the tight tertiary packing even in the native condition.

3.4.3   Investigation of the Interaction of Insoluble Proteins with DPC

1. Record far-UV CD spectrum of the insoluble protein dissolved in MQ H_2O.

2. Record far-UV CD spectra of the insoluble protein with a gradual addition of DPC dissolved in MQ H_2O with the same pH as the protein sample, at a series of molar ratios as we previously described [17] and illustrated in Fig. 2.

3. Calculate the dilution factors for each titration: (original CD sample volume + added DPC volume)/(original CD sample volume).

4. Calibrate each far-UV CD spectrum titrated with DPC by timing the ellipticity values with the corresponding dilution factor.

5. Compare the CD spectra in the presence of DPC at different molar ratios (1:10, 1:20, 1:60, and 1:200). If there are significant changes, this means that DPC interacts with the insoluble protein and triggers the changes of secondary structures.

(a) The contents of secondary structures can be estimated by deconvoluting far-UV CD spectra (dichroweb.cryst.bbk.ac.uk/html/references.shtml).

(b) As exemplified by Fig. 2i, for all titrations, the ellipticity values can be exacted at a certain wavelength (i.e., 222 nm for helical conformation) and plot them vs. the molar ratios. Consequently a transition curve can be generated which gives a clear picture about how addition of DPC will trigger the formation of the secondary structure.

3.5 Two-Dimensional ¹H-¹⁵N NMR Single Heteronuclear Quantum Correlation (HSQC) Spectroscopy

^{1}H-^{15}N HSQC spectroscopy is a very powerful method for a rapid assessment of the structural properties of a protein. Briefly, HSQC peaks are resulting from the backbone amide protons of all 20 amino acids except for proline as well as side chain amide protons. In a well-folded protein which has a tight tertiary packing, each amide proton is located in the unique environment, and consequently the HSQC spectral dispersions at both ^{1}H and ^{15}N dimensions are large, as exemplified by the HSQC spectrum of the wild-type MSP domain with dispersions of ~3.2 ppm and 27 ppm, respectively, at ^{1}H and ^{15}N dimensions (Fig. 3a), which adopts a seven-strand immunoglobulin-like β which ad (Fig. 3b). By contrast, in the partially folded or highly disordered protein whose tight tertiary packing is disrupted, amide protons sense similar surrounding and thus give rise to crowded HSQC peaks, as illustrated by the HSQC spectrum of the insoluble P56S mutant of the MSP domain solubilizing in unsalted water, with dispersions of only ~1.1 ppm and 21 ppm, respectively, at ^{1}H and ^{15}N dimensions (Fig. 3c). In the P56S MSP domain, the native ppm isrupted, amide protons sense similar surrounding and thus give rise to crowded HSQC peaks, as illustrated by the HSQC.
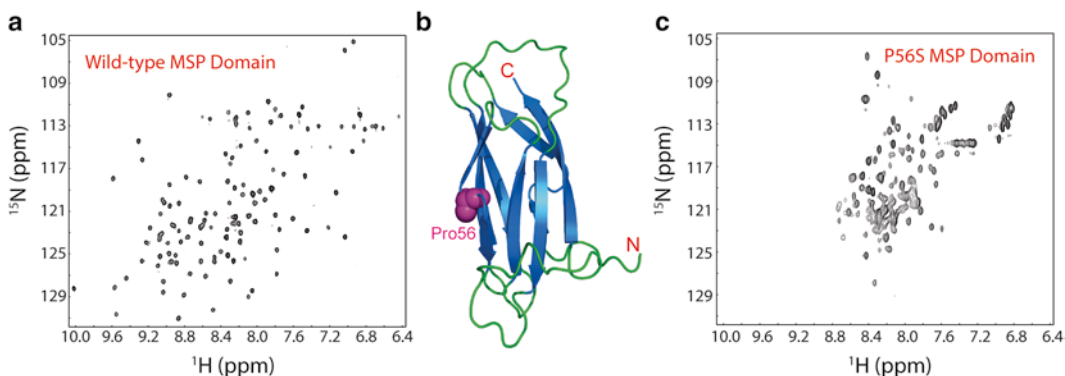
Fig. 3 HSQC assessment of structural properties of a protein. (**a**) HSQC spectrum of the wild-type MSP domain of the human VAPB protein which is uniquely folded, with its crystal structure determined [16] to adopt a seven-strand immunoglobulin-like β-sandwich (**b**). (**c**) HSQC spectrum of the same MSP domain with only Pro56 mutated to Ser, which renders the mutant to become completely insoluble in buffers and to be soluble only in unsalted water, with the native β and to be structure completely eliminated [16]

After the HSQC assessment, the insoluble proteins with separated HSQC peaks (Fig. 1b, d) can be further studied by advanced NMR techniques [20], such as to determine the three-dimensional structure [21], and/or to characterize the backbone dynamics at different time scales [14].

1. Dissolve ^{15}N-labeled protein powder into 500 μL unsalted water with 10 % D_2O for spin lock (*see* **Note 8**). The concentration of protein sample should be at least 100 μM for acquiring high-quality spectra.

2. Check the pH of the protein sample (*see* **Note 9**). Determine the desired pH for the sample by considering that the solution pH should be several units away from the pI of the protein to generate repulsive electrostatic interactions to prevent aggregation. If necessary, to adjust the pH of sample by using very dilute sodium hydroxide solution (NaOH) solution (*see* **Note 10**).

3. Spin the sample and transfer it to a clean and dry NMR tube.

4. Record 1D ^{1}H NMR and 2D ^{1}H-^{15}N HSQC spectra on an 800 MHz spectrometer.

5. Process the NMR data using TopSpin (Bruker), NMRPipe/NMRDraw [8, 14–17, 21], or other suitable software packages.

6. Assess the spectral properties including dispersions, line widths, and number of resolved peaks. This will provide the knowledge whether it is well folded, or lacking of the tight tertiary packing. However, even for an insoluble protein which is only partially folded or highly disordered, further characterization of its conformation and dynamics is feasible as long as the HSQC peaks are sharp and separated, as we have previously conducted [14, 16, 17, 21].

4 Notes

1. Columns for RP-HPLC are selected based on the size and amount of the protein to be purified. For small amount or initial screening analytical columns are used and for large-volume purifications semi-preparative columns are preferred. Proteins with large size are purified using C4 column and C8 column is used for relatively small proteins while C18 for purification of peptides and oligonucleotides.

2. Trifluoroacetic acid is by far the most commonly used ion pairing agent because of its excellent separation capabilities, complex formation with oppositely charged ionic groups to enhance RP retention (ion pairing), pH control (buffering),

low UV absorbance, and high volatility for easy removal in peptide isolation.

3. Pellet should be washed with wash buffer containing detergent before proceeding for purification; else the lipids from *E. coli* membrane may contaminate the protein and also will create problems in protein purification.

4. If pellet is not completely dissolved in lysis buffer, it can be homogenized by sonicating for 1–2 min.

5. Urea precipitates at 4 °C. So the binding of beads to protein should be done at RT.

6. It is important that the concentration of protein is accurate. For secondary structure measurements, sample concentrations may range from 10 to 50 μM depending on the path length of the cell. Measure the protein concentration at 280 nm in 8 M urea. Make the blank in one cuvette with 8 M urea and another cuvette with concentrated protein in 8 M urea. Note down the dilution factor and measure concentration using the formula

$$C\left(\text{concentration}\right) = \frac{A280 \times \text{Molecular weight of protein}}{\varepsilon_0 \left(\text{extinction coefficient at } 280\text{nm} \times 1\left(\text{path length}\right)\right)}.$$

7. Cuvettes for CD measurements must be clean and dry. Quartz cells can be cleaned by soaking in mild detergent solutions, a mixture of 30 % concentrated HCl and 70 % ethanol, or concentrated nitric acid.

8. NMR tube should be clean and dry so as to prevent contamination. Poor-quality tubes or tubes with scratches may result in poor shimming. The tubes can be cleaned with mild detergents or soaking ON in concentrated nitric acid. After cleaning, tubes should be thoroughly washed with water so that no chemical traces are left behind; else it may cause problems in collecting the data. Dry the tubes completely after washing with water.

9. As the lyophilized protein still contains some TFA leftover from the HPLC solvent, the pH of the protein sample is usually around 3.5–5.0 after the protein powder is dissolved in MQ H_2O. Adjust the pH of sample to guarantee several units away from the pI of the protein. However, at pH values higher than the neutral, the HSQC peaks will be undetectable as amide protons exchange with water very rapidly. Therefore, usually we choose to adjust the water pH to be several units lower than the protein pI.

10. The pH should be adjusted slowly by adding NaOH drop by drop. If not done carefully, protein sample may precipitate.

Acknowledgment

References

1. Li H, Helling R, Tang C et al (1996) Emergence of preferred structures in a simple model of protein folding. Science 273:666–669

2. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208

3. Xue B, Dunker AK, Uversky VN (2013) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 30:137–149

4. Ross CA, Poirier MA (2004) Protein aggregation and neurodegenerative disease. Nat Med 10:S10–S17

5. Luheshi LM, Dobson CM (2009) Bridging the gap: from protein misfolding to protein misfolding diseases. FEBS Lett 583:2581–2586

6. Christendat D, Yee A, Dharamsi A et al (2000) Structural proteomics: prospects for high throughput sample preparation. Prog Biophys Mol Biol 73:339–345

7. García-Fruitós E, Vázquez E, Díez-Gil C et al (2012) Bacterial inclusion bodies: making gold from waste. Trends Biotechnol 30:65–70

8. Li M, Liu J, Ran X et al (2006) Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. Biophys J 91:4201–4209

9. Song J (2009) Insight into "insoluble proteins" with pure water. FEBS Lett 583:953–959

10. Song J (2013) Why do proteins aggregate? "Intrinsically insoluble proteins" and "dark mediators" revealed by studies on "insoluble proteins" solubilized in pure water. F1000Res 2:94

11. Delak K, Harcup C, Lakshminarayanan R et al (2009) The tooth enamel protein, porcine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. Biochemistry 48:2272–2281

12. Aguado-Llera D, Goormaghtigh E, De Geest N et al (2010) The basic helix-loop-helix region of human neurogenin l is a monomeric natively unfolded protein which forms a "fuzzy" complex upon DNA binding. Biochemistry 49:1577–1589

13. Nielsen SB, Franzmann M, Basaiawmoit RV et al (2010) The basic helix-loop-helix region of hummulated by heparin but inhibited by amphiphiles. Biopolymers 93:678–689

14. Liu J, Song J (2009) Insights into protein aggregation by NMR characterization of insoluble SH3 mutants solubilized in salt-free water. PLoS One 4:e7805

15. Shaveta G, Shi J, Chow VT et al (2010) Structural characterization reveals that viperin is a radical S-adenosyl-L-methionine (SAM) enzyme. Biochem Biophys Res Commun 391:1390–1395

16. Shi J, Lua S, Tong JS et al (2010) Elimination of the native structure and solubility of the hVAPB MSP domain by the Pro56Ser mutation that causes amyotrophic lateral sclerosis. Biochemistry 49:3887–3897

17. Qin H, Wang W, Song J (2013) ALS-causing P56S mutation and splicing variation on the hVAPB MSP domain transform its β-sandwich fold into lipid-interacting helical conformations. Biochem Biophys Res Commun 431:398–403

18. Fasana E, Fossati M, Ruggiano A et al (2010) A VAP B mutant linked to amyotrophic lateral sclerosis generates a novel form of organized smooth endoplasmic reticulum. FASEB J 24:1419–1430

19. Adler AJ, Greenfield NJ, Fasman GD (1973) Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods Enzymol 27:675–735

20. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog NMR Spectrosc 34:93–158

21. Li M, Liu J, Song J (2006) Nogo goes in the pure water: solution structure of Nogo-60 and design of the structured and buffer-soluble Nogo-54 for enhancing CNS regeneration. Protein Sci 15:1835–1841

Chapter 22

Methods for Characterization of Protein Aggregates

Witold Tatkiewicz, Elisa Elizondo, Evelyn Moreno, Cesar Díez-Gil, Nora Ventosa, Jaume Veciana, and Imma Ratera

Abstract

Physicochemical characterization of protein aggregates is important on one hand, due to its large impact in understanding many diseases for which formation of protein aggregates is one of the pathological hallmarks. On the other hand, recently it has been observed that bacterial inclusion bodies (IBs) are also highly pure proteinaceous aggregates of a few hundred nanometers produced by recombinant bacteria supporting the biological activities of the embedded polypeptides. From this fact arises a wide spectrum of uses of IBs as functional and biocompatible materials upon convenient engineering but very few is known about their physicochemical properties.

