Advances in Experimental Medicine and Biology 794

Marco Fioroni
Tamara Dworeck
Francisco Rodríguez-Ropero

ß-barrel Channel Proteins as Tools in Nanotechnology

Biology, Basic Science and Advanced Applications

Springer

Advances in Experimental Medicine and Biology

Volume 794

For further volumes:
http://www.springer.com/series/5584

Advances in Experimental Medicine and Biology

Recent Volumes in this Series

Volume 782
PROGRESS IN MOTOR CONTROL
Michael A. Riley, Michael J. Richardson and Kevin D. Shockley

Volume 783
THE NEW PARADIGM OF IMMUNITY TO TUBERCULOSIS
Maziar Divangahi

Volume 784
KISSPEPTIN SIGNAING IN REPRODUCTIVE BIOLOGY
Alexander S. Kauffman and Jeremy T. Smith

Volume 785
CROSSROADS BETWEEN INNATE AND ADAPTIVE IMMUNITY IV
Bali Pulendran, Peter D. Katsikis and Stephen P. Schoenberger

Volume 786
TRANSCRIPTIONAL AND TRANSLATIONAL REGULATION OF STEM CELLS
Gary Hime and Helen Abud

Volume 787
BASIC ASPECTS OF HEARING
Brian CJ Moore

Volume 788
NEUROBIOLOGY OF RESPIRATION
Mieczyslaw Pokorski

Volume 789
OXYGEN TRANSPORT TO TISSUE XXXV
Sabine Van Huffel, Gunnar Nualaers, Alexander Caicedo, Duane F. Bruley and David K. Harrison

Volume 790
VIRAL ENTRY INTO HOST CELLS
Stefan Pöhlmann and Graham Simmons

Marco Fioroni • Tamara Dworeck
Francisco Rodríguez-Ropero

ß-barrel Channel Proteins as Tools in Nanotechnology

Biology, Basic Science and Advanced Applications

Springer

Marco Fioroni
Sustainable Momentum
Aachen, Germany

Francisco Rodríguez-Ropero
Center of Smart Interfaces
Technische Universität Darmstadt
Darmstadt, Germany

Tamara Dworeck
Department of Biology
RWTH Aachen University
Aachen, Germany

ISSN 0065-2598
ISBN 978-94-007-7428-5 ISBN 978-94-007-7429-2 (eBook)
DOI 10.1007/978-94-007-7429-2
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013949493

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Abbreviations

ABC	ATP-binding cassette
AFM	Atomic force microscopy
ATP	Adenosine triphosphate
Bam	β-Barrel assembly machine
BCA	Bicinchoninic acid
BNPA	2-Bromo-2-(2-nitrophenyl)acetic acid
CC	Coupled cluster
CD	Circular dichroism
CG	Course grained
CLSM	Confocal laser scanning microscopy
CM	Chemical modification
cmc	Critical micellar concentration
COSY	Correlation spectroscopy
CPU	Central processing unit
CRAMPS	Combined rotation and multiple pulse spectroscopy
Da	Dalton
DAGK	Diacylglycerol kinase
DDAO	Dimethyldecylamine-N-oxide
DFT	Density functional theory
Dh	Hydrodynamic diameters
DHPC	1,2-Dihexanoyl-sn-glycero-3-phosphocholine
DLS	Dynamic light scattering
DMPC	Dimyristoyl-phosphatidylcholine
DNA	Deoxyribonucleic acid
DNPC	1,2-Dinervonyl-sn-glycero-3-phosphocholine
dNTP	Deoxyribonucleotide triphosphate
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine
DTT	Dithiothreitol
ECD	Electronic circular dichroism
EDTA	Ethylenediamine-tetra-acetic acid
EE	Ethylethylene
EIA	Enzyme immunoassays
EM	Electron microscopy
em	Electromagnetic
EO	Ethyleneoxide
ER	Endoplasmic reticulum
ESEM	Environmental scanning electron microscopy

FCS	Fluorescence correlation spectroscopy
FET	Field-effect transistor
FhuA	Ferric hydroxamate uptake component A
FT-NMR	Pulse-Fourier-transformation-NMR
GdmCl	Guanidinium chloride
GF	Gel filtration
GFP	Green fluorescent protein
GPI	Glycosylphosphatidylinositol
GST	Glutathione-S-transferase
GTP	Guanosine triphosphate
GUV	Giant unilamellar vesicle
HF	Hartree-Fock
hIMP	Human integral membrane protein
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HMM	Hidden Markov model
HPH	High pressure homogenisator
HRP	Horse radish peroxidase
HSQC	Heteronuclear single-quantum correlation spectroscopy
IBI	Iterative Boltzmann inversion
IDP	Intrinsically disordered protein
IEXC	Ion exchange chromatography
igG	Immunoglobulin G
IMAC	Immobilized metal-affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
L	Loop
LDAO	Lauryldimethyl amine oxide
LMPG	1-Myristoyl-2-hydroxy-sn-glycero-3-[phosphorac-(1-glycerol)]
LPS	Lipopolysaccharides
LUV	Large unilamellar vesicles
MAD	Multiwavelength anomalous dispersion
MAS	Magic angle spinning
MBP	Maltose binding protein
MD	Molecular dynamics
MM	Molecular mechanics
MP	Membrane protein
MP	Møller-Plesset
MRE	Mean residue ellipticity
mRNA	Messenger ribonucleic acid
MRW	Mean residue weight
ms	Millisecond
MSP	Membrane scaffold protein
NHS	N-hydroxysuccinimide
NLP	Nano-lipoprotein particles
NMR	Nuclear magnetic resonance
NN	Neural network
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy

NRMSD	Normalised root mean square deviation
nS	Nanosiemens
NVOC-Cl	6-Nitroveratryloxycarbonyl chloride
NW	Nanowire
octyl-POE	n-Octylpolyoxyethylene
OES	n-Octyl-2-hydroxyethyl sulfoxide
OG	Octyl-β-D-glucoside
OM	Outer membrane
OMP	Outer membrane protein
OmpA	Outer membrane protein A
ORF	Open reading frame
P2	Poly(Pro)II type structure
PBD-PEO	Polybutadiene-polyethyleneoxide
PCDDB	Protein circular dichroism data bank
PCM	Polarizable continuum model
PCR	Polymerase chain reaction
PCS	Photon correlation spectroscopy
PDB	Protein data bank
PDI	Polydispersity index
PDMS	Poly(dimethylsiloxane)
PEG	Poly(ethylene glycol)
PEtOz	Poly(2-ethyl-2-oxazoline)
PH	Pleckstrin homology
PIB	Polyisobutylene
PME	Particle mesh-Ewald summation
PMF	Proton motive force
PMOXA	Poly(2-methyloxazoline)
POTRA	Polypeptide translocation associated
ps	Picosecond
PS-b-PAA	Polystyrene-block-poly(acrylic acid)
PSI	Protein structure initiative
Pyr	3-(2-Pyridyldithio)-propionic-acid
QELS	Quasi-elastic light scattering
QM	Quantum mechanics
RA	Rotational alignment
RBS	Ribosome binding site
RF	Radio frequency
rHDLs	Reconstituted high density lipoprotein particles
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ROESY	Rotating frame nuclear Overhauser effect spectroscopy
SCF	Self-consistent field
SCRF	Self-consistent reaction field
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SIMPLS	Simultaneous partial least squares
SiNW	Silicon nanowires
SLS	Static light scattering

SPP	Single protein production
SR	Signal recognition particle receptor
SRCD	Synchrotron radiation in circular dichroism
SRP	Signal recognition particle
SSM	Site-specific mutagenesis
ssNMR	Solid-state NMR
STREP	Streptavidin-binding peptide
SUV	Small unilamellar vesicles
SVM	Support vector machines
T	Turn
TAT	Twin arginine translocon
TEM	Transmission electron microscopy
TEV	Tobacco etch virus
TFE	2,2,2-trifluoroethanol
THF	Tetrahydrofuran
TIC	Chloroplast inner membrane translocon
TMB	3,3′,5,5′-Tetramethylbenzidine
TMP	Transmembrane protein
TOC	Chloroplast outer membrane translocon
Toc	Translocase of the chloroplast outer membrane
TOCSY	Total correlation spectroscopy
Tom	Translocase of the mitochondrial outer membrane
TRCD	Time resolved CD spectroscopy
ULV	Unilamellar vesicle
UV	Ultraviolet
VDAC	Voltage dependent ion channel
WT	Wild type
X-FEL	X-ray free electron laser

Introduction 1

Though there is some dispute about whether barrels were first invented by the Egyptians, or rather by Greeks and Romans (some even claim the achievement for the Celts), earliest finds have been dated back to as early as 2500 B.C. and there is no discussion on that the barrel geometry facilitates transport by maximizing cube utilization allowing tight loading of ships and wagons [1] and that even more importantly casks generally hold good things, such as wine, oil, beer, honey or in case of a "barrel of laughs" also fun.

Keeping in mind the macroscopic wooden barrel's great transport and packing-component potentials, as well as its importance for the civilization and technology development, this book will call to the reader's attention a completely different kind of barrel: the nano-sized so-called β-barrel membrane protein (see Fig. 1.1).

β-barrel membrane proteins that regardless of their nanoscopic size have as great potential as their large wooden counterparts These potentials lie within their structure that forms channels in hydrophobic membranes and that it is for a protein exceptionally robust imparting outstanding nano-material properties.

1.1 Nano-technology – An Overview

Originally the term nano-technology was used for anything technologically applicable and smaller than microscopic. More recently the term is associated with the bottom-up construction of nano-scale components purposefully built to be assembled to form nano-materials. Nano-technology thus operates at the first level of organization, means at the level of atoms and molecules (10^{-9} m) and it promises the ability to build precise machines and components at the molecular size scale. In theory the feasibility of nano-technology was envisioned and prophesied by the physicist Richard Feynman [2] as early as 1959 in his famous talk entitled "There's plenty of room at the bottom" [3] (available online at http://www.zyvex.com/nanotech/feynman.html).

As nano-structures can be either derived from non-biological or biological components and as nano-materials can be applied to a broad range of fields like electronics including opto-electronics, medicine, pharmaceutical drug development, water-purification, food-technology to name but a few, nano-technology research necessitates the cooperation of scientists from various disciplines, such as physicists, engineers, chemists, and biologists.

1.2 Biological Nano-materials

The aim of nano-technology is as mentioned the design of new functional materials and devices through controlling their organization at the atomic and molecular level. One common strategy when designing a new nano-device is to survey naturally occurring biological structures with the ability to perform the desired process and use them as nano-material components or

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_1,

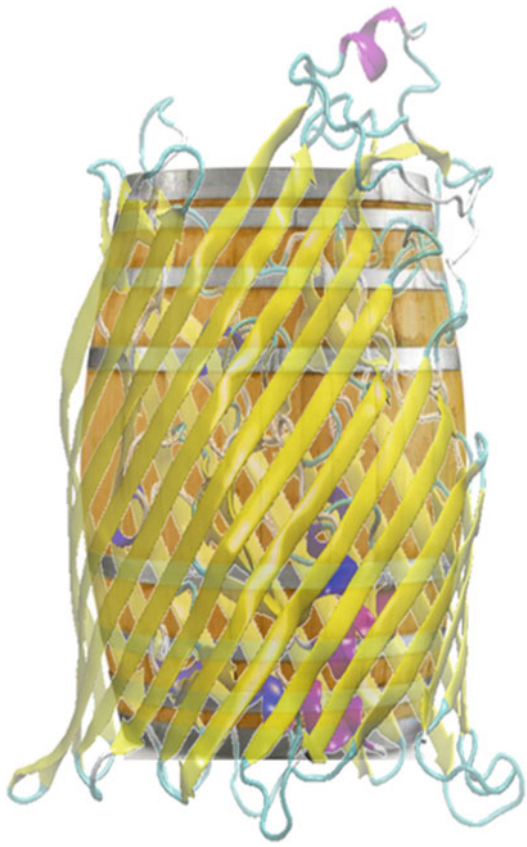

Fig. 1.1 β-barrel shaped membrane protein "in the light of" its macroscopic geometrical counterpart

scaffolds based on which to build the new nano-construct. This approach opens the promising field of bio-nanomaterial design.

In this sense the intrinsic nano-scale architecture and rich chemistry of proteins, as well as their catalytic activity in the case of enzymes, may be exploited to build a wide array of specific components in sophisticated nano-sized devices such as nano-motors, nano-reactors or stochastic nano-sensors. The protein based nano-material development is backed up by the vast progress that has been made in the molecular biology and biotechnology field, specifically in the development and optimization of advanced genetic engineering techniques allowing a tailoring of proteins towards a specific technical application [4].

Many of the possible applications of protein nano-materials (*i.e.* nano-reactors, functionalized nano-compartments, nano-sensors, drug-release systems) require channel shaped nano-components that allow the controlled transport of matter or the detection and analysis of an analyte that interacts with the channel interior by channel conductance measurements (nano-sensing elements).

Having said this, a rather obvious protein choice is the class of transmembrane proteins that reside within the various biological membranes, as many of these proteins form channels and pores to facilitate the passive or active transport of solutes, nutrients or cellular waste over the membrane. From the two types of channel shaped transmembrane protein classes, *i.e.* α-helical bundle and β-barrel proteins, the β-barrel structure stands out due to its versatility, flexibility, exceptional robustness and stability. Moreover β-barrel membrane proteins have the ability to refold *in vitro* and to reconstitute or insert into artificial lipid and polymer flat membranes or lipid and polymer vesicle (*i.e.* liposome and polymersome) membranes.

Due to the mentioned robustness of the β-barrel membrane proteins, they are easily modified by genetic engineering without loss of overall structure or function allowing the resulting protein nano-channels to be adapted to the non-biological synthetic polymer environment, rendering them competitive with artificial non-biological nano-pores. An example of the conception of a β-barrel nano-channel employing polymersome release system is given in Fig. 1.2.

Therefore the β-barrel and its serviceability for the nano-material design will be the chief topic of the present book and on the example of the *Escherichia coli* FhuA (ferric hydroxamate uptake component A) the design of a set of protein nano-channels with tailored geometry (diameter, length), conductance and functionality will be reported as a case study.

In order to make this book a valuable source of information for both biotechnologists and other

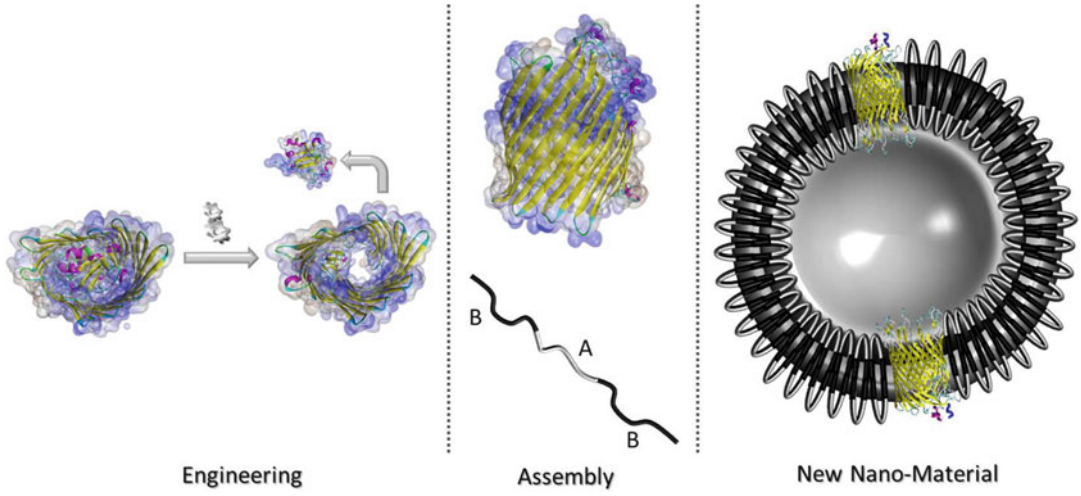

Fig. 1.2 Schematic conception work-flow of a hybrid bio-polymer nano-system on the example of a synthetic block copolymer based (*A* – hydrophophilic block, *B* – hydrophobic block) polymersome functionalized with a tailored β-barrel nano-channel for controlled compound release; involving protein engineering, protein polymer assembly and the resulting finished nano-release system

scientists interested in bio-nanotechnology an overview of the different steps involved in the nano-channel protein design and production will be reported, including concept design, theoretical considerations, genetic engineering and large scale production, as well as the system assembly and biophysical characterization.

In the following section a brief outline on the individual chapters will be given.

1.3 Outline

1.3.1 Chapter 2

To enter the membrane protein nano-channel material design topic, Chap. 2 will introduce to the biological basics of natural membranes and membrane proteins in general.

It will give some elementary information on the lipid bilayer composition and function, the main models to describe the bilayer membrane as well as on the lipids that assemble to form the membrane.

The various classes of membrane proteins (*i.e.* integral *vs.* peripheral) and their key differences will be mentioned. The two main structural classes of integral membrane proteins (*i.e.* α-helical and β-barrel) will be presented with regard to their biological origin, functional and structural features (considering primary, secondary, tertiary and quaternary sequence/structure), their biogenesis and their usability as bio-based nano-material components. For both protein classes several well-known and well-studied literature examples will be given.

In the light of the present book's title: *β-Barrel Channel Proteins as Tools in Nano-technology*, the main focus however will be on the characteristics and unique structural and functional properties of the β-barrel outer membrane proteins (OMPs). Their relevance for the nano-channel material design will be emphasized by presenting recent examples of the nano-technological use of lipid or polymer reconstituted β-barrel channel proteins.

The *E. coli* outer membrane iron transporter FhuA, which is a member of the TonB protein-dependent transporters and is one of the largest known β-barrel membrane proteins, will be introduced. Its unique biological characteristics (regarding structure and function) will be outlined. Since the FhuA and its engineered variants have been successfully employed as a model for the

transformation of a β-barrel outer membrane protein into a custom-made nano-channel, Chap. 2 aims to use the FhuA example to view β-barrel proteins not only within their biological context but to already introduce the reader to the relevance of members belonging to this membrane protein class for the nano-material sciences and specifically to their use for the design of biological nano-channels that can be reconstituted into artificial lipid or polymer membranes.

1.3.2 Chapter 3

The biophysical characterization of β-barrel outer membrane proteins (OMPs) opens the master way for the OMP behavioral understanding under the alien conditions they experience in a new environment such as a polymersome membrane.

The OMP characterization, furthermore, grounds the basics for a rational protein engineering necessary to solve the problems related to their functional reconstitution and widens their use and applications in technologies such as drug delivery, stochastic sensors and bio-nanoelectronics (see Chap. 6).

The main tools for the structural characterization of channel proteins and OMPs in particular is given in Chap. 3, where the state of the art of the main biophysical techniques used in β-barrel membrane protein analysis will be introduced, specifically mentioning X-ray crystallography, Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR).

The main problems, difficulties and limitations for each of the single techniques will be reported and most importantly, their alliances with other sophisticated techniques *i.e.* synchrotron radiation, to obtain structural, kinetic as well as thermodynamical informations will be reported.

Obviously the chapter is not intended as an exhaustive description of the single techniques, where existing reviews and books have been cited alongside the text, but has the aim to show how the X-ray diffraction measurements of protein crystals, CD spectroscopy and NMR have been and are used to study channel proteins further illustrating the direction toward which modern methodology and applications are facing.

However in the reported case of studies, many of the characterization problems associated with channel proteins and OMPs are intertwined with the difficulties deriving from the expression and purification methods that can be, quite often, the determining step in membrane proteins studies.

A particular stress will be given to Circular Dichroism. CD is a qualitative undervalued technique which gives general content in secondary structure of a protein and, in some cases, clues on the tertiary/quaternary structural changes, with no atomistic detail. However its use due to its quite user friendly experimental procedure and rapid data acquisition and processing, can rapidly uncover basic information on the protein stability within different environments not always easily obtainable or even impossible to obtain for example, by NMR analysis.

The obvious aim of each scientist working in the channel protein field is to obtain complete knowledge on the system with which he/she is working, however the fast development and number of bio-technologically modified β-barrel proteins to be, for example, embedded in liposomes or polymersomes, result in a need of a rapid characterization, a need that CD can promptly answer to, giving first clues on the obtained mutant secondary structure and stability.

1.3.3 Chapter 4

Chapter 4 will describe different theoretical and computational approaches that are useful to model and design new nano-systems based on β-barrel membrane proteins. Computational power in terms of continuously growing hardware facilities together with the continuous development of theoretical methodologies makes biomolecular modeling and bioinformatics a basic tool to complement experiments in many ways. In this sense computational simulations can provide valuable information on the specific system under study. Chapter 4 will moreover stress on the importance of computational simulations and bioinformatic

tools to interpret experimental observables, to rationally guide further improvements providing data unaccessible or difficult to access by experimental studies.

The chapter has been divided in two parts. The first part aims to provide the reader with a broad scope about the most common and useful simulation tools in biomolecular modeling stressing on the advantages and limitations of each methodology. At first will be presented the basis of Quantum Mechanics (QM) based methods, including an analysis on the case studies to which QM can be fruitfully applied.

Even though QM represents matter at the most accurate level of description, QM based methods are computationally very expensive and their practical application in the context of the present book will be limited only to small parts of the system. In this sense the present chapter will illustrate the importance of QM methods to design new functional groups that can be used to functionalize β-barrel membrane proteins for nano-technological applications. Secondly atomistic Molecular Dynamics (MD) simulations and their suitability to simulate β-barrel membrane proteins will be presented. Atomistic MD simulations represent a system in the Molecular Mechanics framework, where atoms are represented as spheres and bonds as harmonic springs, and allow the simulation of systems to be analyzed at the nanosecond timescale with a resolution at the atomistic level.

However many important processes such as refolding or self-assembly of lipid/copolymer bilayers occur at the microsecond timescale, far beyond the timescale reachable by atomistic MD simulations. To that end will be introduced a third level of biomolecular simulations, *i.e.* the so-called Coarse Grained (CG) simulations. In CG simulations many atoms are grouped into a single bead and consequently the total number of particles is drastically reduced compared to the total number of particles present in a full atomistic description, which allows reaching longer timescales.

The second part of the chapter will briefly introduce some common bioinformatic structure prediction tools used to perform a first 2D and 3D structure prediction of the β-barrel membrane proteins focusing on the challenges and problems one might encounter when applying these tools to predict the structure of non-natural (engineered) β-barrel membrane proteins. The basic logic algorithms used by these tools will be summarized and reviewed.

1.3.4 Chapter 5

Chapter 5 will assist the experimentalist that plans to work with membrane proteins and in particular with bacterial β-barrel outer membrane proteins in the context of the development of new nano-material components (*i.e.* nano-channels). It will offer an overview on genetic engineering methods as well as expression, extraction and purification procedures considering standard techniques, new alternatives as well as methods compatible with scale-up and high yield production purposes, including difficulties that the beginner might encounter and tips and suggestions on how to overcome these difficulties.

In a narrower sense Chap. 5 will explain in detail how to practically transform a β-barrel membrane protein into a nano-channel with desired geometrical and functional features, starting from the gene design, when planning an entirely novel protein variant. Site-specific mutagenesis for slight changes or the introduction of amino acid residues suitable for chemical modification purposes will be furthermore discussed. An overview on commonly used protein chemical modifications (useful to introduce triggers and switches to OMP channels) will be given focusing on outer membrane protein examples. Here the FhuA protein of *E. coli* will serve as the main example, as it has great potentials as a starting-template to develop a set of engineered nano-channels with tailored geometry and controllable channel function and as the reader will see considerable steps toward this goal have been already made during the last decade.

Chapter 5 will then summarize the conventional means of production and purification of bacterial OMPs (outer membrane proteins), stressing on the problems and challenges of

over-expressing proteins belonging to these class into the Gram-negative bacterial outer membrane and providing the reader with ways to overcome these problems by using alternative methods. These alternatives (*i.e.* expression into inclusion bodies and cell-free expression) will be explained and examples for the successful OMP production will be given.

Different ways of analyzing OMP samples regarding yield, purity and correct folding will be considered briefly and the chapter will discuss distinctive experimental adaptations to scale-up the production of OMPs and genetically engineered OMP variants.

The chapter will then close with a brief discussion on artificial β-barrel structures that are developed for the use in nano-technological applications, as biological OMPs can be a valid alternative material to these chemically derived structures.

1.3.5 Chapter 6

The reconstitution of OMPs is a challenging "art", lacking a general applicable protocol or set of protocols. This lack is mainly due to the different stability and behavior of membrane proteins affecting the functional reconstitution into liposome or polymersome membranes. For example, in some protocols, the presence of organic solvents during the reconstitution process can perturb strongly the protein structure denying its functionality once inserted in the liposome/polymersome membrane.

In general there is also a lack of a detailed understanding on how microscopically β-barrel based membrane proteins reconstitute into a liposome/polymersome membrane, though some information have been translated from the α-helical membrane proteins and membrane inserting peptides.

The failure of a functional reconstitution of an OMP is generally recognized to be due to some variables where the hydrophobic mismatch between the embedded protein and the hydrophobic thickness of the polymer membrane, in particular, plays a fundamental role. Therefore the hydrophobic mismatch term will be defined and explained.

Other variables will be also considered such as the chemical nature of the detergent to solubilize OMPs and its miscibility with the polymer constituting the polymersome membrane, mechanical properties of the membrane as well as the polymer polydispersity, a further "obvious" variable that should be always considered when working with polymers.

A case study will be shown based on a genetically engineered β-barrel membrane protein (FhuA) so to obtain a longer hydrophobic portion, to overcome the hydrophobic mismatch with the used block copolymer membrane.

Further sections will describe some of the characterization methods used in the proteo-liposomes and proteo-polymersome systems, such as dynamic light scattering (DLS), spectroscopic flux assay, patch clamp, electron microscopy (EM) and tryptophan fluorescence.

At the end of the chapter a summary of the actual OMP based nano-channel applications will be given (*i.e.* drug-delivery, stochastic nano-sensing elements and bio-nanoelectronics).

1.3.6 Chapter 7

In Chap. 7 final considerations and future perspectives will be given and outlined.

References

1. Twede D (2005) The cask age: the technology and history of wooden barrels. Packag Technol Sci 18:253–264
2. Santamaria A (2012) Historical overview of nanotechnology and nanotoxicology. Methods Mol Biol 926: 1–12
3. Feynman R (1960) There's plenty of room at the bottom. Caltech Eng Sci 23:22–36
4. Pompe W, Rödel G, Weiss H-J, Mertig M (2012) Bionanomaterials: designing materials inspired by nature. Wiley-VCH, Berlin

Biology

2

This chapter will be an introductory section on the biology of membrane proteins in general, with regard to the different structural as well as functional classes of these proteins (*i.e.* α-helix and β-barrel). It will furthermore give some basic information on biological membranes and the lipids they consist of. The focus will be on the characteristics and unique structural and functional features of β-barrel outer membrane proteins. This chapter will then introduce to the *E. coli* outer membrane iron transporter FhuA (Ferric hydroxamate uptake component A), which as a member of the TonB protein-dependent transporters belongs to the largest known β-barrels. The FhuA has been successfully employed as a model for the transformation of a β-barrel protein into a custom-made nano-channel. By means of the FhuA example this chapter will therefore look at β-barrel proteins not only within the biological context but it will already introduce to the relevance and use of proteins belonging to this class for the nano-material sciences and specifically for the design of biological nano-channels.

2.1 Membranes and Membrane Proteins

Biological membranes, whose functional unit is the phospholipid, are barriers that shield the interior of pro- and eukaryotic cells as well as of organelles and compartments of eukaryotes from their external environment. At the same time membrane spanning proteins or proteins that are located on the membrane surfaces, allow communication with the outside and play a key role in cellular adhesion, molecular recognition as well as in the transport of nutrients and excretion processes; functionalizing the membrane and leading to its selective permeability.

Biological membranes consist of amphiphilic lipid molecules (mainly glycerophospholipids, but also sphingolipids and steroids) that self-assemble in water to form bilayers. Further components of biological membranes are proteins that interact with these bilayers. Single-membrane assembly occurs upon hydrophilic interaction of the phospholipid hydrophilic head regions and hydrophobic interactions between the hydrophobic fatty acid tails. The hydrophilic membrane part is in contact with the aqueous outer medium (*i.e.* cytoplasm, aqueous environment) and the hydrophobic part interacts with the hydrophobic region of a second single-membrane resulting in the formation of a double-membrane. In this way, the nonpolar fatty acid regions are sequestered from contact with water, maximizing hydrophobic interactions. Both membrane halves are aligned to each other in a parallel fashion. This double membrane character of all biological membranes be it Gram-negative bacterial outer membrane, plasma-membranes or internal membrane systems has been perceived as early as 1925 [1] leading to the term lipid-bilayer. Since then numerous models of the membrane structure have been proposed starting out from a suggested very thin lipoid film with a protein film adsorbed upon it [2], as shown in

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_2,

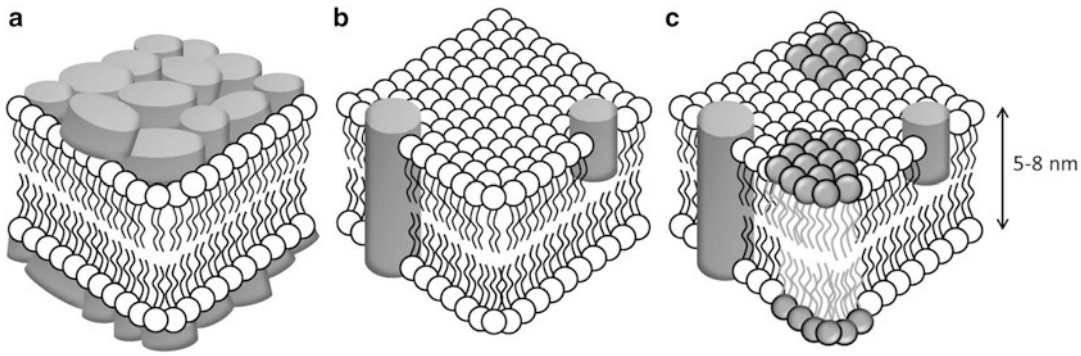

Fig. 2.1 (**a**) Schematic representation of the proposed lipoid film with adsorbed protein film; (**b**) fluid-mosaic model with a homogeneous array of membrane lipids and globular integral or peripheral proteins; (**c**) Schematic representation of the lipid-raft model with a fluid membrane in which phase-separation occurs within sphingolipid and cholesterol rich, raft-like micro-domains (*dark grey*)

Fig. 2.1a, eventually resulting in the now still commonly used fluid mosaic model [3]. The fluid mosaic model states each half of the membrane as a homogeneous lipid environment behaving as a two-dimensional fluid. Incorporated in or floating on this fluid lipid membranes one finds globular proteins and glycoproteins that alternate the membrane's phospholipids in a mosaic fashion (Fig. 2.1b) and that float around more or less freely in the two-dimensional plane of a uniform membrane. However at the time the fluid mosaic model was proposed (1972) not much was known about the actual organization of biological membranes. In 1997 an important addition to the fluid mosaic model led to the understanding that membranes are fluid but not homogeneous in their lipid composition and their molecular dynamics. Based on findings on eukaryotic cell and organelle membranes Rietfield et al. and Brown et al. suggested that within relatively homogeneous, crystalline phases one finds phase-separated domains rich in sphingolipids and cholesterol with proteins that bind specifically to these lipids, so called lipid rafts (Fig. 2.1c) [4]. These rafts are somewhat thicker than the rest of the membrane due to the bulkiness of the sphingolipid head group. While the raft hypothesis with sphingolipid and cholesterol rich patches within the membrane remains controversial, the occurrence of phase separation within biological membranes is a fact.

Membrane lipids can be classified into three separate groups: *phospholipids* (*e*.g. glycerophospholipids, and certain sphingolipids), *glycolipids* (*e.g.* glycosphingolipids) and *sterols* (*e.g.* cholesterol). Most abundant are glycerophospholipids, with phosphatidylcholine, consisting of a glycerol backbone and a two fatty acid chains tail, being the most prevalent glycerophospholipid in biological membranes. The polar head group is bound to the third carbon atom of glycerol and is composed of an alcohol (*e.g.* glycerol, ethanolamine, choline, inositol and serine) [5]. The different membrane lipid classes vary in size, charge, hydration extend and hydrogen donor/acceptor number of their head groups as well as length and conformational freedom (as function of saturation and branching) of their tails. Therefore the flexibility, thickness and organization of a membrane are strongly dependent from its lipid composition [6]. For a very complete lipid information collection with many useful links, the interested reader might want to check Dr. Claude Leray's website "The Cyberlipid Center" (http://www.cyberlipid.org). The composition of the two plasma membrane layers differ from each other with phosphatidylethanolamines and phosphatidylserines in the inner half and phosphatidylcholines and sphingomyelins in the outer half [7], leading to the curving of the membrane. Furthermore it is known that

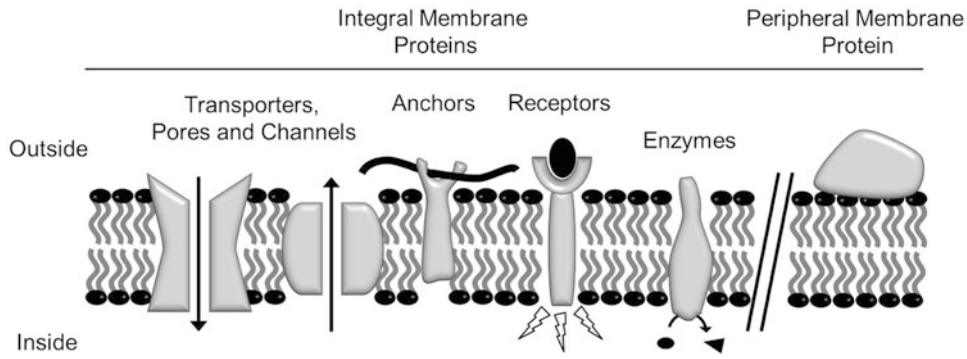

Fig. 2.2 Schematic representation of lipid-bilayer with a set of membrane proteins with different functions

membranes are highly dynamic structures with diffusion of proteins and membrane lipids that may either diffuse laterally or even pass from one layer of the membrane to the other [8].

The overall composition of biological membranes is flexible and dynamic; it differs from organism to organism, organelle to organelle and changes according to the physiological status of a cell and due to its environment [6]. Changes in temperature for instance lead to changes in the membrane composition to maintain a constant fluidity. The ratio of saturated and unsaturated fatty acids influences the membranes solid-to-liquid transition temperature, as different lipids have a different transition temperature. In general it can be said that the more saturated and trans-unsaturated fatty acids or those with long hydrocarbon chains are present, the higher the transition temperature [9, 10]. Furthermore osmotic stress may influence the ratio of different fatty acids present in a membrane [11].

However flexible biological membranes may be they still have a function as barriers to maintain the integrity of what they enclose. Due to their lipidic nature membranes are impermeable for most polar substances, while apolar compounds pass the lipid bilayer. The barrier function of the membrane can be overcome when needed.

The proteins that either span the membrane, are anchored to one of the membrane layers or that are bound to its outer or inner surface by electrostatic or Van-der-Waals interactions or covalently, can be considered the main "communication devices" to allow contact with the environment and are thus a key feature of each membrane. In fact these proteins closely attached or bound to the membrane can make up more than half of the membrane's dry weight [12], in bacteria up to 25 % of all cellular proteins are membrane proteins (MP) and up to 20–30 % of an organisms genes code for MPs [13, 14]. In addition MPs make up more than 50 % of all drug targets [15]. The number of proteins present in and on a membrane differs from organism to organism and depends on where the membrane is localized (*i.e.* Gram-negative bacterial outer membrane, plasma membrane or membrane of an organelle) and on the membranes function.

In 2013 there were approximately 79,000 (accessed 20-04-2013) protein structures solved by X-Ray diffraction analysis saved within the RCSB (Research Collaboratory for Structural Bioinformatics; www.rcsb.org) Protein Data Bank (PDB). Only about 300–400 of these structures belong to integral MPs, showing the current lack of structural understanding on the majority of MPs. About their function is known considerably more. Some various vital cellular functions of membrane proteins are summarized in Fig. 2.2; reaching from receptor functions to transport, anchoring, and enzymatic functions in case of integral membrane proteins, while most peripheral MPs act as regulators for integral proteins [16].

As already mentioned membrane proteins can be integral and span the lipid bilayer, they can be anchored to one membrane layer or they can be located on the membrane surface as peripheral proteins.

Peripheral MPs (also called monotopic membrane proteins) can be bound directly to the membrane by electrostatic interactions between positively charged amino acid residues of the protein and negatively charged lipid head groups (like phosphatidic acid, phosphatidylglycerol or phosphatidylserine), van der Waals interactions with lipid head groups [17] or through a hydrophobic anchor. They can furthermore be indirectly bound to the membrane through electrostatic or hydrophobic interactions with integral MPs. Peripheral proteins can often be stripped from the membrane by increasing the ionic strength of the surrounding solution, by changing the pH or by the addition of mild detergents, pointing out the reversible character of the bond. As peripheral proteins are localized always on the outer membrane surfaces facing in case of the plasma membrane either the cytosol or the outside aqueous medium, they behave like soluble proteins.

Peripheral proteins that bind to the membrane by interaction with the membrane lipids do so via a hydrophobic hydrocarbon chain (*i.e.* palmitoyl, myristoyl or farnesyl) or a lipidic glycosylphosphatidylinositol (GPI) anchor that is attached post-synthetic. The palmitoylation of cystein residues of peripheral proteins is a reversible process adding to the dynamic changes of the membrane composition [18, 19]. Many peripheral MPs contain certain domains that recognize different membrane lipid head groups, the so called PH (pleckstrin homology) domain for instance specifically binds to the phosphoinositide headgroup [20, 21]. Peripheral proteins often interact with integral MPs and are involved in many important cellular functions such as cell-adhesion, regulatory mechanisms, signaling cascades [22], cell-cell recognition, keeping contact with the cytoskeleton or metabolic functions in case of peripheral MPs with enzymatic functions. The peripheral MP's relevance for the development of nano-materials can be seen mainly in their binding interaction with the membrane surface, however as this interaction is reversible use as for example targeting devices in drug-release systems is limited. And though peripheral MPs (especially the ones that act as regulators of integral MPs) make interesting targets for study and genetic modification, the methodology used to study and alter soluble proteins can be applied. Therefore peripheral proteins will not be discussed further.

The second main class of membrane proteins, integral or transmembrane proteins (TMP), is characterized by a membrane spanning, highly hydrophobic region that strongly interacts with the membrane lipids hydrocarbon chains and hydrophilic extra-membrane portions reaching into the aqueous outside. All integral MPs therefore have "to cope" with a heterogeneous solvent with varying dielectric properties (consisting of an aqueous part outside of the membrane, an ionic part on the surface of the membrane and the hydrophobic inner part of the membrane) very different from the conventional aqueous solvent of all soluble cytosolic proteins [23]. While cytosolic, water-exposed proteins show a variety of different folds, the folding of integral MPs is restricted to two main folding classes (α-helical proteins and β-barrel proteins). This limitations in folding diversity can be explained by the physical and chemical constraints imposed on transmembrane proteins by the lipid bilayer, as well as by the interface character of the surrounding environment [24].

Integral MPs (referred to as membrane proteins or MPs in the following) can be obtained only after disrupting the membrane by the use of detergents. As they are not soluble in water, in order to isolate, purify and study them, one needs to provide them with a solvent similar to the membrane environment, satisfying their high hydrophobicity. Lipids, detergents, organic solvents or polymers can be used for this purpose, however reconstitution of purified membrane proteins can be rather challenging [25]. In order to study MPs one often needs greater amounts of pure protein that can only be obtained by over expression in a suitable host [26]. The difficulties of MP production and purification will be discussed in detail in Chap. 5 (Sect. 5.2.1). Furthermore the MPs hydrophobicity leads to the formation of protein aggregates by hydrophobic interactions upon isolation and they generally show poor X-ray diffraction when crystallized. These handling difficulties result in

the mentioned under-representation of membrane protein solved structures. However a continuous and ever growing effort to optimize or find new methods to study and learn more about the structure of these intriguing class of proteins is leading to an almost exponential growth of the number of solved MP structures [27].

As said before MPs fulfill important cellular functions. As receptors they receive chemical signals from the outside resulting in an answer inside (signaling cascades) and as anchors they attach cells to each other or to the substrate. However the most interesting group of integral MPs forms "holes" in the membrane they reside in. These channel-forming proteins act as transporters and pores that allow the passive transport (along a concentration gradient) or active transport (energy consuming) to translocate substances like ions or solutes that due to their nature and characteristics (*i.e.* size, charge or other) cannot pass the plain membrane (see Fig. 2.2). The channel opening may occur upon an external stimulus, a change in membrane current or upon mechanical stimulation, *i.e.* pressure on the membrane [28].

The two known structural classes of integral MPs are: the ubiquitous *α-helical proteins* that can be found in all types of membranes apart from the outer membrane of Gram-negative bacteria (single transmembrane helix, helical bundle or discontinuous helices) and the more specialized *β-barrel proteins* that are a unique feature of the outer membrane of Gram-negative bacteria and can be found also in mitochondrial and chloroplast membranes. Both are named after their main tertiary structure elements. The characteristics, biological facts and advantages/disadvantages for the development of new nano-materials will be discussed for both classes in the following sections.

2.2 Structural Classes: α-Helical Proteins

About 25 % of the proteome of all organisms consists of integral membrane proteins with α-helical secondary structure [29]. These proteins can be found in the plasma membrane, the membrane of the endoplasmic reticulum and other organelles. Each transmembrane helix is highly hydrophobic and on average consists of more than 25 amino acids, while the membrane spanning part of each helix is 15–20 amino acids long [30]. The hydrophobicity is due to an over-abundance of hydrophobic amino acid residues, while polar residues (especially negatively charged ones) occur rarely [31]. The most simple α-helical MPs contain just one membrane spanning helix, as it is the case for receptor tyrosine kinases [32] or glycophorin [33].

These single-helix proteins often form oligomers and oligomerization can be essential to form an active structure. For instance in the potassium channel KcsA, four monomers form the central potassium conductive pore [34]. More complex bundle shaped proteins such as the well-studied bacteriorhodopsin proton pump of *Halobacteria spec.* consist of seven transmembrane helices connected by short external loops [35]. This seven transmembrane helix motif is a common structural element in α-helical MPs, found in signal receptors. For a long time these regular helix bundles were the only known structures among α-helical integral MPs until proteins with discontinuous helix structures were found that may include kinked membrane spanning helices [36], reentrant regions entering and exiting the bilayer on the same side thus not completely spanning it [37, 38] or coils in the membrane [39]. One well-known example for a helical membrane protein containing several irregularities like disrupted helices, non-helical trans-membrane regions and reentrant loops is the glutamate transporter homologue [40] (see also Fig. 2.3). Many helical bundle proteins, as well as the ones that contain discontinuous helices, function as ion channels or transporters.

By the use of selected examples the following section will give a more detailed discussion of the structure and function of α-helical MPs, giving an overview on their amino acid composition and resulting secondary and tertiary structures, on their quaternary structure and thus their geometrical features, as well as on their various functions. Structural families and function examples will be introduced.

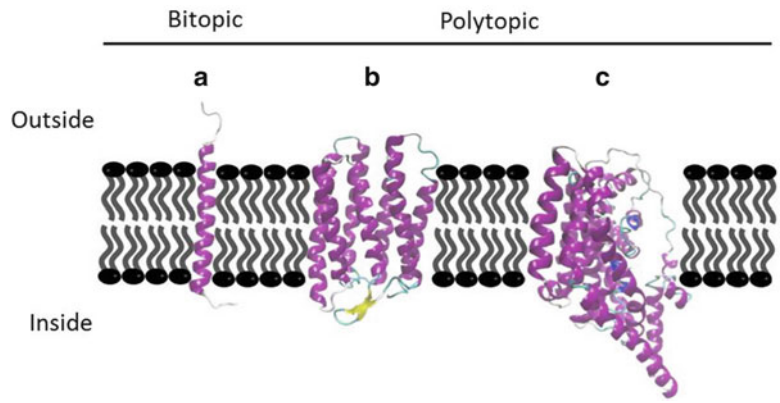

Fig. 2.3 Schematic representation of a lipid membrane with different types of integral membrane proteins with α-helical transmembrane regions in New Cartoon representation (made with VMD). (**a**) Transmembrane helix of growth factor receptor ErbB2 (2JWA; [32]); (**b**) Bacteriorhodopsin (1C3W; [42]); (**c**) Glutamate transporter homologue monomer (3V8G; [43])

2.2.1 α-Helical Membrane Proteins – Structure, Geometrical Features and Function

Integral MPs with α-helical transmembrane regions can be divided into bitopic proteins, which span the membrane once, having only one helix and polytopic proteins, which span the membrane several times [41] as opposed to the monotopic, peripheral proteins (see Fig. 2.3).

When considering the structure (from primary to quaternary) of MPs in general, it can be seen as an adaptation to their heterogeneous environment (aqueous outside, interface, "oily" membrane) thus the structure needs to match the different solvents dielectric properties and it has to align to the lipid hydrocarbon chain packing with hydrogen bonds satisfied. Statistical analysis of the *primary sequences* of α-helical MPs revealed that residues like Ala, Cys, Phe, Gly, Ile, Leu, Met, Ser, Thr, Val, Trp and Tyr are common in the membrane spanning helices. This would be expected for amino acids like Phe, Ile, Leu, Met, Val, Ala, Gly and Trp as their side chains are nonpolar. And in fact a simple comparison of the ErbB2 growth factor receptor transmembrane helix primary sequence, as an example for a typical transmembrane helix with one of the ovine interferon helices (helix 73–95), as an example for a soluble helix shows an amount of polar amino acids of 56 and 21 % respectively (see Table 2.1).

It was found however that the amino acids Ser and Thr in spite of their polar nature prefer to reside within the membrane, also. The same is true for all aromatic amino acids [44]. Originally it has been assumed that Pro as a "helix breaker" will not be found [45], while later studies revealed that proline can be found quite frequently within transmembrane helices [46], which is true also for the membrane spanning helix of the ErbB2 growth factor (Table 2.1). As will be explained when considering the α-helical MPs tertiary structure the amino acid Pro distorts the helix and makes it more flexible, thus playing an important role in membrane helix packing [47].

In conclusion it has to be pointed out that the simple approach of just checking an amino acids polarity can give only a very first clue on whether it is likely to be found within a transmembrane region. A more forward approach is to look at its hydrophobicity value. The hydrophobicity of an amino acid is related to its free energy of transfer from a polar medium (such as the cytoplasm or water) to an apolar medium (like a membrane or octanol). Hydrophobicity scales are scales that give numerical values to define an amino acid's hydrophobicity. Most of these scales are experimental scales derived from

Table 2.1 Primary sequences of ErbB2 growth factor receptor transmembrane helix and ovine interferon helix 73–95

Protein name, UniProt ID	Primary sequence
ErbB2 growth factor receptor transmembrane helix (P04626)	Gly Cys Pro Ala **Glu Gln Arg** Ala **Ser***Pro* Leu **Thr Ser** Ile Ile **Ser** Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Phe Gly Ile Leu Ile
Ovine Interferon helix 73–95 (P56828)	**Glu** Met Val **Glu** Gly **Asp Gln** Leu **Gln Lys Asp Gln** Ala Phe Pro Val Leu **Tyr Glu** Met Leu **Gln Gln Ser** Phe **Asn** Leu Phe **Tyr Thr Glu His**

Polar amino acids are bold and underlined. The "helix breaker" Pro within the ErbB2 growth factor receptor is written in italics

measuring the concentration of the molecule of interest in the two different phases. One of the first scales applied to α-helical MPs has been developed in 1986 [48]. The problem with applying such scales to amino acids is that free amino acids obviously behave very different from amino acids within a peptide or protein. Exposed amino acids will interact stronger with a given solvent than buried ones that are not readily accessible to the solvent. Furthermore structural changes within the protein may occur upon solvent change [49]. It is therefore much more accurate to take into account crystallographic data, and the broadly used scale proposed by Kyte and Doolittle rises from a combination of crystallographic and solution-based methods [50]. The experimental scale by Wimley and White focusses on the transition at the interphase between the aqueous environment and the membrane bilayer and is therefore well-fitted to be applied to membrane proteins [51].