In this chapter we present methods for the characterization of protein aggregates as particulate materials relevant to their physicochemical and nanoscale properties.

Specifically, we describe the use of infrared spectroscopy (IR) for the determination of the secondary structure, dynamic light scattering (DLS) for sizing, nanosight for sizing and counting, and Z-potential measurements for the determination of colloidal stability. To study their morphology we present the use of atomic force microscopy (AFM). Cryo-transmission electron microscopy will be used for the determination of the internal structuration. Moreover, wettability and nanomechanical characterization can be performed using contact angle (CA) and force spectroscopic AFM measurements of the proteinaceous nanoparticles, respectively.

The physical principles of the methods are briefly described and examples of data for real samples and how that data is interpreted are given to help clarify capabilities of each technique.

Key words Protein aggregates, Inclusion bodies, Physicochemical properties, Nanoscale properties, Recombinant bacteria, Proteinaceous nanoparticles, Biomaterial

1 Introduction

1.1 Determination of the Secondary Structure

Spectroscopic methods are widely used for structural characterization and biophysical analyses of various assemblies of protein aggregates. Specifically, infrared spectroscopy (IR) is a technique used to investigate the secondary structure of proteins giving information of the intramolecular vibrations of compounds. As in protein aggregates specific interactions between aggregation-prone species are present, characteristic bands in FTIR spectra can be found.

Their presence is indicative of most protein amyloid fibrils [1] and it is common in artificially engineered aggregates from various proteins such as VP1LAC, hDHFR, VP1GFP, or Aβ42(F19D)-BFP [2].

1.2 Determination of IB Size and Concentration

Characterizing the state of aggregation in IBs is of paramount importance when trying to understand its morphology and stability. Product quality, can be highly influenced by the state of protein aggregation. IBs span a broad size range, from small oligomers (nanometers) to insoluble micron-sized aggregates that can contain millions of monomer units. Protein aggregation can occur at all steps in the manufacturing process (cell culture, purification, and formulation), storage, distribution, and handling of products. It results from various kinds of stress such as agitation and exposure to extremes of pH, temperature, ionic strength, or various interfaces (e.g., air–liquid interface). High protein concentrations (as in the case of some monoclonal antibody formulations) can further increase the likelihood of aggregation. Therefore, IB aggregation state needs to be carefully characterized and controlled during development, manufacture, and subsequent storage. Similarly, by monitoring the state of aggregation, modification or optimization of the production process can be achieved. There are two principal techniques that allow either static size determination or monitoring IB aggregation state analysis: dynamic light scattering, DLS (also known as photon correlation spectroscopy, PCS) and a recently developed laser-based nanoparticle tracking analysis (NTA) system. DLS has been applied to structural studies of IBs [3] and also multiple amyloidogenic proteins, including insulin, calcitonin [4], and the model protein barstar [5].

1.3 Determination of Zeta Potential

IBs can also be characterized by means of their zeta potential, which is related to their surface charge and can provide information about their stability and interaction with different dispersing media. Zeta potential, which gives information about the overall charge that a particle acquires in a certain medium, indicates the degree of repulsion between adjacent, similarly charged particles in dispersion and provides, as a result, important information about the stability of disperse systems. Samples with high absolute zeta potential values (normally above 30 mV) are electrically stabilized, while those with low zeta potentials are not stable and tend to coagulate or flocculate. Zeta potential measurements can also be used for studying interactions of particles, in this case IBs, with different molecules present in the dispersing medium by monitoring changes in their surface charge [3].

1.4 Determination of the Morphology

AFM is a technique of choice to investigate morphology and conformation of protein aggregates [6]. This method has been used to visualize details of fibril morphology [7], chirality [8], or

conformational changes during incubation [9]. It enabled in situ investigation of insulin aggregation [10, 11] and other proteins [12] and insight into the mechanism of aggregate formation [13–15]. Advanced studies on interaction of amyloid aggregates and lipid bilayers also take profit of the capabilities of this technique [16].

1.5 Determination of the Internal Structuration

Cryo-transmission electron microscopy (cryo-TEM) is a technique for characterizing the morphology and also the internal structuration of IBs [17]. One of the main advantages of this type of electron microscopy is that IBs do not need to be fixed, stained, or removed from the suspension to be visualized.

1.6 Determination of the Stiffness

Using force spectroscopy AFM it is also possible to get information about the stiffness of a material. It is known that the mechanical properties of a substrate biomaterial, like IBs, can critically affect relevant features like the mammalian cells growing on it [18]. Specifically, it can be very relevant for the cell morphology, proliferation, and differentiation [19]. Therefore, the study of the mechanical properties of different protein aggregates is crucial for their future application in tissue engineering. During the last decade, AFM, and particularly force spectroscopic atomic force microscopy (FSAFM), has proved its value not only for imaging biological samples, but also for measuring inherent properties of biological structures and aggregates, like local interaction forces, mechanical properties, or dynamics in natural (physiological) environments, all of them very interesting properties from the biomedical point of view.

1.7 Determination of the Wettability

Contact angle is an easy and straightforward technique that is normally associated with investigation of surface properties of materials. However it can also be used to cast light on more complex phenomena such as protein adsorption, i.e., the interaction of different IBs with surfaces with different wettabilities enabled to establish that the genetic background of the producing bacteria influences the interaction between the protein aggregates and the surface [17]. This is due to the fact that chemical charge and wettability are very important factors among surface's chemical properties that influence macroscopic behavior of surfaces modified with these biomaterials.

2 Materials

The preparation of protein aggregates or bacterial IBs and its purification procedures can be obtained following different strategies, some of them also described in this book (Chapter 15).

1. Protein aggregates or IB resuspension buffer (*see* **Note 1**): Phosphate saline buffer (PBS), pH 7.4, 94 mM NaCl, 3.1 mM Na_2HPO_4, 0.9 mM NaH_2PO_4 800 mL MQ H_2O, adjust to pH 7.4 with HCl, gauge to 1 L.

2. Zetasizer Nano Series dynamic light scattering analyzer from Malvern Instruments, UK.

3. NTA technology, Nanosight Instruments, UK.

4. Mica surfaces (always cleaved just before the drop cast of the protein suspension) (*see* **Note 2**).

5. Perking Elmer Spectrum One ATR-FTIR.

6. Atomic force microscope (AFM) PicoSPM 5100 from Molecular Imaging Agilent Technologies, Inc., Santa Clara, CA, USA: AFM cantilever with a monolithic supersharp silicon SSS-NCH-50 (Nanosensors, Inc.) tip, a radius of 2 nm, a nominal spring constant of 10–130 N/m, and a resonance frequency of 204–497 kHz was used.

7. Free software Gwyddion, WSxM, or equivalent.

8. AFM (MFP-3D-SA, Asylum Research, Santa Barbara, USA) equipped with a close loop tracking system and working on liquid environment. Pyramidal NSC35/AlBS silicon tips (Mikromasch, USA) having nominal spring constants of 0.28 N/m were used. All solutions were prepared using PBS buffer solution, pH 7.4 and were stored at 4 °C (unless indicated otherwise).

9. Jeol JEM-2011 transmission electron microscope (JEOL, LTD., Tokyo, Japan).

10. Advantage A10 Water System Production Unit or similar to supply MQ H_2O (18.2 MΩ).

11. Mixed SAMs: 1-undecanethiol and 11-mercapto-1-undecanol.

12. A piranha solution: Mix sulfuric acid with 35 % aqueous H_2O_2 in proportion 3:7. (Handle with extreme caution!)

13. The gold substrates (SSens, Hengelo, The Netherlands).

14. Tweezers, glassware, etc. for safe and clean sample handling.

15. OCA 15+ (Dataphysics, Germany) contact angle goniometer. Software SCA20 for data treatment and angle determination.

3 Methods

3.1 Infrared Spectroscopy (IR) for the Secondary Structure

Attenuated total-internal reflection Fourier transform infrared (ATR-FTIR) spectroscopy is a powerful method for recording IR spectra of biological samples. The main advantage of ATR-FTIR compared to the traditional FTIR is that the signal gets amplified many times. In the traditional FTIR, the sample and the IR light interact only once. However, in ATR-FTIR due to the repeated

reflections of the incident beam on the trapezoid plate, the sample has more chance to absorb the incident IR light resulting in greater absorption. This leads to high signal-to-noise ratio. Further, the background due to the solvent is also minimized. Comparison between native, water-soluble species proteins and aggregated forms of the same proteins leads to the conclusion that the 1,621 and 1,691 cm^{-1} bands are indicative of extended antiparallel pleated beta-sheet structures. These bands have also been applied to study kinetics of protein aggregation during IB's formation [20]. Another band commonly found in proteic FTIR spectra is around 1,651 cm^{-1} and it has been attributed to the presence of disordered structures [21]. The experimental procedure to measure the ATR-FTIR consists in the following steps:

1. Spread a drop (5–50 μL) of the IB's suspension on a germanium or diamond plate and wait for a slow evaporation of the solvent, typically under a stream of dry nitrogen gas. This enables minimal contribution from the solvent in the spectra.

2. Record and analyze the spectra of the dried aggregates directly on the spectrometer (e.g., Perkin Elmer Spectrum One ATR-FTIR).

3. It is advisable to acquire a blank spectrum before measuring the samples in order to quantify the background signal and for baseline correction.

4. Acquire ~20 interferograms in the range 1,550 and 1,700 cm^{-1} (at resolution of 2 cm^{-1}) and average them to obtain the spectra. For spectra aligning *see* **Note 3**.