While older experimental studies use short model peptides to study the phase transition behavior of single residues, recent studies on MPs of eukaryotic cells try to give a more complete picture by using the cells own transmembrane helix recognition system, the Sec61 translocon that will be discussed in detail in Sect. 2.2.2. In this way Hessa et al. (2005) offered a set of specifically designed peptide sequences to the translocon and measured how well these peptides were inserted into the membrane [52]. The results of this study are consistent with the values set by a purely database knowledge-based hydrophobicity scale for polytopic MPs, published in 2009 [49].

Further information that can be obtained by statistical analysis of the protein primary structure is the occurrence of significantly recurrent sequence motifs. In fact several motifs have been found in α-helical MPs. The first such pattern found within the membrane spanning regions has been named the "GxxxG" motif as it always consists of two Gly separated by three other amino acids (x) or GpA motif after the protein in which it was noticed first, the glycophorin A [53, 54]. The GpA motif as well as the patterns GxxxGxxxA and AxxxGxxxA play a role in the interaction between transmembrane helices and thus the packing of helices within the membrane [37, 55]. Walters and DeGrado found that many recurrent sequence motifs within the transmembrane protein regions consist of small chain amino acids like Gly, Ala and Ser separated by three other residues (x) [56]. Since then several other motifs have been identified that are unique for membrane protein, such as a dimerization motif (A- [AS]- [FIV]- [NR]-A-P-L- [AT]-G) found in voltage gated chloride channels or the loop sequence motif [WY]-x(2)-Y-P-P-L that occurs at the interphase between membrane and water [57].

The oil-like nature of the hydrocarbon interior of bilayer membranes limits the possible *secondary structures* that are able to overcome the thermodynamic costs to integrate into this environment to two types: α-helices and β-sheets that assemble to a β-barrel tertiary structure. These secondary structure elements have in common that the peptide amide nitrogen and carbonyl oxygen form hydrogen bonds with each other in a regular fashion. The number of available solvent hydrogen donors is strongly reduced in comparison to what soluble proteins encounter within their aqueous environment. It seems that the helical structure, in which all backbone carbonyl oxygens and amide protons are intrahelically hydrogen-bonded to reduce the energetic cost of exposing unsatisfied hydrogen bond donors/acceptors to

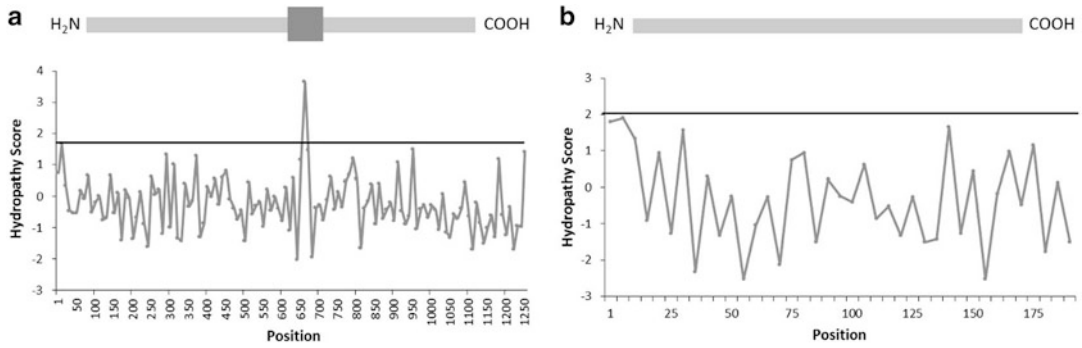

Fig. 2.4 Hydropathy plots for (**a**) the ErbB2 growth factor receptor (P04626) and (**b**) ovine interferon (P56828). Amino acid hydrophobicity values were obtained using the Kyte & Doolittle scales [50]. The single transmembrane helix of the ErbB2 receptor is indicated by the *dark grey box* above the plot

the hydrophobic membrane core [58, 59], is especially suited to answer to the demands the membrane as a solvent makes. For this reason most known transmembrane elements are α-helices consisting of 15–20 mainly apolar amino acids, to accommodate to the lipid hydrocarbon chains length of 14–24 CH_2. A comparison of known structures revealed that in a typical transmembrane helix, hydrogen bonds are as said intrahelical, with interactions between residue i and residue $i+4$. Based on this knowledge it is possible to roughly predict a membrane spanning α-helices from a protein sequence and to distinguish them from non-transmembrane secondary structure elements. Predictions are based on the mentioned hydrophobicity scales, plotting the values for each amino acid of the analyzed protein against the residue number. The protein is scanned from the N- to the C-terminus. If the average hydrophobicity of a certain sequence region lies above a base value this region might be a transmembrane secondary structure element [60]. Figure 2.4 shows such plots for the ErbB2 growth factor receptor (left) and the ovine interferon (right) based on the Kyte & Doolittle scales. Only the growth factor receptor sequence shows one short region above the hydrophobicity value threshold of 1.8 corresponding to its single transmembrane helix (see also Table 2.1 and Fig. 2.3). Further information on the membrane protein structure prediction can be found in Chap. 4.

The membrane spanning helices of polytopic proteins are generally connected by loop shaped secondary structure elements with various lengths, containing mainly polar amino acids, as these loop regions protrude into the aqueous non-membrane environment.

As pointed out a MPs *tertiary structure* is, as is its secondary structure, a function of the amino acid preference depending on whether a residue resides within the membrane, the interphase or the aqueous environment. Therefore one finds amino acids with nonpolar side-chains mainly on the surface of membrane spanning regions, standing to reason, as the membrane interior is nonpolar. Outside-facing polar residues within the transmembrane regions of a protein are rare, as they lower the structural stability [31, 59, 61]. Polar groups that are found in the MPs hydrophobic region generally interact with other hydrophilic residues in the protein, build the surface of internal water filled cavities and channels or if close to the hydrophilic lipid head groups they can reach out toward these groups. Amino acid distributions in the protein parts that reside outside the membrane and that are in contact with water, resemble distributions found in soluble proteins, while on the interphase between the watery outside and the lipid membrane a higher number of tyrosines and tryptophanes may be found [29].

This rather simple generalization on the tertiary structure led to the assumption, that

α-helical MPs form simple bundles, with straight membrane spanning helices that are connected by extra membrane loops. In fact proteins like the seven-helix bundle bacteriorhodopsin (see Fig. 2.3) follow this cut down rules, with seven helices that traverse the membrane in a more or less perpendicular fashion, showing a "knobs-into-holes" geometry [62], in which the nonpolar sidechain positions of the first helix, the "knobs," superimpose between the nonpolar sidechain positions in the second helix, the "holes", to maximize helix-helix packing. In this way two helices interact via van der Waals interactions between their side chains, as well as by polar interactions. This resembles the coiled-coil structure of soluble helical proteins [63]. As the rise/residue of an α-helix is 3.6, approximately every third residue should be polar for two interacting helices and two of three residues in case multiple helices interact with each other. These considerations on helix association also apply to the formation of quaternary structures of MPs.

The very common four-helix bundle proteins follow the straight forward "knobs-into-holes" geometry as well. In the four helix bundle four α-helices are organized into a bundle with an angle of about 20° between the helical axes. The packing of helices is so tight that the hydrophobic bundle core is void of water. The helices in the four-helix bundle can be either parallel as in the human growth hormone [64] or anti parallel as found in cytochrome c [65].

However since then the picture became more and more complex as many proteins have been found that differ from the "knobs-into-holes" structure arrangement. Proteins like aquaporin or the glutamate transporter homologue (Fig. 2.3c) show variations such as "reentrant" regions that do not fully traverse the membrane, but span only part of it to exit on the same side of the membrane as they enter [55]. Other differences from the general model are transmembrane helices that are kinked, interrupted or tilted [36]. Kink-like distortions within the helix structure are known to be the result of the amino acid proline that as a "helix breaker" is rarely found in the helices of soluble proteins [66, 67]. The helix distortions by proline occur to avoid a steric hindrance between the proline ring and the adjacent carbonyl group within the protein backbone. Furthermore less hydrogen bonds are formed leading to an increased flexibility within the kinked region [47], thus affecting the packing of helix pairs and leading to hinge structures that play a role in signal transduction. The transmembrane helices of polytopic proteins are connected by loop structures that differ in their amino acid composition, depending on whether they face the inside or outside of the cell. This "positive-inside" rule may be explained with the mechanism involved in the insertion of transmembrane domains and the electrochemical potential over the respective membrane that leads to positively charged residues (*i.e.* arginine and lysine) staying on the cytoplasmic side more often than not [68], especially in case of *E. coli* outer MPs [69]. Extramembranous tertiary structure elements that are not just connecting loops, often are related to a proteins function and may be involved in ATP (adenosine triphosphate)- or cation binding.

The *functions* of α-helical membrane proteins are diverse and include receptor functions such as is the case for the acetylcholin receptor [70], light driven transporters as Bacteriorhodopsin [35], plant and bacterial photosynthetic light-harvesting complexes [71, 72], influx and efflux transporters belonging to the ABC (ATP-binding cassette) superfamily [73], voltage gated ion transporters [74], ammonia and urea transporters [75, 76] or bacterial vitamin uptake transporters [77]. For readers interested in an extensive database on functional and phylogenetic classification of transporter proteins, visit the "Saier Lab Transporter Classification Database" (http://www.tcdb.org/).

Two structurally and functionally well-studied examples of α-helical MPs are the *Streptomyces lividans* KcsA potassium channel [34, 78] and the ubiquitous aquaporin water channels whose structures and functions have been examined extensively, both resulting in a Nobel Prize in chemistry in 2003 awarded to Professor Peter Agre for his discovery of the aquaporins [79] and to Professor Roderick MacKinnon for his studies on the function of the potassium channel.

The bacterial tetrameric KcsA potassium channel is a model ion channel, as many of its features are analogous to eukaryotic and also mammal potassium channel features. Its opening is activated by acidic pH values. Its narrow and highly selective pore, acts as a selectivity filter leading to a 10,000 times higher selectivity towards potassium than towards sodium ions. The selectivity filter is formed by the amino acid sequence: Thr, Val, Gly, Tyr, Gly. Below the filter lies a water-filled cavity. The KcsA channel transports 10^8 K^+ ions per second, which is close to the diffusion limit [34].

The six transmembrane helix bundle-containing aquaporins or water channels are widely conserved and have been found so far in bacteria, plants, amphibians, insects and mammalian cells and have been isolated first from human kidney cells [80, 81]. As aquaporins allow the passage of water molecules (and in some cases also of small uncharged water soluble compounds as ammonia, urea or glycerol) they increase a membrane's water permeability, by conserving the membrane's impermeability towards ions, as they do not permit the passage of charged species. While aquaporins in monomeric form act as functional water channels, they have been found to form tetramers of four active channels within the membrane [82].

Both the aquaporins and ion channels as the KcsA potassium channel have among others great impact on the application of α-helical proteins with pore structure as functional nano-materials, as will be discussed in Sect. 2.2.3.

2.2.2 α-Helical Membrane Proteins – Genetics, Biogenesis, Folding and Insertion

As far as DNA to mRNA transcription and the subsequent translation to an amino acid chain are concerned, the involved mechanisms are the same for soluble proteins and MPs. In both eukaryotes and prokaryotes translation takes place in the cytoplasm and while soluble proteins fold while they are still synthetized at the ribosome, things are more complex for membrane proteins. The fact that MPs are synthetized within the cytoplasm leads to the questions on how they are transported to the respective membrane, how they insert into this membrane and how and when they obtain their correct folding. In pro- and eukaryotes there exist co- and post-translational pathways for the translocation of a protein to and the insertion into a membrane.

The most abundant plasma membrane protein insertion mechanism is the so called Sec-pathway:

In case of prokaryotes the Sec pathway that depends on the signal recognition particle (SRP) that binds to a short N-terminal signal sequence characteristic for MPs, while the protein is still synthetized and in contact with the bacterial ribosomes. The bound SRP then interacts with the signal recognition particle receptor (SR) within the plasma membrane. As the SR is located in the proximity of the secretase translocon, or SecYEG-translocon (a complex of three proteins Sec Y, E, G), the nascent polypeptide chain can exit through the ribosome exit gate to be transferred into the translocon channel. Meanwhile translation will still proceed. Protein parts passing the translocon channel that show sufficient length and high hydrophobicity will then exit the channel laterally to enter the lipid bilayer [62, 83]. Hydrophobicity plays a key role as transmembrane helices seem to exit the translocon to be inserted into the membrane by means of a process, in which hydrophobic helices will stay in the lipid bilayer, while less hydrophobic elements prefer the translocon channel. Such a mechanism explains also the evident amino acid distributions one finds in helical MPs (see Sect. 2.2.1). Furthermore a helix hydrophobic interaction with other helices that are already positioned within the membrane has been shown to trigger insertion [84].

Correspondingly one finds in eukaryotes a SRP/SR dependent Sec pathway for transport and membrane insertion of proteins that are targeted toward the membrane of the endoplasmic reticulum (ER). Again the SRP will recognize and bind to an N-terminal signal sequence within the MP primary sequence as soon as this sequence exits

the ribosome. In eukaryotes this leads to a chain elongation stop until the ribosome contacts the translocon. After which translation continues and translocation/insertion occurs as described [85].

Apart from the MP insertion via Sec pathway one finds several other mechanisms in the membranes of some eukaryotic organelles (mitochondria, chloroplasts and peroxisomes). Chloroplasts own different translocons for the inner and outer membrane (TOC/TIC) and the so called twin arginine translocon (TAT) that besides its main function to facilitate secretion of proteins over the thylakoid membrane is involved in the insertion of MPs [86]. The complex chaperon-dependent mechanism involved in the import of proteins into mitochondria, seems to be the same that is responsible for the membrane insertion of proteins into the mitochondrial membranes [87], while many peroxisomal proteins originate from ER vesicles [88].

A simple approximating thermodynamic model considers the folding of α-helical MPs as a two-step process that occurs co-translationally [89]. The foundation of this model is that the free energy minimum of a helical MP native conformation depends on the interactions between solvent molecules, intramolecular interaction within the protein itself, and interactions between solvent and protein molecules. During the proposed first step, hydrophobic regions within the protein primary sequence associate with the lipid bilayer upon which insertion occurs. The secondary structure formation takes place simultaneously. In the second step the individual helices associate by interacting with each other to form more complex tertiary structures [90] (see Fig. 2.5). While the two step model can be applied to the insertion of short single helix peptides, MP folding is more complex and the simple two-step model does not take into account all the different problems and effects that come with membrane insertion, protein-lipid interaction, partitioning of membrane components and helix packing within the membrane [91] (For a more detailed discussion on MP reconstitution into artificial lipid or polymer membranes see Sect. 6.1.1). It seems that large structural rearrangements during folding occur quite commonly and that some protein elements as reentrant loops might be inserted into the bilayer post-translational [92].

Nevertheless a number of MPs can be denatured and refolded in vitro, including mainly β barrel proteins and proteins belonging to the class of light-harvesting complexes, the first of which has been bacteriorhodopsin [93].

2.2.3 α-Helical Membrane Proteins – Relevance for Nano-material Development

Channel-type MPs in general have a huge impact on the development of biologically derived or bio-mimetic nano-materials, as they are rather robust and show a natural affinity towards hydrophobic environments such as lipidic- or polymeric membranes. Among the class of α-helical MPs two examples reveal the significance of these proteins as nano-devices with the potential to functionalize membrane systems or vesicular nano-containers.

The *ion channel* protein family is certainly an interesting target to study for various reasons, such as their medical relevance, for mutations in human ion channel coding genes are related to many diseases [94] or their relevance as drug-targets. Within the nano-technology field they were found suitable as biosensors to substitute or complement enzymatic sensors for analytes such as drugs, toxins, viral antigens or DNA to name just a few [95]. The main feature that enables ion channels to act as biosensors is that binding of specific chemicals (ligands) to an extracellular domain, current-change or mechanical stimulation result in a protein conformational change and thus in channel opening and change in channel conduction. Ion channels therefore contain a natural inbuilt "switch" [96]. As the KcsA ion channel is selective for potassium ions, there are ion channels that selectively pass single Na^+, $Ca2^+$ or Cl^- ions. The ion flux results in a change in current that can be measured by several methods of which the most widely-used surely is the so called "patch-clamp" technique [97] and automated patch-clamp variants (see

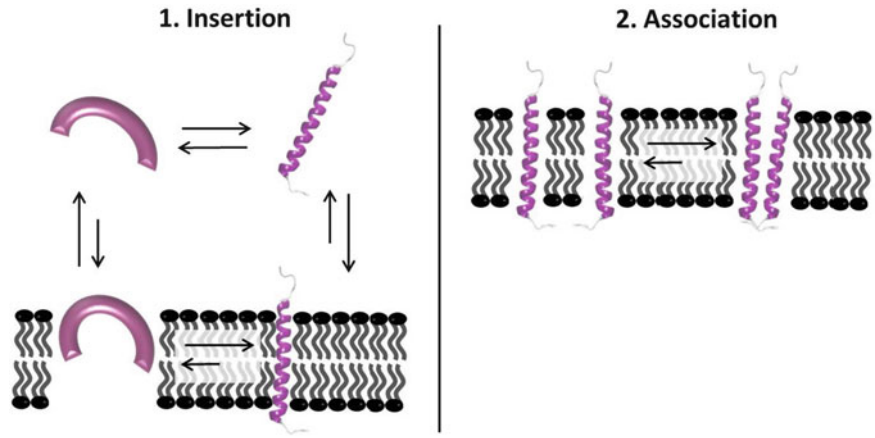

Fig. 2.5 Two-step folding model for α-helical TMPs. Unfolded primary sequence chains fold into secondary structure elements and insert into the membrane simultaneously during step 1. It is unfavorable for an α-helix to unfold in the bilayer as in this apolar environment hydrogen bonds need to be satisfied. In step 2 these helical structure elements associate to form tertiary structures (New Cartoon representations of helices made with VMD)

Sect. 6.1.3). The great advantages of ion channels as biosensors are their great specificity for one single molecule, their sensitivity in which one single channel can detect a single analyte molecule and their ability to transduce and amplify a small chemical signal into an easily measurable response. Furthermore as proteins they can be modified by use of protein-engineering tools (a feature that applies to all MPs with nanotechnological application). In 2010 for instance a constitutively open KscA potassium channel mutant had been reported [98]. Examples for ion channel-based biosensors are potassium or chloride channels used to screen for channel affecting drugs [99, 100].

And recently much effort has been put into the development of new membrane carriers, such as silicone chips [101] or by suspending a membrane patch on a glass surface [102], as the lipid bilayer is rather unstable and not suited for long-term use.

The *aquaporin* water channels obtained great interest for their potential use as nano-sized water-treatment filters, as the majority of aquaporins allows the specific passage of water molecules at rates near to the diffusion limit [103]. This specificity leads to filtered water of great purity and as isolated aquaporins have been successfully reconstituted into synthetic polymer membranes, the development of stable filter systems is possible [104]. This even caught the attention of NASA that is currently testing in collaboration with the Danish company Aquaporin A/S (http://www.aquaporin.dk/) whether aquaporin-functionalized membranes can be utilized to supply the International Space Station (ISS) Extravehicular Mobility Unit space suit cooling system with the necessary ultra-pure water [105].

However there are certain disadvantages to the α-helical ion channels as nano-materials. They are structurally complex and often reveal a narrow pH and temperature range and they require certain ionic conditions to work properly. Problems that can be avoided by using the more stable and structurally simple β-barrel proteins (see Sect. 2.3.3).

2.3 Structural Classes: β-Barrel Proteins

The second structural class of integral membrane proteins, the β-barrel membrane proteins or "outer membrane proteins" (OMP) can be found solely in the outer membranes (OM) of Gram-negative bacteria, and in membranes of the eukaryotic organelles with an outer and inner

membrane (mitochondria and chloroplasts). Approximately 2–3 % of the genes of Gram-negative bacteria code for β-barrel proteins. β-barrel MPs cannot be found in the inner membrane or cytoplasmic membrane, although physically their assembly into these membranes should be possible. A probable reason for the β barrel protein absence in the cytoplasmic membrane is their effective pore creating ability that would dissolve the proton motive force [106]. Indeed a special class of barrel shaped MPs, such as the α-hemolysine of *Staphylococcus aureus*, are bacterial toxins that form pores within host energy transducing membranes [107, 108].

All proteins with barrel structure contain an even number of meandering all-next-neighbor antiparallel sheets, with each transmembrane strand being at least seven amino acids long to span the hydrophobic membrane core [109]. Strand residue numbers can be up to 25 and are on average 12.5 [110]. Up to now the smallest known β-barrel membrane protein contains eight β-strands, while the largest consists of 24 strands; water soluble β-barrel proteins show instead a narrow range in size from six to eight strands only [111]. While the water soluble barrels expose their polar residues to the outside aqueous environment and keep nonpolar residues within their hydrophobic core, transmembrane barrels show an inside-out structure where polar residues form a solid core or, from 12 strands on, a water-filled channel. Though the β-barrel is a stable and even rigid structure, barrel shaped MPs greatly vary in the number of included sheets and thus size, they can either be monomeric or exist as oligomers. Furthermore they show many different pore/channel opening and closing mechanisms, as well as a variety of functions, including passive nutrient uptake, active transport of ions, membrane anchor function or enzymatic activity [112]. Table 2.2 lists the features of eight exemplary barrel shaped MPs, clearly showing the variation in number of strands/size, oligomeric states and functions.

Monomeric, solid bacterial outer membrane anchors such as the Outer membrane protein A (OmpA), which can be found in ∼100,000 copies/*E. coli* cell [110], consist of an 8 stranded β-barrel structure and represent the β-barrel outer membrane protein minimal construction [113, 114]. An increase in β-strands leads to increased diameters and formation of a pore or channel as can be seen in the OpcA adhesin of *Neisseria meningitidis* [115] or the Tsx nucleoside transporters of Gram-negative bacteria [116]. The construction maximum in nature seems to be reached with 22–24 β-strands as can be found in the iron transporting, Ton-B dependent "Ferric hydroxamate uptake component A" (FhuA), with a height of approximately 69 Å and an average elliptical cross section of 39×46 Å [109, 117] and the PapC translocation core [118]. Both proteins can be found in *E. coli*. The TonB-dependent transporters are active transporters that bind to their substrate with very high affinity (nM range), allowing transport of low concentration nutrients against a concentration gradient. The β-barrel active transport mechanism will be introduced on the example of the FhuA protein in Sect. 2.4.1. Many of the monomeric barrels large enough to form a pore contain an N-terminal "plug"-domain closing the channel reversibly. A FhuA based "plug-less" barrel with 24 strands has been created artificially [119] and will be discussed extensively in Chap. 5 (see Fig. 2.6, structures 1–5). General porins such as the OmpF protein or the Malto-porin of *E. coli* [120] are formed by 16–18 β-strands. These proteins trimerize and form passive diffusion channels. These water filled channels allow unspecific diffusion of small molecules up to 600 Da without binding to the transported substrate [121] or facilitate the passive but specific uptake of abundant nutrients (see Fig. 2.6, structure 7 and 8).

Proteins like the *S. aureus* heptameric α-hemolysin [122] are members of a special class of cell-toxic, pore-forming β-barrel proteins. These barrels consist of several portions belonging to different subunits. Only when all subunits assemble, a functional pore is formed (see Fig. 2.6, structure 6).

Due to their structural robustness and the fact that many of them are known to refold after denaturation β-barrel MPs are very suitable for nano-biotechnological applications. The fact that geometry and architecture of all proteins of that structural family follow the same general rules,

Table 2.2 Features of eight exemplary β-barrel MPs that differ among other in size, number of β-strands and function

Protein name	PDB code	Nr of β strands	Oligomeric state	Nr of pores	Protein family	Function	Organism
OmpA	1BXW	8	1	–	Membrane anchors	Linking OM to peptidoglycane	*E. coli*
OpcA	1K24	10	1	1	Adhesins	Adhesion to host cell	*Neisseria meningitidis*
Tsx	1TLY	12	1	1	Transporters	Nucleoside transporter	Gram negative
α-haemo-lysin	7AHL	14	7	1	Pore-forming toxins	Exotoxin	*S. aureus*
OmpF	3K1B	16	3	3	Porins	Diffusion pore	*E. coli*
Malto-porin	1MAL	18	3	3	Porins	Maltose uptake	*E. coli*
FhuA	1BYA	22	1	1	TonB-dependant auto-transporters	Iron uptake, signal transduction, phage binding	*E. coli*
PapC translocation core	3FIP/2VQI	24	2	1	Transporters Usher	Export/assembly of pili subunits across OM	*E. coli*

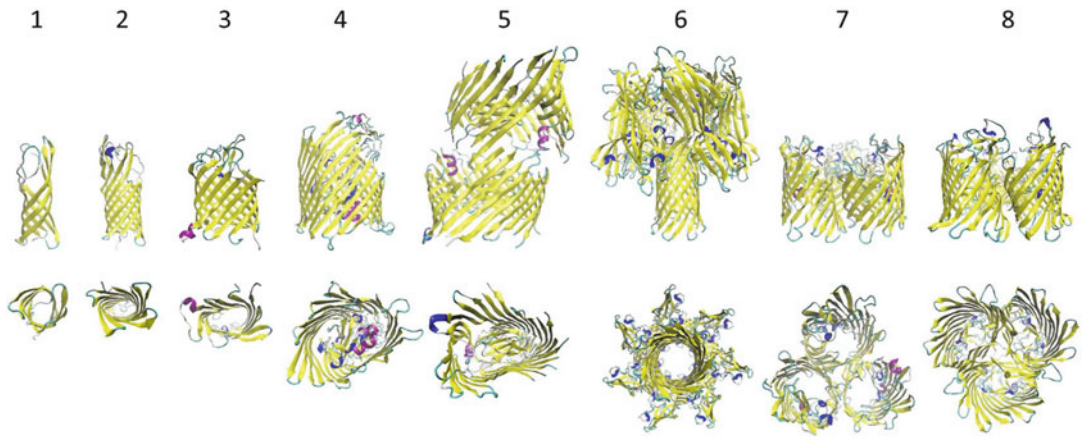

Fig. 2.6 New Cartoon representation of tertiary structures of eight exemplary β-barrel shaped membrane proteins (made with VMD), showing their structural diversity. 1 – OmpA (1BXW), 2 – OpcA (1K24), 3 – Tsx (1TLY), 4 – FhuA (1BYA), 5 – PapC translocation core (3FIP/2VQI), 6 – α-haemolysine (7AHL), 7 – OmpF (3K1B), 8 – Malto-porin (1MAL)

which will be discussed in the following section, adds to their suitability to be used as nanochannels and makes it easy to modify them for the respective application. Furthermore these general architectural principles allow the construction of synthetic artificial β-barrel structures [123] or the design of peptide building-blocks that self-assemble into barrel structures [124].

2.3.1 β-Barrel Membrane Proteins – Structure, Geometrical Features and Function

The β-barrel is the second regular structural element found in membrane spanning portions of MPs, in which hydrogen bonds are formed between protein backbone atoms. Membrane barrels are structurally much less variable than their α-helical counterparts (see Fig. 2.6) and follow clearly definable geometrical principles, while their sequences may vary greatly.

The membrane spanning minimal length of a β-strand is ∼7 residues and in trimeric porins for instance each strand consists of 11 residues on average, while strands of monomeric barrel proteins are on average 13–14 residues long [125, 126]. In a β-strands *primary sequence* one finds repeated patterns of amino acid pairs of which in the secondary structure one amino acid will face to one plane of the sheet, while the other will face the other plane. In a β-barrel membrane spanning portion therefore one half of the residues faces the barrel inside, while the other half faces the outside [127]. Generally one would assume that while the inside facing residues are always hydrophilic, as in a pore-forming barrel they reach into the water-filled channel; the outside facing residues are hydrophobic, as they are in contact with the membrane. This would mean that one should find alternating polar and non-polar residues in the primary sequence. This has been found partly true, but it has been found to be a too simple generalization also, as a barrel interior must not necessarily be hydrophilic. Furthermore the alternating inside- outside-facing amino acid pairs have been found in soluble protein β-strands, too. Table 2.3 shows the primary sequences of representative β-strands of the OmpA and the soluble *Aequorea Victoria* "Green fluorescent protein" (GFP) with corresponding loop and turn sequences. Both strand sequences show ∼42 % of polar amino acids. Therefore the prediction of membrane β-strands or in general β-barrel proteins from their amino acid sequence is not trivial and simple methods as the common hydrophobicity scales applicable to predict membrane helices do not apply to barrel proteins (For details see Sect. 4.2).

Table 2.3 Primary sequences of OmpA strand 4, with loop 2 (beginning) and turn 2 (end) and the soluble "Green fluorescent protein" (GFP) strand 6, with short turns (beginning and end)

Protein name, UniProt ID	Primary sequence
OmpA (P0A910)	**Arg** Met Pro **Tyr Lys** Gly **Ser** Val **Glu Asn** Gly Ala **Tyr Lys** Ala – **Gln** Gly Val **Gln** Leu **Thr** Ala **Lys** Leu Gly **Tyr** Pro – Ile **Thr Asp Asp**
GFP (P42212)	Glu **Asp** – **Thr** Leu Val **Asn Arg** Ile **Glu** Leu **Lys** Gly Ile – **Asp** Phe

Polar amino acids are bold and underlined

Recently a unified hydrophobicity scale applicable to both α-helical and β-barrel MPs, based on theoretical considerations and statistical methods has been reported [49]. Protein statistical analysis revealed that the aromatic amino acids Tyr, Phe, Ile, Thr, Trp, and Val occur frequently within β-sheets, while Gly and Pro are rare. Near to a strand end however Pro may be found with a higher rate resulting in a bulge within the barrel (see Table 2.3). The β-strand connecting periplasmic turns (T) that extend into the hydrophilic environment mainly contain polar amino acids, with a preference for Asn, Gly, Pro, Asp, Ser and also the outside loop (L) structures mostly contain polar residues [128]. The amino acid Cys is in general rarely found in transmembrane β-barrels and if so then never in a β-strand. When discussing the membrane barrel's tertiary structure further position dependent amino acid preferences will be introduced.

The β-sheet *secondary structure* element is not immediately obvious from a proteins primary sequence and not easy to predict from sequence, as it involves the hydrogen bond formation between backbone atoms of adjacent β-strands that within the primary sequence are positioned far from each other (see Fig. 2.7a). Strands can be parallel (running in the same direction) or antiparallel (running in opposite directions). Two interacting, antiparallel β-strands, connected by a short loop or turn sequence (β-turn) of on average two to six residues on one end, form a so called β-hairpin structure [129], shown in Fig. 2.7a. Several hairpins may then form super secondary structure sheets with varying numbers of strands. In β-barrels adjacent β-hairpins are connected by longer loop regions. A sheet may also be formed of adjacent strands that are not connected by turns, but that are in each other's vicinity. Sheets may curl up or twist right- or left-handedly, with a predominance of right-handed twists [130]. In a sheet as good as all backbone amide-groups/carboxylic-groups are hydrogen bonded intra-molecular, exceptions are the amide- and carboxylic-groups on the terminal strands ends of a β-sheet structure. In soluble non-barrel proteins these groups may form hydrogen bonds with water molecules, whereas in barrels these groups hydrogen bond to each other to close and form the cylindrical tertiary structure. While in soluble proteins one finds β-sheets consisting of two or more strands, the membrane barrel sheets are made of two strands only (one hairpin each). Transmembrane β-barrel protein strands are always anti-parallel to each other, whereas soluble proteins contain parallel, anti-parallel or mixed sheets [131].

As the hydrogen bonds between backbone amino- and carboxylic groups of antiparallel strands and resulting sheets are more linear than in parallel sheets, anti-parallel sheets are more stable [132]. Though the strands in antiparallel sheets are actually shifted or "sheared" against each other in a way that backbone carboxylic- and amino-groups are not exactly opposite from each other, due to a displacement of the C_α along the backbone direction [133]. Shearing can be negative or positive depending on the shearing direction (see Fig. 2.7b).

The β-barrel protein *tertiary structure* construction follows several universal rules first compiled by G. E. Schulz [134]:

1. The N- and C-termini are located at the periplasmic barrel end, restricting the β-strands number to an even number.
2. Due to the β-sheet twist the β-strands tilt at an angle of ∼20–45°
3. The shear number of an n-stranded barrel is positive
4. All β-strands are antiparallel.

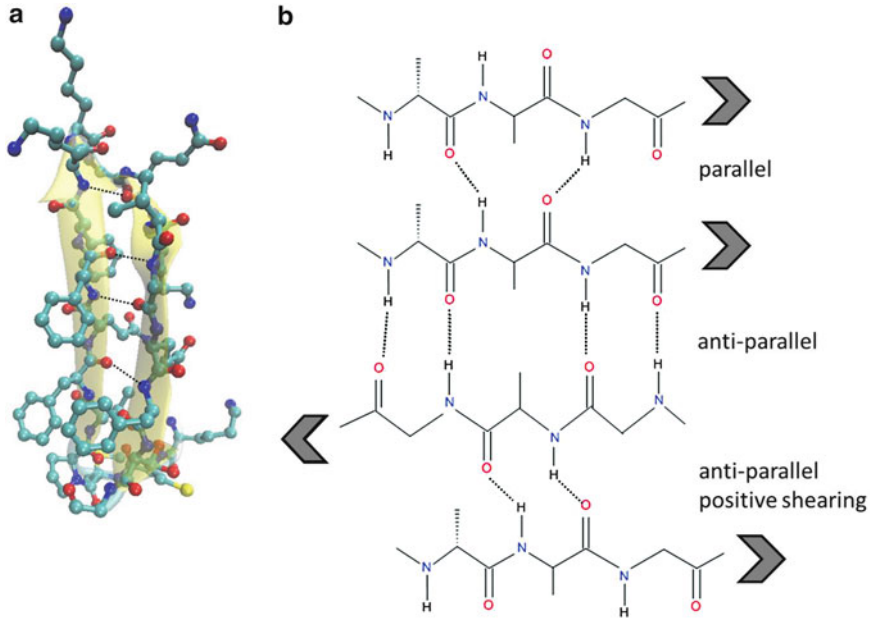

Fig. 2.7 (**a**) "CPK" representation of two antiparallel β-strands forming one β-hairpin, by backbone atom hydrogen-bonding (*dotted lines*) (made with VMD). (**b**) Hydrogen bonds (*dotted lines*) between backbone atoms of parallel, antiparallel and antiparallel sheared β-strands (structures drawn with Accelrys Draw (http://accelrys.com/products/informatics/cheminformatics/draw/))

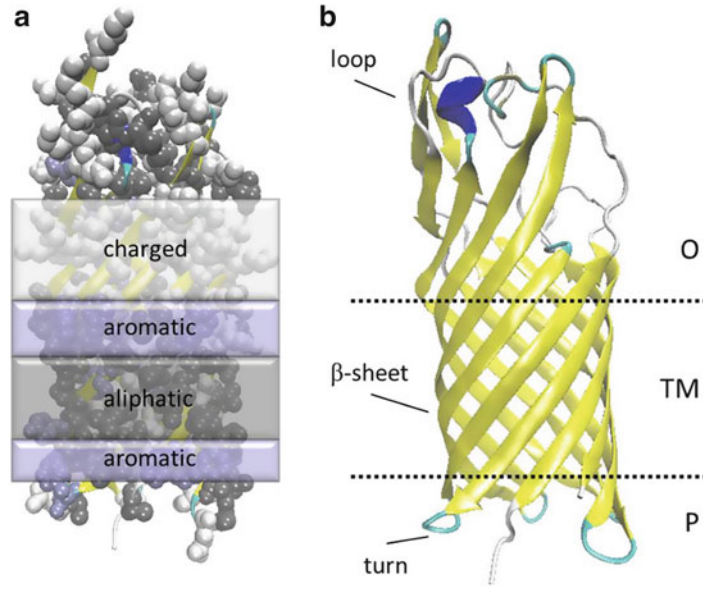

Fig. 2.8 (**a**) Schematic view of the general β-barrel protein architecture with aliphatic ring, between two aromatic bands and external, charged residues, shown on the example of the OpcA protein (1K24). (**b**) Cartoon representation of OpcA 3-D structure, showing outside (*O*) loops, transmembrane (*TM*) β-sheets and periplasmic (*P*) turns (made with VMD)

5. The strands on the periplasmic barrel end are connected by β-turns (T1, T2 etc.), with a length of one to 12 residues (Fig. 2.8b).
6. The strands on the external barrel end are connected by longer loop regions (L1, L2 etc.), with a length of two to 46 residues. These loops show exceptionally high sequence variability, contain mainly polar, charged amino acid residues and are usually flexible and mobile (Fig. 2.8b).

7. The β-barrel surface in contact with the hydrophobic membrane consists of aliphatic side-chains, forming an on average 22 Å wide, hydrophobic ring (Fig. 2.8a).
8. Above and below of the non-polar ring can be found two bands with an overabundance of aromatic side-chains (here up to 40 % of all lipid-exposed residues may be aromatic) [135]. These bands with intermediate polarity are positioned at the interfaces between lipid-bilayer and aqueous environment (Fig. 2.8a).
9. Protein parts in contact with the aqueous environment contain mainly polar, charged residues.
10. Above a number of 10 β-sheets the barrel interior will form a water-filled pore.

Each β-barrel protein can be characterized by a set of geometric parameters such as the number of strands (n), the shear number (S) and the resulting β-sheet tilt and barrel diameter. By giving values to these parameters it is possible to get an idea on the size and shape of a barrel protein. As mentioned before membrane β-barrels have an even number of antiparallel strands and the strands are inclined at an angle to the transmembrane axis. This results in a shift in the H-bonded residues, termed the shear number. The shear number is a measure for the β-strand inclination angle against the barrel axis. A shear number of $+1$ means that the H-bonded partner of the residue at position a, is at position $b+1$ rather than j (see Fig. 2.7b). Shear numbers of transmembrane barrels are positive and generally even, they range between 8 and 26. As a rule of thumb the shear number can be defined as $n+2$ [136].

Apart from these principles some peculiarities of the β-barrel tertiary structure element can be explained by the introduced β-sheet secondary structure characteristics. The extensive interstrand hydrogen bonding between backbone atoms leads to an exceptional stability of the formed barrel and barrel proteins do not easily unfold in their membrane environment [137]. However with extended stability comes an increase in structural rigidity, also [138, 139].

Regarding local amino acid preferences within a transmembrane barrel protein it has to be noted that apart from the already mentioned aliphatic and aromatic girdles the barrel interior mainly contains polar amino acids such as Ser, Tyr, Asn, Gln or Thr, as well as the small Gly, while non-polar residues are rarely found [135]. Some OMPs contain N-terminal, globular cork- or plug domains that are mobile and may close the channel, *e.g.* in the TonB-dependent auto-transporters like the FhuA protein (Fig. 2.6, structure 5). The barrel plugs are often comprised of β-sheet structures as well as α-helical portions connected by short loops and as melting temperature calculations show in some cases the plugs play a role in stabilizing otherwise unstable transmembrane parts [140]. Apart from cork-structures there can be other non β-barrel regions, such as α-helices that are packed against the β-strands, as reported for the antimicrobial peptide resistance and lipid A acylation protein PagP, where an N-terminal helix associates with two β-strands on the periplasmic protein end [141]. In case of the PagP the N-terminal helix can be deleted without losing protein function. However as a deletion decreases the protein's stability, seemingly the helix conserves the native protein structure by acting as a locking clamp after the protein is folded and inserted into the bilayer [142].

In the Gram-negative bacterial outer membrane the trimeric porins are most abundant. All porins known so far contain three identical β-barrel monomers that form a trimeric *quaternary structure*. Non-specific porins are comprised of 16-stranded barrels, while barrels of substrate specific porins show 18 strands [134]. The interface region between the trimerized barrels contains mainly apolar amino acid residues. Eloffson et al. reported a higher than average sequence conservation for residues in this region [127]. Porins are extremely robust and do not fully denature in presence of 5 M guanidium hydrochloride or at 70 °C, partly due to the stabilizing effects of the interface region between the three monomers. There are several rules apart from the general β-barrel construction rules summarizing their general construction features [110].

1. Upon trimerization, a nonpolar core is formed at the threefold interface axis so that the porin's central part resembles a water-soluble protein.
2. The porin barrel center is constricted by an inserted long loop L3.

In the α-hemolysin pore forming toxins the functional protein consists of seven identical monomers that assemble to form one 14 stranded barrel, with each monomer contributing one β-hairpin [122].

As mentioned Gram-negative bacteria outer membrane β-barrel proteins serve a great number of different cellular functions, ranging from membrane anchors, as the OmpA, physically linking the *E. coli* outer membrane to the peptidoglycan layer below [143]; non-specific passive diffusion pores, such as the OmpF porin allow [144] the diffusion of small, polar molecules (600–700 Da in size), *e.g.* water, ions, monomeric sugars or cellular waste products; substrate-specific passive diffusion pores like Malto-Porin for the up-take of maltose; enzymes such as proteases (OmpT) or lipases (OmPlA); bacterial pore-forming toxins (α-hemolysins) and active transporters (FhuA, Tsx). The FhuA protein structure and function will be discussed in detail in Sect. 2.4.

Examples for mitochondrial and chloroplast β-barrel MPs are the pore domains of the respective protein translocase complexes (translocase of the mitochondrial outer membrane – Tom, translocase of the chloroplast outer membrane – Toc) that translocate pre-proteins from the cytoplasm into the organelle's intermembrane spaces. These pore domains are formed in mitochondria by the 19 strand Tom40 protein [145] and in chloroplasts by Toc75 with 16 β-strands [146].

2.3.2 β-Barrel Membrane Proteins – Genetics, Biogenesis, Folding and Insertion

For the protein OmpA it was reported that it cooperatively folds and inserts into liposomes in spontaneous manner [147]. The same has been demonstrated for other small β-barrel proteins and several pore-forming toxins [148–150]. However as outer membrane proteins have to reach their target membrane prior to insertion a wholly spontaneous mechanism is not likely for *in vivo* systems, as will be discussed in the following, with regard to the Gram-negative bacterial OMPs (a detailed review on mitochondrial and chloroplast OMP assembly can be found in [151]).

The Gram-negative bacterial cell envelope consists of two membranes, the inner and the outer one. The two membranes enclose the periplasmic space with a thin (1–2 nm) stabilizing peptidoglycan layer (murein) between them [152]. The inner membrane is a common phospholipid bilayer, as introduced in Sect. 2.1, showing the ubiquitous α-helical TMPs. The outer membrane (OM) instead is a robust asymmetrical bilayer containing phospholipids (70–80 % phosphatidylethanolamine and 20–30 % phosphatidylglycerol and cardiolipin) in the inner leaflet and lipopolysaccharides (LPS) in the outer leaflet (Fig. 2.9) [153, 154]. The LPS have saturated fatty acyl chains with an average of 14 carbon atoms and complex acidic polysaccharide chains linked to a glucosamine disaccharide backbone. Due to the saturated fatty acyl chains the OM is rather rigid and due to acidic sugars its net-charge is negative. The outer membrane is restricted in permeability and thus a selective barrier that protects the cells. One possible reason for the absence of α-helical MPs in the outer membrane is that OMPs have to pass the inner membrane to reach their destination. Proteins containing hydrophobic helices would most probably be retained in the inner membrane [155]. This stresses the necessity of bacterial barrel OMPs to pass from the cytoplasm through the inner membrane to reach their destination.

While in eukaryotes mitochondrial and chloroplast OMPs are encoded in the nucleus, expressed as pre-proteins in the cytoplasm and need to be trafficked to the respective organelle membrane. In Gram-negative bacteria the different OMPs are encoded by 2–3 % of all genes. They are expressed with an N-terminal signal sequence.

The signal sequences or signal peptides of different Gram-negative bacteria show little

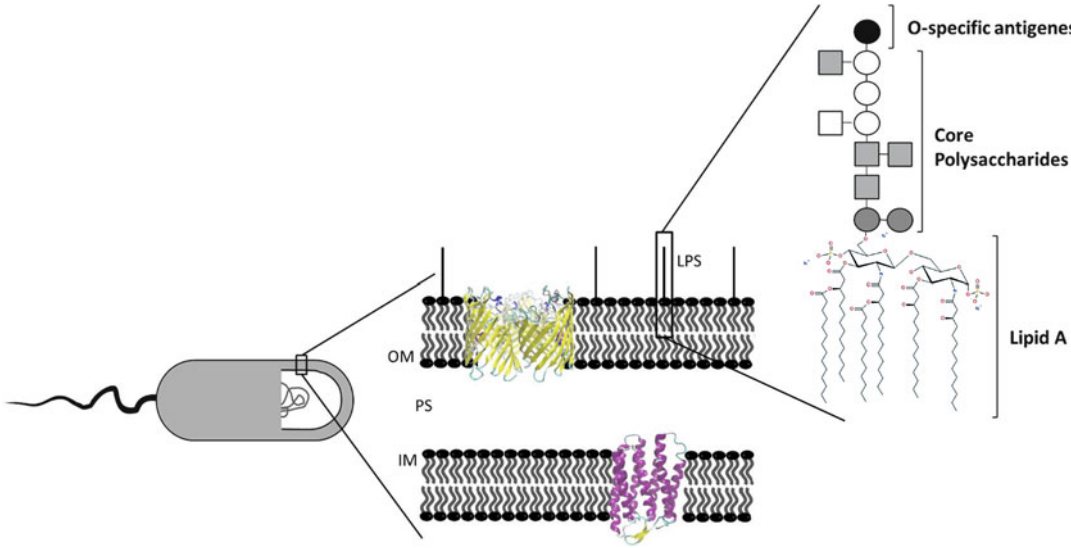

Fig. 2.9 Schematic representation of *E. coli* cell, outer and inner membrane with α-helical and β-barrel membrane protein respectively and LPS structure

Table 2.4 Amino acid sequence of *E. coli* FhuA

Amino acid sequence (N → C)
Met Ala Arg* *Ser* Lys* *Thr* Ala *Gln* Pro Lys* His *Ser* Leu Arg* Lys* Ile Ala Val Val Val Ala *Thr* Ala Val *Ser* **Gly** Met *Ser* Val *Tyr* Ala *Gln* Ala

Underlined – hydrophobic amino acids; italics – polar, neutral amino acids; bold – boundary between hydrophobic and C-terminal region
*Positively charged amino acids

sequence identity, however they reveal several common features, such as a short N-terminal region of 1–5 positively charged residues, followed by a hydrophobic region consisting of 7–15 amino acids and a C-terminal region of 3–7 neutral, polar amino acids that is recognized by a signal peptidase. The boundary between the hydrophobic and the signal peptidase recognition region is often marked by a Pro or Gly amino acid [156]. The 33 amino acid long signal peptide sequence of the *E. coli* FhuA (Table 2.4) shows positively charged residues (*) on the N-terminal side, hydrophobic residues (underlined) mainly in the sequence middle, while polar, neutral amino acids (in italics) are distributed over all the sequence parts. However the three different regions can be identified approximately (highlighted in grey). The hydrophobic and signal peptidase regions are divided by Gly (bold).

Signal sequence deletion on gene level leads to OMP accumulation in inclusion bodies upon over-expression. Inclusion bodies can then be isolated and OMPs can be refolded by choosing suitable detergents [157] or diblock-copolymers [158] (see Sect. 5.2.2).

After expression OMPs have to be transported to the outer membrane and therefore first pass the inner membrane. The inner membrane passage requires energy provided by an exergonic reaction provided by the SecYEG-translocon (see also Sect. 2.2.2). The pre-protein has to be kept in an unfolded state, as correct folding can only occur at the outer membrane and as will be discussed folding occurs together with membrane insertion. Unfolded OMPs tend to aggregate when in aqueous environment and furthermore translocation through the inner membrane is possible only if the protein is still

in an unfolded state. Therefore the cytoplasmic chaperon SecB binds to the newly synthetized OMP, keeps it in open chain conformation and thus fit for translocation [159]. The necessary energy for the passage of the inner membrane via the SecYEG-translocon is provided by SecA, an ATPase, binding to the N-terminal signal peptide [160, 161]. The N-terminus passes the Sec-translocon first after which the whole protein enters the periplasm, a process during which the signal peptide is cleaved by peptidases localized at the inner half of the inner membrane. As the periplasm is a water-filled cell compartment, the still unfolded outer membrane proteins are again prone to aggregation after leaving the translocon pore. Therefore periplasmic chaperones as the *E. coli* SurA bind to the OMPs, keep them in an unfolded state, preventing aggregation and assist in the OMP folding. If OMPs are left chaperon-unbound (for instance in certain deletion mutant strains), misfolded aggregates will form that are toxic for the cells or the OMPs might partly assemble into the inner membrane, triggering stress-response [162]. SurA is able to assemble all OMPs, however deletion of the SurA gene does not affect the correct folding and membrane insertion of most OMPs [163]. In SurA absence the chaperon Skp (seventeen kDa protein) together with the chaperon/protease DegP facilitate the assembly of most OMPs [164, 165].

Now that the OMPs have passed the inner membrane, reached the periplasm and have been recognized and bound by periplasmic chaperones, they are ready to fold and insert into the outer membrane.

As for the protein insertion into the low-dielectric-constant bilayer lipid phase the complete saturation of main-chain hydrogen bonds in the transmembrane regions is necessary. Folding and insertion have to be simultaneous, concerted events. This is due to the hydrogen bonding between backbone residues in neighboring strands instead of, as in helical membrane proteins, between residues next to each other in one helix. Therefore a two-step model as proposed for the folding of α-helix membrane proteins, in which each helix folds and then inserts into the membrane, is not applicable to OMP folding/insertion. The assembly of an OMP is thus following an "all-or-none" mechanism [139]. Since the periplasmic space lacks ATP β-barrel folding has to be either a spontaneous act or a guided process in which cellular components at the outer membrane are involved. The fact that in *vitro* some OMPs are able to refold from a completely denatured state [166] speaks for the spontaneously occurring folding. However the *in vitro* folding happens too slowly as to be relevant for *in vivo* folding [162], which shows the need for an *in vivo* folding assistant. This assistant has been found with the discovery of the Bam (β-barrel assembly machine) multi-protein complex in the *E. coli* outer membrane [167]. The Bam complex has been identified in several other Gram-negative bacteria [168]. The Bam complex main component is the 16-stranded OMP BamA, through whose pore the nascent OMPs might reach the outer membrane, as it was found that denatured *E. coli* OMPs increased the BamA channel activity by direct interaction, while periplasmic proteins have no such effect [169]. In earlier studies a conserved sequence at the C-terminus of most OMPs has been identified. This sequence that is supposed to be recognized by BamA consists of a Phe or Trp at the terminal, a hydrophobic residue or Tyr on the third position and hydrophobic residues at positions 5, 7 and 9, as counted from the C-terminal end [170]. BamA contains an N-terminal region with five so called POTRA domains (polypeptide translocation associated) P1–P5 that reach into the periplasm, allowing interaction with four lipoproteins BamB, C, D and E [171]. BamA homologues have been identified in the outer membranes of mitochondria and chloroplasts [172]. While BamB, C and E can be deleted without lethal effect in *E. coli*, deletion of both BamA and BamD lead to lethal OMP assembly defects [173]. Furthermore suppressor mutations in BamB led to a defective LPS assembly, so BamB might directly assist in LPS assembly or

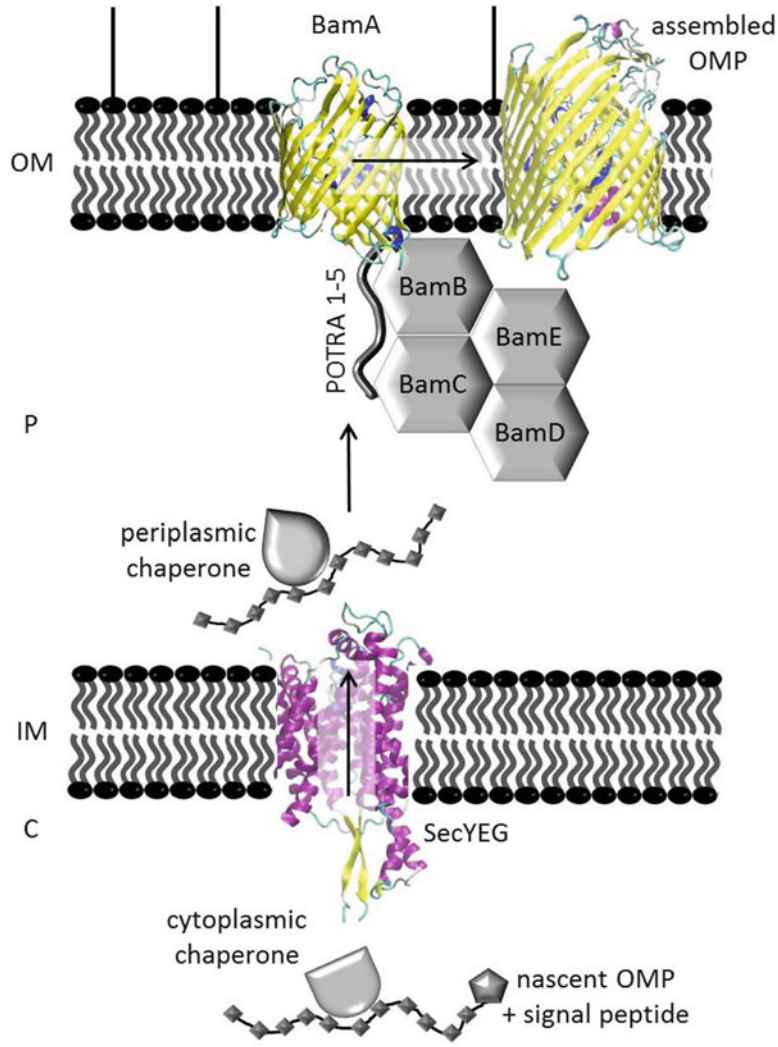

Fig. 2.10 Schematic representation of OMP translocation from the cytoplasm (*C*) through the inner membrane (*IM*) into the periplasm (*P*), via the SecYEG translocon (1RH5). In the periplasm the nascent OMP reaches the Bam complex, which facilitates the assembly into the outer membrane (*OM*). Cytoplasmic and periplasmic chaperones guide translocation and keep OMP in unfolded, non-aggregated state (Protein structures in New Cartoon representation made with VMD)

in assembly of components necessary for the LPS-assembly [167].

How exactly the Bam complex facilitates β-barrel-assembly remains a secret up to day. However the recent success of reconstituting the purified, functional components of the *E. coli* OMP assembly machine into proteoliposomes [174, 175] and the successful crystallization of BamBCDE [176–178] might lead to a much better understanding of *in vivo* mechanisms involved in β-barrel membrane protein folding and membrane insertion. Figure 2.10 gives an overview on the translocation and assembly of a bacterial OMP.

2.3.3 β-Barrel Membrane Proteins – Relevance for Nano-materials Development

As the potentials, limitations and engineering-strategies of bacterial OMPs application for the design of nano-materials and especially nano-pores will be the main subject of Chaps. 5 and 6, here only some general point will be addressed.

β-barrel OM proteins and bacterial pore-forming toxins are suited for various nano-pore applications such as drug-release [179–181], nano-reactors with controlled permeability

[179–182], nano-sensors for the detection of biomolecules [183–185], the sequencing of short RNA- or DNA-fragments [186], and hybrid-biocatalysts [187]. Up to now the pore-forming toxin α-hemolysin is the most widely used biological nano-pore sensor for the detection of single molecules, as it lacks any mobile domains that might interfere with channel conductance. Furthermore the α-hemolysin has been successfully engineered to obtain specificity for a certain analyte [184]. The OMPs relevance towards the design of nano-materials is mainly due to the exceptional robustness of the β-barrel structure [110], the ability of OMPs to spontaneously refold from a denatured state, when in presence of detergent micelles, lipid membranes or diblock-copolymers, opening the possibility to purify them from inclusion bodies, including solubilization in urea and subsequent refolding [166], as shown for FhuA [158], OmpA [188], OmpF [189] or FepA [190]. In some cases refolding occurs even after harsh treatment, *i.e.* heating to 75 °C for 30 min in presence of 6M guanidine hydrochloride and isopropanol, as it is necessary to extract the *E. coli* OmpF [110]. Furthermore correctly folded OMPs, as obtained from purified outer membrane fractions, spontaneously insert into lipid or block-copolymer membranes and vesicles or nano-discs [179, 182, 191], allowing for the design of new hybrid materials and better ways of characterization. Moreover it has been shown that the β-barrel scaffold tolerates extensive deletions and protein engineering (*e.g.* deletion of plug-domains and loops regions or mutations changing the channel geometry) without losing its overall structure and channel functionality, or leading to a change in function, as shown for the *E. coli* FhuA [119, 181, 183, 192–194] and OmpA [195]. The wide range of different applications and approaches to use OMPs as nano-materials is also due to their good temperature and overall stability, which has been determined by solvent-stability tests and differential calorimetry [126]. It has been reported for instance an exceptional heat resistance for proteins like OmpA [126], or FhuA with and without plug-domain. The FhuA containing the plug-domain proved to be more stable, up to temperatures of 75 °C, while the plug deletion decreased its stability to 62 °C [196]. However certain limits have to be overcome when working with OMPs, as their production especially in larger scales is not trivial. As with all membrane proteins the general production strategy involves expression into the respective membrane, the membrane then has to be isolated from all other cellular compounds and the protein has to be solubilized, commonly by the addition of detergents [25, 26]. The yield and purity are often rather poor, as the membrane offers a limited space for production and other proteins residing in the same membrane might be solubilized likewise [158]. These production issues can be overcome by the mentioned expression into inclusion bodies and subsequent solubilization/refolding or by the use of cell-free expression systems [197] (see Sect. 5.2). The hydrophobic residue exposure characteristic for all OMPs makes it impossible to work in aqueous solution, instead a hydrophobic or rather a mixed (water/hydrophobic) environment has to be offered to conserve OMP structure and function, limiting employable OMP carrying-, or immobilization-systems. The overexpression of certain OMP variants can have toxic effects on the expression host, for instance barrel mutants that act as passive diffusion pores for larger molecules might lead to osmotic imbalance and cell death, when using conventional media.

Since Chap. 5 will introduce the *E. coli* FhuA protein and variants thereof as main models to explain the β-barrel channel geometry modification by protein engineering, the following sections will give an overview on the general FhuA features.

2.4 *E. coli* Iron Transporter: FhuA

The transition metal iron is a vital trace-element for all living organisms as it, by changing its oxidative state, acts as oxidation/reduction reagent. Therefore its presence as a cofactor is required for many essential cellular processes, such as electron transport, synthesis of desoxyribonucleotides (dNTPs) [198] or the

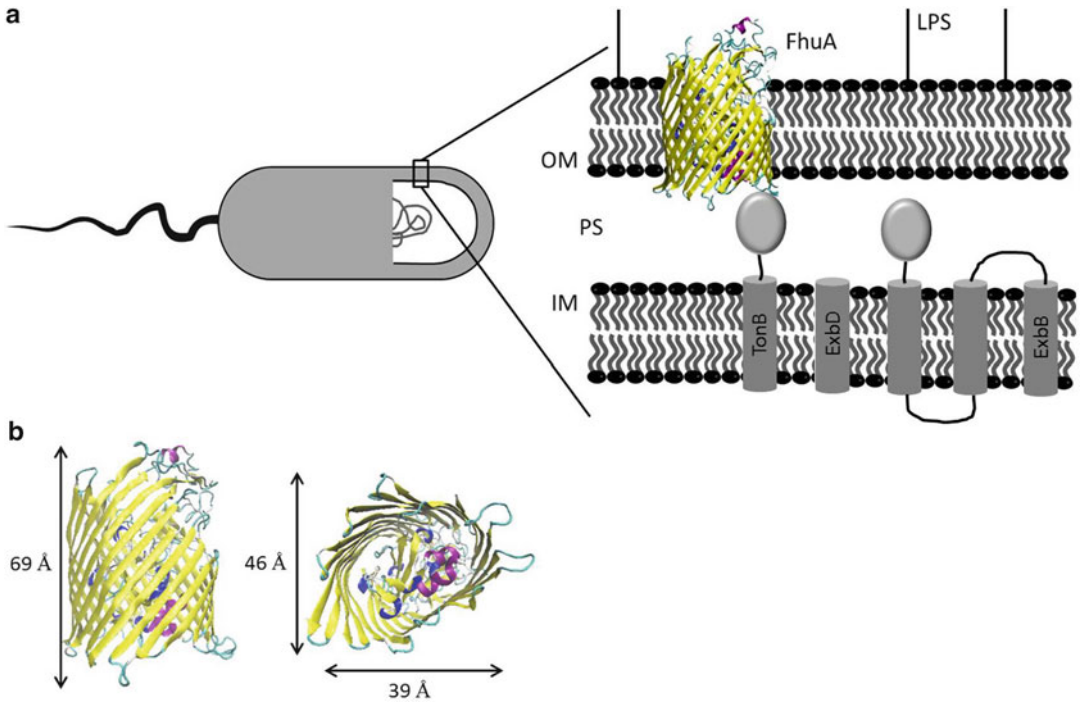

Fig. 2.11 (**a**) Schematic representation of *E. coli* cell, inner membrane with TonB-ExbDB complex and outer membrane with FhuA protein in cartoon representation. (**b**) Cartoon representation of FhuA protein in side-view (*left*) and top-view (*right*); protein height and elliptical cross-sections are indicated by *arrows* (Protein structures made with VMD)

catalase reaction [199] and bacteria require a minimal concentration of 10^{-7} M in order to grow. Despite the relative abundance of iron in nature, the concentration of biologically available, soluble Fe^{3+} is $\sim 10^{-9}$ M [200]. This is due to the formation of insoluble Fe^{3+}-hydroxide polymers in presence of oxygen. These hydroxides have a rather poor solubility of 10^{-18} M at pH 7 [201]. In order to obtain the iron necessary to survive bacteria secrete low molecular weight Fe(III)-chelating organic compounds with a very high iron-affinity, when under the effect of iron depletion. These so called siderophores can be classified into three different structural groups, catechols (*e.g.* enterobactin), hydroxycarboxylates (*e.g.* citrate) and hydroxamates (*e.g.* the fungal ferrichrome) [202]. The complex of siderophores and iron is then transported actively into the cell, as it happens generally with all essential components that are present in low concentrations. In Gram-negative bacteria specific transporter OMPs could be identified for all three siderophore classes, often expressed in parallel. These transporter OMPs are dependent on the inner membrane protein TonB that in complex with the proteins ExbB and ExbD transduces the electrochemical potential of the cytoplasmic membrane into energy needed for the active transport, as the periplasmic space is void of ATP (see Fig. 2.11a).

From *E. coli* TonB-dependant OMP receptors for all three kinds of iron-binding siderophore types were isolated, *i.e.* FepA, specific for ferric enterobactin [203], FecA, specific for diferric dicitrate [204] and FhuA, specific for among others ferrichrome [205]. Though *E. coli* itself secretes only the enterobactin, it is able to utilize siderophores secreted by other organisms, such as ferrichrome, secreted by fungi [202]. Crystal structures have been obtained for all three receptors [206–208] and structures show a high similarity. All three siderophore transporters consist

of a C-terminal 22-stranded β-barrel and an N-terminal, mobile plug (or cork) domain, located inside the barrel and that is able to reversibly obstruct the channel interior [110]. The cork domain contains the so-called TonB-box close to the N-terminus, interacting with the TonB C-terminus [209].

2.4.1 FhuA – General Overview and Unique Features

FhuA, Ferric hydroxamate uptake component A (Mw 78.9 kDa) is a monomeric outer membrane protein composed of 723 amino acids. It always attracted much attention, since it has been first isolated in 1973 by Brown et al. as the second pure protein obtained from the *E. coli* outer membrane ever [210]. Originally it had been named TonA, as it had been found that T1 phages are unable to binds to certain *E. coli* mutants, namely tonA and tonB (ton for T one), while T5 phages do not bind to mutant tonA [211]. Two years later its ability to bind and transport ferrichrome was revealed [212], leading to the new name "Ferric hydroxamate uptake component A" [213]. In the following years several other components transported by FhuA had been identified, such as siderophores structurally close to ferrichrome (*i.e.* ferricrocin) antibiotics albomycin, rifamycin or CGP 4832 and its receptor function was found to range from phages T5, T1 and φ 80 and UC-1 to the toxic protein colicin M protein and the toxic peptide microcin J25 [206]. The FhuA's main physiological function however is the uptake of ferrichrome bound iron.

More recent structural studies revealed that amino acids 1–160 compose the protein's plug region, while the β-barrel domain is comprised by amino acids 161–723. The barrel consists of 22 antiparallel β-strands that are connected by short periplasmic turns (T1-T10) and longer loops (L1-L11) reaching into the outside environment [206]. The plug domain consists of a mixed 4-stranded β-sheet and five α-helices, it is connected with the barrel by at least 60 hydrogen bonds and 9 salt bridges [117]. The plug β-sheet plane is inclined by ∼45° to the membrane normal, thus obstructing the barrel interior, prohibiting passive siderophore diffusion [202]. Plug-related channel obstruction was further supported by the finding that FhuA reconstituted into planar lipid bilayers shows no channel conductance [214]. Deletion of the plug domain instead results in a passive diffusion channel, allowing the passage of ferrichrome, maltodextrins and antibiotics into the periplasm [193, 215, 216]. As depicted in Fig. 2.11b the FhuA barrel is 69 Å in height and has an elliptical cross section of 46–39 Å in the part delimited by the β-turns, while the loop side has a smaller cross section of 40–35 Å as deduced from its crystal structure. It is thus one of the largest known β-barrel proteins [206]. Co-crystallization of FhuA with its ligand ferrichrome revealed that its binding pocket mostly contains aromatic and hydrophilic amino acids and that the first four residues in the plug domain are involved in ferrichrome binding. Structures showed however that ferrichrome binding did not lead to considerable changes in conformation and especially had no effect on the position of the cork domain [117]. The plug residues 7–11 comprise the FhuA TonB-box [217]. Its structure and the interaction with the TonB C-terminus had been revealed due to the co-crystallization with TonB. The obtained structures showed furthermore that the TonB C-terminus forms a four stranded β-sheet to which the FhuA TonB-box adds a fifth strand [218]. Currently it is thought that upon ferrichrome binding the TonB-box reaches out into the periplasm to interact with TonB. The proteins ExbB and ExbD then help to energize TonB via the cytoplasmic membrane's proton motive force (PMF). ExbB and ExbD are homologous to the bacterial flagellar motor proteins MotA and MotB [219].

As ferrichrome binding did not seem to result in the plug-domain exiting the channel interior, the question remained how the necessary channel for ferrichrome transport was formed. In FhuA this channel has to be at least 10 Å in diameter to allow ferrichrome translocation, a pore this large can be formed only by the plug leaving the barrel interior or by major conformational changes in the plug-domain [209]. Several in vivo studies

on either FhuA or FepA addressing this question came to different conclusions, either showing that the FhuA plug-domain remains within the barrel [220] or that in case of FepA it leaves the barrel entirely [221]. In 2012 Udho et al. reported an in vitro study, where the FhuA was reconstituted into planar lipid bilayers, showing that 67 plug residues could be cleaved off of the reconstituted OMPs in presence of 4 M urea and trypsin. They concluded that urea pulls part of the plug out of the barrel, making it accessible to trypsin. Furthermore they showed that though the separately expressed and purified plug-domain is unfolded in solution, it is still able to seal-off reconstituted barrel domains without cork [222].

Much has been discovered about the FhuA transport activity since the protein was first isolated from the *E. coli* outer membrane, however the exact substrate translocation and energy transducing mechanisms remain still unclear and further studies are necessary.

Regarding the interaction with bacteriophages it has been found that in case of phage T5, the receptor binding protein pb5 at the phage tail end binds to the FhuA protein in an irreversible way. Upon interaction with FhuA the phage structure undergoes changes in conformation that ultimately lead to capsid opening and DNA-release. As the host cell envelope is permeabilized at the same time, phage DNA is transferred into the cell [223]. The same has been reported for *in vitro* systems, using liposome reconstituted FhuA [224].

2.4.2 FhuA – Relevance for Nano-channel Design

It's well studied robust β-barrel structure, solved crystal structure and large diameter make the FhuA an ideal starting point for the design of new nano-channel materials. The following section will give a short introduction to this topic, while Chaps. 5 and 6 will deal with it in detail.

The most striking FhuA feature apart from its channel diameter is, as for all β-barrel membrane proteins, its exceptional robustness and thermostability. As mentioned before, for the FhuA wild-type (WT) barrel domain it could be shown that it maintains its folded state up to temperatures of 75 °C, while the plug-deleted variant shows a decreased temperature stability, however maintaining the folded state up to 62 °C, as determined by differential scanning calorimetry [196]. This can be explained by the extensive hydrogen bonding between backbone atoms in adjacent β-strands. As FhuA WT and FhuA variants have been successfully reconstituted into triblock-copolymer membranes, with copolymers solubilized in organic solvents like ethanol [193] or tetrahydrofuran (THF) [181], it can be deduced that the FhuA structure tolerates lower to middle concentrations of these solvents.

A molecular dynamics simulation study on the FhuA ferrichrome-free and ferrichrome-bound states showed that while the loop and turn regions at both channel ends are extremely flexible, the hydrophobic barrel part maintains a rigid geometry [225]. Of further importance is that the FhuA structure proved to tolerate drastic mutations such as deletions (*e.g.* cork-domain deletion or deletion of loop regions) [183, 192, 193] and protein engineering to modify its geometry [119, 181] without losing its overall structure and channel functionality Other mutations led to a functional change *i.e.* the creation of a passive diffusion channel upon cork deletion.

Of further advantage is that procedures to obtain pure FhuA WT or variants from purified fractions of *E. coli* outer membrane are well established [192, 193, 226]. The addition of an internal Hexa-His-tag opens the possibility to further purify isolated FhuA protein by metal affinity chromatography [227]. In addition a cork-deleted FhuA variant was successfully expressed into inclusion bodies in *E. coli* upon deletion of the N-terminal signal sequence. Inclusion body material was solubilized in urea and the protein could be refolded into polyethylene-poly(ethylene glycol) diblock copolymer solutions by dialysis [158]. Another variant lacking the cork domain and four major, channel occluding extracellular loops (FhuAΔC/Δ4L), was equally expressed into inclusion bodies. Inclusion body material

was solubilized, affinity purified and refolded on-column by the addition of the detergents n-dodecyl-D-maltoside [192], n-octyl-β-d-glycopyranosid or 1-lauroyl-2-hydroxy-sn-glycero-3-phosphocholin [183].

The FhuA and its variants tolerate diverse hydrophobic environments, apart from conventional detergents. Membrane obtained and purified FhuA WT [228] and different FhuA variants [119, 180] for instance were successfully reconstituted into lipid vesicles. The variant FhuAΔC/Δ4L was reconstituted into planar lipid membranes [192]. FhuA missing the cork domain was furthermore reconstituted into ABA triblock copolymer vesicles of the PMOXA-PDMS-PMOXA type [PMOXA = poly(2-methyloxazoline); PDMS = poly(dimethylsiloxane)] [193] and a plug lacking variant with extended hydrophobic part was reconstituted into thick membranes of triblock-copolymer PIB_{1000}-PEG_{6000}-PIB_{1000} (PIB = polyisobutylene, PEG = polyethyleneglycol) [181].

Since the development of β-barrel membrane protein based nano-channels necessitates a good structural understanding of the involved proteins, Chap. 3 will give an overview on the main protein structural characterization techniques with a focus on the circular dichroism (CD) spectroscopy method.

References

1. Gorter E, Grendel F (1925) On bimolecular layers of lipoids on the chromocytes of the blood. J Exp Med 41:439–443
2. Danielli JF, Davson H (1935) A contribution to the theory of permeability of thin films. J Cell Comp Physiol 3:495–508
3. Singer SJ, Nicholson GL (1972) The fluid mosaic model of the structure of cell membranes. Science 175:720–731
4. Rietveld A, Simons K (1998) The differential miscibility of lipids as the basis for the formation of functional membrane rafts. Biochim Biophys Acta 1376:467–479
5. McElhaney RN, Tourtellotte ME (1971) The relationship between fatty acid structure and the positional distribution of esterified fatty acids in phosphatidylglycerol from Mycoplasma laidlawii B. Biochim Biophys Acta 202:120–128
6. van Meer G, Voelker D, Feigenson G (2008) Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9:112–124
7. Rothman JE, Lenard J (1977) Membrane asymmetry. Science 195:743–753
8. Muller P, Herrmann A (2002) Rapid transbilayer movement of spin-labeled steroids in human erythrocytes and in liposomes. Biophys J 82:1418–1428
9. Marr AG, Ingraham JL (1962) Effect of temperature on the composition of fatty acids in Escherichia coli. J Bacteriol 84:1260–1267
10. Silvius JR (1982) Thermotropic phase transitions of pure lipids in model membranes and their modifications by membrane proteins. In: Jost PC, Griffith OH (eds) Lipid-protein interactions. Wiley, New York
11. Dove SK, Cooke FT, Douglas MR, Sayers LG, Parker PJ, Michell RH (1997) Osmotic stress activates phosphatidylinositol-3,5-bisphosphate synthesis. Nature 390:187–192
12. Guidotti G (1972) Membrane proteins. Annu Rev Biochem 41:731–752
13. Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G (2005) Global topology analysis of the Escherichia coli inner membrane proteome. Science 308:1321–1323
14. Miller JP, Lo RS, Ben-Hur A, Desmarais C, Stagljar I, Noble WS, Fields S (2005) Large-scale identification of yeast integral membrane protein interactions. Proc Natl Acad Sci USA 102:12123–12128
15. Arinaminpathy Y, Khurana E, Engelman DM, Gerstein MB (2009) Computational analysis of membrane proteins: the largest class of drug targets. Drug Discov Today 14:1130–1135
16. Luckey M (2008) Membrane structural biology – with biochemical and biophysical foundations. Cambridge University Press, New York
17. Hedin LE, Illergard K, Elofsson A (2011) An introduction to membrane proteins. J Proteome Res 10:3324–3331
18. Bijlmakers MJ, Marsh M (2003) The on-off story of protein palmitoylation. Trends Cell Biol 13:32–42
19. Magee AI, Gutierrez L, McKay IA, Marshall CJ, Hall A (1987) Dynamic fatty acylation of p21N-ras. EMBO J 6:3353–3357
20. Dowler S, Currie RA, Campbell DG, Deak M, Kular G, Downes CP, Alessi DR (2000) Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. Biochem J 351:19–31
21. Ferguson KM, Lemmon MA, Schlessinger J, Sigler PB (1995) Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. Cell 83:1037–1046
22. Kholodenko BN, Hancock JF, Kolch W (2010) Signalling ballet in space and time. Nat Rev Mol Cell Biol 11:414–426
23. Yeagle P (1992) The structure of biological membranes. CRC Press, Boca Raton

24. Ubarretxena-Belandia I, Engelman DM (2001) Helical membrane proteins: diversity of functions in the context of simple architecture. Curr Opin Struct Biol 11:370–376
25. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta 1666:105–117
26. Junge F, Schneider B, Reckel S, Schwarz D, Dötsch V, Bernhard F (2008) Large-scale production of functional membrane proteins. Cell Mol Life Sci 65:1729–1755
27. Von Heine G (2011) Introduction to theme "membrane protein folding and insertion". Annu Rev Biochem 80:157–160
28. Haswell ES, Phillips R, Rees DC (2011) Mechanosensitive channels: what can they do and how do they do it? Structure 19:1356–1369
29. Granseth E, Daley D, Rapp M, Melen K, von Heijne G (2005) Experimentally constrained topology models for bacterial inner membrane proteins. J Mol Biol 352:489–494
30. Bowie JU (1997) Helix packing in membrane proteins. J Mol Biol 272:780–789
31. Arce J, Sturgis JN, Duneau JP (2009) Dissecting membrane protein architecture: an annotation of structural complexity. Biopolymers 91:815–829
32. Bocharov EV, Mineev KS, Volynsky PE, Ermolyuk YS, Tkach EN, Sobol AG, Chupin VV, Kirpichnikov MP, Efremov RG, Arseniev AS (2008) Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J Biol Chem 283:6950–6956
33. Smith SO, Bormann BJ (1995) Determination of helix-helix interactions in membranes by rotational resonance NMR. Proc Natl Acad Sci USA 92:488–491
34. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. Science 280:69–77
35. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. J Mol Biol 259:393–421
36. Yohannan S, Faham S, Yang D, Whitelegge J, Bowie J (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA 101:959–963
37. Fu D, Libson A, Miercke L, Weitzman C, Nollert P, Krucinski J, Stroud R (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290:481–486
38. Viklund H, Granseth E, Elofsson A (2006) Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. J Mol Biol 361:591–603
39. Kauko A, Illergard K, Elofsson A (2008) Coils in the membrane core are conserved and functionally important. J Mol Biol 380:170–180
40. Elofsson A, von Heijne G (2007) Membrane protein structure: prediction versus reality. Annu Rev Biochem 76:125–140
41. Blobel G (1980) Intracellular protein topogenesis. Proc Natl Acad Sci USA 77:1496–1500
42. Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK (1999) Structure of bacteriorhodopsin at 1.55 A resolution. J Mol Biol 291:899–911
43. Verdon G, Boudker O (2012) Crystal structure of an asymmetric trimer of a bacterial glutamate transporter homolog. Nat Struct Mol Biol 19:355–357
44. Gromiha MM (1999) A simple method for predicting transmembrane a helices with better accuracy. Protein Eng 12:557–561
45. Barlow DJ, Thornton JM (1988) Helix geometry in proteins. J Mol Biol 201:601–619
46. Reiersen H, Rees AR (2001) The hunchback and its neighbours: proline as an environmental modulator. Trends Biochem Sci 26:679–684
47. Cordes FS, Bright JN, Sansom MSP (2002) Proline-induced distortions of transmembrane helices. J Mol Biol 323:951–960
48. Engelman DM, Steitz TA, Goldman A (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 15:321–353
49. Koehler J, Woetzel N, Staritzbichler R, Sanders CR, Meiler J (2009) A unified hydrophobicity scale for multi-span membrane proteins. Proteins 76: 13–29
50. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
51. Wimely WC, White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol 3:842–848
52. Hessa T, Kim H, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Hejne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433:377–381
53. Brosig B, Langosch D (1998) The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. Protein Sci 7:1052–1056
54. Russ WP, Engelman DM (2000) The GxxxG motif: a framework for transmembrane helix-helix association. J Mol Biol 296:911–919
55. Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y (2000) Structural determinants of water permeation through aquaporin-1. Nature 407:599–605
56. Walters RF, DeGrado WF (2006) Helix-packing motifs in membrane proteins. Proc Natl Acad Sci USA 103:13658–13663
57. Marsico A, Henschel A, Winter C, Tuukkanen A, Vassilev B, Scheubert K, Schroeder M (2010) Structural fragment clustering reveals novel structural

and functional motifs in α-helical transmembrane proteins. BMC Bioinformatics 11:204–224
58. White SH, Ladokhin AS, Jayasinghe S, Hristova K (2001) How membranes shape protein structure. J Biol Chem 276:32395–32398
59. White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 28:319–365
60. Rees DC, DeAntonio L, Eisenberg D (1989) Hydrophobic organization of membrane proteins. Science 245:510–513
61. Beuming T, Weinstein H (2004) A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. Bioinformatics 20:1822–1835
62. Von Heijne G (2006) Membrane-protein topology. Nat Rev Mol Cell Biol 7:909–918
63. Langosch D, Heringa J (1998) Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. Proteins 31:150–159
64. Chantalat L, Jones ND, Korber F, Navaza J, Pavlovsky AG (1995) The crystal-structure of wild-type growth-hormone at 2.5 Angstrom resolution. Protein Pept Lett 2:333–340
65. Dobbs AJ, Anderson BF, Faber HR, Baker EN (1996) Three-dimensional structure of cytochrome c′ from two Alcaligenes species and the implications for four-helix bundle structures. Acta Crystallogr D Biol Crystallogr 52:356–368
66. Chang DK, Cheng SF, Trivedi VD, Lin KL (1999) Proline affects oligomerization of a coiled coil by inducing a kink in a long helix. J Struct Biol 128:270–279
67. Woolfson DN, Williams DH (1990) The influence of proline residues on alpha-helical structure. FEBS Lett 277:185–188
68. Von Heijne G (1987) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J 5:3021–3027
69. Johansson M, Nilsson I, Von Heijne G (1993) Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes. Mol Gen Genet 239:251–256
70. Miyazawa A, Fujiyoshi Y, Unwin N (2003) Structure and gating mechanism of the acetylcholine receptor pore. Nature 423:949–955
71. Amunts A, Drory O, Nelson N (2007) The structure of a plant photosystem I supercomplex at 3.4 Å resolution. Nature 447:58–63
72. Brunisholz RA, Zuber H (1992) Structure, function and organization of antenna polypeptides and antenna complexes from the three families of Rhodospirillaneae. J Photochem Photobiol B 15:113–140
73. Davidson AL, Maloney PC (2007) ABC transporters: how small machines do a big job. Trends Microbiol 15:448–455
74. Akopian AN, Sivilotti L, Wood JN (1996) A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. Nature 379:257–262
75. Andrade SL, Einsle O (2007) The Amt/Mep/Rh family of ammonium transport proteins. Mol Membr Biol 24:357–365
76. Minocha R, Studley K, Saier MH Jr (2003) The urea transporter (UT) family: bioinformatic analyses leading to structural, functional, and evolutionary predictions. Receptors Channels 9:345–352
77. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J Biol Chem 278:41148–41159
78. Schrempf H, Schmidt O, Kümmerlen R, Hinnah S, Müller D, Betzler M, Steinkamp T, Wagner R (1995) A prokaryotic potassium ion channel with two predicted transmembrane segments from Streptomyces lividans. EMBO J 14:5170–5178
79. Agre P, Bonhivers M, Borgnia MJ (1998) The aquaporins, blueprints for cellular plumbing systems. J Biol Chem 273:14659–14662
80. Agre P, Preston GM, Smith BL, Jung JS, Raina S, Moon C, Guggino WB, Nielsen S (1993) Aquaporin CHIP: the archetypal molecular water channel. Am J Physiol 265:463–476
81. Kaldenhoff R, Bertl A, Otto B, Moshelion M, Uehlein N (2007) Characterization of plant aquaporins. Methods Enzymol 428:505–531
82. Gonen T, Walz T (2006) The structure of aquaporins. Q Rev Biophys 39:361–396
83. Driessen AJM, Manting EH, van der Does C (2001) The structural basis of protein targeting and translocation in bacteria. Nat Struct Biol 8:492–498
84. Hedin L, Ojemalm K, Bernsel A, Hennerdal A, Illergard K, Enquist K, Kauko A, Cristobal S, Von Heijne G, Lerch-Bader M, Nilsson I, Elofsson A (2009) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. J Mol Biol 1:221–229
85. Goder V, Spiess M (2001) Topogenesis of membrane proteins: determinants and dynamics. FEBS Lett 504:87–93
86. Lee P, Tullman-Ercek D, Georgiou G (2006) The bacterial twin-arginine translocation pathway. Annu Rev Microbiol 60:373–395
87. Rehling P, Brandner K, Pfanner N (2004) Mitochondrial import and the twin-pore translocase. Nat Rev Mol Cell Biol 5:519–530
88. Emanuelsson O, Elofsson A, Von Heijne G, Cristobal S (2003) In silico prediction of the peroxisomal proteome in fungi, plants and animals. J Mol Biol 330:443–456
89. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29:4031–4037
90. Hunt JF, Earnest TN, Bousche O, Kalshathi K, Reilly K, Horvath C, Rothschild KJ, Engelman DM (1997) A biophysical study of integral membrane protein folding. Biochemistry 36:15156–15176

91. White SH (2009) Biophysical dissection of membrane proteins. Nature 459:344–346
92. Kauko A, Hedin L, Thebaud E, Cristobal S, Elofsson A, Von Heijne G (2010) Repositioning of transmembrane alpha-helices during membrane protein folding. J Mol Biol 397:190–201
93. Booth PJ, Curran AR (1999) Membrane protein folding. Curr Opin Struct Biol 9:115–121
94. Waxman SG (2007) Channel, neuronal and clinical function in sodium channelopathies: from genotype to phenotype. Nat Neurosci 10:405–409
95. Turner APF (1997) Switching channels makes sense. Nature 387:555–557
96. Hucho F, Weise C (2001) Ligand-gated ion channels. Angew Chem Int Ed 40:3100–3116
97. Neher E, Sakmann B (1992) The patch clamp technique. Sci Am 266:44–51
98. Cuello LG, Jogini V, Marien Cortes DM, Somporpisut A, Purdy MD, Wiener MC, Perozo E (2010) Design and characterization of a constitutively open KcsA. FEBS Lett 584:1133–1138
99. Beacham DW, Blackmer T, O'Grady M, Hanson GT (2010) Cell-based potassium ion channel screening using the FluxOR assay. J Biomol Screen 15:441–446
100. Galietta LV, Jayaraman S, Verkman AS (2001) Cell-based assay for high-throughput quantitative screening of CFTR chloride transport agonists. Am J Physiol Cell Physiol 281:C1734–C1742
101. Peitz I, Voelker M, Fromherz P (2007) Recombinant serotonin receptor on a transistor as a prototype for cell-based biosensors. Angew Chem Int Ed 46:5787–5790
102. Mach T, Chimerel C, Fritz J, Fertig N, Winterhalter M, Futterer C (2008) Miniaturized planar lipid bilayer: increased stability, low electric noise and fast fluid perfusion. Anal Bioanal Chem 390:841–846
103. Savage DF, Stroud RM (2007) Structural basis of aquaporin inhibition by mercury. J Mol Biol 368:607–617
104. Kumar M, Grzelakowski M, Zilles J, Clark M, Meier W (2007) Highly permeable polymeric membranes based on the incorporation of the functional water channel protein aquaporin Z. Proc Natl Acad Sci 104:20719–20724
105. Hill TR, Taylor BW (2012) Use of aquaporins to achieve needed water purity on the International Space Station for the Extravehicular Mobility Unit Space Suit System. Conference proceedings, conference on environmental systems (ICES), American Institute of Aeronautics and Astronautics, San Diego, 15–19 July 2011
106. Bishop RE (2008) Structural biology of membrane-intrinsic β-barrel enzymes: sentinels of the bacterial outer membrane. Biochim Biophys Acta 1778:1881–1896
107. Montoya M, Gouaux E (2003) Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. Biochim Biophys Acta 1666:250–263
108. Saier MH Jr (2000) Families of proteins forming transmembrane channels. J Membr Biol 175: 165–180
109. Schulz GE (2002) The structure of bacterial outer membrane proteins. Biochim Biophys Acta 1565:308–317
110. Koebnik R, Locher KP, Van Gelder P (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol 37: 239–253
111. Newcomer ME, Jones TA, Aqvist J, Sundelin J, Eriksson U, Rask L, Peterson PA (1984) The three-dimensional structure of retinol-binding protein. EMBO J 3:1451–1454
112. Galdiero S, Galdiero M, Pedone C (2007) β-barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. Curr Protein Pept Sci 8:63–82
113. Vogt J, Schulz GE (1999) The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. Structure 7:1301–1309
114. Clavel T, Germon P, Vianney A, Portalier R, Lazzaroni JC (1998) TolB protein of Escherichia coli K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. Mol Microbiol 29:359–367
115. Prince SM, Achtman M, Derrick JP (2002) Crystal structure of the OpcA integral membrane adhesin from Neisseria meningitidis. Proc Natl Acad Sci USA 99:3417–3421
116. Ye J, Van den Berg B (2004) Crystal structure of the bacterial nucleoside transporter Tsx. EMBO J 23:3187–3195
117. Locher KP, Rees B, Koebnik R, Mitschler A, Moulinier L, Rosenbusch JP, Moras D (1998) Transmembrane signaling across the ligand-gated FhuA receptor: Crystal structures of free and ferrichrome-bound states reveal allosteric changes. Cell 95: 771–778
118. Huang Y, Smith BS, Chen LX, Baxter RH, Deisenhofer J (2009) Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. Proc Natl Acad Sci USA 106:7403–7407
119. Krewinkel M, Dworeck T, Fioroni M (2012) Engineering of an E. coli outer membrane protein FhuA with increased channel diameter. J Nanobiotechnol 9:33
120. Schirmer T, Keller TA, Wang YF, Rosenbusch JP (1995) Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. Science 267:512–514
121. Buchanan SK (1999) Beta-barrel proteins from bacterial outer membranes: structure, function and refolding. Curr Opin Struct Biol 9:455–461
122. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science 274:1859–1866

123. Sakai N, Mareda J, Matile S (2008) Artificial β-barrels. Acc Chem Res 41:1354–1365
124. Lim Y, Lee M (2011) Toroidal β-barrels from self-assembling β-sheet peptides. J Mater Chem 21:11680–11685
125. Gromiha MM, Majumdar R, Ponnuswamy PK (1997) Identification of membrane spanning β strands in bacterial porins. Protein Eng 10: 497–500
126. Tamm LK, Hong H, Liang B (2004) Folding and assembly of β-barrel membrane proteins. Biochim Biophys Acta 1666:250–263
127. Seshadri K, Garemyr R, Wallin E, Von Heijne G, Elofsson A (1998) Architecture of β-barrel membrane proteins: analysis of trimeric porins. Protein Sci 7:2026–2032
128. Chou PY, Fasman GD (1977) Beta-turns in proteins. J Mol Biol 115:135–175
129. Sibanda BL, Thornton JM (1986) β-Hairpin families in globular proteins. Nature 316:170–174
130. Chothia C (1973) Conformation of twisted β-pleated sheets in proteins. J Mol Biol 75:295–302
131. Gellmann SH (1998) Minimal model systems for β-sheet secondary structure in proteins. Curr Opin Chem Biol 2:717–725
132. Petsko GA, Ringe D (2004) Protein structure and function. Sinauer Associates, Sunderland
133. Stickle DF, Presta LG, Dill KA, Rose GD (1992) Hydrogen bonding in globular proteins. J Mol Biol 226:1143–1159
134. Schulz GE (2000) β-Barrel membrane proteins. Curr Opin Struct Biol 10:443–447
135. Wimley WC (2002) Toward genomic identification of β-barrel membrane proteins: composition and architecture of known structures. Protein Sci 11: 301–312
136. Kleinschmidt JH (2006) Folding and stability of monomeric β-barrel membrane proteins. In: Tamm LK (ed) Protein-lipid interactions. Wiley VCH, Weinheim
137. Rosenbusch JP (2001) Stability of membrane proteins: relevance for the selection of appropriate methods for high-resolution structure determinations. J Struct Biol 136:144–157
138. Arora A, Abildgaard F, Bushweller JH, Tamm LK (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. Nat Struct Biol 8:334–338
139. Wimley WC (2003) The versatile β-barrel membrane protein. Curr Opin Struct Biol 13:404–411
140. Naveed H, Jackups R, Liang J (2009) Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins. Proc Natl Acad Sci 106:12735–12740
141. Evanics F, Hwang P, Cheng Y, Kay L, Prosser R (2006) Topology of an outer-membrane enzyme: measuring oxygen and water contacts in solution NMR studies of PagP. J Am Chem Soc 128: 8256–8264
142. Huysmans G, Radford S, Brockwell D, Baldwin S (2007) The N-terminal helix is a post-assembly clamp in the bacterial outer membrane protein PagP. J Mol Biol 373:529–540
143. Koebnik R (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. Mol Microbiol 16:1269–1270
144. Cowan SW, Garavito RM, Jansonius JN, Jenkins JA, Karlsson R, Konig N, Pai EF, Pauptit RA, Rizkallah PJ, Rosenbach JP (1995) The structure of OmpF porin in a tetragonal crystal form. Structure 3:1041–1050
145. Hill K, Model K, Ryan MT, Dietmeier K, Martin F, Wagner R, Pfanner N (1998) Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins. Nature 395:516–521
146. Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R (2003) Characterization of the translocon of the outer envelope of chloroplasts. J Cell Biol 160:541–551
147. Kleinschmidt JH, Tamm LK (2002) Secondary and tertiary structure formation of the b-barrel membrane protein OmpA is synchronized and depends on membrane thickness. J Mol Biol 324:319–330
148. Wimley WC, Hristova K, Ladokhin AS, Silvestro L, Axelsen PH, White SH (1998) Folding of b-sheet membrane proteins: a hydrophobic hexapeptide model. J Mol Biol 277:1091–1110
149. Ramachandran R, Heuck AP, Tweten RK, Johnson AE (2002) Structural insights into the membrane-anchoring mechanism of a cholesterol-dependent cytolysin. Nat Struct Biol 9:823–827
150. Bayley H (1997) Toxin structure: part of a hole? Curr Biol 7:R763–R767
151. Misra R (2012) Assembly of the β-barrel outer membrane proteins in Gram-negative bacteria, mitochondria, and chloroplasts. ISRN Mol Biol 2012:1–15
152. Holtje JV (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. Microbiol Mol Biol Rev 62:181–203
153. Beveridge TJ (1999) Structures of gram-negative cell walls and their derived membrane vesicles. J Bacteriol 181:4725–4733
154. Tamm LK, Arora A, Kleinschmidt JH (2001) Structure and assembly of beta-barrel membrane proteins. J Biol Chem 276:32399–32402
155. Walther DM, Rapaport D, Tommassen J (2009) Biogenesis of β-barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. Cell Mol Life Sci 66:2789–2804
156. Paetzel M, Strynadka NCJ (2001) Signal peptide cleavage in the E. coli membrane. CSBMCB/SCBBMC Bull 2001:60–65
157. Bannwarth M, Schulz GE (2003) The expression of outer membrane proteins for crystallization. Biochim Biophys Acta 1610:37–45
158. Dworeck T, Petri AK, Muhammad N, Fioroni M, Schwaneberg U (2011) FhuA deletion variant Δ1-