3.2 Dynamic Light Scattering (DLS) for Sizing

The principle of dynamic light scattering is that fine particles and molecules that are in constant random thermal motion, called Brownian motion, diffuse at a speed related to their size. Thus, smaller particles diffuse faster than larger particles. The speed of Brownian motion is also determined by the temperature and therefore precision temperature control is essential for accurate size measurement. To measure the diffusion speed, a speckle pattern produced by illuminating the particles with a laser is observed. The scattering intensity at a specific angle will fluctuate with time, and this is detected using a sensitive avalanche photodiode detector (APD). The intensity changes are analyzed with a digital autocorrelator which generates a correlation function. This curve can be analyzed to give the size and the size distribution. Measurement range is between 0.3 nm and 10 μm. The experimental procedure to measure the particle size or monitoring the aggregation state of a suspension of IBs is done using a Zetasizer Nano Series dynamic light scattering analyzer (Malvern Instruments, UK). Figure 1 shows an example of IB size distribution of IBs aged for 3 h and isolated from different bacterial strains. The experimental procedure to measure the particle size or monitoring the aggregation state of a suspension of IBs using a Zetasizer Nano Series dynamic
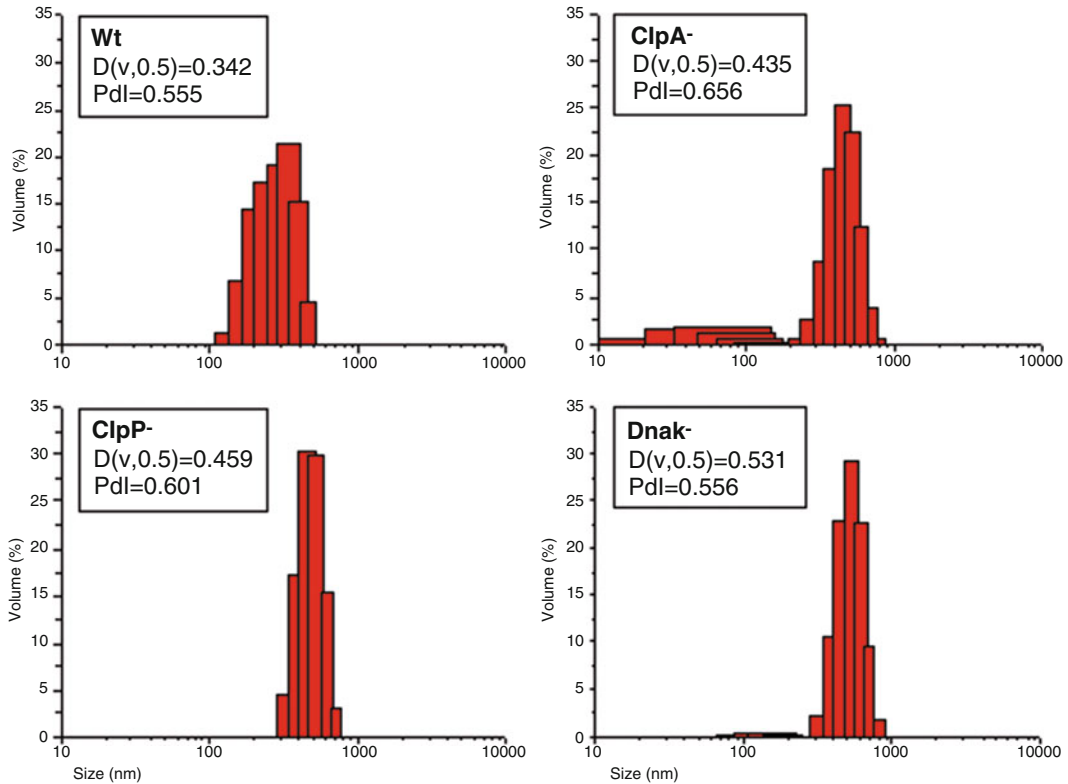
Fig. 1 IB size distribution of IBs aged for 3 h and isolated from different bacterial strains. IB D(v, 0.5) is the volume median particle diameter [nm] and the PdI is defined as [D(v, 0.1)/D(v, 0.9)] × 100. Reproduced from ref. 3 with permission from Wiley-VCH

light scattering analyzer (Malvern Instruments, UK) comprises the following steps:

1. Load the appropriate cell (disposable polystyrene cells for aqueous samples) previously cleaned with the sample solvent, with a minimum of 1 mL of IBs' suspension. Ideally the sample should cover 1 cm of the cell height. Avoid air bubble formation in the cell by slowly introducing the IBs' suspension with the aid of a pipette. Remove any spillage and cover the cell.

2. Insert the cell in the instrument and allow temperature equilibrium to be established. Make sure that the weld line faces towards the front of the instrument.

3. In the Zetasizer software introduce parameters regarding sample, sample measurement, and data processing into the software. Those include refractive index and absorption of IBs, viscosity, dielectric constant and refractive index of the dispersant, type of cell (disposable cuvettes), temperature of measurement (25 °C), measurement angle (173° Backscatter, NIBS default), measurement duration (automatic) number of measurements (3), and data processing mode (general purpose).

4. Start the measurement.

3.3 Nanoparticle Tracking Analysis for Sizing and Counting

Nanoparticle tracking analysis (NTA) is a method for the direct and real-time visualization and analysis of nanoparticles in liquids. Based on a laser-illuminated microscopical technique, Brownian motion of nanoparticles is analyzed in real time by a CCD camera. Each particle is simultaneously but separately visualized and tracked by a dedicated particle tracking image analysis program. Because each and every particle is visualized and analyzed separately, the resulting estimate of particle size and particle size distribution does not suffer from the limitation of being intensity weighted as it happens with the *Z*-average distribution which is normal in conventional ensemble methods like dynamic light scattering (DLS) of particle sizing in this size regime. The ability of NTA to simultaneously measure particle size and particle scattering intensity allows heterogeneous particle mixtures to be resolved and, importantly, particle concentration can be estimated directly—the particle size distribution profile obtained by NTA is a direct number/frequency distribution. Measurement size range is between 10 nm and 1 μm. The experimental procedure to measure the particle size using NTA technology (Nanosight Instruments, UK) comprises the following steps:

1. Check that the sample chamber is clean prior to a measurement. To do this, check that the solvent which is being used to dilute the sample is totally free from nanoparticles. If not, use water or a water-ethanol mixture to clean the sample chamber.

2. Load the sample (0.5 mL approx.) into the sample chamber of the laser module (LM) viewing unit using a syringe (without needle). To avoid generating pressures which might result in the sample bypassing the seals in the sample chamber or damaging the window, care must be taken to introduce the sample slowly. Similarly, care should be taken to avoid the introduction of bubbles at this stage. The sample unit should be tilted such that the syringe is injecting vertically upwards, allowing the chamber to be filled slowly against gravity. A properly prepared and loaded sample will appear as a clear sample through which the laser beam can be seen as a thin line passing through the sample chamber.

3. Place the LM unit onto the microscope stage and adjust the position and height of the microscope objective to be used to obtain a clear image of particles present within the beam. Once an image can be seen either by viewing via the oculars or by the camera, an analysis is ready to be performed.

4. Capture a video using the NTA software by choosing the appropriate camera settings (Fig. 2). Those include the following:

   (a) *Gain and shutter*. To see all particles set the shutter and gain to maximum. Slowly reduce shutter keeping an eye on the dimmest particles until they start to disappear and

Fig. 2 Example of an optimal image to allow accurate particle centering by NTA

then increase the shutter a little. Next, carefully decrease the gain but only when you see large and saturated blobs. Remember not to lose any particle.

(b) *Brightness.* Set the brightness so as to detect very dim particles, avoiding that larger particles appear as overexposed bright white discs.

(c) *Alternate gamma.* Use this option when very small particles are present and with polydisperse systems.

(d) *Capture time.* Set longer capture duration times for larger particles with slow Brownian motion and shorter times for small particles, typically 30 s for particles <200 nm.

5. Using the first video frame adjust those filtration settings to identify the center of the particles on the screen without highlighting any areas of background noise.

(a) Adjust gain and brightness as described before.

(b) Adjust blur to smooth the intensity profile for each particle and remove false centers from the analysis.

(c) Adjust detection threshold such that all clearly recognizable particles have been centered with a red cross.

(d) Introduce the maximum expected particle size derived from previous knowledge or by determining the maximum distance a particle jumps in the sample using the "Max Jump Toggle" tool. This will automatically define the minimum track length parameter.

(e) Adjust the temperature and viscosity.

6. Start video processing.

To measure particle concentration the following must be considered. It is likely that the sample will need diluting in order to bring it down to a suitable concentration for the NTA technique. Firstly assess your sample by eye. In a standard 5 mL clear container, if the sample is not totally transparent it is probably too

concentrated. If this is the case, dilute 100× and reassess. When the sample appears transparent or very near transparent you can assess using NTA by getting a 1 sec NTA video to count the number of particles, which ideally should lie between 20 and 60. Repeat the analysis 2–3 times to make sure that the measurement results are repeatable but avoid the settling issue. When you are recording the same sample several times, do not change capture and analysis setting significantly as it can give different concentration results.

3.4 Z-Potential for the Colloidal Stability

Zeta potential is measured by applying a voltage across a pair of electrodes at either end of a cell containing the particle dispersion. Particles will migrate towards the oppositely charged electrode with a velocity proportional to the magnitude of their zeta potential. This velocity is measured and expressed as particle velocity in a unit electric field (electrophoretic mobility) and can be conveniently investigated by electrophoretic light scattering using a dynamic light scattering analyzer. In order to convert electrophoretic mobility (U_E) to zeta potential (ζ) through Henry's equation, it is important to know the viscosity (η) and dielectric constant (ε) of the dispersing medium. The experimental procedure to measure the zeta potential of a suspension of IBs using a Zetasizer Nano Series dynamic light scattering analyzer (Malvern Instruments, UK) comprises the following steps:

1. Load the zeta potential cell with circa 1 mL of IBs' suspension. In order to do so, invert the cell and slowly inject the sample from a syringe (without needle) into the cell through one of the ports until filling the U tube to just over half way. Check no air bubbles form in the cell. Turn the cell upright and continue injecting slowly until liquid reaches the bottom of the port. Remove the syringe and place the cell stoppers. Remove any spillage from electrodes.

2. Insert the cell in the instrument. Make sure that the weld line faces towards the front of the instrument.

3. Introduce parameters regarding sample, sample measurement, and data processing into the software. Those include refractive index and absorption of IBs, viscosity, dielectric constant and refractive index of the dispersant, Henry's function F (Ka) (Smoluchowski approximation), type of cell (zeta potential cell), temperature of measurement (25 °C), type of measurement (automatic), number of measurements (3), and data processing mode (auto mode).

4. Start the measurement.

3.5 Atomic Force Microscopy (AFM) for the Morphology

AFM images the topography of samples adsorbed on atomically flat smooth surfaces, typically mica. A cantilever tip scans the surface contour of the specimen and, upon contact, a repulsive force in the pico Newtons–nano Newtons (pN–nN) range bends the cantilever

upward. A laser beam focused on the end of the cantilever detects the extent of bending and the deflection of the laser is translated into force units by a photodetector. By keeping the force constant while scanning across the surface, the vertical movement of the tip generates the surface contour, which is recorded as the topography map of the sample. AFM has been modified for specific applications and can be used in different modes. In contact mode, the tip is permanently in contact with the sample, whereas in "tapping mode" (also referred to as intermittent-contact or dynamic-force mode), a stiff cantilever oscillates close to the sample. Part of the oscillation extends into the repulsive regime between the tip and the analyzed surface, so that the tip intermittently touches, or "taps," the surface. This mode provides good resolution on soft samples and is useful for investigation of soft materials like prefibrillar species. Figure 3 shows an example of AFM characterization of GFP IBs aged for 3 h. The experimental procedure to measure the topography of protein aggregates or IBs using an atomic force microscopy (PicoSPM 5100 from Molecular Imaging Agilent Technologies, Inc., Santa Clara, CA, USA) comprises the following steps:

1. IBs were resuspended in a buffer solution to obtain a working suspension with concentration of 0.4 mg/mL.

2. Working suspension was ultrasonicated for 10 min to ensure appropriate dispersion and disaggregation of the protein aggregates.

3. Surface preparation: 1×1 cm mica surfaces were cut from a large mica plate with common scissors.
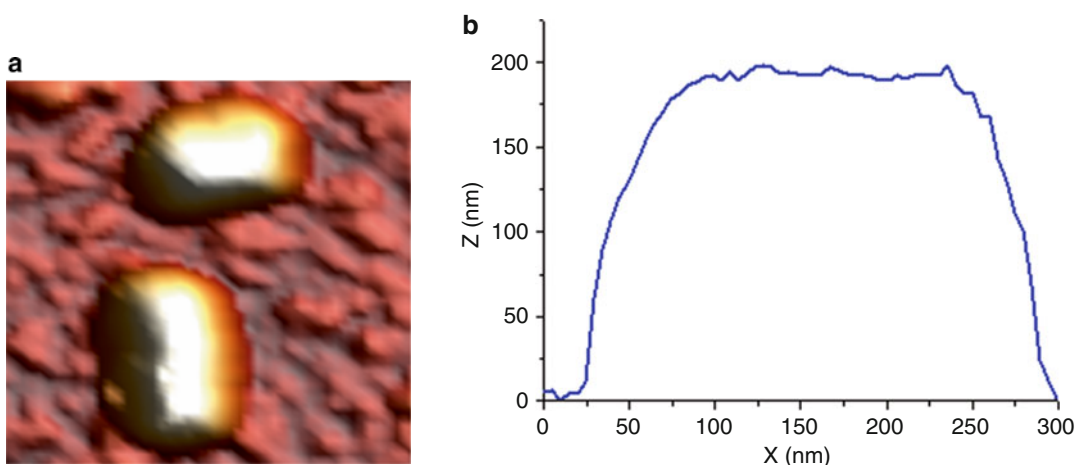
Fig. 3 AFM characterization of GFP IBs aged for 3 h. (**a**) 600×600 nm topography 3D image of randomly deposited wt IBs on a mica surface. (**b**) Topography cross section of an isolated IB particle. Reproduced from ref. 3 with permission from Wiley-VCH

4. Mica surfaces were cleaved using regular adhesive tape immediately before use.

5. A few drops of IBs in the same PBS buffer (20 micro g /mL) working suspension were deposited onto the freshly exposed mica and incubated for 5 min (*see* **Note 2**).

6. AFM measurements were carried out over the protein aggregate deposited on the substrate. Samples were analyzed in dynamic mode working at 8 kHz of frequency and in a liquid environment (PBS buffer, pH 7.4) in order to mimic the cytoplasmic environment of the cell. Pyramidal NSC35/AlBS silicon tips (Mikromasch, USA) having nominal spring constants of 0.28 N/m were used.

7. AFM images were numerically treated with the Gwyddion software in order to remove artifacts, sample curvature, etc. and to present the results in a convenient way (*see* **Notes 2** and **4**).

3.6 Cryo-Transmission Electron Microscopy (cryo-TEM) for the Internal Structuration

By cryo-TEM, the morphology and internal structuration of the sample are studied at cryogenic temperatures. The material is preserved in a frozen hydrated state by a rapid freezing process, usually in liquid ethane, and maintaining it at liquid nitrogen temperature or colder during the imaging with the electron microscope. The experimental procedure for analyzing the structure of IBs in suspension by cryo-TEM comprises the following steps:

1. Place 3 μL of IBs' suspension on the microscope sample support, a copper grid coated with a perforated polymer film.

2. Dry the excess of sample by carefully blotting it with filter paper.

3. Submerge the grid into liquid ethane at a temperature just above its freezing point (−179.1 °C).

4. Rapidly place the frozen sample in the holder, which is refrigerated with liquid nitrogen.

5. Introduce the sample holder into the microscope and keep it cold during sample imaging by adding liquid nitrogen when necessary.

6. Visualize the sample and take images of IBs. The recommended voltage is 120 kV. IBs are not easy to visualize since they do not present a very defined structure. A too long exposure might damage the sample and the polymer grid covering the copper grid.

3.7 Force Spectroscopy AFM for the Stiffness

AFM indentation measurements can be done with a commercial AFM (MFP-3D-SA, Asylum Research, Santa Barbara, USA) equipped with a close loop tracking system and working on liquid environment. Figure 4 shows histogram representations of the number of force spectroscopy curves vs. Young modulus obtained for each indentation performed on IBs produced in different
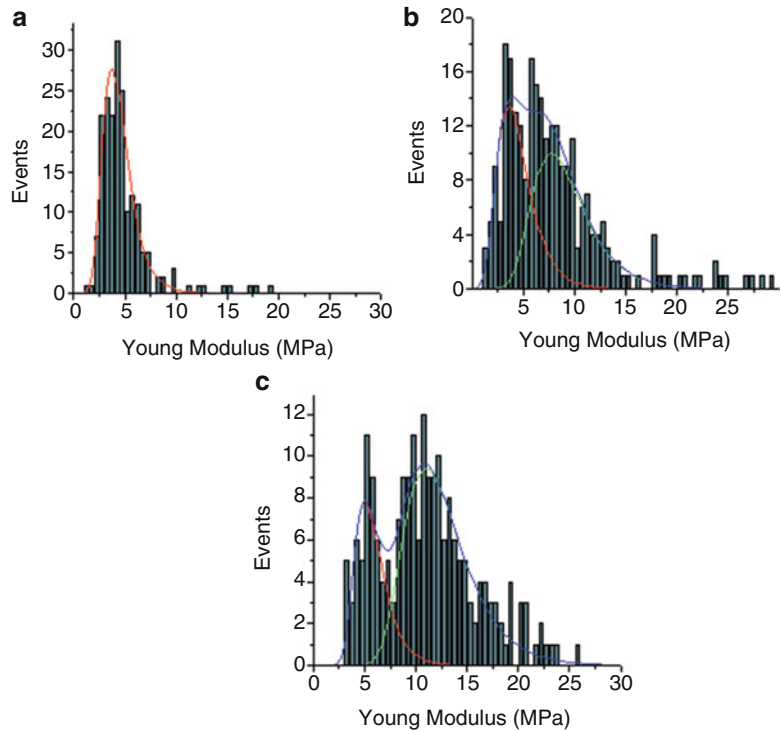
Fig. 4 Histogram representation of the number of events vs. Young modulus for IBs produced in bacterial mutants. (**a**) Wt IBs showing only one peak at 3.73 MPa; (**b**) DnaK IBs show two overlapped Young modulus distributions which centered at 3.56 and 7.75 MPa; (**c**) ClpA IB shows the presence of two different young modulus distributions, at 5.01 and 10.99 MPa. Reproduced from ref. 18 with permission from Elsevier

bacterial mutants. The experimental procedure for analyzing the stiffness of IBs by FS-AFM comprises the following steps:

1. Surface preparation: 1×1 cm mica surfaces were cut from a large mica plate with common scissors.

2. The spectroscopic calibration: The new cantilevers need to be calibrated by measuring force vs. distance curves, on liquid media (PBS buffer solution, pH 7.4) on the freshly cleaved mica surface. The force curves consist of 2,048 data points imposing a maximum applied force of 50 nN at a frequency of 8 kHz. Pyramidal NSC35/AlBS silicon tips (Mikromasch, USA) having nominal spring constants of 0.28 N/m were used.

3. Protein aggregate suspension: 20 mg/mL protein aggregate suspensions were obtained by the addition of the appropriate amount of PBS buffer solution with a micropipette. The suspension was vigorously shacked for 3 min to ensure appropriate dispersion and disaggregation of the protein aggregates.

4. Sample preparation was done by drop casting four drops of the protein aggregate suspension over freshly cleaved mica.

5. The mica surface was scanned using noncontact AFM until an isolated physisorbed protein aggregate was found.

6. Force spectroscopy measurements were carried out over the protein aggregate using the same conditions used in the calibration process (**step 2**).

7. AFM mechanical properties data were calculated from force distance curves according to the procedures described in ref. 22.

3.8 Contact Angle (CA) for the Wettability

Wettability of self-assembled monolayer of mixed thiols on gold, with different hydrophilicities is determined before and after being in contact with IBs with an OCA 15+ (Dataphysics, Germany) contact angle goniometer. Data treatment and angle determination were carried out with the software SCA20 (Dataphysics, Germany). Four sets of static contact angles, at different positions on each sample, were measured. The experimental procedure for performing CA comprises the following steps:

1. To prepare mixed SAMs with different degree of hydrophobicity, solutions with different molar ratios of hydrophilic (–OH terminated) and hydrophobic (–CH$_3$ terminated) alkanethiols were prepared (~1 mM). Different ratios of each alkanethiol (e.g., 1:0, 0.75:0.25, 0.5:0.5, 0.25:0.75, 0:1) in ethanol are used.

2. To prepare the IB suspension, it is needed to resuspend the IB suspension in PBS to obtain a concentration of 20 μg/mL. Then, it is needed to sonicate for 10 min. If stored, sonicate again for 5 min immediately before use.

3. Prepare a piranha solution: Mix sulfuric acid with 35 % aqueous H$_2$O$_2$ in proportion 3:7. (Handle with extreme caution!)

4. Clean gold substrates:
   (a) Sonicate for 5 min in ethanol.
   (b) Dry with a nitrogen flow.
   (c) Immerse in piranha solution for 30 s.
   (d) Rinse with copious amounts of MQ H$_2$O.
   (e) Dry with a nitrogen flow.

5. Deposition of the mixed SAMs on the gold substrate:
   (f) Immerse the gold substrates into the previously prepared mixed thiol solutions (**step 1**) for 24 h at RT.
   (g) Rinse the gold-functionalized substrate with ethanol.
   (h) Sonicate the gold-functionalized substrate in ethanol for 5 min.
   (i) Dry it under a nitrogen flow.

6. Immerse the functionalized gold substrates (**step 5**) in the IB suspension (**step 2**) for 2 h in order that the IBs deposit on the surface.

7. Rinse gently with MQ H_2O and dry with nitrogen.

8. Perform contact angle measurements on the different IB's covered substrates and on the referenced surfaces without IBs. In order to obtain better statistics, it is preferable to triplicate the measurements.

9. Plot the contact angle obtained for each substrate before and after being functionalized with the IBs (*see* **Note 5**).

4   Notes

1. Other buffers (such as PBS, Tris, MQ H_2O, DMSO, H_2O–EtOH, NaCl solutions) can also be used as long as they resuspend aggregates and do not interact with substrate.