159 overexpression in inclusion bodies and refolding with Polyethylene-Poly(ethylene glycol) diblock copolymer. Protein Expr Purif 77:75–79
159. Bechtluft P, Nouwen N, Tans SJ, Driessen AJ (2010) SecB – a chaperone dedicated to protein translocation. Mol Biosyst 6:620–627
160. du Plessis DJF, Nouwen N, Driessen AJM (2011) The Sec translocase. Biochim Biophys Acta 1808:851–865
161. Kusters I, Driessen AJ (2011) SecA, a remarkable nanomachine. Cell Mol Life Sci 68:2053–2066
162. Ricci DP, Silhavy TJ (2012) The Bam machine: a molecular cooper. Biochim Biophys Acta 1818:1067–1084
163. Vertommen D, Ruiz N, Leverrier P, Silhavy TJ, Collet JF (2009) Characterization of the role of the Escherichia coli periplasmic chaperone SurA using differential proteomics. Proteomics 9:2432–2443
164. Rigel NW, Silhavy TJ (2012) Making a beta-barrel: assembly of outer membrane proteins in gram-negative bacteria. Curr Opin Microbiol 15: 189–193
165. Schäfer U, Beck K, Müller M (1999) Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. J Biol Chem 274:24567–24574
166. Booth PJ, Templer RH, Meijberg W, Allen SJ, Curran AR, Lorch M (2001) In vitro studies of membrane protein folding. Crit Rev Biochem Mol Biol 36:501–603
167. Wu T, Malinverni J, Ruiz N, Kim S, Silhavy TJ, Kahne D (2005) Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. Cell 121:235–245
168. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. Science 299:262–265
169. Robert V, Volokhina EB, Senf F, Bos MP, Van Gelder P, Tommassen J (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. PLoS Biol 4:1984–1995
170. Struyve M, Moons M, Tommassen J (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J Mol Biol 218:141–148
171. Kim S, Malinverni JC, Sliz P, Silhavy TJ, Harrison SC, Kahne D (2007) Structure and function of an essential component of the outer membrane protein assembly machine. Science 317:961–964
172. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E (2010) Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J Biol Chem 285:6848–6856
173. Malinverni JC, Werner J, Kim S, Sklar JG, Kahne D, Misra R, Silhavy TJ (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. Mol Microbiol 61:151–164
174. Hagan CL, Kim S, Kahne D (2010) Reconstitution of outer membrane protein assembly from purified components. Science 14:890–892
175. Hagan CL, Kahne D (2011) The reconstituted Escherichia coli Bam complex catalyzes multiple rounds of β-barrel assembly. Biochemistry 50:7444–7446
176. Kim KH, Aulakh S, Paetzel M (2011) Crystal structure of β-barrel assembly machinery BamCD protein complex. J Biol Chem 286:39116–39121
177. Kim KH, Paetzel M (2011) Crystal structure of Escherichia coli BamB, a lipoprotein component of the β-barrel assembly machinery complex. J Mol Biol 406:667–678
178. Kim KH, Kang HS, Okon M, Escobar-Cabrera E, McIntosh LP, Paetzel M (2011) Structural characterization of Escherichia coli BamE, a lipoprotein component of the β-barrel assembly machinery complex. Biochemistry 50:1081–1090
179. Güven A, Fioroni M, Hauer B, Schwaneberg U (2010) Molecular understanding of sterically controlled compound release through an engineered channel protein (FhuA). J Nanobiotechnol 8:14
180. Güven A, Dworeck T, Fioroni M, Schwaneberg U (2011) Residue K556-A light triggerable gatekeeper to sterically control translocation in FhuA. Adv Eng Mater 13:B324–B329
181. Muhammad N, Dworeck T, Fioroni M, Schwaneberg U (2011) Engineering of the E. coli outer membrane protein FhuA to overcome the hydrophobic mismatch in thick polymeric membranes. J Nanobiotechnol 9:8
182. Nardin C, Meier W (2002) Hybrid materials from amphiphilic block copolymers and membrane proteins. Rev Mol Biotechnol 90:17–26
183. Mohammad M, Iyer R, Howard KR, McPike M, Borer PN, Movileanu L (2012) Engineering a rigid protein tunnel for biomolecular detection. J Am Chem Soc 134:9521–9531
184. Howorka S, Siwy Z (2009) Nanopore analytics: sensing of single molecules. Chem Soc Rev 38:2360–2384
185. Chen M, Khalid S, Sansom MSP, Bayley H (2008) Outer membrane protein G: engineering a quiet pore for biosensing. Proc Natl Acad Sci 105:6272–6277
186. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Sean Ling X, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA (2008) The potential and challenges of nanopore sequencing. Nat Biotechnol 26:1146–1153
187. Onoda A, Fukumoto K, Arlt M, Bocola M, Schwaneberg U, Hayashi T (2012) A rhodium complex-linked β-barrel protein as a hybrid biocatalyst for phenylacetylene polymerization. Chem Commun 48:9756–9758
188. Surrey T, Jähnig FJ (1992) Refolding and oriented insertion of a membrane protein into a lipid bilayer. Proc Natl Acad Sci USA 89:7457–7461

189. Surrey T, Schmid A, Jähnig F (1996) Folding and membrane insertion of the trimeric β-barrel protein OmpF. Biochemistry 35:2283–2288
190. Klug CS, Feix JB (1998) Guanidine hydrochloride unfolding of a transmembrane h-strand in FepA using site-directed spin labeling. Protein Sci 7:1469–1476
191. Banerjee S, Huber T, Sakmar T (2008) Rapid incorporation of functional rhodopsin into nanoscale apolipoprotein bound bilayer (NABB) particles. J Mol Biol 377:1067–1081
192. Mohammad M, Howard KR, Movileanu L (2011) Redesign of a plugged β-barrel membrane protein. J Biol Chem 286:8000–8013
193. Onaca O, Sarkar P, Roccatano D, Friedrich T, Hauer B, Grzelakowski M, Güven A, Fioroni M, Schwaneberg U (2008) Functionalized nanocompartments (synthosomes) with a reduction-triggered release system. Angew Chem Int Ed 47: 7029–7031
194. Létoffé S, Wecker K, Delepierre M, Delepelaire P, Wandersman C (2005) Activities of the Serratia marcescens heme receptor HasR and isolated plug and β-barrel domains: the β-barrel forms a heme-specific channel. J Bacteriol 187:4637–4645
195. Koebnik R (1999) Membrane assembly of the Escherichia coli outer membrane protein OmpA: exploring sequence constraints on trans-membrane b-strands. J Mol Biol 285:1801–1810
196. Bonhivers M, Desmadril M, Moeck GS, Boulanger P, Colomer-Pallas A, Letellier L (2001) Stability studies of FhuA, a two-domain outer membrane protein from Escherichia coli. Biochemistry 40:2606–2613
197. Klammt C, Schwarz D, Fendler K, Haase W, Dötsch V, Bernhard F (2005) Evaluation of detergents for the soluble expression of a-helical and b-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. FEBS J 272:6024–6038
198. Briat JF (1992) Iron assimilation and storage in prokaryotes. J Gen Microbiol 138:2475–2483
199. Alfonso-Prieto M, Biarnés X, Vidossich P, Rovira C (2009) The molecular mechanism of the catalase reaction. J Am Chem Soc 131:11751–11761
200. Ratledge C, Dover LG (2000) Iron metabolism in pathogenic bacteria. Annu Rev Microbiol 54:881–941
201. Neilands JB (1995) Siderophores: structure and function of microbial iron transport compounds. J Biol Chem 270:26723–26726
202. Ferguson AD, Deisenhofer J (2002) TonB-dependent receptors – structural perspectives. Biochim Biophys Acta 1565:318–332
203. Lundrigan ML, Kadner RJ (1986) Nucleotide sequence of the gene for the ferrienterochelin receptor FepA in Escherichia coli. Homology among outer membrane receptors that interact with TonB. J Biol Chem 261:10797–10801
204. Pressler U, Staudenmaier H, Zimmermann L, Braun V (1988) Genetics of the iron dicitrate transport system of Escherichia coli. J Bacteriol 170:2716–2724
205. Coulton JW, Mason P, DuBow MS (1983) Molecular cloning of the ferrichrome-iron receptor of Escherichia coli K-12. J Bacteriol 156:1315–1321
206. Ferguson AD, Hofmann E, Coulton JW, Diederichs K, Welte W (1998) Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science 282:2215–2220
207. Buchanan SK, Smith BS, Venkatramani L, Xia D, Esser L, Palnitkar M, Chakraborty R, van der Helm D, Deisenhofer J (1999) Crystal structure of the outer membrane active transporter FepA from Escherichia coli. Nat Struct Biol 6:56–62
208. Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J (2002) Structural basis of gating by the outer membrane transporter FecA. Science 295:1658–1659
209. Chakraborty R, Storey E, van der Helm D (2007) Molecular mechanism of ferricsiderophore passage through the outer membrane receptor proteins of Escherichia coli. Biometals 20:263–274
210. Braun V, Schaller K, Wolff H (1973) A common receptor protein for phage T5 and colicin M in the outer membrane of Escherichia coli B. Biochim Biophys Acta 323:87–97
211. Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511
212. Hantke K, Braun V (1975) Membrane receptor-dependent iron transport in Escherichia coli. FEBS Lett 49:301–305
213. Kadner RJ, Heller K, Coulton JW, Braun V (1980) Genetic control of hydroxamate-mediated iron uptake in Escherichia coli. J Bacteriol 143:256–264
214. Bonhivers M, Ghazi A, Boulanger P, Letellier L (1996) FhuA, a transporter of the Escherichia coli outer membrane, is converted into a channel upon binding of bacteriophage T5. EMBO J 15:1850–1856
215. Braun M, Killmann H, Maier E, Benz R, Braun V (2002) Diffusion through channel derivatives of the Escherichia coli FhuA transport protein. Eur J Biochem 269:4948–4959
216. Braun M, Killmann H, Braun V (1999) The beta-barrel domain of FhuADelta5-160 is sufficient for TonB-dependent FhuA activities of Escherichia coli. Mol Microbiol 33:1037–1049
217. Killmann H, Herrmann C, Torun A, Jung G, Braun V (2002) TonB of Escherichia coli activates FhuA through interaction with the β-barrel. Microbiology 148:3497–3509
218. Pawelek PD, Croteau N, Ng-Thow-Hing C, Khursigara CM, Moiseeva N, Allaire M, Coulton JW (2006) Structure of TonB in complex with FhuA, E. coli outer membrane receptor. Science 312:1399–1402

219. Braun V, Braun M (2002) Iron transport and signaling in E. coli. FEBS Lett 529:78–85
220. Eisenhauer HA, Shames S, Pawelek PD, Coulton JW (2005) Siderophore transport through Escherichia coli outer membrane receptor FhuA with disulfide-tethered cork and barrel domains. J Biol Chem 280:30574–30580
221. Ma L, Kaserer W, Annamalai R, Scott DC, Jin B, Jiang X, Xiao Q, Maymani H, Massis LM, Ferreira LC, Newton SM, Klebba PE (2007) Evidence of ball-and-chain transport of ferric enterobactin through FepA. J Biol Chem 282:397–406
222. Udho E, Jakes KS, Finkelstein A (2012) TonB-dependent transporter FhuA in planar lipid bilayers: partial exit of its plug from the barrel. Biochemistry 51:6753–6759
223. Bertin A, de Frutos M, Letellier L (2011) Bacteriophage-host interactions leading to genome internalization. Curr Opin Microbiol 14:492–496
224. Böhm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL (2001) FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol 11:1168–1175
225. Faraldo-Gómez JD, Smith GR, Sansom MSP (2003) Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states. Biophys J 85:1406–1420
226. Locher KP, Rosenbusch JP (1997) Oligomeric states and siderophore binding of the ligand-gated FhuA protein that forms channels across Escherichia coli outer membranes. Eur J Biochem 247:770–775
227. Ferguson AD, Breed J, Diederichs K, Welte W, Coulton JW (1998) An internal affinity-tag for purification and crystallization of the siderophore receptor FhuA, integral outer membrane protein from Escherichia coli K-12. Protein Sci 7:1636–1638
228. Plancon L, Chami M, Letellier L (1997) Reconstitution of FhuA, an Escherichia coli outer membrane protein, into liposomes. Binding of phage T5 to Fhua triggers the transfer of DNA into the proteoliposomes. J Biol Chem 272:16868–16872

Biophysical Characterization

3

This chapter will give a state of the art review on the most commonly used biophysical characterization techniques used in β-barrel outer membrane protein (OMP) structural analysis such as X-ray diffraction analysis of protein crystals, Circular Dichroism (CD) spectroscopy and Nuclear Magnetic Resonance (NMR). For each of the aforementioned techniques the alliances with complementary as well as with other sophisticated techniques to obtain structural, kinetic and thermodynamical information will be reported. Though X-Ray diffraction measurements and NMR studies result in valuable tertiary structure information, while CD spectroscopy can give information on the secondary structure mainly, CD will be particularly analyzed as a reliable and user friendly but under-represented method with which to obtain rapid secondary structural information on wild-type and engineered β-barrel OMPs. Protein structural information studies are especially important for genetically engineered β-barrel OMP channel variants with novel structural/geometrical features. For the use as nano-material components the correct protein folding into the desired structure is a prerequisite for channel function and reconstitution ability, therefore a particular focus should be given to structure determination.

3.1 Crystallization

Crystal structure determination is the "battle horse" of structural chemistry and biology. The high resolution molecular three-dimensional (3D) structure of biomolecules determined by X-ray [1], neutrons [2] or electrons [3] poses the fundamental knowledge to understand how to resolve the everlasting "structure to function" relationship problem. Though the methodology of these techniques is well established and defined, due to their long time averaged measurements, as static high resolution structures [4] opposite to the dynamic high resolution obtained, for example by NMR, their use is always considered a fundamental step toward the general knowledge of the analyzed system. However *Time Resolved Crystallography* is a crystallographic method that gives precious information on the picosecond scale structural changes, thus offering valuable insight on the structure-function relationship and catalytic compound intermediates in enzymes [5].

The following sub-section will give a comparison of the main crystallization methods for membrane proteins in general, listing some of the known examples of membrane proteins crystallized by each of the different

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_3,

methods. Furthermore specializations regarding the crystallization of β-barrel membrane proteins will be mentioned and briefly explained.

3.1.1 Comparison of Membrane Protein Crystallization Methods

X-ray crystallography by itself when referred to protein structure determination already has a long standing tradition. Its history started almost 60 years ago when the structures of water soluble proteins like myoglobin [6] and hemoglobin [7] were first resolved. The constraints posed on the selection of the protein to be crystallized were, at the early days of X-ray crystallography, mainly based on the easy process to obtain a relatively high amount of the sample with high purity by protein expression and protein purification. Water soluble proteins were obviously the first choice. A time span of nearly 30 years [4] had to pass before the first crystal structure of a transmembrane protein was obtained, represented by the publication of the crystal structure of the photosynthetic reaction center from *Rhodopseudomonas viridis* [8, 9].

This long time span stresses how the inherent technical problematic encountered in membrane protein crystallization made their structural resolution rather complex and difficult. However with the successful crystallization of the *Rhodopseudomonas viridis* photosynthetic reaction center a considerable shift of interest toward membrane protein structure determination could be observed. The mentioned difficulties in membrane protein crystallization originate from a "*multi-scale*" level because of the nearly inextricable steps of expression, purification and crystallization. Furthermore membrane protein crystals once obtained often show rather poor X-ray diffraction.

However during the last two decades a powerful alliance between molecular biology and evolved crystallographic techniques based on robotic systems, able to probe thousands of different crystallization conditions varying parameters such as ionic strength, concentration, temperatures, etc. [10] drastically increased the number of resolved membrane proteins structures. As a consequence all the three aforementioned principal steps *i.e.* protein expression, protein purification, and crystal growth took advantage of the relatively fast development of each of the specific method.

Due to the fundamental importance of membrane proteins within the physiology of both prokaryotic and eukaryotic organisms, a continuous research and development is underway. For example in humans 27 % of the total genome is supposed to express α-helical trans-membrane proteins [11] stressing the fundamental and general importance of this class of proteins.

Just to mention few of the physiological roles covered by membrane proteins that are analyzed by X-ray crystallography, a new series of channels and transporters have been recently reported [12] while on the receptor side agonist-bound structures of G protein-coupled receptors [13] have been discovered. Another important success has been obtained by analyzing the protein-membrane interactions of the P-type ATPase family, functioning as cation pumps and lipid flippases maintaining the electrochemical gradients and asymmetric lipid distributions across membranes [14]. As a last example, on the prokaryotic side, in Gram-negative bacteria the pilus biogenesis [15] has been found to be based on the large β-barrel outer membrane transporter PapC consisting of 24 β-strands. PapC is up to now, to the best of our knowledge, the largest known naturally occurring β-barrel membrane protein (see also Sect. 2.3). For more general and detailed information on the structure-function relationship and importance of the α- and β-barrel membrane proteins see Chap. 2, while Chap. 5 will consider the challenges in expression and purification of bacterial outer membrane β-barrel proteins.

Though the number of crystallized membrane proteins has increased considerably over the last decade, "only" 393 distinct resolved structures as reported on the website of S.H. White "Membrane proteins of known 3D structure" (http://blanco.biomol.uci.edu/mpstruc/listAll/list, accessed on 03-04-2013) have been published, a rather low number if compared to

the total number of X-ray diffraction resolved protein structures, *i.e.* ~79,000 RCSB PDB entries (Research Collaboratory for Structural Bioinformatics, www.rcsb.org, accessed on 20-04-2013). This under-representation of membrane proteins is due to the nowadays still up-to-date problems of membrane protein expression and purification in order to obtain the necessary amounts of highly pure protein (on the hundreds of milligrams) for subsequent crystallization [16].

However once large amounts of the pure target protein have been obtained, the protein solution has to be brought to supersaturation where spontaneous nucleation and crystal growth can take place. Pre-requisite is that the protein is in solution. For water soluble proteins solubility will generally increase as salt is added to an aqueous solution, then protein solubility will start to decrease as the salt concentration is increased and high enough to compete with the protein for hydration. Supersaturation can then be achieved by various vapor diffusion, slow evaporation or dialysis techniques (*e.g.* hanging- or sitting-drop vapor diffusion or micro-dialysis). Within the thermodynamically instable supersaturation state of protein solubility (as a function of protein and salt concentration) three distinct zones can be defined as shown in the protein solubility plot in Fig. 3.1. Zone I may sustain crystal growth, it will generally not easily promote nucleation, zone II will promote both growth and nucleation, while zone III instead will lead to protein precipitation.

The main parameters affecting crystallization are therefore apart from target protein purity the protein and salt concentration; moreover temperature and pH, as well as certain additives that can influence crystal growth and initial nucleation. For further reading on protein crystallization [17] can be suggested.

When attempting to crystallize membrane proteins however the situation is not so straightforward as these proteins are not soluble in neat water and require an amphiphilic environment (*e.g.* detergent solutions instead) adding a further parameter to the already complex circumstances. Different and specialized methodologies have therefore been developed for the crystallization of α-helical [18] and β-barrel membrane proteins [18, 19] with a trend oriented toward crystallization high-throughput screening and automated imaging of membrane crystals using electron microscopy techniques [20].

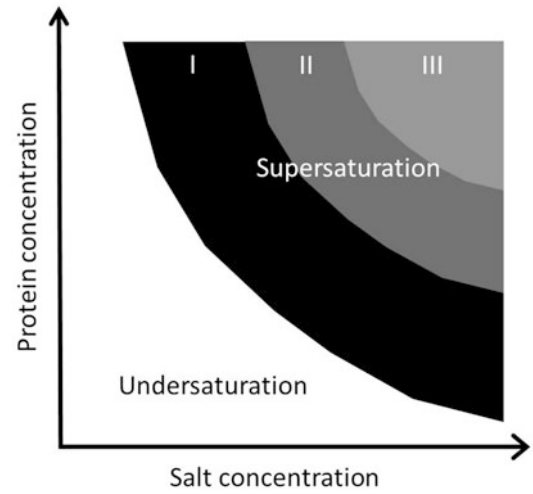

Fig. 3.1 Protein solubility curves, plotting protein concentration over salt concentration. The three zones of supersaturation are indicated (I–III). Zone I promotes crystal growth but not nucleation (so crystal seeds are necessary to start crystallization), zone II promotes crystal growth and nucleation and zone III leads to protein precipitation

The general membrane protein crystallization experiment work flow (including the most common crystal types and how to obtain these) is depicted in Fig. 3.2.

The first step corresponds to the extraction of the membrane proteins embedded in the lipid bilayer. This step is accomplished by the use of so called “mild detergents”, important not only in the extraction step but also to solubilize the membrane protein, as they keep the protein stable in an aqueous solution.

Detergents must be selected with regard to the protein type to be extracted and solubilized. A detergent that efficiently extracts or solubilizes a certain protein can be often also applied to other proteins of the same class. A rather huge variety of detergents is used and is represented by four main classes characterized by their chemical structures: ionic, non-ionic, zwitterionic and bile acid salts [21, 22] (For more information on specific outer membrane

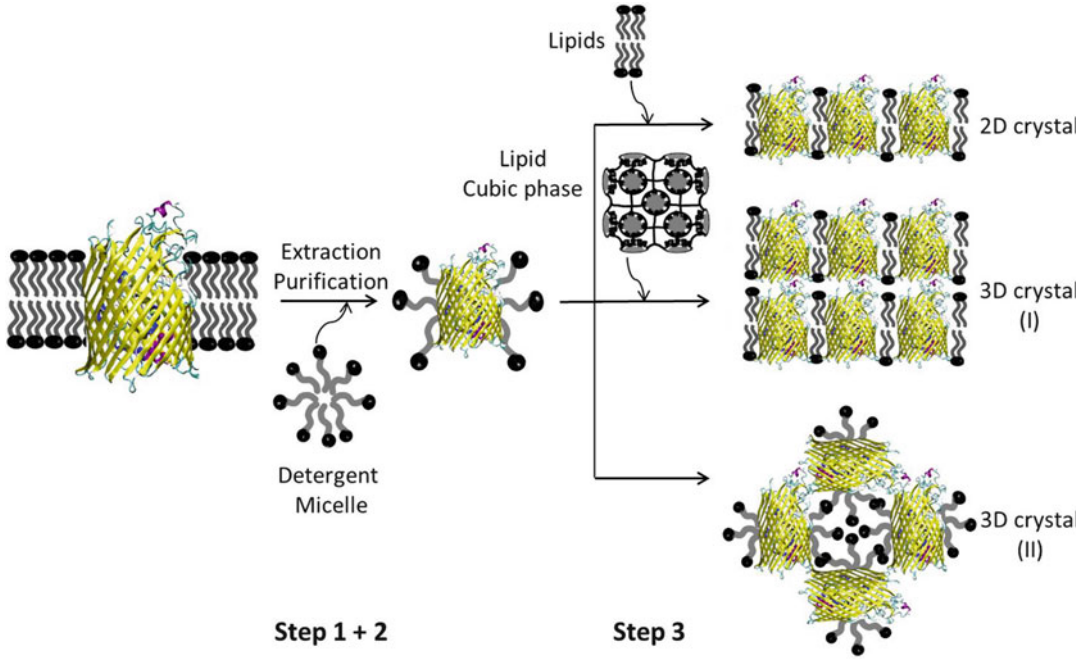

Fig. 3.2 Schematic representation of the most commonly used membrane protein crystallization techniques. The target protein has first to be isolated from its respective membrane and solubilized by a suitable detergent (step 1 and 2). In step 3 crystals are obtained either by exchanging detergent by lipids resulting in 2D crystals (*top*); in lipid cubic phase obtaining 3D crystals of type I (*middle*) or in detergent leading to 3D crystals of type II (*bottom*)

protein extraction/solubilization methods and the involved detergents see Sect. 5.2.1). Each class has some determined physico-chemical characteristics and a balance must be reached considering their different ability to be denaturants or to break down selectively, the lipid-lipid, lipid-protein and protein-protein interactions [21, 22].

Because detergents are used in great excess to extract/solubilize the protein, the next step is their removal by single methods each one taking advantages of the different chemical-physical characteristics of the detergents like critical micelle concentration (cmc), charge and aggregation number.

Apart from "pure" detergents, mixtures of lipids and detergents have been found useful trying to reproduce a local environment embedding the protein similar to the natural lipid bilayer environment.

From this point on there are three possible routes to obtain crystals of the target membrane protein. The first is trying to crystallize the complex detergent-protein itself obtaining a 3D crystal (Type II 3D crystal). A second possible route is based on the detergent exchange with a lipid reconstituting the protein in an "artificial" lipid bilayer defined as a 2D crystal generally used in electron microscopy diffraction studies while the last considers the insertion of the extracted membrane protein within a cubic lipid phase denominated Type I 3D Crystal.

Each of the aforementioned possible pathways is a quite painstaking "art" to obtain a high quality crystal to be used in X-ray diffraction studies and a very active research is under development facing the crystallization problem at the "multiscale" level previously mentioned.

A great effort is also applied to develop a powerful software basis for the interpretation of X-ray data. For example the CCP4 (Collaborative Computational Project, Number 4) software suite is a database/collection of programs with associated data and software libraries that can be used in macromolecular structure determination

by X-ray crystallography [23]. Another example is Xsolve, a software environment for automated structure determination able to execute the necessary crystallographic data processing and MAD (Multiwavelength Anomalous Dispersion) structure determination steps [24] or PHENIX, a highly automated software for macromolecular structure determination giving an initial partial structure model, when fed with moderate resolution and good quality data [25].

When considering crystallization techniques that avoid the influence of detergents that can interfere with the protein function by *e.g.* destabilizing the structure, extraction protocols based on a detergent-free approach to membrane protein solubilization using "nanodiscs" or nanolipoprotein particles have been developed [26]. By definition "nanodiscs" or nanolipoprotein particles consist of a raft of lipid bilayer surrounded by two copies of an engineered amphiphilic helical section of apolipoprotein, called a membrane scaffold protein (MSP) [26].

However, many of the X-ray resolved structures of the membrane proteins have been obtained by crystallizing proteins in detergents [27]. In general the use of detergents that involves the embedding of the protein within micelles [28] is thought to influence the membrane protein conformation and flexibility due to its inhomogeneity and due to the destruction of protein-protein contacts within the crystal. To avoid this problem a partial solution was found by stabilizing the protein-protein contacts by addition of antibody fragments [22].

To reproduce a more natural environment, the use of lipids in the crystallization procedure targets the reproduction of the original environment, though complications can be brought by the complex phase diagrams that lipids show [29].

On expression and purification level instead, a new approach consisting of an automated discontinuous-batch protein synthesis robot, using a wheat-germ cell-free translation system was developed [30]. The same technique able to overcome the problems related to the protein expression in an organism, the wheat-germ cell-free expression technique has been successfully applied to produce the human stearoyl-CoA desaturase complex [31] (For more information on the cell-free expression of bacterial outer membrane proteins see Sect. 5.2.3). Special purification protocols for proteins expressed in the yeast *Saccharomyces cerevisiae* have been developed [32] to enhance and increase the quantity and quality of eukaryotic transmembrane proteins available for structure determination by X-ray crystallography. Shifting the attention to the membrane proteins of prokaryotic organisms, a newly engineered strain of *Escherichia coli* able to accept a strong over-expression of membrane proteins without too many negative effects for the cells has been developed [33].

As it will be reported in Sect. 5.1.2.2, a general problem related to the over-expression of membrane proteins in bacteria like *Escherichia coli*, especially if dealing with channels or transporters, is the strong deterioration of the physiological parameters of the cell, depressing or even killing the cell itself practically disabling the production of the desired protein.

Several general protocols for the crystallization of membrane proteins for X-ray structural investigation have been developed as can be found in Newby et al. [34] where ten main points for a successful crystallization process are reported. A more general protocol that is not only applied to membrane proteins, defined as a "minimal crystallization screen" is based on an analysis of 340,000 individual crystallization trials. This analysis led to the development of a new minimal coarse screen (GNF96), which is highly effective in identifying targets which crystallize easily and provides leads for the optimization of crystallization conditions [35].

The "*in meso*" approach, *i.e.* using lipidic mesophases, is generally not widely used due to the inherent difficulties related to the handling of the media. However there are a series of advantages due to the fact that the target protein is taken out of the potentially harmful environment of a detergent micelle originally used to solubilize the protein itself, and is instead placed in a more natural environment similar to the natural one [36].

Recently a new method developing the lipidic cubic phase crystallization concept was extended toward a "lipidic sponge phase" [37] that can be used for all types of bacterial and eukaryotic membrane proteins in an ordinary hanging- or sitting- drop vapor diffusion experiment and in combination with high-throughput crystallization approaches. The sponge phase method has been also coupled with sophisticated techniques. For example a micrometer-sized lipidic sponge phase crystal of the *Blastochloris viridis* photosynthetic reaction center was delivered into an X-FEL (X-ray free electron laser) beam using a sponge phase micro-jet obtaining interpretable diffraction data with a resolution of 0.82 nm [38].

A new "*in meso*" method has been developed, where rationally designed lipidic mesophases have been used to crystallize proteins applicable to a wide range of integral membrane protein classes [29]. The "*in meso*" method accounts for ∼10 % of the published X-ray structures of integral membrane proteins and though it is not yet a "user friendly" technology it surely shows great potential for future developments [36].

Another level of approach is to reduce the amount of protein using micro to nanocrystals as successfully carried out for the cyanobacterial Photosystem I (PSI), which is the largest and most complex membrane protein crystallized to date [39]. By the use of femtosecond X-ray protein nanocrystallography based on the high-energy X-ray free electron laser of the LCLS at SLAC National Accelerator Laboratory, a liquid jet of fully-hydrated Photosystem I nanocrystals was brought into the interaction region of the pulsed X-ray source. Successively, diffraction patterns were recorded for millions of individual Photosystem I nanocrystals and data from thousands of different, randomly oriented crystallites were integrated using Monte Carlo integration of the peak intensities obtaining a resolution of 0.85 nm.

A continuously updated website based on the Protein Structure Initiative (PSI) Nature Structural Biology Knowledgebase is the Membrane Protein Hub (http://sbkb.org/kb/membprothub.html) where the state of the art on methods and structural information regarding membrane proteins is reported.

3.1.2 State of the Art in β-Barrel Protein Crystallization

Outer membrane proteins (OMPs) found in the outer membrane of Gram-negative bacteria or of the eukaryotic organelles mitochondria and chloroplasts (also referred to as β-barrel membrane proteins) possess their own expression, purification and crystallization protocols (see Sect. 5.2). Bacterial porins, in particular have been considered the main example and precursor of these class of proteins, with their specifically developed protocols considered as "master templates" to be used as first attempt in β-barrel crystallization procedures [22] (A detailed discussion on OMP structural and functional features, biogenesis and on their relevance for the development of new nano-materials can be found in Sect. 2.3).

The website http://blanco.biomol.uci.edu/mpstruc/listAll/list dedicated to membrane proteins reports as a subset the known β-barrel protein structures mainly resolved by X-ray crystallography or by NMR.

Some procedures specialized on OMPs have been published in the literature. For example Tanabe and Iverson developed a general formulation to be applied for crystallization of β-barrel proteins based on 96 screening formulations (BetaMem™) [40]. Another screening based on 48 formulations (MemPlus™) has been developed by Iwata and colleagues [41].

The crystallization of OMP is assisted by the possibility to be over-expressed in *Escherichia coli* as inclusion bodies with considerably high yields. In contrast to α-helical membrane proteins OMPs readily fold back from solubilized inclusion body material when detergent micelles are present (see Sect. 5.2.2).

An extension of the techniques previously reported using bicelles or lipidic cubic phases to β-barrel membrane proteins has been reported. Using the bicelles technique [42], high resolution

X-ray crystal structure of mouse VDAC (Voltage Dependent Ion Channel) a β-barrel transmembrane protein has been obtained though lipidic cubic phases are more frequently used and successfully applied [36].

Quite often, however, β-barrel membrane proteins are associated in complexes or machineries rendering the crystallization procedure quite challenging. For example the Bam (β-barrel assembly machine) complex found in the *Escherichia coli* outer membrane is responsible for the assembly and insertion of outer membrane proteins and is constituted by an integral β-barrel outer membrane protein BamA and four accessory lipoproteins BamB, BamC, BamD and BamE [43]. While the four lipoproteins have been successfully and recently crystallized the BamA β-barrel protein is still unresolved [43]. Further information on the Bam complex and the role it plays in the OMP assembly into the outer membrane can be found in Sect. 2.3.2.

Other protocols based on Type V secretion proteins have been recently published [44]. The type V secretion pathway is constituted by large substrate proteins (passenger domains or TpsA) and membrane proteins (β-barrel domains or TpsB). The obtaining of the crystal structure of Type V secretion proteins was possible by overcoming the challenges in protein production and crystallization [45].

In conclusion, the development of efficient crystallization protocols applicable to β-barrel membrane-proteins is strongly assisted by the rapidly appearing new crystallization techniques for membrane proteins in general.

3.2 Circular Dichroism

In recent years the Circular Dichroism (CD) technique experienced a "*Renaissance*" especially as a tool to obtain protein secondary structure information. By itself CD is a powerful and elegant absorption spectroscopy mainly used to extract secondary structure content information of proteins though it is widely applied to a vast range of optically active organic, inorganic and organometallic molecules [46].

Here will be given some basic CD spectroscopy theory including an explanation of the different spectra obtained for the various *pure* protein secondary structure elements. The different data de-convolution methods will be mentioned and a short explanation on the analysis of liposome or polymersome reconstituted outer membrane protein channels by CD spectroscopy will be given.

3.2.1 Basic Theory

The CD-spectra are a result of electronic transitions induced by circularly left and right polarized light within optical active (asymmetric) or symmetric chromophores the last immersed in asymmetric environments. The difference in absorbance between the left and right polarized light will result in a CD signal.

The kinds of asymmetries leading to optical activity in biopolymers can be generalized into three main categories:

1. the primary structure is asymmetric. A classic example are the C_α in amino acids where the electronic transitions involving the C_α itself of a peptide bond are asymmetric due to the four different substituents (with the exception of amino acid Gly) and the static electric field they generate is asymmetric;
2. the secondary structure is characterized by a well-defined three dimensional organization where the incident light or electromagnetic field induces a magnetic as well as an electric dipole with the electric and magnetic moments opposite depending on if they are excited by circularly polarized left or right light. The α-helix protein secondary structure element is the prototype of such phenomenon, where the induced electric and magnetic moments are parallel to the α-helix axis;
3. the tertiary structure of a biopolymer of a symmetric group experiencing an asymmetric environment.

As an example can be reported the electronic transitions involving the π-electron in the aromatic Tyr ring generally weakly optically active, but as a consequence of an asymmetric electric

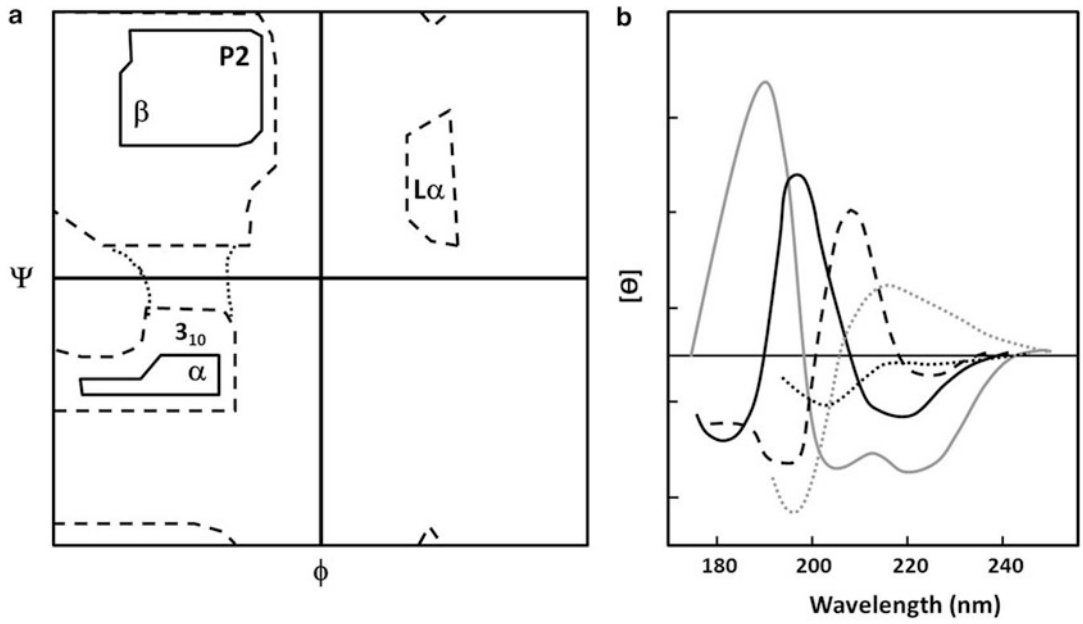

Fig. 3.3 (**a**) Ramachandran Plot with characteristic, defined φ and ψ angles, showing spaces for the α, β, 3_{10}-helix and poly(Pro) II helix structures; (**b**) Far UV CD-spectra associated with various types of secondary structure. *Solid grey line* – α-helix; *solid black line* – antiparallel β-sheet; *long dashed line* – type I β-turn; *dotted grey line* – extended 3_{10}-helix or poly(Pro) II helix; *dotted black line* – irregular structure

field built by the protein environment distorting the electron displacement, it induces a strong optical activity. In proteins all amino acids with the sole exception of Gly are asymmetric in their C_α and, as a consequence, optically active.

The use of CD spectroscopy in protein and peptide characterization deals with:

1. conformational changes under different chemical-physical conditions like pH, temperature, solvent environment, ionic strength;
2. quantitative determination of secondary structure;
3. quantitative analysis of thermodynamical parameters in protein/peptide folding;
4. tertiary structure changes (generally local) between wild type and mutant proteins;
5. binding studies related to conformational changes;
6. conformational changes kinetics (ms timescales).

In an actual CD experimental measurement [47], the sample is alternately irradiated by a right and left rotating polarized light and data are generally represented by two different units: the differential absorbance (ΔA) correlated to the molar differential extinction coefficient Δε (M^{-1} cm^{-1}), and the mean residue ellipticity [θ], measured in degrees. As the absorption is monitored in a range of wavelengths the CD-spectra is a plot of [θ] or Δε *vs.* wavelengths, showing a characteristic shape depending on the respective secondary structure present (Fig. 3.3).

The Δε values can be directly deduced by considering the Lambert-Beer relation, Eq. 3.1:

$$A = \varepsilon C l \tag{3.1}$$

where ε = extinction coefficient,
C = sample concentration,
l = length of the measuring cell.

Due to the chirality of the molecular sample or the environment itself, a differential absorption ΔA between the left and right circularly polarized light result, Eq. 3.2:

$$\Delta A = A_L - A_R = (\varepsilon_L - \varepsilon_R)\, C l \tag{3.2}$$

where formula symbols are the same as in Eq. 3.1 with the subscript L and R referred to left or right.

Finally from Eq. 3.2 the molar differential extinction coefficient $\Delta\varepsilon$ can be deduced, Eq. 3.3:

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{(A_L - A_R)}{Cl} \tag{3.3}$$

When considering the ellipticity unit, the mean residue ellipticity $[\theta]_{\mathrm{MRW},\lambda}$ (deg cm^2 dmol^{-1}) at a determined wavelength λ is given by Eq. 3.4:

$$[\theta]_{(MRW,\lambda)} = \frac{MRW * \theta_{(\lambda)}}{10lC} \tag{3.4}$$

where θ is the measured ellipticity in degrees, l the cell path-length in cm, C the concentration in g/ml and MRW is the Mean Residue Weight as a result of the data normalization where in case of polymers like proteins is selected the repeating unit, which is in case of proteins the peptide bond.

The peptide bond MRW is equal to $MRW = \frac{M}{N-1}$, where M is the molecular mass of the protein in Da, N is the number of amino acids and $N-1$ the number of peptide bonds. For most proteins the MRW is 110 ± 5 Da.

When referring to the protein molar concentration (M) the molar ellipticity at a determined wavelength λ is, Eq. 3.5:

$$[\theta]_{(M,\lambda)} = \frac{100 * \theta_{(\lambda)}}{lM} \tag{3.5}$$

where $[\theta]_\lambda$ and l have the same meaning as in the previous formula.

The units for both the mean residue ellipticity and molar ellipticity are deg cm^2 dmol^{-1}.

The relationship between $[\theta]$ and $\Delta\varepsilon$ is given in Eq. 3.6

$$[\theta]_{(MRW)} = 3,298 * \Delta\varepsilon \tag{3.6}$$

It should be underlined that in most of the biological studies the detected CD signals are extremely small with the measured ellipticities on the scale of ~10 mdeg, corresponding to tiny differences in absorbance (ΔA) of approximately $3*10^{-4}$.

Proteins can be well defined as heteropolymers made by secondary and tertiary amides when proline is involved. Their CD-spectra is a function of the extent of their secondary structure and the respective secondary structure content.

As a consequence there is a correspondence between the secondary structure and the CD-spectra. In recent years [48] the Ramachandran plot, showing the correlation between secondary structures and the dihedral space ϕ and φ has been expanded including additional structures like P2 (poly(Pro)II type structure; see below). Such new structures posed some further challenges to the definition of an overall CD-spectra of a protein based on single pure structures.

In general α-helices and β-sheets are the most important secondary structure elements in proteins and they are stabilized by intra- (α-helices) and inter-chain hydrogen bonds (β-sheet) (For more details check Chap. 2). The Ramachandran dihedral angles φ and ψ for the α are defined as $-57°$, $-47°$ and for the β as $-120°$, $+120°$ of each of the amide bonds throughout the polypeptide chain, forming a well ordered and oriented structure toward the direction of chain propagation. Amide groups in α-helices build a cylindrical surface with intra-chain hydrogen bonds parallel to the helix axis while those in the β-sheets build a planar surface. β-sheets exist in two different forms with the polypeptide strands organized in a parallel or antiparallel fashion. A further important secondary structure is the β-turn, built by at least three residues and stabilized by a hydrogen bond between the first and the third amide group reversing the direction of chain propagation.

Projecting the β-turns in the Ramachandran space they are characterized by different dihedral angles for the successive $i+1$ and $i+2$ residues. An important secondary structure with backbone-backbone dihedral angles defined within the zone (φ, ψ: $-78°$, $150°$) present in proteins to some limited extent (~5 %) is the poly(Pro)II type structure (P2).

This classification originates from proline-rich proteins, such as collagen and is a result of the constraints posed by the proline side chain that can be defined as a cyclic imino acid. However P2 structures built by Pro and non-Pro residues do exist in proteins.

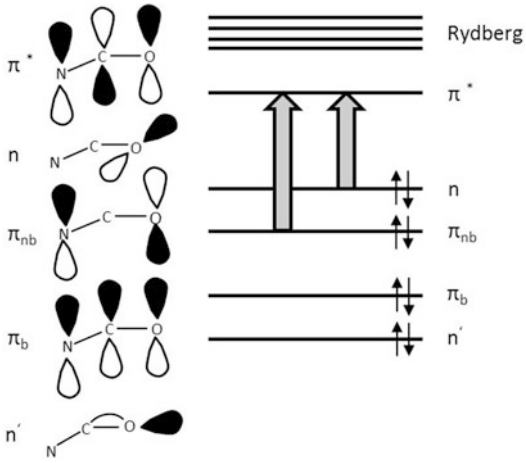

Fig. 3.4 Molecular orbitals representation of the amide group reporting the bonding, nonbonding and antibonding π orbitals (π_b, π_{nb} and π*) and two lone pairs localized on the oxygen atom (n and n′). Electronic transitions are in the far-UV region

Amino acid residues that do not form any of the aforementioned well defined and ordered secondary structures fall under the definition of "random coils". Originally the P2 conformation was not explicitly identified and categorized as random coil.

In Fig. 3.3a is reported a Ramachandran plot of the secondary structures with a characteristic and well defined φ and ψ angles space with the α, β and random coil structures leading to characteristic CD-spectra for each of the structural elements (Fig. 3.3b).

The theoretical basis of protein or peptide CD-spectra [49] is funded on the amide or peptide bond chromophore electronic transitions which dominate the CD-spectra under 250 nm leaving two typical footprints originated by a $n \rightarrow \pi^*$ $(215 < \lambda < 230$ nm) and $\pi_0 \rightarrow \pi^*$ $(185 < \lambda < 200$ nm) transition, Fig. 3.4.

The $n \rightarrow \pi^*$ transition is electrically prohibited and magnetically permitted and it is the main responsible for the negative bands shown in the α-helix spectrum at 222 nm, and the same blue-shifted at 216–218 nm to characterize the β-sheet spectrum.

The second transition $\pi_0 \rightarrow \pi^*$ is primarily responsible for one positive band centered at 193 nm coupled as well as a negative band at 208 nm both characterizing the α-helix spectrum, while in a β-sheet spectrum the $\pi_0 \rightarrow \pi^*$ shows only one negative band at 198 nm. Depending on the local protein folding, each transition is differently perturbed resulting in a characteristic footprint, Fig. 3.3b.

Specifically β-rich proteins exhibit two kinds of CD-spectra; a first one showing characteristics reminiscent of the β-sheets with a positive $\pi \rightarrow \pi^*$ band centered at ~195 nm followed by a negative $n \rightarrow \pi^*$ band at ~215 nm and a second one characteristic of unordered polypeptides, reporting a negative $\pi \rightarrow \pi^*$ band centered at ~200 nm, resulting in a classification of βI and βII proteins [50].

Other chromophores than peptide bonds contributing to the protein CD-spectra include aromatic amino acid side chains, characterized by an absorption range of 260–320 nm and S–S bonds with a weak broad absorption band centered at 260 nm.

Each single side chain of the aromatic amino acids as for the Cys report a defined spectrum depending on its mobility and further perturbed by the local characteristics of the environment like the presence of hydrogen-bonding, nearby polar groups and media polarizability [51, 52]:

1. Phe shows a sharp fine structure in the range 255–270 nm with peaks close to 262 and 268 nm ($\Delta\varepsilon_M \pm 0.3\ M^{-1}\ cm^{-1}$);
2. Tyr has a maximum in the range 275–282 ($\Delta\varepsilon_M \pm 2\ M^{-1}\ cm^{-1}$) with a possible shoulder ~6 nm to the red;
3. Trp shows a fine structure above 280 nm in the form of two ^1L_b bands (one at 288–293 and one some 7 nm to the blue, with the same sign ($\Delta\varepsilon_M \pm 5\ M^{-1}\ cm^{-1}$)) and a ^1L_a band (around 265 nm) with little fine structure ($\Delta\varepsilon_M \pm 2.5\ M^{-1}\ cm^{-1}$).
4. Cys CD starts at long wavelength in the blue region (>320 nm) reporting one or two broad peaks > 240 nm ($\Delta\varepsilon_M \pm 1\ M^{-1}\ cm^{-1}$) with the long wavelength peak often negative.

In proteins including cofactors or prosthetic groups like Fe based heme groups (reporting a strong band at 410 nm and a wide range of bands between 350 and 650 nm function of the Fe spin and oxidation states [53]),

pyridoxal-5V-phosphate (330 nm) and flavins (300–500 nm being function of the oxidation states) their strong spectral signatures can be used to follow the secondary and tertiary structure changes.

Though the CD technique is mainly a qualitative spectroscopy to measure the secondary structure content due to inherent difficulties in the calculation of the theoretical spectra (see Sect. 3.2.2), the tertiary structural changes can be monitored by following the CD signals of the aromatic amino acids [54].

Such analyses are based generally on a comparison of the differences in the aromatic CD fingerprints between the wild type and mutants of the same protein. Another phenomenon that can be analyzed by following the aromatic amino acid CD signal is the presence of a "molten globule" state in the protein, which arises from the changed mobility of the aromatic side chains.

It must be underlined that in a "classical" CD experiment, the obtained spectra are a product of an averaged measurement performed on a large ensemble of molecules in thermodynamic equilibrium, furthermore recorded in time intervals where interesting structural transitions can occur.

As a consequence to further push the CD technique toward a more sophisticated spectroscopy pursuing the target of obtaining protein structure quantitative data, the use of Synchrotron Radiation in Circular Dichroism (SRCD) has enlarged the CD horizon by shifting the absorption spectrum to lower wavelength $\leq$180 nm, reaching a working range down to 160–120 nm [55]. Spectral information originating from these lower wavelengths includes additional electronic transitions allowing for a unique characterization of the single secondary structures contributing to a protein spectrum [56].

Furthermore some other factors like the high signal to noise ratio of the synchrotron radiation beam, open the possibility to use small amount samples and expand the CD toward Time Resolved CD spectroscopy (TRCD) [57]. As an example due to its sensitivity and high ratio of signal to noise SRCD has been successfully used to put in evidence the very tiny differences of proteins in complexes as compared to the non-complexed proteins that are not detectable by classical CD experiments, as in the case of carboxypeptidase A in complex with its inhibitor latexin [58].