2. AFM measurements can also be performed with other substrates including Si wafers, glass, and surfaces with evaporated gold.

3. Second derivatives of the amide I band spectra can also be used to determine the frequencies at which the different spectral components are located.

4. For the measurements in liquid media, in case the density of protein aggregates of the working suspension is low enough and does not interfere with the cantilever it is possible to image the suspension directly and the cantilever can be left during scanning. In the opposite case, it is recommended to change the media for buffer/solution without protein aggregates after incubation for 5 min. Care must be taken to ensure that liquid does not evaporate during observations as most of the setups permit to use small amounts of media (usually a few mL) and scanning times are usually long.

5. Samples should be as fresh as possible. Any delay can influence the properties of the proteic material.

Acknowledgement

References

1. Carrio M, Gonzalez-Montalban N, Vera A et al (2005) Amyloid-like properties of bacterial inclusion bodies. J Mol Biol 347:1025–1037

2. García-Fruitós E, Gonzàlez-Montalbán N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Factories 4:27

3. García-Fruitós E, Rodriguez-Carmona E, Díez-Gil C et al (2009) Surface cell growth engineering assisted by a novel bacterial nanomaterial. Adv Mater 21:4249–4253

4. Avidan-Shpalter C, Gazit E (2006) The early stages of amyloid formation: biophysical and structural characterization of human calcitonin pre-fibrillar assemblies. Amyloid 13:216–225

5. Kumar S, Mohanty SK, Udgaonkar JB (2007) Mechanism of formation of amyloid protofibrils of barstar from soluble oligomers: evidence for multiple steps and lateral association coupled to conformational conversion. J Mol Biol 367:1186–1204

6. Li H, Rahimi F, Sinha S et al (2009) Amyloids and protein aggregation–analytical methods. In: Meyers RA (ed) Encyclopedia of analytical chemistry. John Wiley & Sons, New York

7. Stine WB, Snyder SW, Ladror US et al (1996) The nanometer-scale structure of amyloid-i3 visualized by atomic force microscopy. J Protein Chem 15(2):193–203

8. Rubin N, Perugia E, Goldschmidt M et al (2008) Chirality of amyloid suprastructures. J Am Chem Soc 130:4602–4603

9. Apetri MM, Maiti NC, Zagorski MG et al (2006) Secondary structure of a-synuclein oligomers: characterization by raman and atomic force microscopy. J Mol Biol 355:63–71

10. Mauro M, Craparo EF, Podestà A et al (2007) Kinetics of different processes in human insulin amyloid formation. J Mol Biol 366:258–274

11. Jansen R, Dzwolak W, Winter R (2005) Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. Biophys J 88:1344–1353

12. Ortega-Vinuesa JL, Tengvall P, Lundstrom I (1998) Aggregation of HSA, IgG, and fibrinogen on methylated silicon surfaces. J Colloid Interface Sci 207:228–239

13. Liu R, McAllister C, Lyubchenko Y et al (2004) Residues 17–20 and 30–35 of beta-amyloid play critical roles in aggregation. J Neurosci Res 75:162–171

14. Hoyer W, Cherny D, Subramaniam V et al (2004) Rapid self-assembly of a-synuclein observed by in situ atomic force microscopy. J Mol Biol 340:127–139

15. Goldsbury C, Green J (2005) Time-lapse atomic force microscopy in the characterization of amyloid-like fibril assembly and oligomeric intermediates. Methods Mol Biol 299:103–128

16. Lashuel HA, Lansbury PT (2006) Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? Q Rev Biophys 39:167–201

17. Cano-Garrido O, Rodríguez-Carmona E, Díez-Gil C et al (2013) Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. Acta Biomater 9:6134–6142

18. Díez-Gil C, Krabbenborg S, García-Fruitós E et al (2010) The nanoscale properties of bacterial inclusion bodies and their effect on mammalian cell proliferation. Biomaterials 31:5805–5812

19. Tatkiewicz WI, Seras-Franzoso J, García-Fruitós E et al (2013) Two-dimensional microscale engineering of protein-based nanoparticles for cell guidance. ACS Nano 7:4774–4784

20. Ami D, Natalello A, Gatti-Lafranconi P et al (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett 579:3433–3436

21. Dong AD, Huang P, Caughey WS (1990) Protein secondary structures in water from second-derivative amide I infrared spectra. Biochemistry 29:3303–3308

22. Parra A, Casero E, Lorenzo E et al (2007) Nanomechanical properties of globular proteins: lactate oxidase. Langmuir 23:2747–2754

Chapter 23

Predicting the Solubility of Recombinant Proteins in *Escherichia coli*

Roger G. Harrison and Miguel J. Bagajewicz

Abstract

We describe a statistical model that uses binomial logistic regression for predicting the solubility of heterologous proteins expressed in *E. coli*. The model is based on a set of proteins reported to have been expressed in *E. coli* in either soluble or insoluble form. The 22 parameters used in the final model based on proteins' amino acid composition are discussed. The overall accuracy of the model developed is 94 %. The way to use this model on the website http://www.ou.edu for the prediction of protein solubility is explained.

Key words Heterologous protein solubility prediction, *Escherichia coli*, Binomial logistic regression model

1   Introduction

The expression of heterologous proteins in *Escherichia coli* is often desirable when posttranslational modifications such as glycosylation are not required, but there are sometimes proteins with the target protein being insoluble as aggregates in "inclusion bodies" within the cell. As an aid to researchers and others involved in producing recombinant proteins in *E. coli*, we have developed a model to predict the solubility of these proteins that uses a statistical technique known as logistic regression [1]. For this model, only the amino acid composition of the protein is required. The overall accuracy of this model is 94 %. The protein database, the type of model, and the parameters used in this model are outlined.

1.1   Protein Database   Proteins were selected for inclusion in the database using literature searches to find studies where the solubility or insolubility of a protein expressed in *E. coli* was determined. Only proteins expressed at 37 °C without fusion proteins or chaperones were considered, and membrane proteins were excluded. The expression of the target protein as a fusion with a solubility-promoting protein or the coexpression of folding chaperones can make an insoluble protein

soluble by helping improve folding kinetics or changing its interactions with solvent [2, 3]. This can give false positives, making an inherently insoluble protein soluble. The temperature chosen is a common temperature for much work done with *E. coli*, and it had to be consistent because temperature has been shown to affect protein solubility and the formation of inclusions bodies [4]. In determining the sequence of each protein expressed, signal peptide sequences that were not part of the expressed protein were excluded due to their hydrophobic nature. The signal peptide sequence of a protein is a short (5–60) stretch of amino acids, and these are found in secretory proteins and transmembrane proteins. The removal of these signal sequences does not affect the prediction of protein solubility because at some point in the folding pathway of these proteins, the signal sequence is removed.

The database for the model contains a total of 160 insoluble proteins and 52 soluble proteins. The solubility or insolubility of the 212 proteins was assigned as follows:

- Proteins that appeared almost entirely in the inclusion body were classified as insoluble proteins.
- Conversely, if a significant amount of the protein appeared in the soluble fraction, the protein was classified as soluble.

The significance of the expression of the protein in the soluble fraction was determined by the sodium dodecyl sulfate polyacrylamide electrophoresis gel electrophoresis (SDS-PAGE) analysis of the proteins expressed when available. Proteins that showed bands in the soluble lanes that were more than faintly visible were identified as having a significant amount of protein in the soluble fraction. When the SDS-PAGE result was unavailable, the protein was classified according to the qualitative information obtained from its described expression in *E. coli*. The reason for assigning the proteins this way was due to their overexpression in *E. coli*. Overexpression causes conditions where even soluble proteins will form inclusion bodies due to the cell becoming overly crowded [5]. Hence, when proteins were expressed in significant amounts in both the soluble fraction and inclusion bodies, it was assumed that the inclusion bodies were formed due to overexpression and that under normal expression the protein would fold correctly and be soluble.

1.2 Model Type

Binomial logistic regression was used to develop this model. This is a form of regression that is used when the dependent variable is a dichotomy (it belongs to one of two nonoverlapping sets) and the independent variables are of any type (continuous or categorical, i.e., belonging to one or more categories without an intrinsic ordering to the categories) [6]. In our case, the dichotomy is soluble/insoluble, and the independent variables are the parameters. Thus, the goal was to develop a model capable of separating data into these two categories, depending on properties of proteins

that could affect positively or negatively the solubility. In this model, we considered interactions between any sets of two variables to determine if the effect of any one variable is not constant over the level of any other variable. We created interaction variables by taking the arithmetic product of pairs of original parameters. The model equation used for logistic regression is

$$\log\left[\frac{p_i}{1-p_i}\right] = \alpha + \sum_{j=1}^{n}\beta_j x_{j,i} + \sum_{k=1}^{n}\sum_{j=1; j\neq k}^{n} \gamma_{k,j} x_{j,i} x_{k,i}$$

where p_i is the probability for a datum to belong to one group, n is the number of characteristic parameters integrated in the model, α is an intercept constant, β_j is a coefficient for the parameter j, $x_{j,i}$ is the value for the parameter j for datum i, $x_{k,i}$ is the value for the parameter k for datum i, and $\gamma_{k,j}$ is the interaction coefficient for parameters j and k. Interactions between the same parameter were not included, so j is not equal to k.

1.3 Model Parameters

When the model was originally developed, a total of 32 protein-related parameters were studied. These included some parameters that depended on the sequence of the amino acids and not just the amino acid composition. However, after determining the model that gave the highest prediction accuracy, it was found that only 22 parameters needed to be included, none of which involved the protein sequence:

- Aliphatic index
- Charge average
- Hydrophilicity index
- Isoelectric pH (pI)
- Molecular weight
- Alanine fraction
- Arginine fraction
- Asparagine fraction
- Aspartic acid fraction
- Glutamic acid fraction
- Glutamine fraction
- Histidine fraction
- Isoleucine fraction
- Leucine fraction
- Lysine fraction
- Methionine fraction
- Phenylalanine fraction

- Proline fraction
- Serine fraction
- Threonine fraction
- Tryptophan fraction
- Valine fraction

These parameters are calculated by the model as follows: The aliphatic index (AI) is calculated using the following equation [7]

$$AI = \frac{n_A + 2.9 n_V + 3.9 (n_I + n_L)}{n_{tot}}$$

where n_A, n_v, n_I, n_L, and n_{tot} are the number of alanine, valine, isoleucine, leucine, and total residues, respectively. The charge average is found by taking the absolute value of the sum of the difference between the positively charged lysine and arginine amino acids and the negatively charged aspartic acid and glutamic acid amino acids and dividing by the total number of amino acid residues. The hydrophilicity index is found by summing the hydrophilicity indices for all the amino acids [8] and dividing by the total number of amino acids. The pI (pH at which the net charge of the protein is zero) and molecular weight are obtained with the aid of the pI/MW calculation tool from the Swiss Institute of Bioinformatics (ExPASy Proteomics Server, website address http://ca.expasy.org). The fraction of each amino acid listed above is obtained by dividing the number of amino acid residues in the sequence by the total number of residues.

1.4 Model

The final model chosen has an overall accuracy of 94 % (calculated based on 96 % and 86 % accuracy for the insoluble and soluble proteins in the database, respectively) [1]. This model involved only γ interaction coefficients and the α intercept constant in the logistic regression equation. The use of this model to predict the solubility in *E. coli* of a protein for which the sequence is known is relatively simple—all that is needed is to substitute the values of the parameters for the protein into the logistic regression equation for the probability of being either soluble or insoluble. For convenient use of this solubility prediction model, we have set up a website at http://www.biotech.ou.edu.

2 Materials

1. Protein database.
2. ExPASy Proteomics Server, website address http://ca.expasy.org.
3. http://www.biotech.ou.edu website.

3   Methods

3.1   Specification of the Protein Sequence

The sequence of the protein needs to be specified where each amino acid is indicated by its one-letter (lower or upper case) abbreviation. Sequences can be copied directly from protein databases such as GenBank (*see* **Note 1**). It is not necessary to remove numbers given with the sequence.

3.2   Determination of Isoelectric pH and Molecular Weight

Paste the sequence obtained in **item 1** of Subheading 2 into the window in p*I*/MW tool from the Swiss Institute of Bioinformatics (ExPASy Proteomics Server, website address http://ca.expasy.org). This will give the p*I* and molecular weight of the protein.

3.3   Prediction of Protein Solubility

Go to the http://www.biotech.ou.edu website for the prediction of protein solubility in *E. coli*. In the windows for the average p*I* and molecular weight for the Diaz et al. model [1], paste the values for these determined in **item 2** of Subheading 2 into their respective windows. Paste the protein sequence found in **item 1** of Subheading 2 into the protein sequence window. Then press "Submit Query." The output gives the percent chance of solubility along with the number of each of the amino acids in the sequence (*see* **Note 2**).

4   Notes

1. It is important that the signal peptide sequence be removed before pasting the sequence into the window on the http://www.biotech.ou.edu website. Signal peptide sequences are usually specified in the protein database used along with the "mature peptide," which is the protein without the signal sequence.

2. The model we have developed can be used to make experimental work involving recombinant protein expression more efficient. Proteins with a high predicted probability of solubility (>75 %) can very likely be expressed in soluble form at 37 °C, without the need for expression using a fusion or chaperone to promote solubility. Proteins with intermediate predicted probability of solubility (50–75 %) are possibly soluble when expressed at temperatures lower than 37 °C. Finally, proteins with a lower predicted probability of solubility (<50 %) may need other measures to obtain a soluble protein, such as expression with a solubility-promoting fusion or the coexpression of a chaperone or chaperones.

Acknowledgment

References

1. Diaz AA, Tomba E, Lennarson R et al (2009) Prediction of protein solubility in Escherichia coli using logistic regression. Biotechnol Bioeng 105:374–383

2. Davis GD, Elisee C, Newham DM et al (1999) New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 65:382–388

3. Walter S, Buchner J (2002) Molecular chaperones—cellular machines for protein folding. Angew Chem Int Ed Engl 41:1098–1113

4. Schein CH, Noteborn MHM (1988) Formation of soluble recombinant proteins in Escherichia coli is favored by lower growth temperature. Bio/Technology 6:291–294

5. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22:1399–1408

6. Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York

7. Idicula-Thomas S, Balaji PV (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. Protein Sci 14:582–592

8. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci U S A 78: 3824–3828

Part X

Insoluble Protein Applications

Chapter 24

Insoluble Protein Applications: The Use of Bacterial Inclusion Bodies as Biocatalysts

Eva Hrabárová, Lucia Achbergerová, and Jozef Nahálka

Abstract

Biocatalysis and biotransformations have a broad application in industrial synthetic chemistry. In addition to the whole cell catalysis, purified recombinant enzymes are successfully used for biocatalysis of specific chemical reactions. In this contribution, we report characterization, immobilization, and application of several model target enzymes (D-amino acid oxidase, sialic acid aldolase, maltodextrin phosphorylase, polyphosphate kinase) physiologically aggregated within inclusion bodies (IBs) retaining their biological activity as immobilized biocatalysts.

Key words Inclusion bodies, Recombinant enzymes, Biocatalysis/biotransformation

1 Introduction

1.1 General Aspects

Phenomenon of intracellular protein aggregation is widely observed in human diseases, biopharmaceutical production, and biological research [1, 2]. Usually overloaded cell translation machinery deposits proteins in such protein aggregates commonly named as inclusion bodies (IBs). A novel concept of purposely "pull-down" enzymatic activity to IBs has been recently introduced predicting them for the use in biocatalysis particularly for pharmaceutical industry [3]. Almost complete cell tolerable activity (the maximal activity of the overexpressed enzyme that can be achieved in cytosol) can be "pull down" into IBs via *N*-terminal fusion with a pull-down module, e.g., cellulose-binding domain (CBD) of *Clostridium cellulovorans* [4].

1.2 Inclusion Bodies: Waste Materials No More

In the past, IB formation was thought to reflect the passive kinetic competition between complete folding and aggregation driven by intermolecular interactions between solvent-exposed hydrophobic patches on the surface of partially folded proteins. In this context, IBs of recombinant enzymes were regarded as nonproductive

protein waste [5]. In light of this, the scientists exploring a protein expression in prokaryotic hosts have been looking for conditions of increasing the solubility [6]. On the contrary, those scientists exploring the application of recombinant enzymes in industry have been currently looking for an alternative to recycle the "expensive" biocatalyst. In this context, the originally soluble enzymes are usually made insoluble by their attachment to the surface of an insoluble carrier, a process called enzyme immobilization [7].

In 2005, Prof. Villaverde and coworkers summarized findings from their original works and also from the previous works in the field of the IB research and declared that the retention of biological function in IBs is not so rare as it was previously thought. In this regard, IBs were proposed as catalysts for industrial purposes [8]. In 2007, Nahálka and Nidetzky proposed to reverse the process of improving solubility of aggregation-prone proteins via fusion to highly soluble protein. Specifically, they proposed to fuse otherwise soluble enzymes to highly aggregation-prone tags to achieve "in vivo enzyme immobilization" by "pulling down" the enzyme activity [3]. At present, biologically active IBs represent an appealing alternative as "in vivo" immobilized enzymes directly applicable and recyclable in industrial processes.

1.3 Isolation and Purification of Inclusion Bodies

Once produced, IBs have to be isolated and purified. However, the simplicity of the purification procedure represents the basic advantage of recombinant protein production in the form of IBs. Among the best systems for performing routine cell breakage at the small- and high–medium-scale lysis of either bacterial or yeast cells to obtain the abovementioned insoluble proteins in homogenized state are mechanical methods—especially the French pressure cell press [9, 10]. The cell lysate is first centrifuged to obtain the pellet fraction of IBs which is then washed with gentle detergent such as Triton X-100 to remove attached material. Insoluble cell wall material can stay as contaminant or can be enzymatically solubilized (e.g., by lysozyme) and wash out by the detergent.

1.4 Examples of Applications of Inclusion Bodies as Biocatalysts

The enzymes of interest are involved in various important industrial reactions. For example, D-amino acid oxidase from the yeast *Trigonopsis variabilis* (*Tv*DAO) catalyzes, with a broad specificity, the O_2-dependent conversion of different D-amino acid substrates into the corresponding α-keto acids, ammonia, and H_2O_2. The basic economically important transformation is the oxidation of cephalosporin C to keto-adipinyl-7-aminocephalosporanic acid which represents the first reaction in the two-step enzymatic production of 7-aminocephalosporanic acid (an intermediate in the production of antibiotics) [11]. The commercialized biocatalytic process is performed on a multiton-per-year industrial scale, and

the initial conversion of cephalosporin C is carried out in aerated stirred tank reactors and employs carrier-bound insoluble *Tv*DAO that is continuously reused [12]. Thus, an "in vivo" immobilized *Tv*DAO is a potential alternative for this process [3].

N-acetyl-D-neuraminic acid aldolase (sialic acid aldolase, SAA) is an industrial enzyme involved in influenza antivirotics production. It is used for aldol condensation of acylmannosamines and pyruvate into sialic acids and their analogs [13]. SAA represents the second enzyme of industrial relevance which was successfully "in vivo" immobilized into active IBs [4].

Alpha-glucan phosphorylases (EC 2.4.1.1) catalyze the reversible phosphorolytic cleavage of α-1-4-linked glucose oligosaccharides with formation of α-D-glucose-1-phosphate (G-1-P). G-1-P is a suitable starting material in various chemical or enzymatic syntheses, representing an example of the industrial biotransformation of the cheap starch. Thermostable maltodextrin phosphorylase from *Pyrococcus furiosus* (*Pf*MP) can be a better alternative to commonly used potato or *E. coli* maltodextrin phosphorylases, regarding the starch solubility and the enzyme stability. "In vivo" immobilized *Pf*MP showed improved technological parameters for repetitive batch conversions [14].

Systems biocatalysis is a new approach consisting of organizing enzymes in vitro to generate an artificial metabolism for synthetic purposes. The controllable construction of metabolic pathways, assigned for the efficient synthesis of valuable chemical products, represents future trend in industrial applications. However, more complex chemistry, achievable by artificial metabolic pathways, includes cofactor-dependent enzymes, especially adenosine 5′-triphosphate (ATP) "burning" enzymes. In light of this, polyphosphate kinases (PPKs) could provide the energy or could be used as starting phosphorylating agents [15]. "In vivo" immobilized PPK from *Silicibacter pomeroyi* (*Sp*PPK) demonstrated such power to drive multi-enzymatic reactions [16].

1.5 The Protocol Objectives

Monitoring the conditions of recombinant protein production as biologically active IBs to explore their biocatalytic applications has been a principal goal as outlined in Nahálka's papers [3, 4, 14, 16–18]. The proposed protocol has been focused on the integration of experiences from characterization, immobilization, and application of several chosen model target enzymes (*Tv*DAO, *Ec*SAA, *Pf*MP, *Sp*PPK) produced as IBs. Biocatalyst characterization (specific and volumetric activity, "in vivo" immobilization efficiency, IB purity) and the subsequent IB immobilization (cross-linking, entrapment, encapsulation) are usually performed because of enhancing operational and storage stability of the biocatalyst and better handling of biotransformation process.

2 Materials

2.1 Biocatalyst Characterization

*Tv*DAO measurement:

1. 10 mM Tris buffer, pH 7.5.
2. 300 mM D-methionine in 10 mM Tris buffer, pH 7.5.
3. Fiber-optic oxygen microoptode connected to transmitter and computer (PreSens GmbH, Regensburg, Germany).
4. Jacketed Double Sidearm Flask 50 mL (CelStir® Spinner, Wheaton).
5. Refrigerated and Heating Circulator.
6. Magnetic stirrer.
7. French press.

*Pf*MP measurement:

1. 50 mM Tris buffer, pH 7.5.
2. 400 mM KH_2PO_4, pH 8.0.
3. 5 % dextrin in the phosphate buffer.
4. 5 mM uridine 5′-triphosphate trisodium salt dihydrate and 10 mM $MgCl_2$, in the Tris buffer.
5. UDP–glucose pyrophosphorylase (GalU, EC 2.7.7.9) in the Tris buffer.
6. CelLytic™ B Cell Lysis Reagent (Sigma, B7435-500mL).
7. HPLC system equipped with NH_2 column (TESSEK Separon SGX NH_2).
8. Mobile phase: 50 mM H_3PO_4, 10 mM $MgCl_2$, pH 6.4 (adjusted with triethylamine).