Many typologies of experimental alliances between CD spectroscopy and other experimental techniques are continuously developed. For example, pump probe laser experiments are able to follow conformational changes occurring in photolyzed carboxymyoglobin [59] or small time resolution to follow early events in peptide/protein folding on nanosecond time scales by T-jumps [60]. Though these techniques allow the use of CD spectroscopy as an insightful method to understand local or global conformational changes out of equilibrium, the complexity of the experimental apparatus and of the data analysis/interpretation to avoid undesirable artifacts [59] and the theoretical difficulty to calculate "*ab-initio*" CD-spectra (see Sect. 3.2.2) limits the main stream of CD spectroscopy application to the qualitative understanding of the conformational changes in secondary and, partially, in the tertiary structures of proteins.

To list a series of advantages and disadvantages, the *pro and the contra* of the CD technique can be categorized as:

Advantages

1. No molecular size limitations;
2. Operatively the experimental set-up is fast;
3. ms acquisition time at single wavelength measurements;
4. Low concentration samples can be analyzed;

Disadvantages

1. No atomic level information is given. As a consequence though it is possible to obtain the amount of secondary structure elements present in a protein or its difference from a previous sample, it is not possible to affirm which specific part of the protein is in a determined secondary structure;
2. The observed spectrum cannot be uniquely determined by a single representation of secondary structures due to the different conformational equilibrium ensemble of

proteins/peptides and to limitations in the theoretical evaluation of the spectra (see Sect. 3.2.3);
3. Tertiary structure can be only determined in cases where special chromophores are involved.

Though CD spectroscopy derived data lacks detailed structural information when compared to X-ray crystallography or NMR spectroscopy, it is an apparently user friendly and easy to handle procedure [61] and can be used for survey studies varying a series of physico-chemical parameters like temperature, ionic strength, solvent environment and pH.

For example a thermal, chemical and pH induced denaturation study of multimeric chick pea β-galactosidase has been conducted. Protein unfolding was monitored by CD spectroscopy completed by fluorescence and enzyme activity measurements [62]. Furthermore targeted protocols based on temperature variation have been developed to extract thermodynamical parameters of protein unfolding, binding interactions [63] or folding kinetics [64].

In conclusion up to now the applications of CD spectroscopy were based on the determination of structural changes of proteins/peptides in solution, complementing and completing other techniques such NMR and X-ray crystallography. However CD spectroscopy reveals its usefulness also in cases where peptides or proteins are embedded in liposomes or polymersomes. Studies on such more complex systems can be performed under specific circumstances *i.e.* avoiding light scattering and low peptide/protein concentrations and a more detailed account will be given in Sect. 3.2.3.

3.2.2 Extracting Secondary Structure Information: Data De-convolution and Algorithm Comparison

As mentioned before the number of membrane protein crystal structures is rather under-represented among the total number of structures (see Sect. 3.1.1). As a consequence CD spectroscopy plays an important role in the investigation of the secondary structure of membrane proteins in solution due to its fast experimental procedure and set-up (see Sect. 3.2.1).

Once the CD-spectra are obtained, two main procedures to analyze acquired data can be adopted: an "*ab-initio*" one where the CD-spectra are calculated based on the expected chemical structure by first principle electronic calculations or the spectrum itself is empirically de-convolved by a series of structural basis-sets contributing to the overall spectrum. The selection of the method, as described in Sect. 3.2.1 is mainly influenced by the dimensions of the sample.

When considering the "*ab-initio*" approach, in the last decades efficient electronic structure methods for Electronic Circular Dichroism (ECD) calculations considering molecules with approximately 10–20 atoms, including as examples helicenes, fullerenes, iso-schizozygane alkaloids, paracyclophanes, β-lactams, and transition metal complexes, are widely used opening the possibility to determine absolute configurations of chiral molecules [65].

From the theoretical point of view the calculation of rotational strengths of proteins by applying quantum chemical methods is, in principle, possible. However when applied to peptides or proteins where the number of atoms and consequently of involved chromophores increases to $>10^2$ atoms, the calculations of CD-spectra from first principles becomes nearly prohibitive. Such calculations would be extremely welcome especially considering that a protein sample is represented by an ensemble of conformations each giving a CD-spectra defined by small differences that can contribute to the final overall CD-spectra as measured, limited by natural broadening due to the Heisenberg uncertainty principle, un-resolved vibronic components and the coupling of the chromophore with its environment represented by other chromophores and the solvent [49].

For example the peptide bond, which is the principal chromophore whose spectrum is detected, exists in many conformations in the Ramachandran ϕ, ψ space function within the protein location and consequentially the

spectrum is a result of an average of the various conformation parameters.

Within the theoretical approach, the matrix method seems a promising one [66, 67]. The matrix method has its fundamentals in the coupled-oscillator and exciton models, by building a secular determinant based on the energy and interactions between the transitions on chromophores [49].

Such theoretical models are today well developed and on-line tools like DichroCalc can be applied to calculate CD-spectra of proteins [68] (http://comp.chem.nottingham.ac.uk/dichrocalc/) using the DichroCalc web interface for dichroism calculations.

If the calculated CD-spectra fail to reproduce the experimental data, the structural information contained in the spectra can be obtained by using empirical methods generally parameterized on a set of known protein structures obtained by crystallography or NMR methods.

The simplest approach to extract the secondary structure content from a CD spectroscopy derived data set can be based on the assumption of the spectrum to be composed of a linear combination of "pure" CD-spectra *i.e.*, "pure" α-helix, "pure" β-strand, "pure" P2 etc. each one statistically weighted by its relative abundance in the polypeptide conformation.

This hypothesis is described by a linear relation (Eq. 3.7):

$$[\theta]_{MRE} = \sum_1^n a_n [\theta]_n \tag{3.7}$$

where the Mean Residue Ellipticity (MRE) $[\theta]_{MRE}$ is the sum of the single contributions $[\theta]_n$ of α-helix, β-sheet, β-turn, random coil etc. weighted by the factor a_n defined as a fractional proportion.

However some major drawbacks affect the linear relation reported in Eq. 3.7:

1. The absence of standard reference CD-spectra related to "pure" secondary structures. In fact the synthetic homo-polypeptides that would be the basic set used to build the reference spectra are in general poor performing models for the secondary structures found in proteins, *i.e.* influence of the residues as additional chromophore contributors to the CD-spectra and no homo-polypeptide has been found to be a good representation of β-sheets;
2. the CD signal of an α-helix is length dependent and sensitive to other neighborhood α-helices with their dipoles in parallel or in antiparallel fashion to the considered α-helix dipole.

In general the fitting methods give the most accurate results when helical secondary structures are considered. To report some of the reasons for this phenomenon, an α-helical structure tends to be highly ordered possessing a restricted Ramachandran dihedrals space thus producing very similar spectra and the CD signal intensity of α-helical segments is very strong. However β-sheet structures possess a wider dihedral angle space in the Ramachandran plot with the existence of parallel and antiparallel orientations of adjacent strands and different twists, varying considerably in their CD-spectra [69] with a lower intensity resulting in negative peaks of only about one-third of the size as compared to the CD-spectra minima caused by α-helical structures.

As a consequence when a protein contains a large amount of α-helix and only a small β-sheet content, the spectral contribution of the latter may be overcome and consequently the accuracy of the derived sheet content is considerably lower, giving a false low de-convolution derived result.

As previously reported in Sect. 3.2.1 the inclusion of the far UV light that is able to analyze further electronic transitions of the peptide bond can partially solve this problem because the very low wavelength data for α-helices and β-sheets have opposite signs [57]. Moreover the inclusion of the high-energy transitions in the ultraviolet CD-spectra of polypeptides and proteins increase the possibility to calculate and determine better poly(Pro) II (P(II)) type of conformations [70, 71].

A further problem originates from protein structure turns that though they show distinct spectra together with less represented secondary structures like 3_{10}-helices, are under-represented in the databases used for de-convolution. The number of examples for a given secondary structure available in any given reference

database may limit the accuracy of de-convolution methods.

Nowadays a common solution to these limitations has been obtained by using a database of reference protein CD-spectra containing known amounts of secondary structure, giving as results a series of methods generally more accurate and reliable than the simple model previously mentioned using only pure secondary structure elements. The new methods are often coupled with highly sophisticated analysis algorithms and the wider the set of secondary structures included in the reference databases, the more accurate will be the obtained result, independently of the empirical analysis method used.

A series of sophisticated algorithms has been developed such as singular value de-convolutions [72], parameterized fits [73], self-consistency [74], convex constraints [75], matrix descriptor [76] and neural networks [77, 78]. Further developments [79] include new computational tools such as support vector machines, simultaneous partial least squares (SIMPLS), principal component regressions, or combinations of several such methods to improve and extend the analyses [80].

Based on this research philosophy the Protein Circular Dichroism Data Bank (PCDDB) sets a freely available database of CD-spectra [81, 82] (http://pcddb.cryst.bbk.ac.uk/home.php) and the web interface DichroWeb (http://dichroweb.cryst.bbk.ac.uk/home/shtml) [80] gives the possibility to analyze obtained CD-spectra by using several open source fitting algorithms like SELCON3 [74, 83], CONTINLL [84, 85], CDSSTR [72, 86, 87], VARSLC [72, 86, 87] and K2D117 [78].

To check the reliability of the output analysis of the secondary structure analysis by CD several parameters should be considered [61]:

1. The fitting parameter NRMSD (Normalised Root Mean Square Deviation) is the best indicator on how well the calculated CD-spectra matches with the experimental data.

   The NRMSD parameter is defined by relation 3.8:

$$NMRSD = \sqrt{\left(\frac{\sum \left(\theta_{exp} - \theta_{calc}\right)^2}{\sum \left(\theta_{exp}\right)^2} \right)} \quad (3.8)$$

   leading to values between 0 (perfect fit) and 1 (no fit). A NRMSD above 0.25 is generally considered as a strong error in the analysis procedure while values of less than 0.1 or lower are aimed at;
2. A good fitting program should always provide the R value reporting the measure on how appropriate is the secondary structure composition derived from CD when compared to its X-ray structure, *i.e.* knowing "*a-priori*" the protein structure. The R value is defined as the sum of all differences between the content fractions of the different secondary structure elements (α-helix, β-sheet, and turns) derived from CD and X-ray analysis of an analyzed protein. Low R values (<0.1) are an indicator for a successful analysis. Considering the R value helps in selecting which one of the several possible algorithms is the most appropriate together with the reference dataset to be used and this consideration should be made for each single spectrum to be de-convoluted;
3. A visual inspection of the calculated and experimental spectra should always be performed in order to check for the presence of systematic differences. In case of membrane protein CD data analysis using a reference dataset of soluble proteins a consistent small wavelength shift in the absorption maxima has been reported. This shift is due to the differences in polarity between the aqueous environment of the reference protein set and the environment of a lipid bilayer or detergent surrounding the analyzed membrane proteins [88];
4. The output of different algorithms should be always compared. Though not an absolute rule the output of a certain de-convolution tool (using a certain algorithm) can be considered reliable if a similarity is obtained with other algorithm outputs.

Furthermore the use of larger reference protein datasets containing both soluble and membrane proteins always leads to improvements in the accuracy of CD spectroscopy derived protein secondary structure estimation. When considering β-barrel OMPs it is extremely important to have a reference set including a wide ensemble of spectral and structural variation of reference proteins. The lack of a membrane proteins-only reference set can cause some problem when analyzing spectroscopic data from a membrane protein analysis.

In the last years a specific reference set for membrane proteins has been compiled to obtain a meaningful and reliable set of results based on CD-spectra of membrane proteins (data set MP180) [89].

Specifically the MP180 data set contains the spectra of 30 membrane proteins including the secondary structure and fold space covered by all known membrane protein structures. In addition a second reference set SMP180 has been created, which includes 98 soluble protein spectra on top of the MP180 included spectra.

The analysis of the CD-spectra including both membrane and soluble protein secondary structures using the SMP180 dataset gives a significant improvement over the use of reference sets that include only soluble protein CD information permitting the determination of the percentage of transmembrane residues, further enhancing and refining the information previously obtainable from CD spectroscopy.

3.2.3 Methodological Considerations: Detergent Solutions or Liposome/Polymersome Samples

Due to the insolubility of the β-barrel membrane proteins in aqueous solutions a solubilization using detergent or a reconstitution into a lipidic or polymeric membrane becomes necessary.

Such a procedure, however, imposes some measurement limits due to the presence of the light scattering phenomenon (see Sect. 6.1.2). In fact to obtain a clear and processable signal from a CD measurement, the solutions must be homogeneous in respect to the wavelength used in the analysis typically within a range of $185 \leq \lambda \leq 250$ nm. As a consequence if particles with a size comparable to the wavelength used in the CD-spectra detection are present in solution, a deterioration of the received signal will be observed. This is a problem well known in the polymersome or liposome nano-container technology, as these nano-containers are on the same size order of the wavelength used in the CD spectroscopy measurements.

In general the presence of light scattering in CD samples is detected by analyzing the wavelength range $240 \leq \lambda \leq 400$ nm. When a gradual increase of the absorbance is measured as the wavelength decreases from 400 to 310 nm, a light scattering phenomenon is present [61]. This absorbance increase can be balanced and corrected by plotting log absorbance over log wavelength, as the obtained line in the non-scattering region from 400 to 310 nm can be extrapolated toward the shorter wavelengths to determine the contribution from scattering to the measured absorbance at these wavelengths [90, 91].

In general when applied to measurements performed in solution containing nano-containers of various nature like liposomes or polymersomes, the best way is to minimize the signal deterioration due to scattering which is directly proportional to the number of nano-containers contained in the CD light path and to maximize the number of proteins reconstituted into the nano-containers membranes.

However though the scattering problem can be a serious issue, several measurements of membrane proteins or membrane inserting peptides have been successfully performed.

To avoid the scattering problem the general approach is to use only sonicated small unilamellar vesicles (SUVs) avoiding the extruded large unilamellar vesicles (LUVs) the latter are supposed to introduce a high scattering level in the signal. However in a recent work the Melittin peptide inserted into LUVs, the fractional helical content of Melittin in LUVs was determined [92].

Always by applying CD spectroscopy on LUVs, a test on the mechanism of antimicrobial,

cytolytic, and amphipathic cell-penetrating peptides in model membranes was set up and the Gibbs free energy of peptide binding to the lipid bilayer from the surface-associated state was derived [93], while following conformational transitions, the effect of the membrane physical properties of the pHLIP peptide during transmembrane helix formation was analyzed and followed by CD spectroscopy [94].

Applying CD spectroscopy to proteins embedded in membranes, differential CD spectroscopy was used to detect conformational changes that occur upon binding of the *Escherichia coli* RNase E catalytic domain to anionic liposomes calculating the binding constants [95].

Some novel methodology to estimate the CD-spectra distortion caused by light scattering when analyzing membrane-bound peptides and proteins with high content of α-helix has been recently developed [96].

Shifting to the polymersome case where, by definition a polymersome is an artificial vesicle with a membrane constituted of synthetic polymers instead of lipids as used in the case of liposomes, direct measurements of the change in the folding state of the Cecropin A peptide in polymersomes constituted by a synthetic triblock copolymer PIB_{1000}-PEG_{6000}-PIB_{1000} (PIB = polyisobutylene, PEG = polyethylene glycol) was analyzed by CD and fluorescence spectroscopy [97].

Further details on the polymersome system analysis will be given in Chap. 6.

3.3 NMR

The Nuclear Magnetic Resonance (NMR) is a dynamic high resolution structure determination technique [4]. Its ability lies in the possibility to probe the fast relaxation of NMR active nuclei after a perturbation by a magnetic field. While the NMR technique has been used since the 1940s to inquire on the structure and dynamics of organic molecules, the first attempt to apply NMR spectroscopy to protein molecules had been made by Kurt Wüthrich in 1980. Since then the protein NMR spectroscopy method has been under continuous evolution and a multitude of protein structures has been solved by NMR. As a consequence NMR (when applied to proteins) is a powerful complementary technique able to give a dynamic overview of the protein in its environment compared to the static picture given by X-ray crystallography, apart from the special cases where time resolved crystallography is used to "trap" intermediates in enzymatic reactions on the picosecond scale [5]. In general the advantage of the NMR technique when compared to other structural techniques is based on its ability to measure the protein conformational dynamics spanning over a wide range of time scales from picoseconds to hours [98] and the protein does not have to be in crystallized form.

In the following section the basics of protein NMR structural determination will be explained focusing on examples of the structural analysis of β-barrel membrane proteins by NMR.

3.3.1 State of the Art on β-Barrel Membrane Protein NMR

While the resolved crystal structure is derived essentially in solid state, the derived protein NMR structures represent the "real" liquid conditions experienced by the protein, though solid state NMR is a further powerful variant to obtain protein structural information.

For example liquid NMR is perfectly suitable for the analysis of intrinsically disordered proteins (IDPs), where the IDP conformational space can be mapped. As IDPs can be crystallized in spite of their extreme flexibility, but the IDPs disordered part cannot be resolved by X-ray diffraction [99].

There is a series of fundamental differences between solution NMR and solid state NMR, concerning sample preparation and pulse sequences used to obtain analyzable signals.

Specialized servers exist reporting the current number of membrane proteins analyzed by NMR methods (liquid and oriented and magic-spinning solid state) (http://www.drorlist.com/nmr/MPNMR.html) giving a total of 98 unique protein coordinates (accessed on 03-04-2013).

A typical NMR experiment, independent from if performed in liquid or in solid state, considers the insertion of the desired sample under an oriented, intense magnetic field. Such a magnetic field aligns the dipoles of all the sample spin nuclei (in biopolymers these are mainly represented by ^{1}H, ^{2}H, ^{13}C, ^{15}N and ^{31}P), splitting the energies of each spin-state of the atoms into resolved energy levels within the microwave range.

Due to the high number of atoms constituting a protein molecule ($>10^3$), the "classical" 1D spectra representation, *i.e.* chemical shift toward intensity used in the analysis of small organic compounds does not lead to sufficient resolution causing strong signal overlapping. To avoid this problem, multidimensional 2D, 3D and 4D NMR experiments have been developed opening the possibility to analyze complex spectra such as the ones derived for proteins.

For each of the nD NMR techniques have been developed a variety of different characteristic pulse protocols. In fact each NMR experiment is based on a time delayed sequence of radio frequency (RF) pulses and the timing, frequencies, and intensities of these pulses define the different NMR experiments [100, 101]. Furthermore to increase the resolution and to reduce the acquisition time or to render the analyzed protein more NMR active, isotopic labeling with ^{13}C (^{12}C is not NMR active) or ^{15}N (^{14}N has a quadrupolar moment preventing high resolution) can be helpful.

Several pulse protocols have been developed for the determination of protein structures. When considering the "homonuclear through-bond correlation method" where the magnetization transfer occurs by a J-coupling of atom *nuclei of the same type* connected by up to a few bonds, using a pulse sequence named correlation spectroscopy (COSY) [101] or total correlation spectroscopy (TOCSY) [102] to mention only a few.

The "heteronuclear through-bond correlation method" analyzes the magnetization transfer occurring between *nuclei of two different types* with pulse sequences like heteronuclear single-quantum correlation spectroscopy (HSQC) [101] and heteronuclear multiple-bond correlation spectroscopy (HMBC) [101].

When considering the "through-space correlation methods" such methods are able to correlate nuclei which are physically close to each other independently from whether there is a bond between them. These methods make use of the Nuclear Overhauser effect (NOE) [101] where nearby atoms within ~0.5 nm undergo cross relaxation by a mechanism related to the spin–lattice relaxation mechanism. The pulse sequences most used in this context are Nuclear Overhauser effect spectroscopy (NOESY) [103, 104] and the Rotating frame nuclear Overhauser effect spectroscopy (ROESY) [101].

Another methodology considers the "resolved-spectrum methods" where unlike correlated spectra, resolved spectra spread the peaks in a 1D-NMR experiment into two dimensions without adding any extra peaks [100, 101].

For even more complex pulse sequences and couplings used in the 3D and 4D NMR specialized literature can be referred to [105–107].

For a list of NMR acronyms, abbreviations, and terms the following website can be very useful: http://www.bmrb.wisc.edu/education/nmr_acronym.php?query_id= (accessed 07.04.2013).

Independent of the NMR methodology, solution or solid state, a four step protocol is always used: Preparation, Evolution, Mixing, and Detection [108].

1. Preparation step: In this step a magnetization coherence is induced by a set of RF pulses.
2. Evolution step: The evolution step is a length of time during which no RF pulses are irradiated on the sample and the nuclear spins are allowed to precess (rotate).
3. Mixing step: During the mixing step the nuclei coherence is perturbed by another series of pulses into a state that leads to an observable signal.
4. Detection step: During the detection the free decay of the signal obtained from the sample is analyzed as a function of time, as done in a conventional 1D FT-NMR (Puls-Fourier-Transformation-NMR).

A general procedure starting from isotope enrichments to the final structural information output is reported in Fig. 3.5.

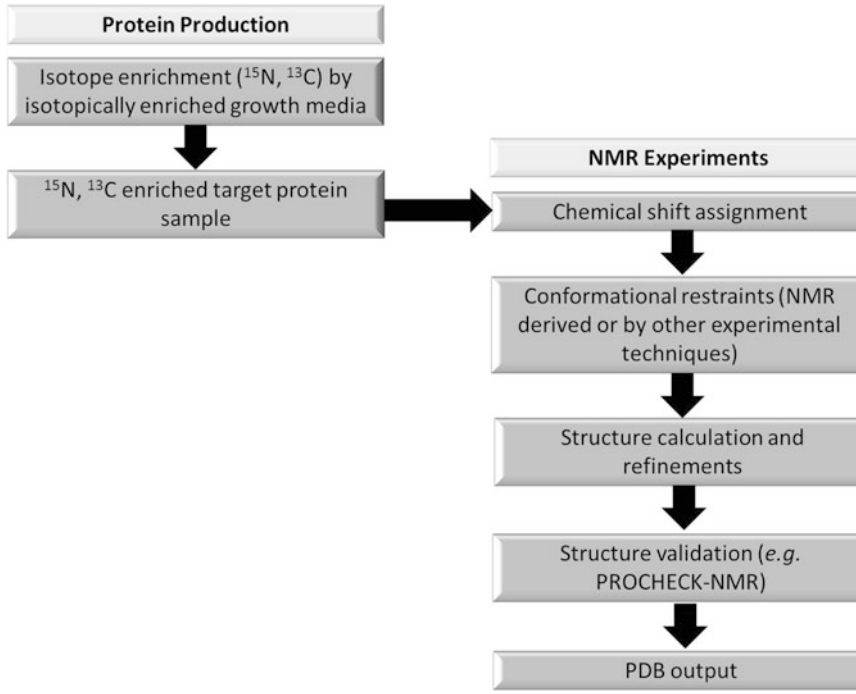

Fig. 3.5 Flow chart representing a typical NMR experiment starting from the protein expression, isotope enrichment and protein purification to the basic NMR steps including measurements, refinements, validation and to the final output

Due to the mentioned difficulties to obtain membrane protein structural information by X-Ray diffraction measurements of crystallized proteins, the use of liquid NMR and solid state NMR resulted in a fundamental contribution to and a broadening of the structural knowledge of this protein class.

However the main problem with both liquid and solid state NMR analysis of membrane proteins is related to the ability to reproduce the ideal environment such as the membrane of the original organism from which the protein has been extracted. Such a problem can seriously narrow the validity of data obtained by NMR techniques and the possibility to obtain any data at all.

Furthermore in liquid state a major obstacle to obtain high resolution spectra of proteins is related to the slow tumbling practically deleting the possibility to average out the anisotropic interactions resulting in broad resonances and consequently giving unusable spectra [109, 110].

Such a spin correlation time has been for decades a major limiting factor in NMR spectroscopy, affecting even more the membrane protein analysis by liquid NMR. In fact as most of the membrane proteins are insoluble in aqueous solutions and must be solubilized by use of detergents or lipid micelles, the aforementioned spin correlation time becomes a major obstacle due to the low tumbling frequencies of the huge objects in which the protein resides.

Several solutions have been developed based on isotope protein-labeling strategies [111, 112], pulse protocols [113], on reducing the solution volume by embedding the protein in micelles and reverse micelles [114, 115] or on increasing the tumbling frequency reducing the viscosity.

Considering the techniques dealing with the media for the embedding of the membrane proteins, the insertion of membrane proteins into bicellar phases has been well established [116] where bicellar phases are based on a mixture of phospholipids with detergents [117]. One of

the fundamental characteristics of such bicellar phases is their morphological variation as function of the detergent content. By increasing the detergent content it is possible to obtain multilamellar vesicles, multilamellar vesicles with toroidal pores lined up by detergents, extended lamellae with magnetic-alignment, magnetically alignable chiral nematic "worm-like" ribbons, isotropically tumbling flat disk-like aggregates and "pure" detergent micelles [116].

Lipid bicelles open up a space between the two model media based on detergent micelles and multilamellar vesicles reporting a quite high flexibility regarding their size with the small bicelles (also known as isotropic bicelles) small enough to tumble quickly on the NMR time scale and consequently used in solution NMR studies.

Larger bicelles, if macroscopically aligned, can be used under static solid-state NMR spectroscopy while magic angle spinning (MAS) NMR experiments (definition see below) can be applied to lipid bicelles [116]. Alternative non-micellar solubilization techniques using amphipoles [118] and lipid bilayer systems apart from bicelles, such as nano-lipoprotein particles (NLPs) have been considered also [119].

Amphipols are amphipathic polymers consisting of polymeric backbones that are covalently modified with a stochastic distribution of hydrophobic and hydrophilic groups [120] and have been successfully used showing conservation of the membrane protein function for proteins like the β-barrel outer membrane proteins OmpA [121], FomA [121] and OmpX [122].

Nano-lipoprotein particles (NLPs), also known as nano-discs or reconstituted high density lipoprotein particles (rHDLs) are considered membrane mimics providing a novel technique for studying membrane proteins in a native-like membrane environment. NLPs are made by a non-covalent assembly of phospholipids organized as a discoidal bilayer, surrounded by amphipathic apolipoproteins [119] (see Fig. 3.6). The NLP technology has been successfully explored to investigate a large number of membrane proteins by a battery of biophysical methods [123].

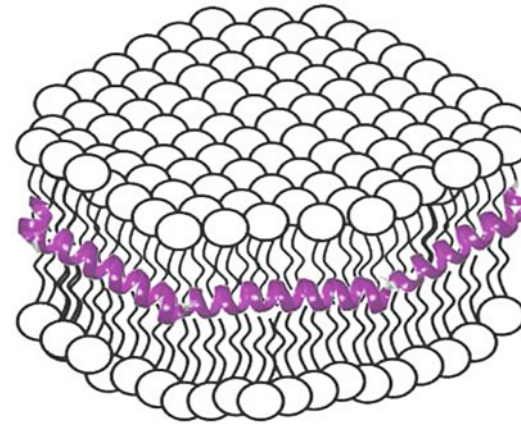

Fig. 3.6 Schematic representation of lipid nano-disc consisting of disc-shaped lipid bilayer part surrounded by amphipathic helical apolipoprotein

Nowadays the upper limit for "routinely" carried out NMR analysis experiments lies at ~100 kDa in molecular weight [124], this upper limit however has continually increased over the last years and is therefore expected to further increase in the future.

However some examples of high resolution NMR spectra of proteins up to 900 kDa have been published recently [125, 126].

The same techniques have been applied to resolve channel proteins such as the α-helical potassium channel KscA [127], the *Escherichia coli* outer membrane proteins OmpX [128] and OmpG [129], the influenza B proton channel [130] and ion channel fragments such as the voltage-sensing region of KvaP [131].

In parallel to the development of special technical protocols applied to solution NMR, a coupling of this method with other techniques permits to increase the level of sophistication.

For example recently [132] the composition and physico-chemical properties of detergent-solubilized integral membrane proteins has been solved primarily by the use of the micro-coil NMR technology and functional assays have been developed and applied to the β2-adrenergic receptor characterized by assigning ^{19}F-NMR. Another study [133] again using micro-coil NMR was able to monitor and determine the optimization conditions for NMR structure

determination of the OmpW β-barrel membrane protein in aqueous solution.

The same micro-coil NMR based technique has been successfully employed to determine the final composition of solutions of the integral membrane protein β(2)-adrenergic receptor (β(2)-AR), reconstituted with a detergent, defining the hydrodynamic properties of the mixed β(2)-AR/detergent/lipid micelles [134]. Such NMR translational diffusion coefficient studies can be used as qualitative checks for reproducible preparation for crystallization trials and solution NMR studies of G-protein coupled receptors.

Another important solution NMR application originates from the fast structure determination of human integral membrane proteins (hIMPs) by using systematically labeled proteins produced via cell-free expression [135] (For details on cell-free expression of membrane proteins and especially OMPs see Sect. 5.2.3) obtaining the backbone determination of six hIMPs solved in only 18 months from a pool of 15 initial targets.

Another recent study [136] aimed to overcome the problem of a suitable membrane mimicking system usable for NMR, made use of specific small diameter phospholipid nano-discs, *i.e.* particles consisting of two copies of apolipoprotein A-I (ApoA-I) wrapped around a patch of phospholipid bilayer. The combination of small nano-disc size and high protein deuteration levels coupled with the use of advanced non-uniform NMR sampling methods enabled to resolve and assign the NMR peaks obtaining a high resolution structural determination of the bacterial outer membrane protein OmpX.

Based on the same line of research trying to understand the interactions between the membrane proteins and detergents and lipids [137] the integral membrane enzyme diacylglycerol kinase (DAGK) has been analyzed in presence of three different micelles composed by lysomyristoylphosphatidylcholine, lysomyristoylphosphatidylglycerol and tetradecylphosphocholine. Liquid NMR measurements revealed significant differences in DAGK-detergent interactions involving lysomyristoylphosphatidylcholine micelles versus micelles composed of dodecylphosphocholine underlining how some integral membrane proteins can still behave as in their native environment when in lipid-free detergent micelles. And ^{14}C-based detergents have been found to be targets for the analysis of membrane proteins.

A coupling between NMR spectroscopy and analytical ultracentrifugation of membrane protein detergent complexes has been developed for the screening and selection of the extraction detergent, all these results particularly important for the finding of initial crystallization conditions [138].

Solution NMR has been also used to establish specificity of weak heterodimerization of membrane proteins [139].

Protein isotope enrichment [140] has been accomplished by using SPP (Single Protein Production), where membrane proteins are produced with high efficiency and assembled into appropriate membrane fractions, practically converting *E. coli* into a bioreactor producing only the targeted membrane protein. Proteins derived by such method are very valuable NMR study targets, due to the label induced signal enhancement (see above).

Apart from some exceptions previously reported, solution NMR poses as mentioned an upper limit of ∼100 kDa with a routinely reached base of ∼30 kDa to the molecular weight of the biomolecules to be analyzed.

To overcome the limit of 100 kDa, solid-state NMR (ssNMR) can be used as it does not require the sample to be soluble or able to form crystals [141, 142].

Though solid state NMR applied to biopolymers and in particular to membrane proteins is a quite young technique it is rapidly developing and has therefore great potential to further increase structural insights on membrane proteins [143].

One of the main challenges in a NMR experiment is to produce analyzable signals with peaks narrow enough to reach a suitable resolution eliminating overlapping lines. In fact while in solution NMR the spectra are characterized by well-defined sharp peaks due to the averaging out of the anisotropic (orientation-dependent) NMR interactions by rapid random tumbling, in ssNMR

spectra are very broad as the anisotropic effects are not averaged out as position of the atoms within the solid is "fixed".

As a consequence the ssNMR requests for a number of special techniques and hardware/software to extract the desired information from the sample under analysis. Techniques include magic-angle spinning, cross polarization, special 2D experiments and enhanced probe electronics.

The fundamental works of both E.R. Andrew [144] and I.J. Lowe [145] were able to show that anisotropic dipolar interactions can be suppressed by rotating the sample about an axis oriented at 54.74° with respect to the external magnetic field, today known as magic-angle spinning (MAS) technique. MAS enabled ssNMR to enter in its "golden age".

Other specific methods have been developed to minimize the anisotropic NMR interactions between nuclei increasing the signal to noise ratio in NMR spectra apart from MAS dilution however naturally occurring when considering ^{15}N or ^{13}C nuclei, such as multiple-pulse sequences that combined with rotation gives CRAMPS (combined rotation and multiple pulse spectroscopy) and cross polarization. Cross polarization can be combined with MAS and the polarization from abundant nuclei like ^{1}H, ^{19}F and ^{31}P can be transferred to dilute or rare nuclei like ^{13}C and ^{15}N enhancing the signal to noise ratio [146].

In the late years fundamental developments have been brought to the magic angle spinning solid-state NMR (MAS ssNMR) methods representing an important approach for studying membrane proteins of moderate size. Such methods have already been successfully used to characterize fibrils, globular proteins and membrane proteins embedded in lipids [146].

The environment into which a membrane proteins is embedded is also under ssNMR resolved by using different techniques, for example by applying different bilayer constructs mimicking a cellular membrane, including bicelles, multilamellar and unilammelar vesicles [117, 147] and solid supported membranes deposited on glass layers [148].

An alliance between solution NMR and ssNMR has been used to solve the structure of the *Klebsiella pneumonia* OmpA in DHPC (1,2-dihexanoyl-sn-glycero-3-phosphocholine) detergent micelles (liquid state NMR) and after reconstitution in lipid bilayers, its loop dynamic was assessed by solid state NMR and relaxation studies [149].

As for the solution NMR, ssNMR is coupled with other techniques to obtain the desired structural information. A nice work has been published, reporting crystallization trials that yielded only poorly diffracting microcrystals of the transmembrane domain *Yersinia enterocolitica* adhesin A (YadA). A single, uniformly ^{13}C- and ^{15}N-labeled sample was then used and ssNMR allowed to obtain information on some parts of the protein regarding structural flexibility and mobility [150].

To avoid time consuming protein extraction and purification steps, an interesting technique involves the direct application of "cellular solid-state NMR spectroscopy" on porins from *C. glutamicum* within the bacterial cell envelope. This method is based on the use of a combination of isotope labelling schemes, sample preparation routes and state-of-the-art multidimensional solid-state NMR experiments under MAS conditions at high magnetic field [151].

Another powerful alliance has been found by coupling data from X-ray diffraction analysis with ssNMR data on the example of the integral membrane protein complex DsbB-DsbA. In fact since crystals of biomolecular assemblies based on membrane proteins can diffract weakly and their dimensions likewise limits their study by NMR in solution due to the molecular tumbling, the parallel use of ssNMR restraints and X-ray reflections allows a refining of obtained structural information. In case of the DsbB-DsbA complex based on a 3.7 Å crystal structure the backbone precision could be improved by 0.92 Å in the transmembrane region by ssNMR. This means a 58 % enhancement as compared to the X-ray reflections alone [152].

Finally the so called Rotational Alignment (RA) ssNMR can be used as a general method for

structural determination of membrane proteins in phospholipid bilayers under physiological conditions. The technique relies on the ssNMR spectroscopy and residue-specific structural restraints for membrane proteins that undergo rotational diffusion around the membrane normal, but whose mobility is restricted by interactions with the membrane phospholipids [153].

NMR solved protein structures can be validated using structure validation program suites like PROCHECK-NMR [154].

In conclusion the rapid development of different techniques on a multi-level scale regarding molecular biology and biotechnological methods (expression and purification of membrane proteins, as will be discussed in Chap. 5) and structural characterization techniques such as X-Ray diffraction measurements of protein crystals, CD spectroscopy and NMR, opened the possibility to understand and increase the knowledge on membrane proteins. In the future membrane protein structural analysis methods will become more and more important, as they will promote the engineering and development of entirely novel proteins. As Chap. 5 will show such new proteins (especially when based on β-barrel OMPs) have great potentials for the development of bio-based nano-material components that can be combined with artificial membrane systems, while the following Chap. 4 will focus on theoretical considerations that can be used to validate a new protein engineering concept or to understand an existing system.

References

1. Bijvoet JM, Burgers WG, Hägg G (1969) Early papers on diffraction of X-ray by crystals. A. Oosthoek's Uitgevermaatschappij N.V., Utrecht
2. Schoenborn BP (2010) A history of neutrons in biology: the development of neutron protein crystallography at BNL and LANL. Acta Crystallogr D Biol Crystallogr 66:1262–1268
3. Bill RM, Henderson PJF, Iwata S, Kunji ERS, Michel H, Neutze R, Newstead S, Poolman B, Tate CG, Vogel H (2011) Overcoming barriers to membrane protein structure determination. Nat Biotechnol 29:335–340
4. Robertson JWF, Kasianowicz JJ, Banerjee S (2012) Analytical approaches for studying transporters, channels and porins. Chem Rev 112:6227–6249
5. Hajdu J, Neutze R, Sjögren T, Edman K, Szöke A, Wilmouth RC, Wilmot CM (2000) Analyzing protein functions in four dimensions. Nat Struct Biol 7:1006–11012
6. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 181:662–666
7. Perutz M, Rossmann M, Cullis A, Muirhead H, Will G, North A (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. Nature 185:416–422
8. Deisenhofer J, Epp O, Miki K, Huber R, Michel H (1984) X-ray structure analysis of a membrane protein complex. Electron density map at 3 A resolution and a model of the chromophores of the photosynthetic reaction center from Rhodopseudomonas viridis. J Mol Biol 180:385–398
9. Deisenhofer J, Epp O, Miki K, Huber R, Michel H (1985) Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3 Å resolution. Nature 318:618–624
10. Li D, Boland C, Walsh K, Caffrey M (2012) Use of a robot for high-throughput crystallization of membrane proteins in lipidic mesophases. J Vis Exp 67
11. Pieper U, Schlessinger A, Kloppmann E, Chang GA, Chou JJ, Dumont ME, Fox BG, Fromme P, Hendrickson WA, Malkowski MG, Rees DC, Stokes DL, Stowell MH, Wiener MC, Rost B, Stroud RM, Stevens RC, Sali A (2013) Coordinating the impact of structural genomics on the human α-helical transmembrane proteome. Nat Struct Mol Biol 20:135–138
12. Sciara G, Mancia F (2012) Highlights from recently determined structures of membrane proteins: a focus on channels and transporters. Curr Opin Struct Biol 22:476–481
13. Lebon G, Warne T, Tate CG (2012) Agonist-bound structures of G protein-coupled receptors. Curr Opin Struct Biol 22:482–490
14. Thøgersen L, Nissen P (2012) Flexible P-type ATPases interacting with the membrane. Curr Opin Struct Biol 22:491–499
15. Allen WJ, Phan G, Waksman G (2012) Pilus biogenesis at the outer membrane of Gram-negative bacterial pathogens. Curr Opin Struct Biol 22:500–506
16. Carpenter EP, Beis K, Cameron AD, Iwata S (2008) Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 18:581–586
17. Bergfors TM (2009) Protein crystallization, 2nd edn, UL biotechnology. International University Line, La Jolla
18. Newstead S, Ferrandon S, Iwata S (2008) Rationalizing alpha-helical membrane protein crystallization. Protein Sci 17:466–472

19. Simon SA (2009) Membrane protein crystallization, vol 63, Current topics in membranes. Elsevier, Amsterdam/Boston
20. Ubarretxena-Belandia I, Stokes DL (2012) Membrane protein structure determination by electron crystallography. Curr Opin Struct Biol 22:520–528
21. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta 1666:105–117
22. Iwata S (2003) Methods and results in crystallization of membrane proteins, IUL biotechnology series. International University Line, La Jolla
23. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242
24. van den Bedem H, Wolf G, Xu Q, Deacon AM (2011) Distributed structure determination at the JCSG. Acta Crystallogr D Biol Crystallogr 67:368–375
25. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221
26. Bayburt TH, Sligar SG (2010) Membrane protein assembly into nanodiscs. FEBS Lett 584:1721–1727
27. Schulz GE (2011) A new classification of membrane protein crystals. J Mol Biol 407:640–646
28. Rigaud J, Chami M, Lambert O, Levy D, Ranck J (2000) Use of detergents in two-dimensional crystallization of membrane proteins. Biochim Biophys Acta 1508:112–128
29. Caffrey M, Li D, Dukkipati A (2012) Membrane protein structure determination using crystallography and lipidic mesophases: recent advances and successes. Biochemistry 51:6266–6288
30. Beebe ETMS, Nozawa A, Matsubara Y, Frederick RO, Primm JG, Goren MA, Fox BG (2011) Robotic large-scale application of wheat cell-free translation to structural studies including membrane proteins. Nat Biotechnol 28:239–249
31. Goren MA, Fox BG (2008) Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. Protein Expr Purif 62:171–178
32. Clark KM, Fedoriw N, Robinson K, Connelly SM, Randles J, Malkowski MG, DeTitta GT, Dumont ME (2010) Purification of transmembrane proteins from Saccharomyces cerevisiae for X-ray crystallography. Protein Expr Purif 71:207–223
33. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R (2008) Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci USA 105:14371–14376
34. Newby ZE, O'Connell JD, Gruswitz F, Hays FA, Harries WE, Harwood IM, Ho JD, Lee JK, Savage DF, Miercke LJ, Stroud RM (2009) A general protocol for the crystallization of membrane proteins for X-ray structural investigation. Nat Protoc 4:619–637
35. Lesley SA, Wilson IA (2005) Protein production and crystallization at the joint center for structural genomics. J Struct Funct Genomics 6:71–79
36. Caffrey M, Cherezov V (2009) Crystallizing membrane proteins using lipidic mesophases. Nat Protoc 4:706–731
37. Wadsten P, Wöhri AB, Snijder A, Katona G, Gardiner AT, Cogdell RJ, Neutze R, Engström S (2006) Lipidic sponge phase crystallization of membrane proteins. J Mol Biol 364:44–53
38. Johansson LC, Arnlund D, White TA, Katona G, Deponte DP, Weierstall U, Doak RB, Shoeman RL, Lomb L, Malmerberg E, Davidsson J, Nass K, Liang M, Andreasson J, Aquila A, Bajt S, Barthelmess M, Barty A, Bogan MJ, Bostedt C, Bozek JD, Caleman C, Coffee R, Coppola N, Ekeberg T, Epp SW, Erk B, Fleckenstein H, Foucar L, Graafsma H, Gumprecht L, Hajdu J, Hampton CY, Hartmann R, Hartmann A, Hauser G, Hirsemann H, Holl P, Hunter MS, Kassemeyer S, Kimmel N, Kirian RA, Maia FR, Marchesini S, Martin AV, Reich C, Rolles D, Rudek B, Rudenko A, Schlichting I, Schulz J, Seibert MM, Sierra RG, Soltau H, Starodub D, Stellato F, Stern S, Strüder L, Timneanu N, Ullrich J, Wahlgren WY, Wang X, Weidenspointner G, Wunderer C, Fromme P, Chapman HN, Spence JC, Neutze R (2012) Lipidic phase membrane protein serial femtosecond crystallography. Nat Methods 9:263–265
39. Hunter MS, Fromme P (2011) Toward structure determination using membrane-protein nanocrystals and microcrystals. Methods 55:387–404
40. Tanabe M, Iverson TM (2009) A practical guide to X-ray crystallography of beta-barrel membrane proteins: expression, purification, detergent selection, and crystallization. Membr Protein Cryst 63:229–267
41. Newstead S, Hobbs J, Jordan D, Carpenter E, Iwata S (2008) Insights into outer membrane protein crystallization. Mol Membr Biol 25:631–638
42. Ujwal R, Cascio D, Colletier JP, Faham S, Zhang J, Toro L, Ping P, Abramson J (2008) The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating. Proc Natl Acad Sci USA 105:17742–17747
43. Albrecht R, Zeth K (2011) Structural basis of outer membrane protein biogenesis in bacteria. J Biol Chem 286:27792–27803
44. Hye-Jeong Y (2013) Production and crystallization of bacterial Type V secretion proteins. In: Delcour AH (ed) Bacterial cell surfaces: methods and protocols, vol 966, Methods in molecular biology. Humana Press, New York, pp 205–222
45. Yeo HJ (2013) Production and crystallization of bacterial type V secretion proteins. Methods Mol Biol 966:205–222

46. Rogers DS (2012) Circular dichroism: theory and spectroscopy biochemistry research trends chemical engineering methods and technology. Nova Science, Hauppauge
47. Cantor CR, Schimmel PR (1980) Biophysical chemistry part II: techniques for the study of biological structure and function. W. H. Freeman, San Francisco
48. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 50:437–450
49. Bulheller BM, Rodger A, Hirst DJ (2007) Circular and linear dichroism of proteins. Phys Chem Chem Phys 9:2020–2035
50. Wu J, Yang JT, Wu C-SC (1992) β-II conformation of all-β proteins can be distinguished from unordered form by circular dichroism. Anal Biochem 200:359–364
51. Woody RW, Dunker AK (1996) Aromatic and cystine side-chain circular dichroism in proteins. In: Fasman GD (ed) Circular dichroism and the conformational analysis of biomolecules. Plenum Press, New York, pp 109–157
52. Strickland EH (1974) Aromatic contributions to circular dichroism spectra of proteins. CRC Crit Rev Biochem 2:113–175
53. Barteri M, Fioroni M, Gaudiano MC (1996) Oxidation of Fe(II) horse heart cytochrome c by ultrasound waves. Biochim Biophys Acta 1296:35–40
54. Greenfield NJ (2004) Circular dichroism analysis for protein-protein interactions. Methods Mol Biol 261:55–78
55. Wallace BN, Janes RW (2009) Modern techniques for circular dichroism and synchrotron radiation circular dichroism spectroscopy, vol 1, Advances in biomedical spectroscopy. IOS Press, Amsterdam
56. Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. Q Rev Biophys 42:317–370
57. Miles AJ, Wallace BA (2006) Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. Chem Soc Rev 35:39–51
58. Cowieson NP, Miles AJ, Robin G, Forwood JK, Kobe B, Martin JL, Wallace BA (2008) Evaluating protein: protein complex formation using synchrotron radiation circular dichroism spectroscopy. Proteins 70:1142–1146
59. Hache F (2009) Application of time-resolved circular dichroism to the study of conformational changes in photochemical and photobiological processes. J Photochem Photobiol A 204:137–143
60. Hache F (2012) Ultrafast time-resolved circular dichroism in a pump-probe experiment. In: Rogers DS (ed) Circular dichroism: theory, spectroscopy and advances in nanoscience. Nova, New York
61. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751:119–139
62. Kishore D, Kundu S, Kayastha AM (2012) Thermal, chemical and pH induced denaturation of a multimeric β-galactosidase reveals multiple unfolding pathways. PLoS One 7:e50380
63. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 1:2527–2535
64. Greenfield NJ (2006) Analysis of the kinetics of folding of proteins and peptides using circular dichroism. Nat Protoc 1:2891–2899
65. Warnke I, Furche F (2012) Circular dichroism: electronic. Wiley Interdiscip Rev Comput Mol Sci 2:150–166
66. Bayley PM (1973) The analysis of circular dichroism of biomolecules. Prog Biophys Mol Biol 27:1–76
67. Bayley PM, Nielsen EB, Schellman JA (1969) The rotatory properties of molecules containing two peptide groups: theory. J Phys Chem 73:228–243
68. Bulheller BM, Hirst JD (2009) DichroCalc–circular and linear dichroism online. Bioinformatics 25:539–540
69. Wallace BA, Wien F, Miles AJ, Lees JG, Hoffman SV, Evans P, Wistow GJ, Slingsby C (2004) Biomedical applications of synchrotron radiation circular dichroism spectroscopy: identification of mutant proteins associated with disease and development of a reference database for fold motifs. Faraday Discuss 17:653–661
70. Woody RW (2009) Circular dichroism spectrum of peptides in the poly(Pro)II conformation. J Am Chem Soc 131:8234–8245
71. Woody RW (2010) A significant role for high-energy transitions in the ultraviolet circular dichroism spectra of polypeptides and proteins. Chirality 22:E22–E29
72. Compton LA, Johnson WC (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal Biochem 155:155–167
73. Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. J Biochem 20:33–37
74. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 209:32–44
75. Perczel A, Park K, Fasman GD (1992) Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. Anal Biochem 203:83–93
76. Pancoska P, Janota V, Keiderling TA (1999) Novel matrix descriptor for secondary structure segments in proteins: demonstration of predictability from circular dichroism spectra. Anal Biochem 267:72–83
77. Bohm G, Muhr R, Jaenicke R (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng 5:191–195