Protein concentration measurement:

1. Total Protein Kit (composed of Biuret Reagent and Folin and Ciocalteu's Phenol Reagent) to determine protein concentration.
2. 1 % SDS solution.

2.2 IB Immobilization

Cross-linking:

1. Aqueous 25 or 50 % glutaraldehyde.
2. 50 mM Tris buffer, pH 7.5.

Entrapment to alginate matrix:

1. 1.5 % (w/v) sodium alginate in water (60 % content of D-mannuronic acid).
2. 1.0 % (w/v) $CaCl_2$ in water (weigh $CaCl_2 \cdot 2H_2O$).

3. Syringe.

4. Coaxial nozzle.

5. Compressed gas (nitrogen).

Entrapment to agar matrix:

1. 2 % (w/v) agar (1.5 % gel strength: 650 g/cm^2) solution (60 °C).

2. Vegetable oil (25 °C).

3. 1 % Triton X-100 solution.

4. Magnetic stirrer.

Encapsulation:

1. 10 mM Tris buffer, pH 7.5.

2. 0.9 % (w/v) NaCl solution.

3. Polyanion solution: 0.9 % (w/v) sodium alginate (60 % content of D-mannuronic acid) plus 0.9 % (w/v) cellulose sulfate (sodium salt, Acros Organics, Geel, Belgium) in 0.9 % NaCl solution.

4. Polycation solution: 2.0 % (w/v) PMCG hydrochloride (Scientific Polymer Products, Inc., Ontario, NY 35 % aqueous solution lyophilized after delivery), 1.0 % (w/v) CaCl$_2$, 0.9 % NaCl solution.

5. 50 mM sodium citrate buffer, pH 7.0.

6. Syringe.

7. Coaxial nozzle.

8. Compressed gas (nitrogen).

3   Methods

Before a scale-up of any experiment, one has to provide reproducible data from laboratory experiments.

3.1 Biocatalyst Characterization

A variety of methods are available for the activity measurement of a chosen enzyme. Each biotransformation process usually requires a specific approach. It is important to have in mind that the reactions are performed in IB suspensions; that means the conventional spectrophotometric assays might be influenced by the presence of IB particles in online reaction kinetic measurements. It is useful to couple the reactions to oxygen, pH, or calorimetric probes. For example, we used a fiber-optic oxygen microoptode for the *Tv*DAO activity measurement or a flow calorimeter for the *Ec*SAA activity measurement. Of course, these sensors have to be calibrated, e.g., by conventional off-line methods, e.g., HPLC. In the case that any measurement cannot be performed via online method, then one should consider the requirement of off-line determination of the

enzyme activity for taking a representative sample from a suspension of the biocatalyst. It is because this is the subject of systematic error and the lack of reproducibility, especially in the case of large IB clusters. Regarding spectrophotometric assays, it is better to perform online measurements of a few parallel samples in small volumes (50–100 µL) using a microplate reader. For example, we commonly measure the IB-*Sp*PPK activity in this way applying the kit based on fluorescence dye.

3.1.1 Specific and Volumetric Activity

1. *Tv*DAO online measurement: Fill the jacketed flask with 30 mL Tris buffer, add stir bar and stir at 300 rpm, keep at 30 °C by external refrigerated/heating circulator, switch on the oxygen sensor, add 0.5 mL of D-methionine solution, wait 15 min to achieve steady-state condition, add 200 µL of appropriately diluted biocatalyst, and record initial rate of O_2 consumption (Fig. 1a, *see* **Note 1**).

2. *Pf*MP off-line measurement: Mix 50 µL of appropriately diluted biocatalyst in Tris buffer with 1 mL of 5 % dextrin in the phosphate buffer, perform the reaction at 90 °C, and withdraw and freeze 10 µL of the aliquots at 10 min intervals. Mix 50 µL of appropriately diluted sample in Tris buffer with 50 µL of the Tris solution containing 5 mM UTP and 10 mM $MgCl_2$; after thermal equilibration at 25 °C, add 10 U (5 µL) of UDP–glucose pyrophosphorylase, allow to proceed the reaction for 15 min, and freeze the samples. Measure UDP–glucose by HPLC.

3. Calculate the specific cell activity in U (µmol per min)/g of wet cells or U/mg of dry cells and the volumetric activity in U/L of culture broth (*see* **Note 2**).
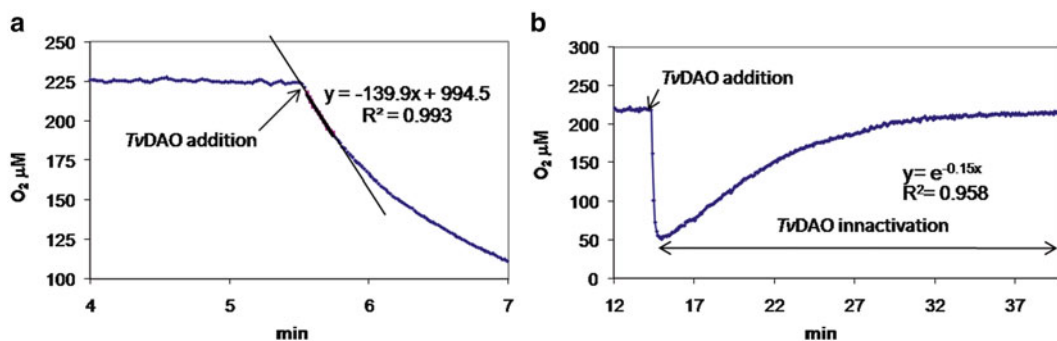
Fig. 1 A typical plot showing the biocatalyst activity (**a**) and the biocatalyst stability (**b**) assays in 30 mL minireactor equipped with oxygen sensor. (**a**) The initial rate is obtained from the part of the curve that exhibits a constant slope ($\Delta O_2/\Delta t$) immediately after addition of the biocatalyst (*Tv*DAO); 5 mM D-methionine, 10 mM Tris buffer. (**b**) The time course of the *Tv*DAO inactivation experiment with bubble aeration, 50 mM D-methionine, 100 mM Tris buffer. The inactivation curve ($y = e^{-0.15x}$) is obtained after reversing the values and their normalization by maximal value ($y = \Delta O_2/\Delta O_{2\,max}$; $x = t$)

3.1.2 "In Vivo"
Immobilization Efficiency

1. Disrupt the biomass by French press or Cell Lysis Reagent; centrifuge the cell lysate (15 min, $13,000 \times g$, 4 °C), and measure the activity of soluble and insoluble cell fractions.

2. Dissolve insoluble fraction in 1 % SDS, and measure the protein concentration of the insoluble cell fraction using the protein assay kit.

3. Calculate "in vivo" immobilization efficiency (%) as [insoluble activity/(soluble activity + insoluble activity)] × 100; calculate specific protein activity—activity per mg of protein (*see* **Note 3**, Fig. 2).

3.1.3 IB Purity (PfMP, SpPPK, TvDAO)

Perform analysis by SDS-PAGE of soluble and insoluble fractions derived from *E. coli* producing active IBs (*see* **Note 4**).

3.2 IB Immobilization

IBs can be directly used as recyclable biocatalyst, but handling of small gelatinous particles, from nano- to micro-size, is difficult in large scale. Large-scale centrifugation process can be a quite expensive step. It is, therefore, convenient to use cellulose-binding domain (CBD) as "in vivo" immobilization-prone module because it can help in filtration process using cheap cellulose filters. Figure 3 depicts the affinity of CBD to crystal and fiber cellulose. Otherwise, additional IB immobilization step will improve the IB separation from a reaction mixture.

3.2.1 Cross-Linking

1. Suspend IBs in Tris buffer at various concentrations (10–50 mg/mL).

2. Add dropwise the solution of glutaraldehyde to various final concentrations (0.1–1 %) and stir at 300 rpm from 10 to
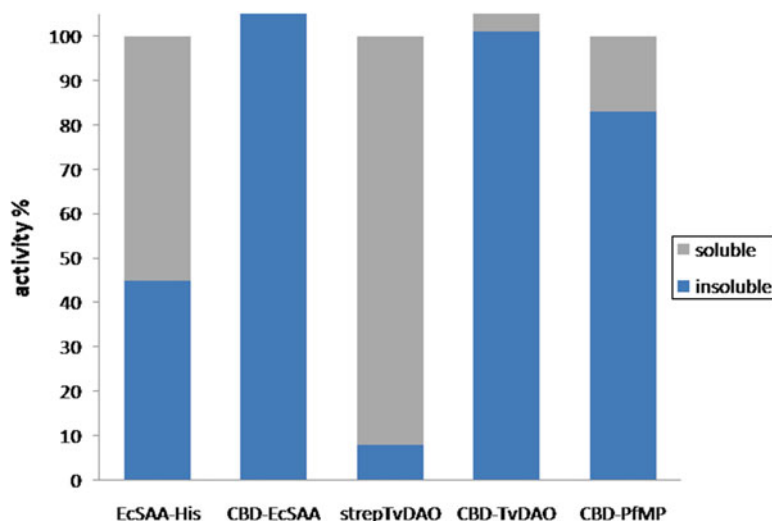
Fig. 2 Activity (%) of soluble and insoluble fractions of the three exemplary enzymes obtained after bacterial cell disruption. 100 % activity means the whole cell activity of His- or Strep-tagged forms
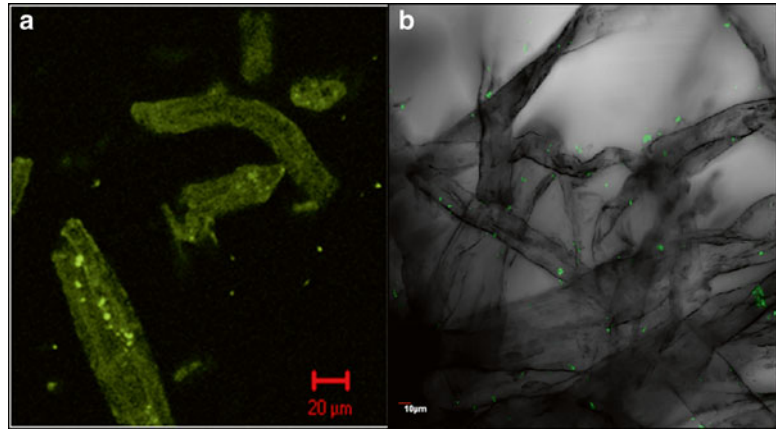
Fig. 3 Demonstration of the form and activity of IBs obtained by "pull-down" domain of fused CBD–GFP protein. The figure depicts the affinity of CBD to crystal (**a**) and fiber (**b**) cellulose. The nano-sized IBs are basically composed of proteins with unfinished folding; therefore, more hydrophobic amino acid residues are exposed on the water–protein interface, and larger clusters are formed by hydrophobic interaction. (**a**) Whole size distribution—"single" IBs to the largest aggregate. (**b**) Cluster fraction filtered through paper

60 min and then wash the IB aggregates by three repetitive dilution (10×, Tris buffer) and centrifugation steps.