78. Andrade MA, Chacon P, Merelo JJ, Moran F (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng 6:383–390
79. Lees JG, Miles AJ, Janes RW, Wallace BA (2006) Novel methods for secondary structure determination using low wavelength (VUV) circular dichroism spectroscopic data. BMC Bioinformatics 7:507–517
80. Whitmore L, Wallace BA (2007) Protein secondary structure analyses from circular dichroism: methods and references databases. Biopolymers 89:392–400
81. Whitmore L, Woollett B, Miles AJ, Klose DP, Janes RW, Wallace BA (2011) PCDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata. Nucleic Acids Res 39:D480–D486
82. Wallace BA, Whitmore L, Janes RW (2006) The Protein Circular Dichroism Data Bank (PCDDB): a bioinformatics and spectroscopic resource. Proteins 62:1–3
83. Sreerama N, Venyaminov SY, Woody RW (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. Protein Sci 8:370–380
84. Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20:33–37
85. van Stokkum IH, Spoelder HJ, Bloemendal M, van Grondelle R, Groen FC (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem 191:110–118
86. Manavalan P, Johnson WC (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem 167:76–85
87. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287:252–260
88. Wallace BA, Lees J, Orry AJW, Lobley A, Jones RW (2003) Analyses of circular dichroism spectra of membrane proteins. Protein Sci 12:875–884
89. Abdul-Gader A, Miles AJ, Wallace BA (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. Bioinformatics 27:1630–1636
90. Wallace BA, Mao D (1984) Circular dichroism analyses of membrane proteins: an examination of differential light scattering and absorption flattening effects in large membrane vesicles and membrane sheets. Anal Biochem 142:317–328
91. Wallace BA, Teeters CL (1987) Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments. Biochemistry 26:65–70
92. Ladokhin AS, Fernández-Vidal M, White SH (2010) CD spectroscopy of peptides and proteins bound to large unilamellar vesicles. J Membr Biol 236:247–253
93. McKeown AN, Naro JL, Huskins LJ, Almeida PF (2011) A thermodynamic approach to the mechanism of cell-penetrating peptides in model membranes. Biochemistry 50:654–662
94. Barrera FN, Fendos J, Engelman DM (2012) Membrane physical properties influence transmembrane helix formation. Proc Natl Acad Sci USA 109:14422–14427
95. Murashko ON, Kaberdin VR, Lin-Chao S (2012) Membrane binding of Escherichia coli RNase E catalytic domain stabilizes protein structure and increases RNA substrate affinity. Proc Natl Acad Sci USA 109:7019–7024
96. Chakraborty H, Lentz BR (2012) A simple method for correction of circular dichroism spectra obtained from membrane-containing samples. Biochemistry 51:1005–1008
97. Muhammad N, Dworeck T, Schenk A, Shinde P, Fioroni M, Schwaneberg U (2012) Polymersome surface decoration by an EGFP fusion protein employing Cecropin A as peptide “anchor”. J Biotechnol 157:7
98. Palmer AG, Williams J, McDermott A (1996) Nuclear magnetic resonance studies of biopolymer dynamics. J Phys Chem 100:13293–13310
99. Ringkjøbing Jensen M, Ruigrok RWH, Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 23(3):426–435
100. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ, Rance M (2007) Protein NMR spectroscopy: principles and practice. Elsevier/Academic, Amsterdam/Boston
101. Keeler J (2010) Understanding NMR spectroscopy, 2nd edn. Wiley, Chichester
102. Griesinger C, Otting G, Wüthrich K, Ernst RR (1988) Clean-TOCSY for ^{1}H spin system identification in macromolecules. J Am Chem Soc 110:7870–7872
103. Kay LE, Clore G, Bax A, Gronenborn A (1990) Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. Science 249:411–414
104. Pervushin K, Wider G, Riek R, Wüthrich K (1999) The 3D NOESY-[1H,15N,1H]-ZQ-TROSY NMR experiment with diagonal peak suppression. Proc Natl Acad Sci USA 96:9607–9612
105. Diercks T, Truffault V, Coles M, Millet O (2010) Diagonal-free 3D/4D HN, HN-TROSY-NOESY-TROSY. J Am Chem Soc 132:2138–2139
106. Huber M, Böckmann A, Hiller S, Meier BH (2012) 4D solid-state NMR for protein structure determination. Phys Chem Chem Phys 14:5239–5246
107. Snyder DA, Zhang F, Brüschweiler R (2007) Covariance NMR in higherdimensions: application to

4D NOESY spectroscopy of proteins. J Biomol NMR 39:165–175
108. Wüthrich K (1995) NMR in structural biology. World Scientific, Singapore
109. Judge PJ, Watts A (2011) Recent contributions from solid-state NMR to the understanding of membrane protein structure and function. Curr Opin Chem Biol 15:690–695
110. Kang C, Li Q (2011) Solution NMR study of integral membrane proteins. Curr Opin Chem Biol 15:560–569
111. Verardi R, Traaseth NJ, Masterson LR, Vostrikov VV, Veglia G (2012) Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. Adv Exp Med Biol 992:35–62
112. Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat Protoc 1:749–754
113. Tian C, Karra MD, Ellis CD, Jacob J, Oxenoid K, Sönnichsen F, Sanders CR (2005) Membrane protein preparation for TROSY NMR screening. Methods Enzymol 394:321–334
114. Kielec JM, Valentine KG, Wand AJ (2010) A method for solution NMR structural studies of large integral membrane proteins: reverse micelle encapsulation. Biochim Biophys Acta 1798:150–160
115. Gong XM, Franzin CM, Thai K, Yu J, Marassi FM (2007) Nuclear magnetic resonance structural studies of membrane proteins in micelles and bilayers. Methods Mol Biol 400:515–529
116. Dürr UHN, Gildenberg M, Ramamoorthy A (2012) The magic of bicelles lights up membrane protein structure. Chem Rev 112:6054–6074
117. Sanders CR, Hare BJ, Howard KP, Prestegard JH (1994) Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules. Prog NMR Spectrosc 26:421–444
118. Zoonens M, Catoire LJ, Giusti F, Popot JL (2005) NMR study of a membrane protein in detergent-free aqueous solution. Proc Natl Acad Sci USA 102:8893–8898
119. Raschle T, Hiller S, Etzkorn M, Wagner G (2010) Non-micellar systems for solution NMR spectroscopy of membrane proteins. Curr Opin Struct Biol 20:471–479
120. Breyton C, Pucci B, Popot JL (2010) Amphipols and fluorinated surfactants: two alternatives to detergents for studying membrane proteins in vitro. Methods Mol Biol 601:219–245
121. Pocanschi CL, Dahmane T, Gohon Y, Rappaport F, Apell HJ, Kleinschmidt JH, Popot JL (2006) Amphipathic polymers: tools to fold integral membrane proteins to their active form. Biochemistry 45:13954–13961
122. Catoire LJ, Zoonens M, van Heijenoort C, Giusti F, Guittet E, Popot JL (2010) Solution NMR mapping of water-accessible residues in the transmembrane beta-barrel of OmpX. Eur Biophys J 39:623–630
123. Borch J, Hamann T (2009) The nanodisc: a novel tool for membrane protein studies. Biol Chem 390:805–814
124. Wang GS (2008) NMR of membrane-associated peptides and proteins. Curr Protein Pept Sci 9:50–69
125. Raschle T, Hiller S, Yu T-Y, Rice AJ, Walz T, Wagner G (2009) Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. J Am Chem Soc 131:17777–17779
126. Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL GroES complex. Nature 418:207–211
127. Yu L, Sun C, Song D, Shen J, Xu N, Gunasekera A, Hajduk PJ, Olejniczak EJ (2005) Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex. Biochemistry 44:15834–15841
128. Fernández C, Hilty C, Wider G, Güntert P, Wüthrich K (2004) NMR structure of the integral membrane protein OmpX. J Mol Biol 336:1211–1221
129. Liang B, Tamm LK (2007) Structure of outer membrane protein G by solution NMR spectroscopy. Proc Natl Acad Sci USA 104:16140–16145
130. Pielak RM, Chou JJ (2011) Influenza M2 proton channels. Biochim Biophys Acta 1808:522–529
131. Shenkarev ZO, Paramonov AS, Lyukmanova EN, Shingarova LN, Yakimov SA, Dubinnyi MA, Chupin VV, Kirpichnikov MP, Blommers MJ, Arseniev AS (2010) NMR structural and dynamical investigation of the isolated voltage-sensing domain of the potassium channel KvAP: implications for voltage gating. J Am Chem Soc 132:5630–5637
132. Liu JJ, Horst R, Katritch V, Stevens RC, Wüthrich K (2012) Biased signaling pathways in β2-adrenergic receptor characterized by 19F-NMR. Science 335:1106–1110
133. Stanczak P, Zhang Q, Horst R, Serrano P, Wuthrich K (2012) Micro-coil NMR to monitor optimization of the reconstitution conditions for the integral membrane protein OmpW in detergent micelles. J Biomol NMR 54:129–133
134. Horst R, Stanczak P, Stevens RC, Wüthrich K (2013) b2-Adrenergic receptor solutions for structural biology analyzed with microscale NMR diffusion measurements. Angew Chem Int Ed 52:331–335
135. Klammt C, Maslennikov I, Bayrhuber M, Eichmann C, Vajpai N, Chiu EJ, Blain KY, Esquivies L, Kwon JH, Balana B, Pieper U, Sali A, Slesinger PA, Kwiatkowski W, Riek R, Choe S (2012) Facile backbone structure determination of human membrane proteins by NMR spectroscopy. Nat Methods 9:834–839
136. Hagn F, Etzkorn M, Raschle T, Wagner G (2013) Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. J Am Chem Soc 135:1919–1925

137. Koehler J, Sulistijo ES, Sakakura M, Kim HJ, Ellis CD, Sanders CR (2010) Lysophospholipid micelles sustain the stability and catalytic activity of diacylglycerol kinase in the absence of lipids. Biochemistry 49:7089–7099
138. Maslennikov I, Kefala G, Johnson C, Riek R, Choe S, Kwiatkowski W (2007) NMR spectroscopic and analytical ultracentrifuge analysis of membrane protein detergent complexes. BMC Struct Biol 7:74
139. Zhuang T, Jap BK, Sanders CR (2011) Solution NMR approaches for establishing specificity of weak heterodimerization of membrane proteins. J Am Chem Soc 133:20571–20580
140. Vaiphei ST, Tang Y, Montelione GT, Inouye M (2011) The use of the condensed single protein production system for isotope-labeled outer membrane proteins, OmpA and OmpX in E. coli. Mol Biotechnol 47:205–210
141. Tycko R (2001) Biomolecular solid state NMR: advances in structural methodology and applications to peptide and protein fibrils. Annu Rev Phys Chem 52:575–606
142. Loquet A, Habenstein B, Lange A (2013) Structural investigations of molecular machines by solid-state NMR. Acc Chem Res (Epub ahead of print)
143. Luca S, Heise H, Baldus M (2003) High-resolution solid-state NMR applied to polypeptides and membrane proteins. Acc Chem Res 36:858–865
144. Andrew ER, Bradbury A, Eades RG (1958) Nuclear magnetic resonance spectra from a crystal rotated at high speed. Nature 182:1659
145. Lowe IJ (1959) Free induction decays of rotating solids. Phys Rev Lett 2:285–287
146. McDermott A (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. Annu Rev Biophys 38:385–403
147. Park SH, Berkamp S, Cook GA, Chan MK, Viadiu H, Opella SJ (2011) Nanodiscs versus macrodiscs for NMR of membrane proteins. Biochemistry 50:8983–8985
148. Marassi FM, Crowell KJ (2003) Hydration-optimized oriented phospholipid bilayer samples for solid-state NMR structural studies of membrane proteins. J Magn Reson 161:64–69
149. Iordanov I, Renault M, Reat V, Bosshart PD, Engel A, Saurel O, Milon A (2012) Dynamics of Klebsiella pneumoniae OmpA transmembrane domain: the four extracellular loops display restricted motion behavior in micelles and in lipid bilayers. Biochim Biophys Acta 1818:2344–2353
150. Shahid SA, Bardiaux B, Franks WT, Krabben L, Habeck M, van Rossum BJ, Linke D (2012) Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. Nat Methods 9:1212–1217
151. Renault M, Tommassen-van Boxtel R, Bos MP, Post JA, Tommassen J, Baldus M (2012) Cellular solid-state nuclear magnetic resonance spectroscopy. Proc Natl Acad Sci USA 109:4863–4868
152. Tang M, Sperling LJ, Berthold DA, Schwieters CD, Nesbitt AE, Niewkoop AJ, Gennis RB, Rienstra CM (2011) High resolution membrane protein structure by joint calculations with solid state NMR and X-ray experimental data. J Biomol NMR 51:227–233
153. Marassi FM, Das BB, Lu GJ, Nothnagel HJ, Park SH, Son WS, Tian Y, Opella SJ (2011) Structure determination of membrane proteins in five easy pieces. Methods 55:363–369
154. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

Theoretical Considerations and Computational Tools

4

This chapter will describe a multi-level approach to the problem of simulating a complex system as a membrane protein in its natural environment or in an artificial environment constituted by a polymer membrane. Due to the impracticality to fold a β-barrel protein "in-silico" due to the long refolding time (from seconds to hours), shortcuts must be taken and evaluated to obtain reasonable results. The chapter will show the benefits and limitations of different levels of resolution for biomolecular modeling. More concretely a review on the use of Quantum Mechanics, all atom Molecular Dynamics simulations and Coarse Grained simulations to face different structural, dynamical and functional features typical in β-barrel membrane protein based systems will be given. Structure prediction tools will be introduced in the context of membrane protein structure prediction underlining the problems of applying prediction methods to β-barrel proteins.

4.1 From MD to CG: A Multi-scale Approach

Throughout the present book will be pointed out the great potential of β-barrel membrane proteins as flexible tools to design a broad new family of nano-devices. The rich chemistry, great availability and enhanced selectivity of β-barrel proteins can be exploited to design highly selective nano-channels with a broad scope of applications in drug delivery systems or nano-sensor techniques. However it should be kept in mind that β-barrel membrane proteins are trans-membrane proteins showing a specific tertiary structure that forms a pore that can be selectively permeated by nutrients, water or ionic species. In nature the activity and function of these membrane proteins is highly modulated by the lipid composition of the membrane hosting the protein [1]. Thus their use in nano-technological applications will always imply their reconstitution into lipid or polymer membranes different from their natural environment. Additionally, for more sophisticated applications it will be necessary to introduce some specific functional groups to the interior of the channel or modify the geometry of the β-barrel membrane protein. These changes in the environment and engineering of the β-barrel membrane proteins may dramatically alter their structural stability, function and dynamics (for details on β-barrel engineering see Chap. 5).

The rational design of a nano-device of this kind starts from the smallest "piece" in this "puzzle", *i.e.* the atoms and individual molecules, following a bottom-up approach. In this sense *in silico* or computational design emerges as a key partner for the rational development of new β-barrel based nano-devices. Computational modeling may help explaining some key features that are not well understood and have a clear impact on the successful preparation and performance of this kind of system, *i.e.* the self-assembly process, the interaction between the β-barrel and its new synthetic environment and the impact

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_4,

of new functionalities introduced to the β-barrel as well as other geometric modifications on the structure and dynamics of the engineered channel protein. In this sense *in silico* or computational design emerges as a key partner for the rational development of new nano-devices such as those focused in this book.

In the following will be described a multi-level approach to the problem of simulating a complex system such as a membrane protein in its natural environment or in an artificial environment constituted by a polymer membrane. Due to the impracticality to fold a β-barrel protein "*in-silico*" due to the long refolding time (from seconds to hours), shortcuts must be taken and evaluated to obtain reasonable results.

4.1.1 Why a Multi-scale Approach?

The increasing importance of computation in chemistry, biotechnology and biology goes hand in hand with the continuous growing of hardware and the availability of more powerful and faster computers. Since the first simulation [2] of a real protein in 1976 computational chemistry has become an essential tool in protein design, polymer science, nano-science, drug design or drug delivery systems design [3]. In present days, computational chemistry/biology is considered an essential tool to complement and interpret experimental measurements as well as to provide data that is difficult, expensive or simply impossible to access experimentally. However some preliminary considerations must be taken into account before facing a new computational modeling problem:

1. What is the length scale or size of the system of interest?
2. What is the characteristic time scale of the process of interest?
3. What are the degrees of freedom or level of resolution needed?

As a general rule the *bigger* is the system, the *longer* is the simulation time and the *smaller* is the level of resolution the *more expensive* is the simulation in terms of CPU (central processing unit) time and hardware resources. These three questions must be answered individually for the specific system and process of interest. Figure 4.1 summarizes the different time and length scales achievable for different levels of resolution and computational methods.

4.1.2 Quantum Mechanics

Quantum Mechanics (QM) offers the most accurate level of description of the matter as it describes its fundamental behavior at the atomic and molecular level. QM methods are used in practice for systems involving only few dozens of atoms and in systems where an accurate description of the electronic structure is needed. QM deals with the motion of electrons under the influence of the electromagnetic force exerted by nuclear charges and requires solving the Schrödinger equation (Eq. 4.1), in which electrons are considered as wave-like particles, using some approximations mostly *ab initio* or Density Functional Theory (DFT) methods:

$$\widehat{H}\psi = E\,\psi \tag{4.1}$$

In the last equation $\widehat{H}$ is the so-called Hamiltonian operator containing the kinetic and potential energy of the nuclei and electrons, E is the energy of the system and ψ is the wave function that describes the molecular orbitals.

4.1.2.1 The Hamiltonian Operator

The general expression of the Hamiltonian operator for a system with M nuclei and N electrons described respectively by position vectors R_A and r_i is given in Eq. 4.2.

$$\widehat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{4.2}$$

where M_A is the relation of the mass of the nucleus A with respect to the mass of the electron,

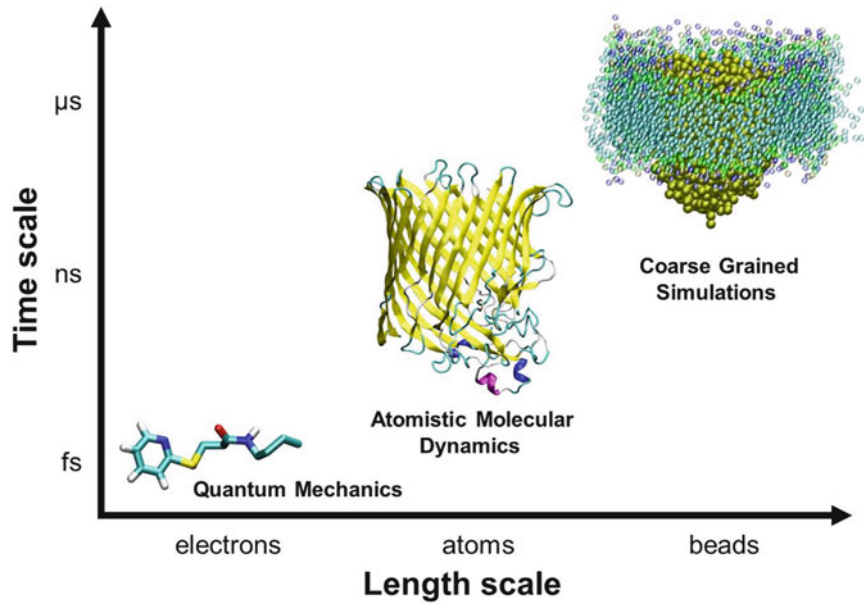

Fig. 4.1 Characteristic time and length scales achievable by Quantum Mechanics (QM), all atom Molecular Dynamics (MD) and Coarse Grained (CG) simulations

Z_A is the atomic number of the nucleus A and ∇_i^2 and ∇_A^2 are the Laplace operators referred to the differentiation between the coordinates of the electron i and nucleus A respectively. In Eq. 4.2 the first two terms are related to the kinetic energy of the electrons and nuclei respectively, the third term defines the electrostatic attraction between nuclei and electrons and the last two terms are the electrostatic repulsion between electrons and nuclei. The Born-Oppenheimer approximation states that nuclei in a molecule are stationary with respect to the electrons because of their higher mass. Consequently the second and fifth terms of Eq. 4.2 can be neglected leading to the so-called electronic Hamiltonian (Eq. 4.3).

$$\widehat{H}_{el} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N} \frac{1}{r_{ij}} \tag{4.3}$$

4.1.2.2 Basis Sets

As previously stated electrons are considered as wave-like particles, being ψ the wave function that describes the molecular orbitals. Molecular orbitals can be represented as linear combinations of m basis functions (Eq. 4.4) following the Linear Combination of Atomic Orbitals (LCAO) approach [4, 5].

$$\psi_i = \sum_{s=1}^{m} c_{si}\varphi_s \tag{4.4}$$

being ψ_i the ith molecular orbital, φ_s the sth basis function and c_{si} weighting coefficients that must be adjusted to get the best molecular orbital. Basis functions can be either atomic orbitals or any set of mathematical functions whose linear combination yields useful representations of the molecular orbitals. The linear combination of basis functions used to generate molecular orbitals is called basis set.

Slater functions (Eq. 4.5) describe very accurately atomic wave functions, but large computer time is required to evaluate them. To notably save computer time, Gaussian functions (Eq. 4.6) may be used instead.

$$\varphi = a \cdot \exp(-br) \tag{4.5}$$

$$\varphi = a \cdot \exp\left(-br^2\right) \tag{4.6}$$

The total wave function ψ is then expressed as a Slater determinant of spin orbitals α and β which for a system containing 2n electrons is expressed according to Eq. 4.7.

$$\Psi = \frac{1}{\sqrt{(2n)!}} \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \psi_2(1)\alpha(1) & \psi_2(1)\beta(1) & \cdots & \psi_n(1)\beta(1) \\ \psi_1(2)\alpha(2) & \psi_1(2)\beta(2) & \psi_2(2)\alpha(2) & \psi_2(2)\beta(2) & \cdots & \psi_n(2)\beta(2) \\ \vdots & \vdots & \vdots & \vdots & & \\ \psi_1(2n)\,\alpha(2n) & \psi_1(2n)\,\beta(2n) & \psi_2(2n)\,\alpha(2n) & \psi_2(2n)\,\beta(2n) & & \psi_n(2n)\,\beta(2n) \end{vmatrix} \tag{4.7}$$

There are mainly two kinds of basis sets. The first family is comprised by the so-called *STO-nG* basis sets, in which a number n of Gaussian like primitive functions as described in Eq. 4.6 are used to get the best approximation to the corresponding Slater function by adjusting the parameters *a* and *b* in Eq. 4.6 [6]. The second big family is called split-valence basis sets, with the Pople and Dunning basis sets the most widely used. Split-valence basis sets make use of one single basis function to describe core electrons while valence electrons are described by two or more basis functions. Pople basis sets are formulated according to the *K-MNG* scheme, where *K* is the number of Gaussian functions used to describe the basis functions of the core electrons while *M* and *N* indicate that each valence orbital is split into two parts represented by M and N Gaussians respectively. Popular examples of this kind of basis sets are the 3-21G [7] and 6-31G [8]. In the same way triple zeta Pople basis set 6-311G indicates that the valence orbital has been split into three parts each of them represented respectively by 3, 1 and 1 Gaussians. To achieve a more realistic electron distribution polarization and/or diffuse functions are added to the basis set. Polarization functions are added by supplementing basis functions with d, p or f orbitals so that the electron distribution is polarized or displaced along a particular direction. When Pople split-valence basis sets are used polarization functions are written in parenthesis. One example is the 6-31G(d,p) basis set, where d and p polarization functions have been added to the 6-31G basis set. On the other hand diffuse functions are Gaussian functions similar to those in Eq. 4.6 but with small coefficients a and b with the aim to better reproduce the behavior of electrons that are far from the nuclei. Diffuse functions are denoted as (+) when added only to heavy atoms and as (++) when additionally added to helium and hydrogen atoms. Diffuse functions are normally used to study molecules with heteroatoms, anions and electronic excited molecules.

Dunning basis sets, also known as correlation consistent basis sets, are the other big family of split-valence basis sets. They are built up by adding shells of functions to a core set of atomic Hartree-Fock functions. Each function in a shell contributes approximately the same amount of correlation energy. Dunning basis sets follow the nomenclature *cc-pVXZ*, with X being the degree of polarization. The addition of diffuse functions to Dunning basis sets is denoted with the prefix "aug" [9–12].

4.1.2.3 Ab Initio Calculations

Ab initio or first principles calculations try to solve the Schrödinger equation (Eq. 4.1) using only basic physical theory without using any empirical fit. The Hartree-Fock (HF) method, also known as Self Consistent Field (SCF) method is the simplest *ab initio* method [13, 14]. The HF wave function is approximated to a single Slater determinant of spin orbitals (Eq. 4.7) while the Schrödinger equation is solved through an iterative process applying the variational principle. Thus energy of any approximate wave function is always larger than the exact energy. Since one electron wave functions are used, electron–electron interactions are neglected and each electron interacts only with the mean field arising from all the remaining electrons, which is described by the Fock operator. The best approximate wave function is obtained by varying iteratively the weighting coefficients of the atomic

orbitals (c_{si} in Eq. 4.4) until the energy expectation value, $E(\psi)$, of the approximate wave function is minimized.

$$\mathrm{E}(\psi) = \frac{\langle\psi|\,\widehat{\mathrm{H}}\,|\psi\rangle}{\langle\psi|\psi\rangle} \tag{4.8}$$

The major inconvenience of the HF method is the poor description of the electron correlation since each electron is considered to move in an electrostatic field whereas in reality electrons repeal each other. This weak point makes the HF method not reliable to get accurate estimates of systems where electron correlation plays an important role such as π-stacking interactions or chemical reactions. Post-HF methods correct this deficiency.

Møller Plesset (MP) [15, 16] and Coupled Cluster (CC) [17, 18] methods are two of the most widespread post-HF methods. The Møller Plesset method, which is based on perturbation theory, makes use of a correction term that handles electron correlation by promoting electrons from occupied to virtual molecular orbitals giving electrons more space to move and thus making it easier for them to avoid one another. The Hamiltonian operator is defined as the addition of a perturbation operator ($\widehat{V}$) to the unperturbed HF Hamiltonian ($\widehat{H}^0$):

$$\widehat{\mathrm{H}} = \widehat{\mathrm{H}}^0 + \lambda\cdot\widehat{\mathrm{V}} \tag{4.9}$$

being λ a dimensionless parameter such that $0 < \lambda < 1$. Consequently both wave function and energy are described by this perturbed Hamiltonian operator.

$$\psi = \psi^{(0)} + \lambda\cdot\psi^{(1)} + \lambda^2\cdot\psi^{(2)} + \cdots + \lambda^{\mathrm{n}}\cdot\psi^{(\mathrm{n})} \tag{4.10}$$

$$\mathrm{E} = \mathrm{E}^{(0)} + \lambda\cdot\mathrm{E}^{(1)} + \lambda^2\cdot\mathrm{E}^{(2)} + \cdots + \lambda^{\mathrm{n}}\cdot\mathrm{E}^{(\mathrm{n})} \tag{4.11}$$

where $\psi^{(n)}$ is the *nth* correction of the wave function related to the number of virtual spin orbitals and $E^{(n)}$ is the n^{th} correction energy term. Since HF energy is the sum of terms $\mathrm{E}^{(0)}$ and $\mathrm{E}^{(1)}$ electronic correlation corrections are taken into account from the second term (MP2) up to the fourth term (MP4). It should be remarked that MPn calculations are computationally very expensive especially as n increases.

The wave function within the Coupled Cluster method is expressed as a sum of the HF ground state determinant plus determinants representing the promotion of electrons to virtual molecular orbitals:

$$\begin{aligned}\psi &= \left(1 + \widehat{\mathrm{T}} + \frac{\widehat{\mathrm{T}}^2}{2!} + \frac{\widehat{\mathrm{T}}^3}{3!} + \cdots + \frac{\widehat{\mathrm{T}}^{\mathrm{n}}}{\mathrm{n}!}\right)\cdot\psi_{\mathrm{HF}} \\ &= \mathrm{e}^{\widehat{\mathrm{T}}}\cdot\psi_{\mathrm{HF}}\end{aligned} \tag{4.12}$$

where $\widehat{T} = \widehat{T}_1 + \widehat{T}_2 + \cdots + \widehat{T}_{\mathrm{n}}$ and the operators $\widehat{T}_n$ are excitation operators and have the effect of promoting n electrons into virtual spin orbitals.

4.1.2.4 Density Functional Theory

Opposite to *ab initio* methods there exist those methods based on the Density Functional Theory (DFT), which is based on the theorems of Hohenberg-Kohn [19]. These theorems postulate that the ground state energy and electronic properties of a non degenerate electronic system are solely defined by its electron density. In consequence DFT does not use the wave function but an electron probability density function $\rho(x,y,z)$, which refers to the probability of finding an electron in a volume element *dxdydz* centered on a point with coordinates (x, y, z). Assuming *r* as the position vector of the point with coordinates (x, y, z) and taking into account Born and Pauli interpretation, *i.e.* the square of a one electron wave function (Kohn-Sham orbitals, ψ_{i} in Eq. 4.13) at any point is the probability density at that point, electron density can be written as follows:

$$\rho(\mathrm{r}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} |\psi_{\mathrm{i}}(\mathrm{r})|^2 \tag{4.13}$$

being *n* the number of occupied molecular orbitals. Electronic energy $E(\rho)$ is calculated as a simple summation of different contributions that depend on the electronic density:

$$\mathrm{E}(\rho) = \mathrm{E}^{\mathrm{T}}(\rho) + \mathrm{E}^{\mathrm{V}}(\rho) + \mathrm{E}^{\mathrm{J}}(\rho) + \mathrm{E}^{\mathrm{XC}}(\rho) \tag{4.14}$$

where $E^T(\rho)$ the kinetic energy, $E^V(\rho)$ the term containing the potential attractive energy electron-nucleus and the repulsive term between nuclei, $E^J(\rho)$ the Coulombic repulsion between electrons and $E^{XC}(\rho)$ the interchange-correlation energy. The latest term is the only one that is not determined directly because of its unknown mathematical formulation. Usually $E^{XC}(\rho)$ is described as a sum of an exchange term $E^X(\rho)$ and another of electronic correlation $E^C(\rho)$.

$$\mathrm{E^{XC}}\,(\rho) = \mathrm{E^{X}}\,(\rho) + \mathrm{E^{C}}\,(\rho) \tag{4.15}$$

The exchange term is normally calculated assuming a homogeneous electron density, such as the Local Density Approximation (LDA) [20] and the Local Spin Density Approximation (LSDA), [21] or by using gradient corrected functionals such as the so called Generalised Gradient Approximation (GGA) methods. Examples of functionals using a homogeneous electron density are the VWN [21] or the local correlation functional of Perdew (PL) [22]. Several GGA based functionals have been developed such as Becke95 (B95) [23], Perdew 86 (P) [24], Perdew-Burke-Ernzerhof (PBE) [25], Perdew-Wang 91 (PW91) [26] and the widely used Lee-Yang-Parr (LYP) [27]. Alternatively the exchange-correlation term can also be calculated using the so called hybrid density functionals, which combine a conventional GGA method with a percentage of Hartree-Fock exchange. Examples of hybrid density functionals include B3LYP [27, 28], B3PW91 [26, 28], MPW1K [29–31], O3LYP [27, 32] and X3LYP [26, 27, 33, 34].

4.1.2.5 Solvent Effects

There are two ways of treating the solvent in molecular simulations, *i.e.* explicitly and implicitly. The first approach includes all the solvent molecules explicitly defined. To properly solvate a solute with explicit solvent molecules, a huge amount of solvent molecules is needed, which increases exponentially the required computational time. To avoid this one must rely on the continuum models, where the solvent is described as an infinite dielectric medium while the solute is treated at the QM level. One of the most popular models is the so called Polarizable Continuum Model (PCM) or MST, developed by Miertuš, Scrocco and Tomasi [35]. PCM is a Self Consistent Reaction Field (SCRF) method, *i.e.* implies the use of a reaction potential to be self-consistently solved together with the solute charge.

The PCM method involves the generation of a solvent cavity from spheres centered at each atom of the molecule and the calculation of virtual point-charges on the cavity surface representing the polarization of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wave function.

The PCM method divides the transfer of a given solute from the gas phase into solution in three steps:

1. Creation of the solute cavity inside bulk solvent.
2. Generation of the van der Waals particle inside the cavity.
3. Generation of the solute charge distribution in solution.

Consequently the net molecular free energy of a system in solution (ΔG_{solv}) is the sum of these three contributions:

$$\Delta \mathrm{G_{solv}} = \Delta \mathrm{G_{cav}} + \Delta \mathrm{G_{vdW}} + \Delta \mathrm{G_{elec}} \tag{4.16}$$

where ΔG_{cav} is the work involved to increase the cavity, ΔG_{vdW} is the contribution due to the van der Waals interactions and ΔG_{elec} is the electrostatic component of the $\Delta \mathrm{G_{solv}}$ and it corresponds to the work required in the polarization process. Interaction potential between solvent and solute is introduced in the solute electronic Hamiltonian as a perturbation operator ($\widehat{V}_R$). Thus the Schrödinger equation, being $\widehat{H}^0$ the solute's Hamiltonian operator, is then expressed as follows:

$$\left(\widehat{\mathrm{H}}^0 + \widehat{\mathrm{V}}_\mathrm{R}\right)\psi = \mathrm{E}\,\psi \tag{4.17}$$

4.1.2.6 Application of QM Calculations to Membrane Proteins

The use of QM calculations, though giving the most accurate and precise description of matter, is strongly limited to small systems. Thus it is clear that the use of QM methods will be limited by nature to small parts of the system where a correct description of the electronic structure is needed. Coming back to the line of the present book, QM calculations will be very useful to design and describe new functional groups to be incorporated in the inner core of the β-barrel channel protein to design for instance a great variety of stochastic sensors [36, 37]. These sensory devices are built upon bioconjugation of a specific binding site for a given analyte in the inner wall of a channel protein. In such cases QM calculations are very useful to evaluate the relative affinity of the binding site towards different species and optimize the chemistry of the sensory site to improve the selectivity towards a specific analyte. For example QM calculations have been used to understand the specificity of crown ether functionalized polythiophene based sensors [38, 39] where the coordination of the binding site with different cations leads to the modification of the electronic properties that explains the experimental affinity in the order $Li^+ > Na^+ > K^+$. In a similar way and within the stochastic sensory applications QM calculations can be useful to explain and optimize enantioselective sensors such as that proposed by Bayley and coworkers [40] where the transmembrane pore of α-hemolysin was functionalized with a β-cyclodextrin in the inner wall of the channel. The resulting sensor was used for chiral discrimination of ibuprofen and thalidomide. Other functionalities that can be preliminarily designed and evaluated by means of QM calculations are pH or photo-switched responsive groups, which have potential applications to design smart nano-channels rendering open or closed states upon an external stimulus. QM calculations are also used to obtain force field parameters to be used in Molecular Dynamics simulations, as it will be discussed in the next section. There are many codes available to run QM calculations. One of the most popular free distributed code is GAMESS (http://www.msg.ameslab.gov/gamess/) [41, 42].

4.1.3 Molecular Dynamics of Membrane Proteins

Molecular Dynamics (MD) simulations calculate the dynamic behavior of a system at the molecular level by means of a classical mechanics approach. Two levels of resolution are typically considered in atomistic MD simulations. The first one is known as all atom MD and uses one individual particle to represent each individual atom present in the system. The second one is called united atom MD and has a lower resolution since hydrogen atoms are lumped together with the heavy atom to which they are connected into one single interaction site. Covalent bonds are represented by springs and each particle is defined by its radius, hardness and net charge.

MD simulation dates back to the late 1950s and its algorithm was first formulated by Alder and Wainwright [43, 44]. First studies were mainly focused on simulation of hard spheres and it was not until the mid-1960th that more realistic systems such as liquid argon were simulated [45]. Major advances were carried out in the 1970s with the first realistic simulation of liquid water [46, 47] and the simulation of the bovine pancreatic trypsin inhibitor, setting the starting point of protein simulations [2].

4.1.3.1 Force Field

Force field is the principal *ingredient* to run a MD simulation and it can be broadly defined as the set of parameters and mathematical functions used to evaluate the potential energy of a system. Force field parameters include both bonded or covalent interactions and non-bonded interactions. The quality of the force field parameters will have a clear impact on the quality of the results obtained from the MD simulation. Just as an example, Eq. 4.18 represents one of the most widely used force fields for biomolecular simulations, the so called AMBER force field [48, 49]:

$$E(r) = \sum_{bonds} k_{stretcht}(r - r_{eq})^2 + \sum_{angles} k_{bend}(\theta - \theta_{eq})^2 + \sum_{dihedrals} k_{torsion}[1 + \cos(n \cdot \phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] + \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_o r_{ij}} \quad (4.18)$$

As can be seen in Eq. 4.18, total energy has five terms which are, respectively:

1. The bond stretching term in which the stretching force constant ($k_{stretch}$) and the distance between two bonded atoms and its equilibrium bond distance (r and r_{eq} respectively) are included.
2. The angle bending term including the bending force constant (k_{bend}) and the angle between three consecutive atoms and its equilibrium value (θ and θ_{eq}).
3. The torsional term contains a dihedral constant ($k_{torsion}$) setting the energy barrier for the rotation profile, the actual dihedral angle (φ) and the equilibrium one (γ).
4. The van der Waals non bonded term, a mathematical model following the Lenard-Jones potential that takes into account the attractive and repulsive forces between two particles being r_{ij} the distance between them. A_{ij} and B_{ij} are parameters that depend on the pair of atoms referring to hard core repulsion and dispersive attraction respectively.
5. The electrostatic interactions are taken into account in the last term, where q_i and q_j are the point charges of atoms i and j respectively, r_{ij} is the distance between them and ε_0 is vacuum permittivity.

There are many force fields available in the literature. Most of them share practically the same terms as those shown in Eq. 4.18 differing mostly in the mathematical functions, the specific parameters and the way to derive them. Most of the force fields are derived from QM calculations and then refined to reproduce some experimental observable such as heat of vaporization, free energy of hydration and solvation, partition coefficients, spectroscopic data, diffusion constants, viscosities, dielectric permittivity or liquid densities. The most widely used force fields for biomolecular simulations are the GROMOS [50, 51], AMBER [48, 49, 52–54], CHARMM [55–57] and OPLS [58, 59] force fields. Sometimes new force field parameters must be derived, as it can be the case of deriving the parameters of a new functional group that is introduced in the inner core of a β-barrel channel protein. To avoid inconsistencies when deriving those new parameters, it is highly recommended to follow the same procedure that has been used to derive the parameters of the chosen force field, *i.e.* the way charges, van der Waals and bonded parameters have been derived. It is also highly recommended to start the parametrization by considering the same force field parameters from similar atom types. For instance, all the aforementioned force fields contain all the parameters for all the natural amino acids. If one has to derive the force field parameters of a new functionality containing a phenyl ring, using the same parameters as those defined for the phenylalanine amino acid is a good starting point. Subsequently the parameters can be refined using QM calculations in the same spirit as the authors of each force field used them to derive their parameters.

Water is normally treated explicitly, but there are some force fields that have been developed to work together with some implicit water models where the solvent is normally treated using the generalized Born continuum electrostatics [60]. All atom implicit solvent simulations are computationally substantially cheaper since all the water related degrees of freedom are not taken into account. However they must be used with caution in systems of high complexity such as channel proteins embedded into lipid or polymer bilayers. There are many water models ready to be used in MD simulations. The interested reader is encouraged to have a look on the review by Guillot entitled "A reappraisal of what we have learnt during three decades of computer simulations on water" for a broad overview of the strong and weak points of several water models [61]. The "top-5" water models used in biomolecular simulations are, in no specific order, TIP3P [62], TIP4P [62], TIP5P [63], SPC [64] and SPC/E [65] water rigid non-polarizable models. In spite

that none of the water models can perfectly reproduce all the thermodynamic properties of water, SPC/E water model and a modification based on the TIP4P coined as TIP4P/2005 [66] seem to be those models that perform better [67].

Again it should be stressed the importance of the selection of the right force field parameters together with the right water model to obtain a reliable expression for the energy of the system under study.

4.1.3.2 Mathematical Formulation of Molecular Dynamics

There are many computer programs available to run MD simulations. Most of them are free code and are optimized to run in parallel, *i.e.* the calculation is shared by a high number of CPUs at the same time. Just to give a short list of popular programs without following any order, GROMACS (www.gromacs.org), NAMD (www.ks.uiuc.edu/Research/namd), AMBER (www.ambermd.org), CHARMM (www.charmm.org), TINKER (dasher.wustl.edu/ffe) or LAMMPS (lammps.sandia.gov) computer programs should be highlighted. Apart from standard MD, some systems may be simulated using Dissipative Particle Dynamics [68, 69], Brownian Dynamics [70] or Stochastic Dynamics where friction terms are added to the Newton equation of motion. However this chapter will focus on the standard MD.

Given an atom i with mass m_i and considering that its position is described by a three dimensional vector r_i, its motion is ruled by Newton's law

$$\frac{dv_i(t)}{dt} = \frac{F_i}{m_i} \tag{4.19}$$

$$v_i(t) = \frac{dr_i(t)}{dt} \tag{4.20}$$

where v_i and F_i are respectively the velocity and the force acting on the atom i in a given moment. F_i can be obtained through Eq. 4.21.

$$F_i = -\frac{\partial E\left(r_N\right)}{\partial r_i} \tag{4.21}$$

where $E(r_N)$ is given by Eq. 4.18 and N is the number of particles in the system.

Equations 4.19 and 4.20 form a system of N coupled differential equations that might be solved numerically. This system can be integrated step by step by the so called leap-frog Verlet algorithm [71] which is founded on a Taylor expansion of the position r_i at time $t_n = t_0 + n\Delta t$, leading to equations 4.22 and 4.23.

$$v_i\left(t_n + \frac{\Delta t}{2}\right) = v_i\left(t_n - \frac{\Delta t}{2}\right) - \frac{\Delta t}{m_i}\frac{\partial E\left(r_N\right)}{\partial r_i} + o\left(\Delta t^3\right) \tag{4.22}$$

$$r_i\left(t_n + \Delta t\right) = r_i\left(t_n\right) + v_i\left(t_n + \frac{\Delta t}{2}\right)\Delta t + o\left(\Delta t^4\right) \tag{4.23}$$

Time step Δt might be small enough to simulate those movements with the highest frequencies, which normally are the bond vibrations. Consequently time steps take values of the order of the femtosecond. Small time steps lead to more expensive simulations.

The CPU time needed for a MD simulation depends on several factors as the number of the explicit particles in the system, the time step or the cut-off, *i.e.* the maximum distance in which non bonded interactions are evaluated. In order to speed up the MD simulation some simplifications might be done such as freezing the fastest modes of vibration by constraining the bonds to hydrogen atoms to fixed lengths, mostly by the SHAKE [72] or LINCS [73] algorithms.

4.1.3.3 Periodic Boundary Conditions and Electrostatics

MD simulations are usually performed in simulation boxes containing the solute and solvent atoms. To represent an infinite sized system the whole system must be replicated periodically in all directions so atoms outside the simulation box are simply images of the atoms simulated in that box. Periodic boundary conditions ensure that all simulated atoms are surrounded by neighboring

atoms, either images or not. This condition guarantees that atoms moving out of the box at one side will enter again inside the box through the opposite side because replicas of each particle in all duplicated boxes move exactly the same way. Minimum image conventions avoid duplicate interactions between atoms *i* and *j* by taking into account only the interaction of atom *i* with the closest *j* atom, either original or copy. Thus, periodic boundary conditions are used to correct potential errors in the van der Waals non bonded term. Regarding the Coulombic non bonded term, the main problem to evaluate it correctly is that a sudden cut-off leads to large errors. This problem is easily solved by the use of the so called Particle Mesh-Ewald summation (PME) [74] which calculates the infinite electrostatic interactions by splitting the summation into short and long range parts. For PME, the cut-off only determines the balance between the two parts, and the long-range part is treated by assigning charges to a grid that is solved in reciprocal space through Fourier transforms. Alternatively electrostatics can be treated by other approaches such as a reaction field [75] or Ewald summation [76].

4.1.3.4 Thermodynamics Ensembles

An ensemble is a collection of all possible systems that have differing microscopic states but belong to a single macroscopic or thermodynamic state. There are four ensembles that are used in MD simulations [77].

1. The canonical or NVT ensemble, whose thermodynamic state is characterized by a fixed number of atoms *N*, volume *V* and temperature *T*.
2. The isobaric-isoenthalpic or NPH ensemble, where the number of atoms *N*, the pressure *P* and enthalpy *H* are fixed.
3. The isobaric-isothermal or NPT ensemble, with fixed values of number of atoms *N*, pressure *P* and temperature *T*.
4. The microcanonical or NVE ensemble, which corresponds to a closed or isolated system since energy *E*, besides the number of atoms *N* and volume *V*, is fixed.

Temperature is mainly obtained from the average velocity or kinetic energy of each particle present in the system. When the temperature must be kept constant, a thermostat must be used to adjust the kinetic energy of the system. The most widely used thermostats in canonical and isobaric-isothermal ensembles are the Berendsen [78], Velocity Rescale [79] and Nosé-Hoover [80, 81] thermostats. The pressure is controlled by readjusting the size of the simulation box. To that end a barostat must be used. The most common barostats are the Berendsen [78] and Parrinello-Rahman [82, 83] barostats.

4.1.3.5 Combined Quantum Mechanics/Molecular Mechanics Calculations

In some specific systems it may be interesting to consider some smaller parts at a higher level of resolution, *i.e.* at the QM level. Imagine a β-barrel protein where a binding site has been introduced to the inner pore. As mentioned in the previous section, electronic structures of that binding site must play a key role in the interaction with some specific analytes such as an ionic species. However the level of resolution of the rest of the system must be kept at the classical level because apart from the electronics at this lower level of resolution the structure and dynamic properties can be correctly evaluated and because treating a very big system at the quantum level of resolution is unaffordable. To that end Quantum Mechanics/Molecular Mechanics calculations can be performed, where a small part of the system is treated according to the quantum Hamiltonian while the rest is treated using a classical Hamiltonian. QM/MM philosophy was first introduced by Warshel and Levitt [84] and further developed by Field et al. [85]. QM/MM methods differ, besides the classical force field and the QM level of theory employed, in the strategy to connect the two parts with different level of resolution. There are two methodologies to connect QM and MM parts:

1. The link atom scheme, which uses a monovalent atom – normally hydrogen – to cap the unsaturated QM atoms [86].
2. Methods based on the use of localized orbitals in the boundaries separating the QM and MM parts. One example is the Local

Self-Consistent Field (LSCF) method [87] in which the bonds connecting both parts are represented by a set of strictly localized bond orbitals (SLBOs) that are determined by calculations on small model compounds and assumed to be transferable.

Within the framework of QM/MM calculations the total Hamiltonian operator is defined as the sum of the QM, QM/MM and MM Hamiltonians:

$$\widehat{H} = \widehat{H}^{QM} + \widehat{H}^{QM/MM} + \widehat{H}^{MM} \tag{4.26}$$

where $\widehat{H}^{QM/MM}$ is expressed as the sum of electron-charge, nuclei-charge and van der Waals interaction potentials between QM and MM atoms:

$$\widehat{H}^{QM/MM} = V_{elec}^{QM/MM} + V_{nucl}^{QM/MM} + V_{vdW}^{QM/MM} \tag{4.27}$$

Equation 4.27 may be expanded for a system containing N QM atoms and M MM atoms as follows:

$$\widehat{H}^{QM/MM} = -\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{q_j}{r_{ij}} + \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{Z_i q_j}{r_{ij}} + \sum_{i=1}^{N}\sum_{j=1}^{M}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] \tag{4.28}$$

Thus energy in a QM/MM system is defined as shown in Eq. 4.29.

$$E = E^{QM} + E^{MM} + E^{QM/MM} \tag{4.29}$$

where $E^{QM/MM}$ is defined as:

$$E^{QM/MM} = \langle\psi| - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{q_j}{r_{ij}}|\psi\rangle + V_{nucl}^{QM/MM} + V_{vdW}^{QM/MM} \tag{4.30}$$

where the first term is the electronic interaction within the Electronic Emboding scheme.

GAMESS-UK (http://www.cfs.dl.ac.uk/gamess-uk/index.shtml) is a very powerful free code to run not only QM calculations but also QM/MM calculations in conjunction to CHARMM MD program [88].

4.1.3.6 MD Simulations of Membrane Proteins

The main drawback of atomistic MD simulations is the timescale reachable with the current computational resources, which is limited to hundreds of nanoseconds for systems as big as membrane proteins inserted in polymer or lipid bilayers. If a huge membrane protein such as the *E. coli* FhuA (ferric hydroxamate uptake component A) protein is considered, simulation of the self-assembly of this protein into a lipid or polymer vesicle or bilayer takes place in the microsecond timescale, which is far beyond the timescale that would be achievable with atomistic MD simulations. Self-assembly processes of systems like this can be tackled at a lower or coarser level of resolution, as will be explained in the next section. However atomistic MD simulations can capture many interesting events that cannot be studied by simulations at the coarse grained level [89]. Atomistic MD simulations have helped to understand the transport mechanism of ions through small ion channels embedded in lipid bilayers and to analyze the interplay between the protein, the water and the lipid bilayer in modulating the transport through the channel [90]. The gating mechanism of mechano-sensitive proteins related to conformational changes in the protein upon modification of the surface tension in the membrane has been also reported by means of MD simulations [91].

This section however will focus on reviewing two articles in which the FhuA protein was simulated by means of atomistic MD simulations, stressing on the information one can obtain from simulations at this level of resolution. The first article is a work by Faraldo-Gómez et al. [92]. In this work the authors performed a 10 ns long MD simulation of a FhuA protein embedded

in a hydrated dimyristoyl-phosphatidylcholine (DMPC) lipid bilayer on both the ligand free (FhuA) and ferrichrome bound states (FhuA-F). Both initial structures were taken from the protein data bank (http://www.rcsb.org/pdb/home/home.do). Simulated systems contained approximately 71,000 atoms. Simulations were used to extract information regarding the structure and dynamic response of the protein embedded in the lipid bilayer. The authors concluded that both ligand free and bound systems do not suffer any dramatic structural change within the simulated timescale and that the secondary structure is kept throughout the whole trajectory. The authors observed that the extracellular loop 8 (L8), in the bounded state, owns an enhanced flexibility that helps to close the binding site. The presence of the plug domain closing the pore in the FhuA protein is expected to reduce the permeability of water throughout the membrane protein. However water permeability was remarkably lower in the FhuA-Ferrichrome system, which was related to the unfolding of the switch helix promoting the displacement of the N-terminal segment and narrowing the pore. This correlation between higher conformational flexibility and reduced water permeability supported the view that high conformational changes in the plug domain of the FhuA protein are needed for the siderophore to either passively diffuse or be translocated into the periplasm (for details on the FhuA protein structure and function see Sect. 2.4.1).

In a more recent study were performed atomistic MD simulations on an engineered FhuA protein embedded into either a N-octyl-2-hydroxyethyl sulfoxide (OES) detergent belt or 1,2-dinervonyl-sn-glycero-3-phosphocholine (DNPC) lipid bilayer [93]. More concretely was considered a FhuA protein where the plug domain was fully removed resulting in an open pore. Therefore were removed the first 159 amino acids (FhuAΔ1-159). Moreover was induced a closed state in the FhuAΔ1-159 engineered protein by labeling six lysine amino acids in the inner pore with six 3-(2-pyridyldithio)propionic-acid moieties (FhuAΔ1-159-Pyr). The size of the analyzed systems ranged from ∼68,000 when the proteins where embedded in detergent belt to ∼83,000 when the proteins where inserted into the lipid bilayers. The GROMOS96 Force field [50] was used and the Pyr labeling moieties were parametrized according to the same force field being the charges obtained from QM calculations. Simulation time was 50 ns for each system. Structural information derived from these simulations stressed on the null role played by the plug domain on the overall stability of the FhuA protein. This initial conclusion is of great importance since it supports the view that the great stability of the FhuA protein is mainly due to the efficient hydrogen bond network between neighboring β-strands and encourages the use of the engineered FhuAΔ1-159 variant for nano-technological applications without fearing a loss of secondary and tertiary structure. The FhuAΔ1-159-Pyr displayed a more elliptical shape when embedded into the OES detergent belt. Figure 4.2 displays the equatorial and axial projection of the final configuration after 50 ns of all the systems embedded into an OES detergent belt. On the other hand for both FhuAΔ1-159 open and FhuAΔ1-159-Pyr closed states a nearly circular section was adopted when embedded into the DNPC lipid bilayer. The FhuAΔ1-159 protein also showed a remarkable anisotropic compressibility as it was deduced from the larger equatorial-longitudinal fluctuations compared to the more rigid axial-latitudinal direction.

By choosing these two examples the authors would like to stress the importance of atomistic MD simulations to obtain information regarding the structure and stability of large membrane proteins such as the FhuA protein and its implications when designing new engineered FhuA based nano-pores. From the biological point of view, detailed structural information can help elucidate transport and actuation mechanism of membrane proteins and their interplay with the surrounding lipid bilayer and hydrated media.

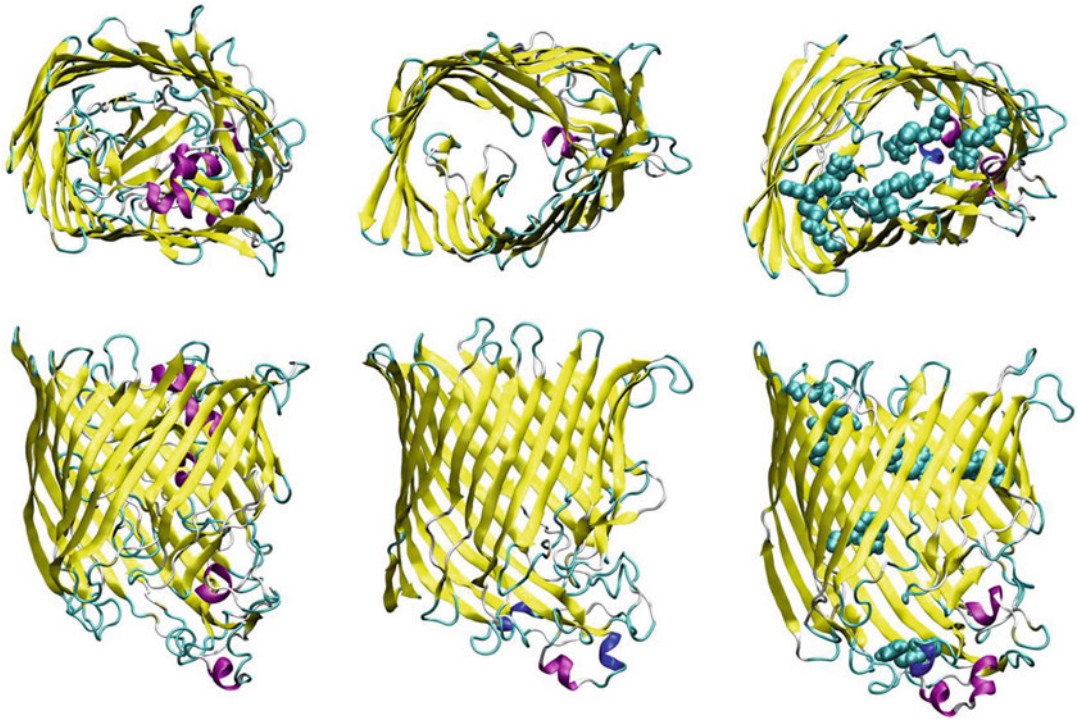

Fig. 4.2 Equatorial (*top*) and axial (*bottom*) projections after 50 ns of the FhuA, FhuA Δ1-159-Pyr and FhuA Δ1-159 proteins embedded in OES detergent belt. For clarity water and detergent atoms have been removed

4.1.4 Coarse Graining of Membrane Proteins

Many relevant phenomena in membrane protein research such as protein folding or membrane self-assembly occur in timescales in the order of several microseconds or even hours. These phenomena cannot be captured at the all atom level of resolution. Thus a lower resolution scale is needed, *i.e.* a Coarse Grained (CG) approach. Simulations using CG potentials follow the same mathematical scheme as presented in the previous section for standard MD, being the main difference in the "form" of the potential energy function or force field definition (Eq. 4.18). The main difference of CG as compared to all atom simulations is that in CG simulations a group of atoms is grouped into a single interaction site. As a consequence of this lower level of resolution the total number of particles in the system is dramatically reduced. Furthermore the interaction potential is highly simplified, *i.e.* Eq. 4.18 adopts an "easier" form or can even be represented in a tabulated form. Tabulated potentials are simple functions in which the interaction potential between two groups is simply a function of the distance between them. Additionally CG simulations present a smoothed potential energy surface which leads to an acceleration of the MD simulation and to a much broader and efficient exploration of the whole conformational space and a better sampling of the thermodynamic ensemble [94]. On the other hand the use of CG simulations implies an inherent loss of the atomistic detail. However the atomistic detail can be, if needed, brought back by using the so called back mapping or reverse mapping strategies, in which the higher level of resolution is recovered from a CG simulation by applying a backmapping operator [95]. The next sub-sections aim to summarize some of the most popular CG approaches to simulate lipid or polymer membranes and membrane proteins.

4.1.4.1 Systematic Derivation of CG Potentials to Simulate Self Assembly of Amphiphilic Block Copolymers

As it has been stressed throughout the whole book, it is desirable for nano-technology applications to substitute the natural lipid bilayer environment where membrane proteins sit by a more robust amphiphilic block copolymer environment, either AB diblock or ABA triblock copolymers (A = hydrophilic block, B = hydrophobic block). These polymers self-assemble in water leading to the formation of different shapes/architectures such as micelles, cylinders, vesicles or bilayers depending on many different factors such as the molecular weight, polydispersity, temperature or hydrophobic/hydrophilic ratio of their blocks. All atom MD simulations fail in capturing such spontaneous self-assembly since the typical timescales of these phenomena are within the microsecond scale and because the systems are too large. It should be remarked at this point the highest molecular weight of amphiphilic block copolymers compared to lipids. This highest molecular weight slows down their diffusion leading to relaxation times that are only captured at long timescales. Thus CG simulations are the only option to reproduce the self-assembly of such polymers in water under the different conditions highlighted few lines before.

At this point there is another important issue to tackle: how to represent the water solvent. In systems like this, water represents a high percentage of the total particles present in the system, so a lot of computational effort is spent on the dynamics of the solvent. Accordingly, there are two alternative ways to treat water in a way that can reduce the computational time:

1. To define water at the CG level grouping in a single bead one or more water molecules or;
2. To treat the water as an implicit solvent.

The second approach is the computationally cheaper one. Decisions on which solvent model should be used are made on basis of the phenomena of interest. The use of implicit solvent may lead to the correct assembled structure, but whenever one is interested in capturing the dynamical properties of these polymers in solution neglecting the hydrodynamics may lead to models unable to properly reproduce such properties.

There are several methods stated in the literature to obtain CG potentials and some of them have been successfully used to simulate polymer melts, polymers in solution or lipid bilayers [96]. Briefly summarizing, a large family of CG potentials are derived so that they can reproduce by construction some target structural or thermodynamic properties from all atom simulations and/or from experiments. One of the most used methodologies to derive CG potentials is the so called Iterative Boltzmann Inversion (IBI), which aims at the construction of a tabulated potential able to reproduce a target radial distribution function from parent atomistic simulations [97]. Two CG beads must be defined; one for the hydrophilic monomers in block A and another for the hydrophobic monomers in block B. The IBI method has been extensively used to obtain coarse-grained potentials of several polymer melts and few implicit solvent potentials for pluronics, which are amphiphilic ABA type block copolymers, resulting in significant savings in computational time [98]. The main disadvantage of this methodology is its limited transferability to be used at thermodynamic state points beyond the one that was used to derive the potential.

On the other hand other methods derive non bonded CG implicit solvent potentials from pair potentials of mean force calculations and without accounting for any experimental data as input. Since these last methods are not biased to reproduce any target property and they are derived from the potential of mean force between two groups as a function of distance, they are expected to reproduce a wider set of properties in a wider range of thermodynamic conditions [99]. However this latter approach has not yet been used to simulate polymer membranes or lipid bilayers, so new methodological developments have to be found to adapt the existing schemes to the systems of interest [99–101]. Bonded potentials to model the flexibility of the degrees of freedom associated with CG bonds, angles and potentials can be obtained in a straightforward manner by Boltzmann-inverting the sampled distribution

of the mapping points obtained from parental all-atom simulations.[100]

4.1.4.2 Deriving CG Potentials for Amino Acids

Deriving CG potentials for polymers using the methods shown in the previous subsection is not an easy task but can be fulfilled since very few CG bead types must be defined. In a block copolymer it would be enough to define two kinds of CG beads, *i.e.* one representing the hydrophilic monomers and another to represent the hydrophobic monomers. Following this strategy to develop implicit solvent models for all the amino acids is cumbersome since there are 20 amino acids and all the effective pair potentials between all the beads defining the 20 amino acids should be calculated.

Additionally derivation of all the necessary bonding parameters would be extremely human and CPU time-consuming since 20 natural amino acids would lead to the consideration of 8,000 local Ramachandran plots to account for three consecutive amino acids ($= 20^3$). This strategy remains valid for short peptides, but it is out of reach for large proteins [102]. However one can take advantage of other developed implicit solvent CG models for amino acids available in the literature, like the one developed by Bereau and Deserno [103]. This model makes use of three beads to represent the backbone – one for the amide group, another for the alpha carbon and a third one for the carbonyl group – while one bead is located in the beta carbon for the side chain. The fine graining in the backbone of the protein was considered by the authors so that a more reliable description of the conformational backbone flexibility that rules the protein secondary structure is acquired. The most drastic approximation in this model is the way side chains are treated. Apart from the position of the center of mass of the only bead representing the side chains (except for glycine, where no side chain is present), they all have the same van der Waals radii. Again this model must be validated and, if necessary, modified to be applied in the systems covered in this book. In spite that this generic protein CG model has not been developed to represent specific secondary structures, its application to both α-bundles and β-barrels led to the correct final structure [104]. Additionally, when considering this approach, a new set of non-bonded potentials between each of the 2 or 3 beads in the amphiphilic polymer membrane and 3 backbone beads plus the 19 side chain beads will have to be developed in a similar way as one aims to develop the implicit solvent non bonded potentials between the beads forming the membrane (see previous section). Thus it appears to be very clear that the use of implicit solvent models to study the self-assembly of this family of systems is not a straightforward task and requires the careful development of a series of potentials.

4.1.4.3 Simplified Models: The Srinivas Approach

As it has been shown before, obtaining realistic and reliable CG potentials for biomolecular simulations is the bottleneck of CG simulations. Interaction CG potentials are not always available and new potentials must be developed for specific cases. However to answer some basic questions such as how the system self-assembles, how the polymer chains in the amphiphilic membrane (or the lipid chains in a lipid membrane) pack around the membrane protein, what is the hydrophobic mismatch (on hydrophobic mismatch see Sect. 6.1.1) or how different thermodynamic conditions influence the system simplified models can be very useful.

In this sense Srinivas and coworkers developed a CG model for the PEO-PE and PEO-PE-PEO block copolymers (PEO = polyethylene oxide; PE = polyethylethylene) using as target properties experimental bulk densities and surface tension values as well as structural information from parental all atom simulations [105]. In their approach three water molecules were lumped into one single CG bead. Their model was used to reproduce the planar membrane formation from 108 $(PEO)_{10}$-$(PE)_9$ diblock copolymers and the formation of a spherical micelle from 48 $(PEO)_{19}$-$(PE)_9$ diblock copolymers. Thus a different spontaneous self-assembly was captured under varying

hydrophilic/hydrophobic ratios and a good agreement with experimental observables such as the hydrophobic thickness was obtained. They extended their study to reproduce a different set of spontaneous self-assembled morphologies upon variation of the hydrophilic fraction (f_{phil}) of the PEO-PE block copolymer [106]. Their results showed that bilayers were obtained at $f_{phil} = 30.9$ %. Cylindrical or worm-like micelles were obtained when f_{phil} was increased up to 51.1 %. Spherical micelles were seen when f_{phil} reached a value of 65.6 %. Additionally other structural information such as the area per polymer and the hydrophobic core thickness/radius was determined from their simulations. The same group of authors studied the insertion of a model membrane protein (based on the *E. coli* OmpF protein) which was modeled as a hydrophobic cylinder with hydrophilic edges. Thus, only two beads were used to represent the protein [107]. Their simulations proved that polymer based membranes can withstand higher hydrophobic mismatches (>22 %) compared to lipid bilayers (2–3 %). In thick membranes it was observed penetration of the hydrophilic PEO block inside the pore, blocking the water permeation. Thus a gating mechanism can be designed by varying the thickness of the polymer membrane or even by functionalizing the amphiphilic polymers. Consequently chain flexibility plays a key role not only to accommodate the membrane protein but also affects the transport across it. In the same work the authors examined the morphology adopted by different PEO-PE-PEO triblock copolymers reproducing the tube-like morphology observed experimentally.

4.1.4.4 Simplified Models: The MARTINI Approach

Another interesting approach is the so-called MARTINI force field [108], which is a general "force-field like" CG model that has been parameterized so that it can reproduce the partitioning free energies between polar and apolar phases of a large number of chemical species. MARTINI force field considers four general types of interaction sites, each one grouping 4 heavy atoms, which interact through a shifted 12–6 Lennard Jones potential and a shifted Coulomb potential (only if the bead is charged). Within this force field the non-bonded parameters σ_{ij} and ε_{ij} are chosen depending on the specific interbead interaction. Bonded interactions are described through weak potentials. The MARTINI force field has been developed for biomolecular simulations, mostly proteins and lipids. However the force field may be refined by considering some structural target properties to obtain hybrid thermodynamics-structural CG descriptions for polymers [109]. Since MARTINI has been also developed to describe proteins [110] and water, [111] parameters derived for amphiphilic block copolymers within the framework of this force field will be compatible to be used together with a MARTINI-like CG description of membrane proteins.

According to the hydrophobicity and hydrophilicity of the molecules, four main types of interaction sites are considered: polar (P), nonpolar (N), apolar (C) and charged (Q), representing hydrophilic groups, hydrophobic groups, mixed groups which are partly polar and partly apolar, and ionized groups, respectively. Furthermore, a number of subtypes are defined to distinguish CG interaction sites with different hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none), or the degree of polarity (from 1 to 5, with 1 representing low polarity and 5 representing high polarity). The MARTINI force field describes each amino acid through 1, 2, 3 or 4 coarser beads [110]. When considering a new functional group, it must be defined in accordance to the MARTINI force field (the same applies for the block copolymers).

The MARTINI force field has been vastly used by Sansom and coworkers [112] to study the insertion of membrane proteins into DPPC lipid bilayer or detergent micelles. They used the following strategy:

1. Download the crystal structure of a given membrane protein from the Protein Data Bank (www.pdb.org);
2. Prepare the initial configuration according to the MARTINI force field mapping scheme by

using the tools provided on the MARTINI web site (http://md.chem.rug.nl/cgmartini/);
3. Surround the protein with randomly distributed DPPC lipids or detergent and water molecules. All these species defined according to the MARTINI model;
4. Run the simulation for at least 200 ns;
5. Extract snapshots of the optimum position of the protein in the lipid bilayer or detergent micelle;
6. Plot the Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) for the membrane protein;
7. Plot the bilayer distortion.

They have currently used this strategy for approximately 140 membrane proteins and all the relevant results are collected in a public data base built by the own group and freely available under http://sbcb.bioch.ox.ac.uk/cgdb/index.php . The final structure obtained for FhuA is depicted in Fig. 4.3.

This strategy can be easily followed to study the reconstitution of membrane proteins in amphiphilic block copolymers and to obtain relevant structural information of the reconstituted systems.

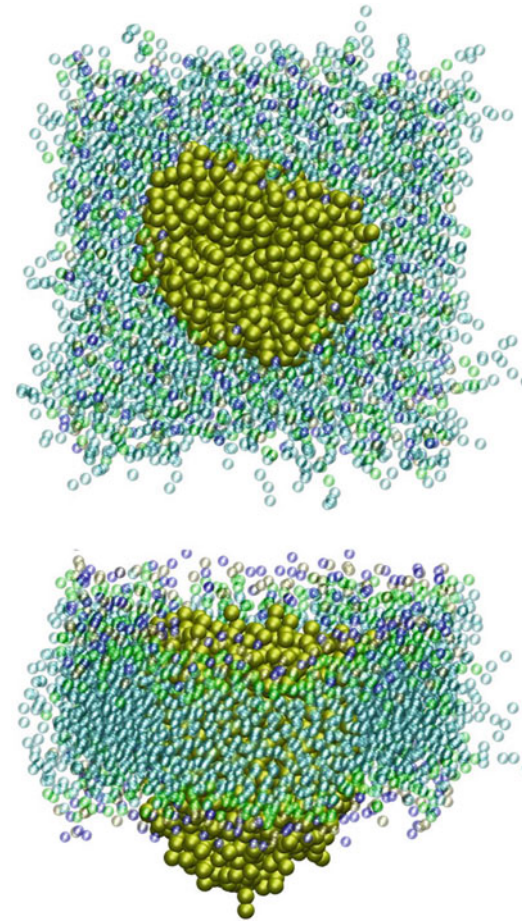

Fig. 4.3 Equatorial (*top*) and axial (*bottom*) projections of the final snapshot after 200 ns of the reconstitution of FhuA in DPPC lipid bilayer. FhuA and DPPC were represented according to the MARTINI force field. For clarity DPPC beads are depicted as transparent spheres. Coordinates have been downloaded from http://sbcb.bioch.ox.ac.uk/cgdb/simtable.php?pdb=2FCP [89]

4.1.4.5 Hybrid Atomistic/CG Simulations

There are several strategies published in the literature to perform hybrid all atom/CG simulations that can be adapted for the described purposes. One of the most popular approaches is running the simulation assuming a resolution at the CG level for the whole system and then "back-map" the trajectory, *i.e.* recover the atomistic resolution from the coarse grained simulation. This approach, which is computationally cheap, will fail however to reproduce specific interactions or the hydrogen bond network supporting the exceptional mechanical behavior of β-barrel membrane proteins. Another interesting approach defines the protein at the CG level while the active site is described at fully atomistic detail, being both CG/atomistic regions linked by a specific interface region [113]. Such approach is conceptually similar to the coupling between MM and QM regions in QM/MM calculations explained earlier in the previous subsection. This methodology can be applied and adapted to the engineered β-barrel membrane proteins where the functional groups will be represented at full atomic resolution and will be connected to the CG protein through an interface region. The same CG potentials that have been developed in the previous section can be used for the rest of the protein and the polymer membrane.

For a full atomistic description of the whole protein interacting with coarse grained membranes two approaches are suggested. The first

one, previously used to study a small ion channel embedded in a lipid bilayer, requires the whole reparametrization of all the atomistic/CG crossed interactions [114]. The second approach makes use of virtual sites to couple both levels of resolution. In this latter approach atomistic parts carry virtual sites that interact with the CG particles through standard CG/CG interactions while atom/atom interactions are kept between particles at the highest level of resolution [115]. This methodology has been proven to give satisfactory results when MARTINI force field and free energy based derived potentials were used to represent the CG interactions. This methodology appears to be highly useful provided the CG potentials for the functional groups used to engineer the β-barrel membrane protein are developed.

4.1.5 Benefits and Limits of Different Simulation Methods

In the previous sections were outlined the benefits and limitations of different simulation strategies. As a short summary the following points are stressed:

1. QM calculations are only useful in the context of membrane protein simulation to design new functional groups where the electronic structure plays a key role. Additionally QM calculations are used to derive new force field parameters for atomistic simulations. QM calculations on large systems are computationally very costly.
2. Atomistic MD simulations are important to obtain structural and dynamic information on systems where the membrane protein is already inserted into the lipid bilayer, polymer membrane or detergent micelle. Reconstitution is not easily captured in the usual timeframes of atomistic MD simulations. Atomistic MD simulations are also useful to elucidate the transport mechanism, gating system and actuation principle of membrane proteins.
3. CG simulations allow longer simulation times at expense of a lower resolution. Obtaining new CG potentials for specific systems is a challenging task. However more simplified strategies such as the MARTINI force field or the one exemplified by Srinivas and coworkers are very useful to study the reconstitution of proteins into lipid/polymer bilayers or detergent micelles and obtain very valid sets of structural, mechanic and dynamic properties that can shed some light onto some biological phenomena which cannot be measured experimentally.

4.2 2D and 3D Structure Prediction of β-Barrel Proteins

The development of transmembrane protein 2D and 3D topology prediction tools has been since many years ago a very active field in bioinformatics research. The goal of these bioinformatics tools is to predict the 2D structure or even the 3D structure of a protein whose crystal structure is unknown from the amino acid sequence. Compared to β-barrel membrane proteins, methods to determine the 2D and 3D structures for α-helical membrane proteins are relatively easier obtained as well as more reliable. The main reason for that is that α-helical membrane proteins generally contain a fixed pattern of highly hydrophobic consecutive residues forming the transmembrane α-helix. Additionally there is a more abundant set of α-helical membrane proteins whose 3D structure has been solved at the atomistic resolution (see also Sects. 2.2.1 and 2.3.1).

4.2.1 Working Principles of the 2D and 3D Structure Prediction Tools

First methods to predict the 2D and 3D structure of β-barrel shaped membrane proteins were based on hydrophobicity analyses. More concretely these primitive methods used sliding windows to identify the alternating patterns of hydrophobic/hydrophilic residues in a sequence of transmembrane strands [116, 117]. Some other methods based on comparing the amino acid sequence with a set of proteins with solved 3D

structure analyze the propensity of the sequence to adopt a given 2D and 3D structure based on propensity scoring criteria [118, 119]. The main disadvantage of these primitive methods developed during the 1980s and 1990s is the reduced protein training set used to derive these methods, which led to a very poor performance when the target protein has a very low similarity with respect to all the proteins in the training set (this can be the case for engineered membrane protein variants, especially when the altered sequence shows significant differences from the parent sequence).

However in the last years many more sophisticated prediction methods relying on a larger training set have appeared. These methods can be mostly divided in three large groups, *i.e.* Neural Networks (NN), Hidden Markov Models (HMM) and Support Vector Machines (SVM) [120].

The most efficient NN based methods are those that combine two consecutive NNs [121, 122]. In this method the first NN aims to build a first estimate of the structure from the amino acid sequence. Sliding windows give a score according to the propensity of the central amino acid to be within or without the membrane. When the sliding windows have screened the whole amino acid sequence the first NN generates as output a score for each amino acid according to their propensity to be within or without the membrane. This output is used as input for the second NN. The second NN refines the propensities assigned by the first NN and is able to capture amphipathicity in helices or length of the helices or β-strands.

HMM based methods are more sophisticated than NN based methods. HMM based methods include in the integrated model the sequence to structure and the sequence to sequence relationships, which are captured in two steps in NN based methods. HMM based methods can capture global patterns in the structure which for β-barrel membrane proteins is especially useful since the repeated pattern β-strand/periplasmatic loop/β-strand/extracellular loop can be potentially captured by such methods. However NN based methods perform better to recognize more local amino acid patterns.

SVM based methods classify patterns into two groups. These methods use a set of sliding windows similar to those used in NN based methods aiming at obtaining a structure from a given sequence.

HMM, NN and SVM methods all have their advantages and disadvantages. In order to obtain the maximum benefit from the advantages of each specific method several ensemble methods have been proposed. These methods use the output of individual predictors based on different methods and combine them in a weighted way. In this way the ensemble based final result will have better accuracy than the result obtained from each individual predictor because the majority vote will tend to cancel the errors of each individual predictor [120].

4.2.2 Tools Comparison

Up to date there are dozens of predictors available based on the aforementioned methodologies. The suitability of each predictor to correctly predict the 2D and 3D structures from a given amino acid sequence may vary, so it is strongly recommended to use more than one and compare the results. Some of the predictors allow the user to specify some conditions that must be satisfied. Examples are discarding β-strands shorter than three amino acids or "forcing" a specific part of the amino acid to adopt a specific 2D structure. Bagos and coworkers evaluated the performance of different predictors based on HMM, NN and SOV methodologies to predict the structure of 20 β-barrel membrane proteins with known crystal structure [123]. The accuracy of each predictor was assessed on basis of the per-residue accuracy; the segments overlap measure and the protein topology. The authors concluded that HMM based methods have a higher success ratio than NN and SVM based methods to predict both 2D and 3D structures. For NN and SVM based methods the use of only transmembrane β-barrel domains rather than the precursor full-length sequences led to a higher reliability of the results. HMM based methods however perform similarly whether the precursor full-length sequences are

Table 4.1 Suggested 2D and 3D predictors for β-barrel transmembrane proteins

Name	Website	Reference
ConBBPRED	http://bioinformatics.biol.uoa.gr/ConBBPRED/	[123]
BOCTOPUS	http://boctopus.cbr.su.se/	[124]
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	[125]
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	[126]
BOMP	http://services.cbu.uib.no/tools/bomp	[127]
PredictProtein	http://www.predictprotein.org/	[128]
TMB-Hunt	http://www.bioinformatics.leeds.ac.uk/betaBarrel/TMB_WEB/	[129, 130]

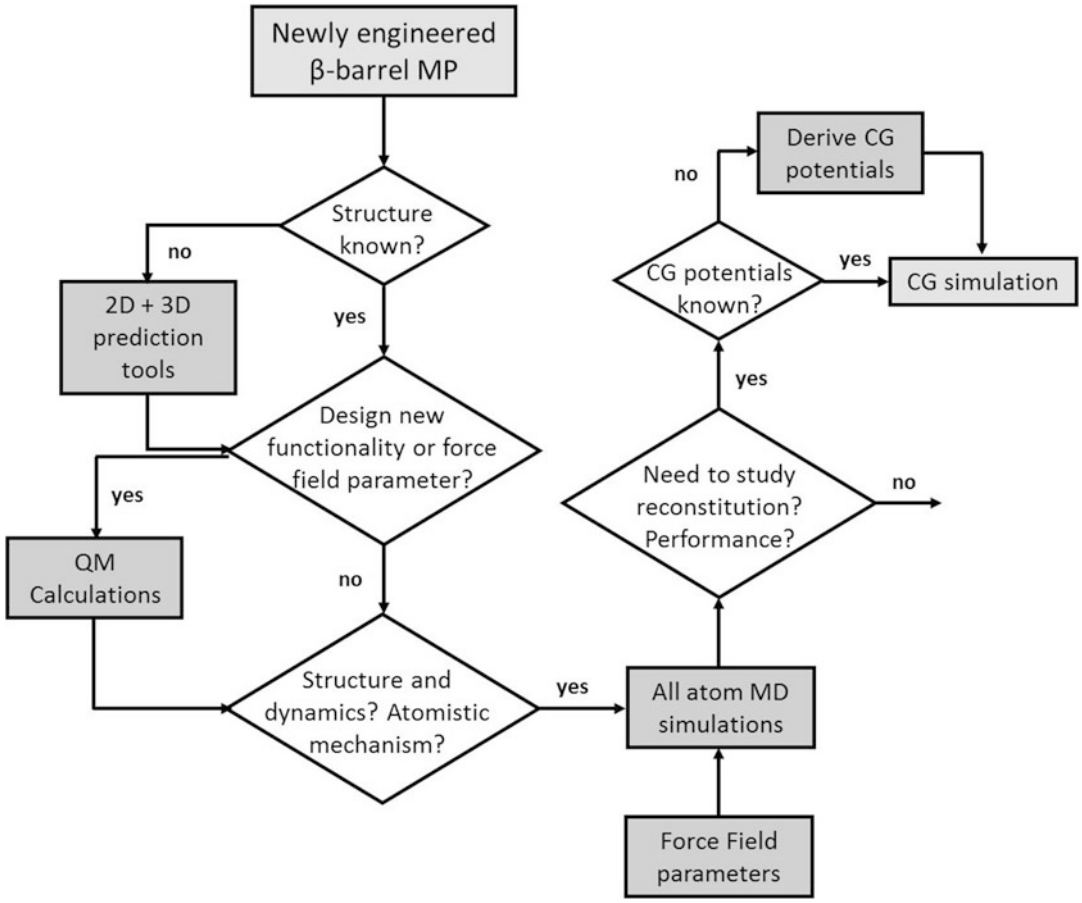

Fig. 4.4 General flowchart to design a new nanodevice based on engineered β-barrel membrane proteins by using bioinformatics tools and biomolecular modeling techniques

present or not. The authors developed a predictor based on an ensemble of various available predictor methods. The new ensemble method performed better than each of the individual predictors evaluated in their work.

As pointed before, there are many predictors with a different performance depending on the studied protein. Thus the use of only one predictor can lead to good or bad predictions depending on the protein. Hence it is strongly recommended to use different predictors. Based on their ability to predict the 2D and 3D structures of β-barrel transmembrane proteins a selection of suggested predictors is presented in Table 4.1.

To close and summarize this chapter, Fig. 4.4 summarizes a general strategy to design a new

nano-device based on β-barrel membrane proteins by using bioinformatics tools and multiscale biomolecular approaches.

In the following Chap. 5 the reader will learn more on how concretely to design novel nano-channel materials based on β-barrel outer membrane proteins and how to practically produce these "designer" proteins in the laboratory. The *E. coli* FhuA will receive particular attention as it can be seen as a model protein channel for protein engineering of nano-channels.

References

1. Jensen MO, Mouritsen OG (2004) Lipids do influence protein function-the hydrophobic matching hypothesis revisited. Biochim Biophys Acta 1666:205–226
2. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
3. VanGunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, Van Der Vegt NFA, Yu HB (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Ed 45:4064–4092
4. Roothaan CCJ (1951) New developments in molecular orbital theory. Rev Mod Phys 23:69–89
5. Hall GG (1951) The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials. Proc R Soc A 205:541–552
6. Hehre WJ, Stewart RF, Pople JA (1969) Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. J Chem Phys 51:2657
7. Binkley JS, Pople JA, Hehre WJ (1980) Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. J Am Chem Soc 102:939–947
8. Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. J Chem Phys 54: 724
9. Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023
10. Kendall RA, Dunning TH, Harrison RJ (1992) Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. J Chem Phys 96:6796–6806
11. Woon DE, Dunning TH (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. J Chem Phys 98:1358–1371
12. Peterson KA, Woon DE, Dunning TH (1994) Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $H + H2 \rightarrow H2 + H$ reaction. J Chem Phys 100:7410–7415
13. Hartree DR (1928) The wave mechanics of an atom with a non-Coulomb central field. Part I – theory and methods. Proc Camb Philos Soc 24:89–110
14. Fock V (1930) Approximation method for the solution of the quantum mechanical multibody problems. Z Phys 61:126–148
15. Binkley JS, Pople JA (1975) Moller-Plesset theory for atomic ground-state energies. Int J Quantum Chem 9:229–236
16. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:0618–0622
17. Coester F, Kümmel H (1960) Short-range correlations in nuclear wave functions. Nucl Phys 17:477–485
18. Paldus J, Shavitt I, Čížek J (1972) Correlation problems in atomic and molecular systems. 4. Extended coupled-pair many-electron theory and its application to BH3 molecule. Phys Rev A 5:50
19. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev B 136:B864
20. Painter GS (1981) Improved correlation corrections to the local-spin-density approximation. Phys Rev B 24:4264–4270
21. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin-density calculations – a critical analysis. Can J Phys 58:1200–1211
22. Perdew JP, Zunger A (1981) Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 23:5048–5079
23. Becke AD (1996) Density-functional thermochemistry.4. A new dynamical correlation functional and implications for exact-exchange mixing. J Chem Phys 104:1040–1046
24. Perdew JP (1986) Density-functional approximation for the correlation-energy of the inhomogeneous electron-gas. Phys Rev B 33:8822–8824
25. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868
26. Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation-energy. Phys Rev B 45:13244–13249
27. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formulation into a functional of the electron-density. Phys Rev B 37:785–789
28. Becke AD (1993) Density-functional thermochemistry. 3. The role of exact exchange. J Chem Phys 98:1372–1377
29. Adamo C, Barone V (1998) Exchange functionals with improved long-range behavior and adiabatic

connection methods without adjustable parameters: the mPW and mPW1PW models. J Chem Phys 108:664–675
30. Lynch BJ, Fast PL, Harris M, Truhlar DG (2000) Adiabatic connection for kinetics. J Phys Chem A 104:4811–4815
31. Lynch BJ, Zhao Y, Truhlar DG (2003) Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. J Phys Chem A 107:1384–1388
32. Schultz NE, Zhao Y, Truhlar DG (2005) Density functionals for inorganometallic and organometallic chemistry. J Phys Chem A 109:11127–11143
33. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098–3100
34. Xu X, Goddard WA (2004) The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. Proc Natl Acad Sci USA 101: 2673–2677
35. Miertuš S, Scrocco E, Tomasi J (1981) Electrostatic interaction of a solute with a continuum – a direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem Phys 55:117–129
36. Movileanu L (2012) Single-molecule detection of proteins using nanopores. In: Barth FG, Humphrey JAC, Srinivasan MV (eds) Frontiers in sensing: from biology to engineering. Springer, Wien/New York, pp 363–381
37. Bayley H, Cremer PS (2001) Stochastic sensors inspired by biology. Nature 413:226–230
38. Casanovas J, Preat J, Zanuy D, Alemán C (2009) Sensing abilities of crown ether functionalized polythiophenes. Chem Eur J 15:4676–4684
39. Casanovas J, Rodríguez-Ropero F, Zanuy D, Alemán C (2010) Microscopic details of the sensing ability of 15-crown-5-ether functionalized poly(bithiophene). Polymer 51:4267–4272
40. Kang X-F, Cheley S, Guan X, Bayley H (2006) Stochastic detection of enantiomers. J Am Chem Soc 128:10684–10685
41. Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic structure system. J Comput Chem 14:1347–1363
42. Gordon MS, Schmidt MW (2005) Advances in electronic structure theory: GAMESS a decade later. In: Dykstra CE (ed) Theory and applications of computational chemistry: the first forty years. Elsevier, Amsterdam, pp 1167–1189
43. Alder BJ, Wainwright TE (1957) Phase transition for a hard sphere system. J Chem Phys 27:1208–1209
44. Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. 1. General method. J Chem Phys 31:459–466
45. Rahman A (1964) Correlations in motion of atoms in liquid argon. Phys Rev A 136:A405
46. Stillinger FH, Rahman A (1974) Improved simulation of liquid water by molecular dynamics. J Chem Phys 60:1545–1557
47. Stillinger FH, Rahman A (1974) Molecular dynamics study of liquid water under high compression. J Chem Phys 61:4973–4980
48. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A 2nd generation force-field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 117:5179–5197
49. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21(2):1049–1074
50. Oostenbrink C, Villa A, Mark AE, VanGunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676
51. VanGunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag, Zürich
52. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24: 1999–2012
53. Kollman PA (1996) Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Acc Chem Res 29:461–469
54. García AE, Sanbonmatsu KY (2002) Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci USA 99: 2782–2787
55. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
56. MacKerellJr AD, Feig M, BrooksIII CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–1415
57. MacKerellJr AD, Banavali N, Foloppe N (2000) Development and current status of theCHARMM

force field for nucleic acids. Biopolymers 56: 257–265
58. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110:1657–1666
59. Jorgensen WL, Maxwell DS, Tirado-Rives J (1995) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
60. Chen J, Im W, Brooks CL (2006) Balancing solvation and intramolecular interactions: toward a consistent generalized born force field. J Am Chem Soc 128:3728–3736
61. Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. J Mol Liq 101:219–260
62. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
63. Mahoney MW, Jorgensen WL (1995) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112:8910–8922
64. Berendsen HJC, Postma JPM, VanGunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342
65. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. J Phys Chem 91:6269–6271
66. Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123:234505
67. Vega C, Abascal JLF (2011) Simulating water with rigid non-polarizable models: a general perspective. Phys Chem Chem Phys 13:19663–19688
68. Hoogerbrugge PJ, Koelman JMVA (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys Lett 19:155–160
69. Koelman JMVA, Hoogerbrugge PJ (1993) Dynamic simulations of hard-sphere suspensions under steady shear. Europhys Lett 21:363–368
70. He S, Scheraga HA (1998) Brownian dynamics simulations of protein folding. J Chem Phys 108:287
71. Verlet L (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 159:98–103
72. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of Cartesian equations of motion of a system with constraints – molecular dynamics of n-alkanes. J Comput Phys 23:327–341
73. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
74. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
75. Tironi IG, Sperb R, Smith PE, VanGunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459
76. Ewald PP (1921) The calculation of optical and electrostatic grid potential. Ann Phys 64:253–287
77. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106:1589–1615
78. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
79. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101
80. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 52:255–268
81. Hoover WG (1985) Canonical dynamics – equilibrium phase-space distributions. Phys Rev A 31:1695–1697
82. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals – a new molecular dynamics method. J Appl Phys 52:7182–7190
83. Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Mol Phys 50:1055–1076
84. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions – dielectric, electrostatic and steric stabilization of carbonium ion in reaction of lysozyme. J Mol Biol 103:227–249
85. Field M, Bash P, Karplus M (1990) A combined quantum-mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11:700–733
86. Maseras F, Morokuma K (1995) IMOMM – a new integrated ab initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition states. J Comput Chem 16:1170–1179
87. Ferré N, Assfeld X, Rivail JL (2002) Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. J Comput Chem 23:610–624
88. Guest MF, Bush IJ, VanDam HJJ, Sherwood P, Thomas JMH, VanLenthe JH, Havenith RWA, Kendrick J (2005) The GAMESS-UK electronic structure package: algorithms, developments and applications. Mol Phys 103:719–747
89. Lindahl E, Sansom MSP (2008) Membrane proteins: molecular dynamics simulations. Curr Opin Struct Biol 18:425–431
90. Jogini V, Roux B (2007) Dynamics of the Kv1.2 voltage-gated K(+) channel in a membrane environment. Biophys J 93:3070–3082
91. Akitake B, Anishkin A, Liu N, Sukharev S (2007) Straightening and sequential buckling of the

pore-lining helices define the gating cycle of MscS. Nat Struct Mol Biol 14:1141–1149
92. Faraldo-Gómez JD, Smith GR, Sansom MSP (2003) Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states. Biophys J 85:1406–1420
93. Rodríguez-Ropero F, Fioroni M (2012) Structural and dynamical analysis of an engineered FhuA channel protein embedded into a lipid bilayer or a detergent belt. J Struct Biol 177:291–301
94. Fritz D, Koschke K, Harmandaris VA, Van Der Vegt NFA, Kremer K (2011) Multiscale modeling of soft matter: scaling of dynamics. Phys Chem Chem Phys 13:10412–10420
95. Ghanbari A, Böhm MC, Müller-Plathe F (2011) A simple reverse mapping procedure for coarse-grained polymer models with rigid side groups. Macromolecules 44:5520–5526
96. Brini E, Algaer EA, Ganguly P, Li C, Rodríguez-Ropero F, Van Der Vegt NFA (2013) Systematic coarse-graining methods for soft matter simulations – a review. Soft Matter 9:2108–2119
97. Reith D, Pütz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem 24:1624–1636
98. Bedrov D, Smith GD, Yoon J (2007) Structure and interactions in micellar solutions: molecular simulations of pluronic L64 aqueous solutions. Langmuir 23:12032–12041
99. Li C, Shen J, Peter C, Van Der Vegt NFA (2012) A chemically accurate implicit-solvent coarse-grained model for polystyrenesulfonate solutions. Macromolecules 45:2551–2561
100. Fritz D, Harmandaris VA, Kremer K, Van Der Vegt NFA (2009) Coarse-grained polymer melts based on isolated atomistic chains: simulation of polystyrene of different tacticities. Macromolecules 42:7579–7588
101. Brini E, Marcon V, Van Der Vegt NFA (2011) Conditional reversible work method for molecular coarse graining applications. Phys Chem Chem Phys 13:10468–10474
102. Villa A, Peter C, Van Der Vegt NFA (2009) Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. Phys Chem Chem Phys 11:2077–2086
103. Bereau T, Deserno M (2009) Generic coarse-grained model for protein folding and aggregation. J Chem Phys 130:235106
104. Bereau T, Globisch C, Deserno M, Peter C (2012) Coarse-grained and atomistic simulations of the salt-stable cowpea chlorotic mottle virus (SS-CCMV) subunit 26–49: β-barrel stability of the hexamer and pentamer geometries. J Chem Theory Comput. doi:10.1021/ct200888u
105. Srinivas G, Shelley JC, Nielsen SO, Discher DE, Klein ML (2004) Simulation of diblock copolymer self-assembly, using a coarse-grain model. J Phys Chem B 108:8153–8160
106. Srinivas G, Discher DE, Klein ML (2004) Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics. Nat Mater 3:638–644
107. Srinivas G, Discher DE, Klein ML (2005) Key roles for chain flexibility in block copolymer membranes that contain pores or make tubes. Nano Lett 5:2343–2349
108. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111:7812–7824
109. Rossi G, Monticelli L, Puisto SR, Vattulainena I, Ala-Nissila T (2011) Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. Soft Matter 7:698–708
110. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ (2008) The MARTINI coarse-grained force field: extension to proteins. J Chem Theory Comput 4:819–834
111. Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ (2010) Polarizable water model for the coarse-grained MARTINI force field. PLoS Comput Biol 6:e1000810
112. Sansom MSP, Scott KA, Bond PJ (2008) Coarse-grained simulation: a high-throughput computational approach to membrane proteins. Biochem Soc Trans 36:27–32
113. Neri M, Anselmi C, Cascella M, Maritan A, Carloni P (2005) Coarse-grained model of proteins incorporating atomistic detail of the active site. Phys Rev Lett 95:218102
114. Shi Q, Izvekov S, Voth GA (2006) Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. J Phys Chem B 110:15045–15048
115. Rzepiela AJ, Louhivuori M, Peter C, Marrink SJ (2011) Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. Phys Chem Chem Phys 13:10437–10448
116. Vogel H, Jahnig F (1986) Models for the structure of outer-membrane proteins of Escherichia coli derived from Raman spectroscopy and prediction methods. J Mol Biol 190:191–199
117. Schirmer T, Cowan SW (1993) Prediction of membrane-spanning beta-strands and its application to maltoporin. Protein Sci 2:1361–1363
118. Gromiha MM, Ponnuswamy PK (1993) Prediction of transmembrane beta-strands from hydrophobic characteristics of proteins. Int J Pept Protein Res 42:420–431
119. Gromiha MM, Majumdar R, Ponnuswamy PK (1997) Identification of membrane spanning beta strands in bacterial porins. Protein Eng 10:497–500
120. Punta M, Forrest LR, Bigelow H, Kernytsky A, Liu J, Rost B (2007) Membrane protein prediction methods. Methods 41:460–474
121. Rost B, Casadio R, Fariselli P, Sander C (1995) Transmembrane helices predicted at 95-percent accuracy. Protein Sci 4:521–533

122. Rost B (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol 525:525–539
123. Bagos PG, Liakopoulos TD, Hamodrakas SJ (2005) Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method. BMC Bioinformatics 6:1471–2105
124. Hayat S, Elofsson A (2012) BOCTOPUS: improved topology prediction of transmembrane beta-barrel proteins. Bioinformatics 28:516–522
125. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40
126. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405
127. Berven FS, Flikka K, Jensen HB, Eidhammer I (2004) BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. Nucleic Acids Res 32:W394–W399
128. Rost B, Yachdav G, Liu J (2003) The PredictProtein server. Nucleic Acids Res 32:W321–W326
129. Garrow AG, Agnew A, Westhead DR (2005) TMB-Hunt: a web server to screen sequence sets for transmembrane beta-barrel proteins. Nucleic Acids Res 33:W188–W192
130. Garrow AG, Agnew A, Westhead DR (2005) TMB-Hunt: an amino acid composition based method to screen proteomes for beta-barrel transmembrane proteins. BMC Bioinformatics 6:56

The following chapter explains how to practically transform a β-barrel membrane protein (MP) into a nano-channel with desired geometrical and/or functional features, starting from the concept-design and design of the respective gene. It will then give an overview on the conventional means of production and purification of bacterial OMPs (outer membrane proteins), stressing on the problems and challenges of over-expressing OMPs into the Gram-negative bacterial outer membrane and of isolating them from the outer membrane. Furthermore the special problems of producing modified β-barrel MPs and ways to overcome these problems by using alternative methods will be named and explained. The different ways of analyzing OMP samples regarding yield, purity and correct folding will be considered briefly and the chapter discusses the distinctive experimental adaptations to scale-up the production of OMPs in general and of modified OMPs especially. As for many industrial applications of OMP derived nano-materials vast amounts of the proteins need to be produced, exceeding the capacities of conventional methods. The chapter will close with a discussion on artificial β-barrel structures to which OMPs are an alternative.