3. Check enzyme activity of the cross-linked IBs and apply the conditions with the largest IB aggregates and with maximal residual activity (*see* **Note 5**).

3.2.2 Entrapment

Entrapment is a gentle process to IBs; the IBs are trapped in insoluble gel beads or microspheres, such as calcium-alginate beads; there is no covalent binding and no chemical reaction that affects the enzyme's active site; and contrary to soluble enzymes, IBs are not leaking from porous gel matrix.

1. Entrapment to alginate matrix: Mix IBs with alginate solution (1–8 mg/mL), pass it through a mesh of similar diameter as the dropping nozzle, and left to stay at 4 °C to get out the air bubbles. Drop the mixture through the nozzle into stirred gelling bath with $CaCl_2$ solution (10 volumes of alginate solution), using coaxial air stream to blow droplets from the needle tip into the gelling bath. After 30 min, decant the alginate beads, dilute them with water, and keep in 20× diluted $CaCl_2$ at 4 °C (*see* **Note 6**, Fig. 4).

2. Entrapment to agar matrix: Suspend IBs in 5 mL of Tris buffer at 25 °C, and mix it with 15 mL of 2 % agar solution at 60 °C, immediately, disperse the mixture in 50 mL of vegetable oil (25 °C) under stirring. After cooling, separate agar beads from the oil phase, and wash them with 1 % Triton X-100 solution and then with water or buffer (*see* **Note 7**, Fig. 4).

Fig. 4 Final form of the biocatalyst: entrapped and encapsulated IBs. (**a**) Entrapped IBs-*Ec*SAA in alginate beads, 5 min hardened with 0.5 % glutaraldehyde; the beads became slightly yellow. (**b**) Entrapped IBs-*Ec*PPK in agar obtained by dispersing and cooling in the vegetable oil. (**c**) Encapsulated IBs-*Tv*DAO; the IBs sediment in liquefied core inside the capsules

3.2.3 Encapsulation

1. Mix IBs with polyanion solution (1 mg/mL), pass it through a mesh of similar diameter as the dropping nozzle, and left to stay at 4 °C to get out the air bubbles.

2. Drop the mixture through the nozzle into stirred gelling bath filled with polycation solution (5 volumes of polyanion solution). Use coaxial air stream to blow droplets from the needle tip into the gelling bath. After 60 ± 5 s, dilute polycation with water (20×).

3. Incubate the beads for 15 min in 10 volumes of sodium citrate buffer; after liquefying of the inner core, hold the hollow capsules in Tris buffer at 4 °C (*see* **Note 8**, Fig. 4).

3.3 IB Applications

"In vivo" enzyme immobilization, cross-linking, entrapment, or encapsulation processes can change known biochemical characteristics of the purified enzyme. It is useful to check how much they are changed, before testing any application of the biocatalyst.

3.3.1 Operational Stability

1. Screen various pH, substrate concentrations, metal-cofactor concentrations, and eventually temperatures regarding the optimal batch yield and biocatalyst stability.

2. Set up the conditions close to required technological conditions, e.g., expected concentration of the product regarding subsequent reaction or isolation step.

3. Run repetitive batch at "optimal" conditions.

4. Calculate residual yield or degree of conversion for each repetitive batch (*see* **Note 9**).

3.3.2 Storage Stability

1. Check residual activity after several repetitive freeze-towing cycles.

2. Check residual activity after lyophilization process.

3. Store the biocatalyst at chosen condition (lyophilized at −20 °C or +4 °C).

4. Measure the activity at appropriately set time intervals (*see* **Note 10**).

4   Notes

1. Measurement of the activity of either soluble or insoluble bio-catalyst in a mini-reactor equipped with biosensor has enormous advantage in the scale-up processes. Besides, it can also be directly used for thermal/operational stability measurements [18]. For example, *Tv*DAO uses O_2 as the substrate, but on the other hand, the protein is quite quickly inactivated on gas–liquid interfacial surface area (Fig. 1b). Online activity measurement in a mini-reactor allows directly to model and set the best biotransformation process conditions (e.g., a comparison of the addition of oxygen by bubble aeration, but in a separate vessel, with guarding the *Tv*DAO by encapsulation).

2. The specific cell activity expressed as (μmol per min)/g of wet cells shows how much is a host cell able to tolerate the cloned protein. Volumetric activity (U/L of culture broth) provides "economic data," giving an idea of the effectiveness of the biocatalyst production. Using Novagen pET expression system, the *E. coli* expression host usually produces the same cloned enzyme to similar level of specific cellular activity, e.g., pET-51 (Strep-tag) and pET-34 (CBD-tag); however, specific protein activity is different. "In vivo" enzyme immobilization usually allows higher volumetric activity because IBs have lower specific protein activity comparing to soluble proteins; the cells grow usually to higher concentrations.

3. It is important to calculate "in vivo" immobilization efficiency (%) because it informs you if "pull-down" strategy works or not. Figure 2 documents "in vivo immobilization" efficiency of the three model enzymes. Specific protein activity of IBs is important for subsequent calculations—to set the biotransformation process.

4. IBs usually are composed mainly of the cloned protein [4], but sometimes the "pull-down module" is hydrolyzed from the fused protein [14]. Moreover, the cloned enzyme (especially membrane proteins) can "pull down" various cytosol proteins into IBs as impurities in this context.

5. It is also very important to carefully consider the cross-linking agent used. Glutaraldehyde, for example, is toxic to enzyme's active center. Thus, it is important to determine the best cross-linking conditions separately for each enzyme [19]. Usually, IBs lose approximately 50 % of their biological activity, although in some cases, they lose all the activity. In such situations, glutaraldehyde can be changed by oxidized dextran (dextran poly-aldehyde). Alternatively, co-cross-linking (addition of a cheap albumin) or addition of silica can also be of help.

6. Entrapment to calcium-alginate gel is well suited for a subsequent IB immobilization. It is important to take into

consideration that 1–3 mm gel beads have better sedimentation and handling properties than small gelatinous IBs or cross-linked IBs. The basic disadvantage of the use of alginate beads is the fact that a biotransformation process has to be run at the absence of Ca^{2+}-chelating agents or at the presence of low concentration of Ca^{2+} ions (5 mM; maintaining the sodium and calcium ratio 25:1) to avoid gel destabilization. Technology for large-scale alginate entrapment is well established (*see* Nisco Engineering or geniaLab®).

7. The procedure is simple and effective, but "mesophilic" proteins denature quite quickly at 60 °C. At large-scale conditions, continuous mixing of agar and enzyme solutions and constant dispersing in the oil will be needed.

8. Before starting the procedure, always try to drop some drops of polyanion enzyme mixture into polycation bath because sometimes the drops are hydrophobic too much and float on the polycation surface. In such cases, decrease the IB concentration in the polyanion solution. The encapsulation process is quite "expensive"; it is, therefore, applicable for "extraordinary" biotransformations. For larger-scale processes, tube chemical reactors were designed to allow a precise control of the reaction time between polyanion drops and polycation solution [20].

9. Sometimes, Mg^{2+} cofactor is blocked by calcium in Ca^{2+}-alginate gels which excessively decrease the activity, e.g., in the case of *Sp*PPK; therefore, another entrapment procedure has to be used. Sometimes negatively charged substrates precipitate in the capsule pores, e.g., in the case of biotransformation of phenoxymethylpenicillinic acid to phenoxyacetic acid and 6-aminopenicillanate. These precipitates excessively decrease the activity after each cycle. Sometimes too much protein is released from alginate beads. In this case, mixing the beads with cross-linking agent 0.5 % glutaraldehyde for 1–5 min could help. Actually, *Ec*SAA alginate beads mixed with 0.5 % glutaraldehyde for 5 min showed the best operational stability.

10. Storage stability will depend on the properties of any protein. Thus, different proteins need different conditions. However, 10 % sucrose is generally a better lyophilizing protectant for IBs compared to 10 % glycerol.

Acknowledgments

References

1. Gatti-Lafranconi P, Natalello A, Ami D et al (2011) Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology. FEBS J 278:2408–2418

2. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332

3. Nahálka J, Nidetzky B (2007) Fusion to a pull-down domain: a novel approach of producing Trigonopsis variabilis D-amino acid oxidase as insoluble enzyme aggregates. Biotechnol Bioeng 97:454–461

4. Nahálka J, Vikartovská A, Hrabárová E (2008) A crosslinked inclusion body process for sialic acid synthesis. J Biotechnol 134:146–153

5. Schein CH (1989) Production of soluble recombinant proteins in bacteria. Bio/Technology 7:1141–1149

6. Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8:1668–1674

7. Cao L, van Langen L, Sheldon RA (2003) Immobilised enzymes: carrier-bound or carrier-free? Curr Opin Biotechnol 14:387–394

8. García-Fruitós E, González-Montalbán N, Morell M et al (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27

9. Palmer I, Wingfield PT (2004) Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. In: Coligan JE, Dunn BM, Speicher DW, Wingfield PT (eds) Current protocols in protein science, supplement 38, chapter 6:unit 6.3. John Wiley & Sons, Hoboken, N. J. pp 6.3.1–6.3.18

10. Peternel Š, Kome R (2010) Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. Microb Cell Fact 9:66

11. Buchholz K, Kasche V, Bornscheuer UT (2005) Case study 1: the enzymatic production of 7-ACA from cephalosporin C. In: Wandrey C (ed) Biocatalysts and enzyme technology. Wiley, New York, pp 381–392

12. Riethorst W, Reichert A (1999) An industrial view on enzymes for the cleavage of cephalosporin C. Chimia 53:600–607

13. Liese A, Seelbach K, Buchholz A et al (2006) Processes. In: Liese A, Seelbach K, Wandrey C (eds) Industrial biotransformations. Wiley, New York, pp 457–460

14. Nahálka J (2008) Physiological aggregation of maltodextrin phosphorylase from *Pyrococcus furiosus* and its application in a process of batch starch degradation to alpha-D-glucose-1-phosphate. J Ind Microbiol Biotechnol 35:219–223

15. Achbergerová L, Nahálka J (2011) Polyphosphate—an ancient energy source and active metabolic regulator. Microb Cell Fact 10:63

16. Nahálka J, Pätoprstý V (2009) Enzymatic synthesis of sialylation substrates powered by a novel polyphosphate kinase (PPK3). Org Biomol Chem 7:1778–1780

17. Nahálka J, Gemeiner P, Bučko M et al (2006) Bioenergy beads: a tool for regeneration of ATP/NTP in biocatalytic synthesis. Artif Cells Blood Substit Biotechnol 34:515–521

18. Nahálka J, Dib I, Nidetzky B (2008) Encapsulation of *Trigonopsis variabilis* D-amino acid oxidase and fast comparison of the operational stabilities of free and immobilized preparations of the enzyme. Biotechnol Bioeng 99:251–260

19. Schoevaart R, Wolbers MW, Golubovic M et al (2004) Preparation, optimization, and structures, of cross-linked enzyme aggregates (CLEAs). Biotechnol Bioeng 87:754–762

20. Anilkumar AV, Lacik I, Wang TG (2001) A novel reactor for making uniform capsules. Biotechnol Bioeng 75:581–589

INDEX