5.1 Outer Membrane Protein Modification

As addressed in Chap. 2 the β-barrel shaped integral OMPs of Gram-negative bacteria are well-suited for the nano-material design. From 10 β-strands on, these proteins form pores and channels that reach to quite substantial dimensions in strand-rich proteins such as the TonB-dependant transporters that harbor 22 strands as for instance the *E. coli* FhuA or 24 as the so-far largest known natural OMP PapC from *E. coli*. The robust barrel structure tolerates mutations and facilitates protein-refolding from the fully or partially denatured state. OMPs spontaneously insert into lipid or polymer membranes opening the possibility to design novel hybrid materials for various applications (see Chap. 6), well studied examples are nano-container systems, in which a protein nano-channel allows controllable compound release or biosensor systems in which a membrane reconstituted protein channel is used to monitor analytes. Analytes are detected by single channel conductance measurements. The special class of bacterial β-barrel pore-forming toxins is quite interesting in this respect, as they self-assemble from several monomers to form stable pores with stable conductance. Especially the heptameric α-hemolysin from *S. aureus* has been successfully employed in the design of nano-pore sensors for the detection of single molecules. Since α-hemolysin does not contain an intrinsic specificity for a certain molecule or a class of molecules and since specificity is the result of genetic engineering it can be utilized very flexibly [1].

In general and as logic dictates the OMP has to be chosen according to the desired application, in case of nano-sensors for instance

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_5,

proteins lacking motile loop-regions that interfere with measurable channel conductance are especially suited. Further characteristics depend on the analyte's nature, a huge and bulky analyte for example might necessitate a protein pore with larger inner diameter. For a hybrid-catalyst system with an OMP as reaction-specificity inducing scaffold for a metal catalyst instead, a more narrow pore might be of advantage, again depending on the substrate to be transformed [2]. In drug-release systems specificity (apart from size-dependent specificity) is generally not an issue as a nano-container enclosed drug simply has to be released at a target. In this case the ability to reversibly open or close an OMP channel is of greater importance [3, 4] and motile loops that might lead to a temporary channel-closing are again undesirable. All of these applications and further ones such as membrane systems or molecular sieves with certain cut-off might ask for larger than the naturally occurring protein channels. Depending on the lipid- or polymer-based hydrophobic materials used into which the OMPs should be reconstituted, the protein characteristics have to be adapted to the material, if an adaptation of the material is not possible or proves disadvantageous (for instance due to cost-ineffectiveness). As these examples show nano-material design often requires the change of a chosen OMP's natural features or the addition of new features. These changes can be either introduced by chemical modifications or by genetic-engineering or a combination of both.

Major changes as an altered geometry however will mostly require the introduction of mutations on gene level that are subsequently translated to protein level, while a triggered opening and closing of an OMP channel can be achieved by chemical labeling of adequate amino acid residues within the channel. It might however be necessary to introduce suitable amino acid residues by site-directed mutagenesis. Genetic engineering to transform an OMP into a nano-channel will generally be based on a rational approach.

A recent example shows the great potential of β-barrel proteins and their units (*i.e.* strands and sheets) for the creation of artificial nano-pores. De Pinto et al. reported the nano-pore design starting from natural β-barrel structures. A multi-alignment of general diffusion porins identified a highly conserved motif coding for two β-strands (obtained from *E. coli* OmpF) as the basic module of the artificial pores. Hexameric repeats of the respective sequence were obtained through cDNA recombinant technology. The coded protein was expressed, purified and reconstituted in planar bilayer membranes and showed channel-forming ability [5].

The following sections will explain how to modify OMPs for nano-channel design purposes starting out from gene-design considerations.

5.1.1 Gene-Design

The design of completely new genetic information is a very flexible and powerful way to obtain engineered or entirely novel proteins, especially so since the designed genes can nowadays be synthetized commercially at an affordable price of ∼0.25€/bp (in 2013) [6]. Furthermore the vast sequence information obtained by the various sequencing projects (as collected on the website of the "International Sequencing Consortium": http://www.intlgenome.org/viewDatabase.cfm), such as the *E. coli* genome project [7], allowed the development of computer based gene-design software tools helping the scientist to optimize a newly written DNA sequence. General design parameters and criteria that apply to the design of genes coding for soluble proteins apply just as well to MPs and OMPs. A good way to check whether a planned design is valid is the use of theoretical methods that can supply valid and important information. A comprehensive overview on computation methods and tools that can be applied to β-barrel membrane proteins and nano-systems based on these proteins can be found in Chap. 4.

Before designing a gene coding for a modified OMP one generally chooses the expression host (organism and strain). Considerations on how to choose the best suited system will be given in Sect. 5.2. After rationally deciding for positions to be mutated, the desired modification will be

introduced "on paper" using the amino acid sequence of the OMP to be engineered as a template. While point-mutations can be introduced by well-established PCR (polymerase chain reaction) techniques, as will be shown in Sect. 5.1.3, vast mutations that will effect a protein's overall geometry generally require the redesign of the respective open reading frame (ORF) and the subsequent synthesis of a synthetic gene. In theory the design of a new gene from a given amino acid sequence is rather simple, as the desired amino acid sequence has "just" to be reverse translated to the corresponding DNA sequence.

However the practical conversion of such an altered "virtual" amino acid sequence to a DNA sequence that is supposed to code for the desired protein is not so trivial, as there are certain prerequisites to be fulfilled:

1. The resulting protein has to be intact, correctly folded and stable. Sequence elements that are crucial for stability, folding or function should be known as they may not be removed or may be newly introduced.
2. The DNA sequence has to optimally fit the selected expression host in terms of codon usage. Each amino acid can be coded by one to six different codons and different organisms differ in their preference for the various codons [8]. Many companies that offer gene synthesis services will also optimize the gene sequence to the expression host codon bias.
3. For prokaryotic hosts: An important component affecting expression levels is the ribosome binding site (RBS) needed for translation initiation between 5 and 15 bases upstream of the AUG start codon. Sequence changes within the RBS might change expression levels substantially [9].
4. mRNA secondary structures have to be avoided, as they might occlude the RBS or start codon in prokaryotic expression hosts inhibiting translation [10].
5. Certain restriction sites might be desired for easy cloning, while others instead might interfere with cloning purposes.
6. The protein has to be produced in sufficient yield. This point can be addressed by using a strong promoter sequence, such as the phage derived T5 or T7 promoters.
7. For purification sequence tags, such as the Hexa-His- [11], glutathione-S-transferase (GST)- [12], streptavidin-binding-peptide (STREP- or STREP II)- [13, 14], FLAG-tag [15] or others can be introduced on DNA sequence level. To enhance solubility the attachment of tags like the maltose binding protein (MBP)- [16] or the N-utilization substance A (NusA) [17] can be useful. A comprehensive review on the different affinity and solubility tags and how they can be combined for best results can be found in [18].

There are various software kits and web platforms available that use information from known DNA sequences to assist the user in optimizing an engineered or newly written DNA sequence. While older tools focused mainly on the organism specific codon bias [19–23], recently developed tools, such as GeMS [24], Gene Composer [25], GeneOptimizer [26], Synthetic Gene Designer [27], GeneDesign [28], Visual Gene Developer [29] or Gene Designer [30] consider a combination of the above mentioned parameters.

Apart from these parameters that apply to the synthetic gene design in general certain issues have to be considered when designing a gene coding for a MP or OMP.

If the synthetic gene derived OMP should be expressed homologous and into its natural environment (the outer membrane), the N-terminal signal sequence that leads to outer membrane targeting (see Sect. 2.3.2) has to be added on gene level, otherwise the protein won't be folded correctly and will be expressed into inclusion bodies. Successful heterologous expression of MPs is often difficult [31], as membrane targeting might function poorly or might not occur at all and the protein might be toxic for the expression host cells [31–34]. To avoid or improve these issues lowering the expression level (*e.g.* by adding a weak promoter sequence) may be considered [34].

Several of the designing parameters listed can be also addressed by choosing the optimal expression plasmid. Especially the larger tags or

promoter sequences will be generally located on the plasmid used.

In conclusion though the design of truly novel genes is becoming more and more facile due to existing sequence knowledge applied to the software tool development, the knowledge on factors that might lead to unsuccessful protein expression (especially when expressing heterologous and especially when producing OMPs) is not yet sufficient to guaranty success in any case. The experimentalist will still have to test the virtually designed gene in laboratory experiments and considerable effort might still be needed to optimize the protein expression conditions (see Sect. 5.2). When a new protein is then effectively expressed it generally needs to be isolated and purified, which again might prove a rather challenging task, especially when working with MPs (see Sect. 5.2.4).

The following Sect. 5.1.2 will explain and give examples on how the OMP geometry can be altered by introducing changes to their "virtual" amino acid sequence in order to derive a synthetic gene from which to express the altered protein. The main example will be the *E. coli* FhuA protein.

5.1.2 Geometry Modification

Outer membrane β-barrel channel proteins like the *E. coli* FhuA can be an alternative to artificial chemically synthesized nano-pores (on artificial β-barrels see Sect. 5.4). In order to be useful as channel structures for nano-technological applications channel proteins must be flexible enough to be modified in their geometry, *i.e.* length and diameter depending on the application and hydrophobic carrier (*i.e.* lipid or polymer membranes or vesicles). Due to the mentioned robustness of the β-barrel structure that tolerates vast sequence mutations; major changes in channel geometry are rendered possible. As these geometry changes have been quite successfully carried out with the FhuA protein, it will serve as an example on how to mutate a TonB-dependent siderophore transporter protein with wide channel diameter to obtain a passive diffusion channel, whose diameter can be increased, whose hydrophobic membrane spanning region can be elongated and whose flexible loop regions can be partly removed leading to a more regular channel structure (For details on the FhuA WT protein structure and function see Sect. 2.4).

5.1.2.1 From Ferrichrome Transport to Passive Diffusion

The first mutations leading to a modification of the FhuA channel geometry were introduced before the protein was recognized as a nano-channel for biotechnological applications. Partial or total deletions of the proteins N-terminal plug-domain (amino acids 1–159) were carried out to reveal the plug function (see Fig. 5.1).

In 1999 Braun et al. created a FhuA deletion variant lacking amino acids 5–160, as they hypothesized that the resulting mutant FhuAΔ5–160 lacking most of the N-terminal plug should form a stable permanently open channel allowing diffusion of substances smaller than its inner diameter. Active transport of ferrichrome instead was thought to be defective, as the TonB-box (amino acids 7–11) thought to be vital for the energy-providing interaction with TonB, as well as the ferrichrome binding domains within the plug were missing [35]. They found that indeed FhuAΔ5–160 formed stable channels within the *E. coli* outer membrane, rendering cells sensitive to the large antibiotics erythromycin, rifamycin, bacitracin and vancomycin, and enabling them to grow on maltotetraose and maltopentaose in the absence of the LamB protein that is involved in the transport of maltose and maltodextrins. Furthermore cells lacking the TonB in the inner membrane but expressing FhuAΔ5–160 were able to grow on media containing high concentrations of ferrichrome, while cells lacking TonB and FhuA or lacking TonB and carrying FhuA WT were not. These findings confirmed the passive diffusion channel hypothesis. Interestingly though FhuAΔ5–160 still facilitated active ferrichrome transport (at 40 % of WT level) in cells with active TonB, leading to the conclusion that apart from the TonB-box amino acids there have to be other regions within the FhuA that interact with TonB [35]. The same had been

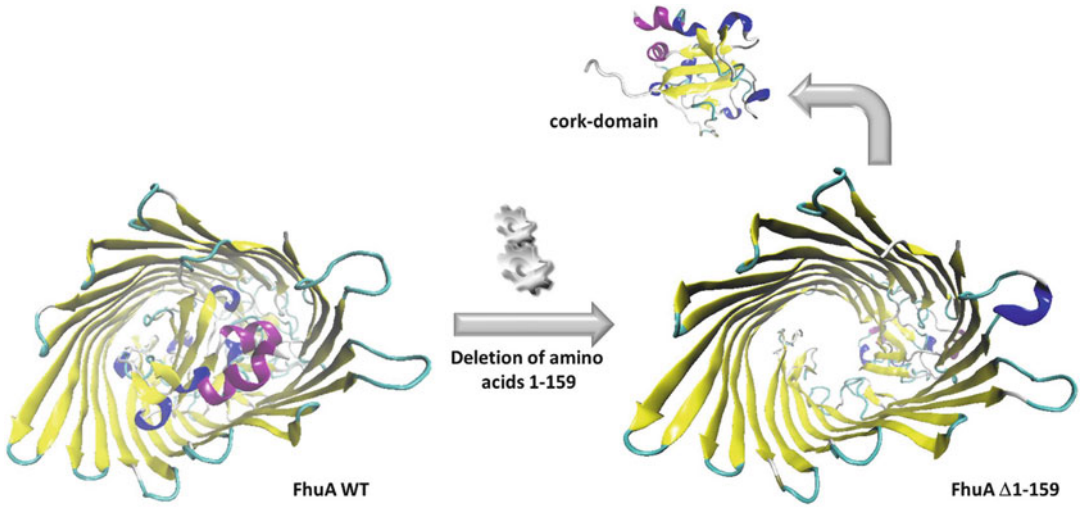

Fig. 5.1 Schematic representation of *E. coli* FhuA genetic engineering concept leading from the cork-closed WT protein to passive diffusion facilitating variant FhuAΔ1–159, by deletion of the cork-domain amino acids 1–159 (Proteins are shown in New Cartoon representation made with VMD)

found for plug-lacking variants of the Gram-negative bacteria *Salmonella paratyphi* B and *Salmonella enterica* serovar Typhimurium FhuA homologues [36] and for a plug-less *E. coli* FepA (TonB-dependant ferric enterobactin transporter; see Chap. 2) variant [37].

To further analyze the FhuAΔ5–160 features the protein was purified and reconstituted into black lipid bilayer membranes formed by diphytanoyl PtdCho/n-decane and the membrane current was recorded by high-resolution, single-channel electrical recordings in 1 M KCl. Results showed that the reconstituted protein led as expected to an increased conductance of 0.5 nS in 1 M KCl, however the recordings showed a high degree of noise, leading to the conclusion that the channels were not permanently open [38], most likely due to the long extracellular, motile loops of the protein that might temporarily close the channel.

In a later study by Nallani et al. the fully or partly cork-depleted FhuA was first recognized as a robust protein channel with wide diameter, facilitating passive diffusion and thus a perfect protein-based nano-material. Here the FhuA N-terminal amino acids 1–159 were deleted (Fig. 5.1) leading to variant FhuAΔ1-159 and the deletion of amino acids 1–129 led to FhuAΔ1-129. Both variants had been successfully reconstituted into ABA triblock copolymer vesicles, where A is poly(dimethylsiloxane) and B is poly(2-methyloxazoline) (PMOXA-PDMS-PMOXA). The passive diffusion of single stranded DNA-oligomers through the protein channels was reported and it was shown that mutant FhuAΔ1-159 revealed a higher translocation efficiency [39]. The FhuAΔ1-159 potential as nano-material for biotechnological applications was further demonstrated when it was found that the protein residing in PMOXA-PDMS-PMOXA polymer vesicles (nano-compartments) can be used to selectively recover and trap negatively charged molecules such as sulforhodamine B within the vesicles, by the use of positively charged poly-lysine traps enclosed in the nano-compartment. It was furthermore reported that the same FhuAΔ1-159 functionalized nano-compartment system can be used for the enzymatic conversion in nano-compartments, shown on the example of 3,3′,5,5′-tetramethylbenzidine (TMB) oxidation by vesicle enclosed horseradish peroxidase (HRP) [40]. Sulforhodamine B and TMB are thus able to diffuse through the large FhuAΔ1-159 pore. In 2009 the same FhuA variant was shown to facilitate diffusion of the

fluorescein derivative calcein again after being reconstituted into PMOXA-PDMS-PMOXA nano-compartments [3].

Though the FhuAΔ1-159 has not been crystallized up to now, analysis by circular dichroism (CD) spectroscopy showed a clear β-structure (49–65 % contribution of β-sheet content) [3, 4, 41], while for the FhuA WT a CD derived β-sheet content of 51 % had been reported [42]. Furthermore a recent molecular dynamics (MD) simulation study on the FhuA WT and FhuAΔ1–159 performed in a DNPC (1,2-dinervonyl-sn-glycero-3-phosphocholine) lipid bilayer and a water/OES (N-octyl-2-hydroxyethyl sulfoxide) detergent solution revealed that the mutant protein shows a remarkable stability in both environments independent from the presence of the cork domain [43].

Though the cork-less FhuA variants usefulness as nano-channel, in nano-compartment systems was further strengthened by the development of a reduction-triggerable [3] or light-triggerable [4] opening mechanism of chemically labeled FhuAΔ1–159 proteins, as will be discussed in more detail in Sect. 5.1.3, the problem of the flexible extracellular loops interrupting channel diffusion properties remained and required for the design of new variants.

5.1.2.2 Smoothing Channel Ends for a Permanently Open Diffusion Channel

As it was known from the FhuA WT crystal structure that one of the long extracellular loops, loop 4 (L4, amino acids 318–339; see Fig. 5.2) partially constricts the channel entrance and decreases its diameter to about half the area of the total cross section [44], it was thought that the removal of the cork domain in mutant FhuAΔ5–160 might lead to an increase in loop-flexibility, explaining the instabilities in single-channel potassium conductance the protein variant revealed. Braun et al. therefore deleted L4 amino acids 322–336 in FhuAΔ5–160, ending up with the double mutant FhuAΔ5–160 Δ322–336. However the variant, as the FhuAΔ5–160 before, increased the conductance of diphytanoyl PtdCho/n-decane lipid bilayer membranes but did not show uniform single-channel conductance. Instead it revealed rapid alternating channel opening and closing [38] and again the active ferrichrome transport mechanism was still intact. Construction of a further variant FhuAΔ5–160 Δ335–355 (deletion of one-third of L4 and half of transmembrane (TM) strand 8) showed uniform single-channel conductance of 2.5 nS in 1 M KCl (2 nS higher than the FhuAΔ5–160 channel conductance), while it did not facilitate active ferrichrome transport [38].

Endriß and Braun consecutively deleted the FhuA extracellular loops in the WT protein and replaced them by the short peptide sequence NSEG(S), to reveal their functions. They found that deletion of L3 or 11 inactivated active ferrichrome transport, deletion of L8 removed receptor activity for colicin M and the phages T1, T5, and φ80, while deletion of L7 removed only colicin M receptor function. Removal of L4 caused resistance against phages T1 and φ80 [45].

While these first FhuA loop-deletion studies were still concerned with the analysis of the protein function, later works were oriented towards the FhuA application in the development of a stochastic single molecule sensing element. Mohammad et al. realized the FhuA's value for the purpose at hand [46]. Though the suitability of OMPs in general had been recognized before, the so far used proteins, *S. aureus* pore-forming toxin α-hemolysin [1], *E. coli* porin OmpG [47] and *E. coli* porin OmpF [48], hold certain disadvantages as they are multimeric proteins making it difficult to engineer them (For more information on these MPs see Sect. 2.3). It was reported that the heptameric character of α-hemolysin leads to many combinations of engineered and WT monomers, complicating the separation of a desired modified single sub-unit [49]. Though recently a mutation study led to the development of a monomeric OmpF variant [50]. Furthermore the crystal structure derived inner channel diameters of all three proteins are rather small at the point of highest constriction; one finds for α-hemolysin: ∼15 Å [51], for OmpG

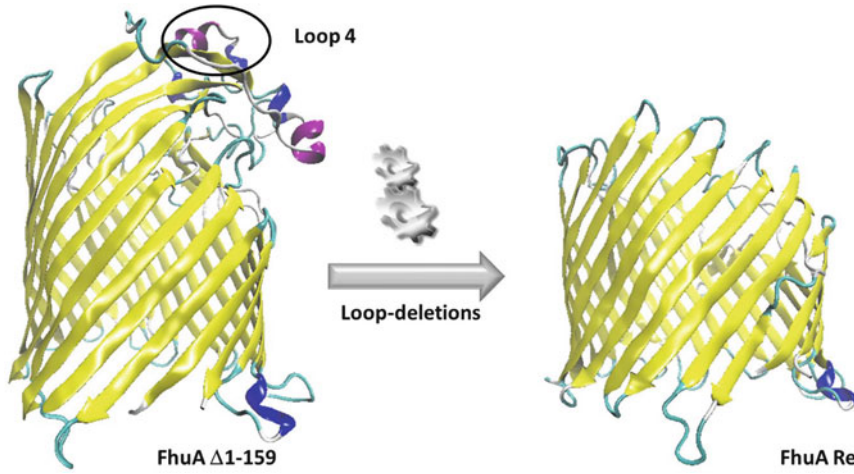

Fig. 5.2 Schematic representation of further genetic engineering concept starting with FhuAΔ1–159 theoretically leading to regular channel variant (FhuA Reg) lacking 10 of the 11 flexible external loops that tend to close the channel. The first ever deleted loop (L4) is marked by a *black circle* (Proteins are shown in New Cartoon representation made with VMD)

∼13 Å [52] and for OmpF: ∼6 Å [53, 54], allowing the passage of small, non-bulky molecules with a maximum molecular weight of ∼700 Da. The FhuA protein combines the general β-barrel advantages with further inborn advantages, such as its monomeric character and much wider elliptic inner channel diameter of 39 × 46 Å [44, 55]. Mohammad et al. therefore deleted apart from the plug amino acids 1–160 four major extracellular loops (L3, L4, L5, and L11) including L4 that were thought to interrupt channel conductance of FhuAΔ1–160. L3 (amino acids 243–274) and L5 (amino acids 394–419) are large flexible loops folding back into the plugless channel interior, L11 (amino acids 482–704) is another long loop that reaches into the pore interior [56]. Loops were deleted and replaced by short turns of the sequence NSEGS. The resulting mutant FhuAΔC/Δ4L revealed a high unitary channel conductance in 1 M KCl of ∼3.9–4.9 nS [46, 57]. Results show that the deleted loop regions actually do interfere with channel conductance, as their removal led to a ∼9.5-times increase in channel conductance as compared to FhuAΔ1–160. Furthermore conductance occurred uninterrupted over a long time span [57] and as the plug-domain deletion seems already a rather huge protein modification, in the mutant FhuAΔC/Δ4L almost one third of all FhuA WT amino acid residues has been removed, indicating to what extend the FhuA structure tolerates protein engineering measures. In a later study the same group showed that the FhuAΔC/Δ4L structure and conductance remained stable (in contrast to the α-hemolysin) under harsh conditions, such as acidic pH, low ion concentration or temperatures up to 65 °C. As a proof of concept that the new FhuA variant may be used as a molecular sensor, it had been reconstituted into planar lipid membranes and was used to monitor the pepsin digest of immunoglobulin G (igG) (fragments of digested igG led to short current interruptions) or to monitor the interaction between the retroviral gag nucleocapsid (NCp7) protein with a DNA-aptamer. While free NCp7 blocks the protein channel, interaction with the DNA-aptamer leads to a channel opening [57].

In an unpublished study an attempt toward an even more drastically modified loop-deletion variant of the cork-less FhuA was made [58]. The engineering concept is shown in Fig. 5.2. The FhuAΔ1–159 sequence was used as a template to design a synthetic gene in which apart from the very short loop L1 all 11 extracellular loop-coding regions were cut leading to variant FhuAΔ1–159 Reg (for *reg*ular channel structure),

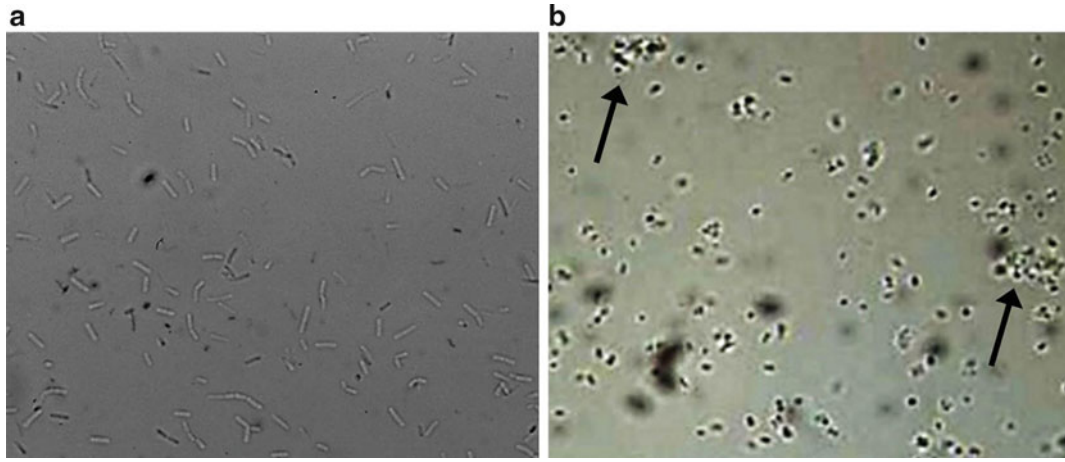

Fig. 5.3 Microscopic images at a 400-times magnification of (**a**) non-induced *E. coli* BE strain BL 21 (DE3) omp8 with plasmid pET22b + −FhuA Reg, showing characteristic rod-shape; (**b**) *E. coli* BE strain BL 21 (DE3) omp8 with plasmid pET22b + −FhuA Reg after induction, showing altered spherical morphology and formation of cell aggregates (*black arrows*)

or short FhuA Reg. Loop amino acid residues were deleted apart from terminal non-β-strand loop residues (3–5 amino acids), these residues were kept as β-strand connections (compare FhuA topology in [44]). Including the removed cork-domain 321 amino acid residues had been deleted, leading to a protein with 413 amino acids and an expected molecular weight of ∼46 kDa.

The synthetic gene had been obtained from GeneArt (Regensburg, Germany) and cloned into *E. coli* expression vector pET22b+ (Novagen, Merck, Darmstadt, Germany). Expression of the protein was attempted as previously described [40] using *E. coli* strain BL 21 (DE3) omp8 (F- hsdSB (rB- mB-) gal ompT dcm (DE3) ΔlamB ompF::Tn5 ΔompA ΔompC). The general OMP expression methods will be discussed in greater detail in Sect. 5.3. However no over-expressed protein with the expected size could be isolated from the outer membrane. It was observed that expressing *E. coli* cells showed an altered spherical cell morphology (Fig. 5.3a), while non-expressing cells showed the typical rod-shape (Fig. 5.3a). In liquid culture expressing cells tended to form large macroscopic aggregates.

These changes in *E. coli* cell morphology are known for cells that encounter defects in their cell division mechanism due to mutations, resulting in cells forming spheres that do not grow or propagate further [59] and also *E. coli* cells lacking the outer membrane turn elliptical or spherical [60]. The so-called L-form *E. coli* are spherical, osmo-sensitive cells due to a partial or total loss of their cell envelope occurring either spontaneous or after treatment with β-lactam antibiotics that inhibit murein synthesis [61]. In the case at hand the morphological changes were most likely due to the FhuA Reg over-expression and they might be connected to osmotic imbalances upon the insertion of the regular and expected to be permanently open channel protein into the outer membrane or to other toxic effects of the protein. A similar effect has been reported for the *Rhodobacter blasticus* porin; its expression into the outer membrane of *E. coli* resulted in cell lysis before the protein could be produced in sufficient amount [62]. The same phenotype as for FhuA Reg had been observed when expressing an FhuAΔ1–159 variant with increased channel diameter, as will be explained in the following. In that case the problem could be overcome by the optimization of culture conditions including a change of growth media. A possible approach to produce FhuA variant FhuA Reg might be its expression into inclusion bodies with subsequent

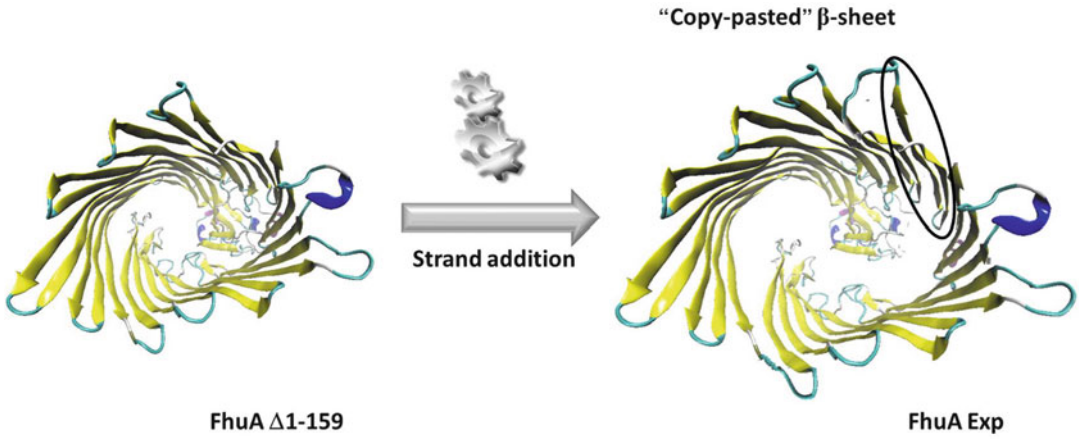

Fig. 5.4 Schematic representation of further genetic engineering concept again starting from FhuAΔ1–159 leading to a channel variant with expanded channel diameter (FhuA Exp), in which the last two N-terminal β-strands have been duplicated (*black circle*) (Proteins are shown in New Cartoon representation made with VMD; the FhuA Exp image has been edited using GNU image modification program GIMP 2.6 to add the duplicated sheet)

refolding to native state or a cell-free expression approach in the future (see also Sect. 5.2.1).

A similar loop deletion study had been carried out recently on the OmpF porin of *Yersinia pseudotuberculosis*. The deletion of three major loops L1, L6 and L8 however did not lead to a change in channel conductivity, but rather had an effect on the antigenic structure of the mutant porins, as revealed by immune-blotting and ELISA. Protein variants were expressed heterologous in *E. coli* and obtained from inclusion bodies [63].

5.1.2.3 Increasing the Number of β-Strands for a 0.4 nm Wider Channel Diameter

As the FhuA and some of its variants had been found useful as nano-channel biosensors or to functionalize artificial lipid or polymer vesicle systems and as its tolerance towards extensive sequence mutations (especially deletions) had been shown in several studies, the attempt to further increase the channel diameter of variant FhuAΔ1–159 by inserting β-strand-forming amino acids seemed not too far-fetched. This was further backed up by a study in which the three-stranded β-sheet of *Borrelia burgdorferi* protein OspA was successfully extended by two further strands by duplicating a beta-hairpin that formed a new sheet, indicating that the addition of new sheets to an OMP should be possible [64, 65]. A so derived increase in inner channel diameter would further widen the spectrum of substances that can be translocated through the protein channel.

The gene was designed considering the following concept: As a proof of concept only one β-sheet (two strands) was to be added, without introducing completely new genetic information, leading to a barrel with 24 strands as it can be found also in nature [66]. Therefore the amino acid sequence of one existing sheet was to be copied from the template protein FhuAΔ1–159 and pasted into the sequence of the new protein FhuAΔ1–159 Exp (for *exp*anded channel diameter), or short FhuA Exp. For this purpose the first two N-terminal strands (30 amino acids) were chosen for duplication as they are connected by the short loop L1, in this way the protein N-terminal sequence is conserved and N- and C-terminus are still expected to close by hydrogen bonding to form the intact barrel [67]. The engineering concept is shown in Fig. 5.4. The resulting protein has 628 amino acids and an expected molecular weight of 66.3 kDa.

When assuming a simple regular polygonal geometry (hendecagon for FhuAΔ1-159 and dodecagon for FhuAΔ1-159 Exp), with constrained side length given by the β-sheets connecting hydrogen bonds, the expected diameter increase can

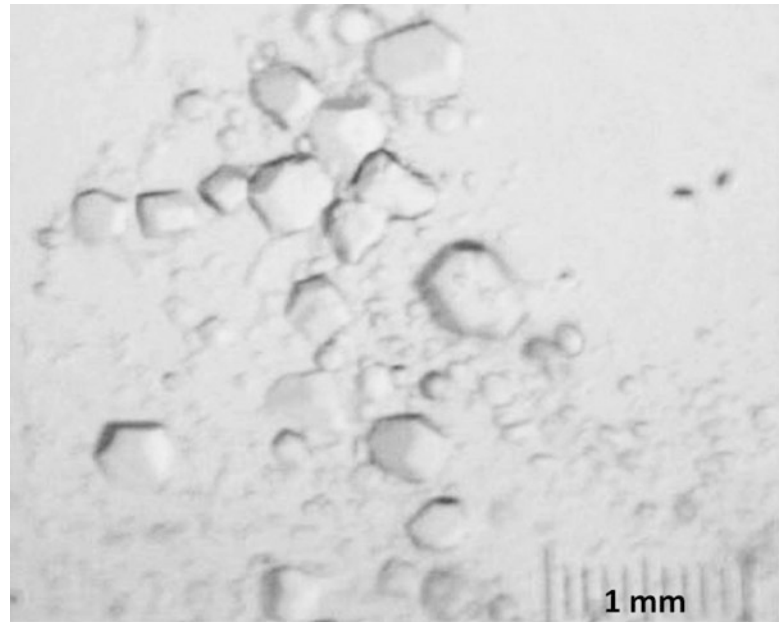

Fig. 5.5 Microscopic image of FhuA Exp pre-crystals

be estimated: The crystal structure derived FhuA WT inner channel diameter is ∼4.2 nm [56]. Based on the FhuA WT apothem, the expected FhuAΔ1-159 Exp inner channel cross section is ∼4.6 nm as calculated from the apothem ratio, resulting in a 16 % increase in channel surface area [67].

The on-paper reverse-translation derived synthetic gene was expressed in *E. coli* BL21 (DE3) omp8 using standard expression conditions as used for FhuAΔ1–159 [40]. However for the FhuA Reg variant no protein of the expected size could be isolated and cells again showed the described, spherical morphology (see Fig. 5.3). Obtained optical densities were poor (unpublished results, [68]). However after optimizing media and growth conditions (see Sect. 5.2.1 for more details) the protein was over-expressed without affecting the expression host cells and could be isolated from the outer membrane.

Channel functionality was verified after protein reconstitution into lipid vesicles, by measuring TMB-conversion enzyme kinetics of vesicle entrapped horseradish peroxidase (HRP), as described in [40]. The channel proved to be functional and TMB conversion occurred ∼17 % faster than with reconstituted FhuAΔ1–159 correlating with the 16 % increase in inner pore surface area. The FhuA Exp structural integrity had been verified by CD spectroscopy, revealing 63 % β-sheet contribution [67]. In a first attempt to crystallize FhuA Ext to analyze its three dimensional structure by X-Ray diffraction, pre-crystals were obtained (see Fig. 5.5) after an intensive screening of more than 600 different crystallization conditions carried out by Prof. E. Mizohata (Division of Applied Chemistry, Graduate School of Engineering, Osaka University) (unpublished results; [68]).

In conclusion the study demonstrated that the FhuA structure tolerated larger insertion mutations as well and that the β-sheet secondary structure information is fully contained in the existing FhuA amino acid sequence, therefore the simple "copy-paste" strategy was successful and led to a protein with increased inner channel cross-section and it seems likely that insertion of further strands should be possible. To the author's knowledge no similar approach of increasing the inner pore diameter of an OMP has been yet followed in other works.

Purified variant FhuA Exp was as mentioned successfully reconstituted into lipid vesicles (of *E. coli* lipid extract) [67] and it was

attempted to reconstitute the protein into BAB triblock copolymer vesicles, formed by the commercially available, cost effective polymer PIB_{1000}-PEG_{6000}-PIB_{1000} (PIB = Polyisobutylene; PEG = Polyethylene glycol), forming impermeable membranes with a hydrophobic thickness of 5 nm [69]. However FhuA Exp did not insert into PIB_{1000}-PEG_{6000}-PIB_{1000} membranes, due to their thick hydrophobic region and the resulting hydrophobic mismatch, the same had been reported for FhuAΔ1–159 [69]. However the use of impermeable, commercially available and biocompatible polymers (*e.g.* PIB_{1000}-PEG_{6000}-PIB_{1000}) can be of great advantage for the design of functionalized nano-compartments, therefore OMP variants that can be reconstituted into such polymer membranes are desirable. A FhuAΔ1–159-based variant was especially engineered for this purpose (see below).

5.1.2.4 Overcoming the Hydrophobic Mismatch

Polymer vesicles or polymersomes can as lipid vesicles (liposomes) be utilized as nano-sized encapsulation devices for applications such as delivery systems, bio-mimetic membranes, biomedical imaging tools, as protection devices for labile substances or as nano-reactors for sealed-in chemical or enzymatic reactions [70]. In contrast to liposomes, polymersomes formed by self-assembling synthetic amphiphilic block copolymers have been reported to possess superior biomaterial properties, such as better chemical and physical stability and impermeability [71]. The polymersome (or polymer membranes in general) functionalization by inserted MPs is therefore a potent way to design new hybrid materials, such as protein-polymer nano-compartment systems.

As mentioned before the FhuAΔ1–159 had been successfully reconstituted into PMOXA-PDMS-PMOXA polymer vesicles [3, 40]. The PMOXA-PDMS-PMOXA polymer had been synthetized to be used for the formation of hollow sphere (polymersome) structures with biomimetic membranes [72, 73]. It had been found suitable for the functional reconstitution of *E. coli* OmpF [74] before it was used in combination with FhuAΔ1–159. Here the standard strategy for the functional reconstitution of MPs into polymeric membranes was followed. This strategy includes the specific design of a polymer membrane's characteristics to fit the protein to be reconstituted. The membrane should be as thin and as fluid as possible, to minimize the energetic penalty when exposing a nonpolar/polar interface. However membranes formed by block copolymers can be rather thick, as they vary in thickness from 5 to 22 nm, while "natural" phospholipid membranes are only 3–4 nm thick. The polymer membrane greater thickness may cause problems in protein insertion, due to the hydrophobic mismatch between MP transmembrane regions and hydrophobic membrane parts [75]. Since block copolymers that assemble to membrane systems with thick membranes or hydrophobic portion might still have otherwise desirable features (such as mechanical stability, biocompatibility, cost effectiveness, or commercial availability), a protein nano-channel that can be tailored to fit the polymer is of superior advantage.

For a first proof of concept the FhuA protein was chosen for its robust structure and the possibility to introduce vast mutations, and a BAB triblock copolymer PIB_{1000}-PEG_{6000}-PIB_{1000} was chosen for its commercial availability and self-assembling ability. The PIB_{1000}-PEG_{6000}-PIB_{1000} polymer membrane hydrophobic part is ∼5 nm thick, while the FhuAΔ1-159 hydrophobic portion is only ∼3 nm thick [69]. Therefore based on FhuAΔ1-159 as a template, a new variant sequence was planned by "copy-pasting" the last five amino acids of each β-strand, leading to a total insertion of 110 amino acids thus increasing the expected hydrophobic length by 1 nm to reduce the hydrophobic mismatch of FhuA insertion into PIB_{1000}-PEG_{6000}-PIB_{1000} polymersomes (design concept shown in Fig. 5.6).

The resulting protein FhuAΔ1-159 Ext (for *ext*ended channel length), or short FhuA Ext, consisted of 665 amino acids with an expected molecular weight of ∼74 kDa. The protein was obtained from the outer membrane after cloning

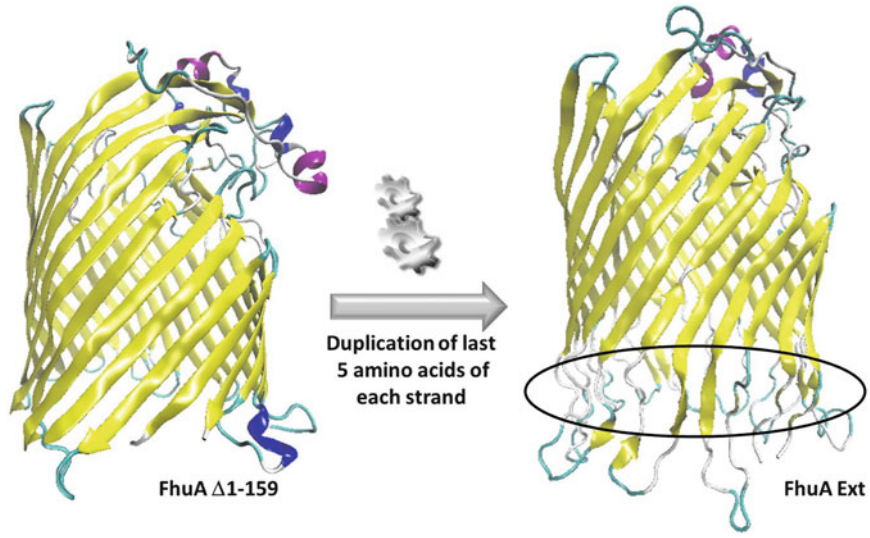

Fig. 5.6 Schematic representation of genetic engineering concept starting from FhuAΔ1–159 leading to channel variant with extended channel length (FhuA Ext), by duplicating the last five amino acids of each β-strand on the periplasmic protein side (*black circle*) (Proteins are shown in New Cartoon representation made with VMD)

the synthetic gene and expressing the protein in *E. coli* BL21 (DE3) omp8. The standard FhuAΔ1-159 membrane extraction and protein solubilization protocol [40] had to be modified due to the proteins increased hydrophobicity [69] (for details see Sect. 5.2.1). Again structural integrity could be verified by determining secondary structure by CD spectroscopy and the determined β-sheet content was 75 %. In contrast to FhuAΔ1-159 and FhuA Exp variant FhuA Ext was successfully inserted into PIB_{1000}-PEG_{6000}-PIB_{1000} polymersome membranes, as determined by the already mentioned TMB conversion kinetics measurements after TMB translocation through the polymer embedded protein channel [69]. Figure 5.7 shows schematically how the elongated FhuA Ext is expected to reside in membranes of the PIB-PEG-PIB type, with the protein TM region thickness overlapping better with the thickness of the hydrophobic membrane portion (top) than it is the case for the shorter FhuAΔ1–159 [69].

In order to determine the PIB_{1000}-PEG_{6000}-PIB_{1000} vesicle size and shape with and without reconstituted FhuA Ext protein, dynamic light scattering (DLS) measurements were carried out after polymersome purification (details on the DLS method can be found in Sect. 6.1.2), showing that plain polymer vesicles were spherical and had an average diameter of ∼242 nm (when in presence of detergent 2-Hydroxyethyloctylsulfoxide – OES, used to solubilize FhuA Ext), while vesicles with reconstituted protein (in 2-Hydroxyethyloctylsulfoxide) had a larger diameter of ∼279 nm on average and were also spherically shaped (unpublished data; [76]). The results are shown in Fig. 5.8. Diameters of polymersomes with reconstituted FhuA Ext were ∼37 nm larger than diameters of plain polymersomes. This finding was rather interesting, as from it might be obtained a first clue on the number of proteins interacting with a polymer vesicle, in case the membrane thickness (and polymer volume) remains constant in presence of or without the protein. As such information cannot be obtained by simple DLS measurements; in the future multi-angle light scattering measurements instead might provide information on the membrane thickness and shape. Other analytical methods such as Cryo TEM may be equally suited to compare the polymersome membrane thickness with and without channel protein (For an overview on the

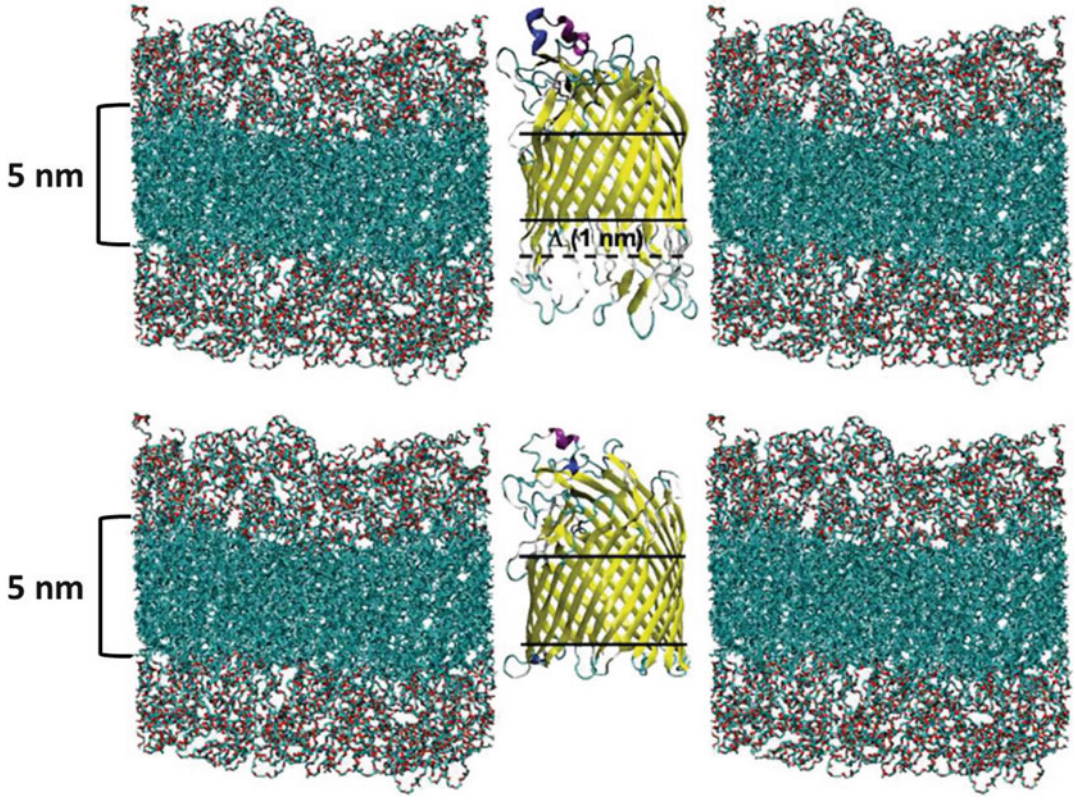

Fig. 5.7 Schematic representation of FhuA Ext (*top*) and FhuAΔ1-159 (*bottom*) within triblock copolymer PIB_{1000}-PEG_{1500}-PIB_{1000} membranes. The hydrophobic TM regions of FhuAΔ1-159 Ext (4 nm) and FhuAΔ1-159 (3 nm) are indicated by *lines*; the duplicated part of FhuAΔ1-159 Ext is indicated by a *broken line* (Proteins are shown in New Cartoon representation made with VMD) [69]

state of the art analytical methods to characterize protein functionalized membrane systems see Chap. 6, a focus on DLS methods is given in Sect. 6.1.2).

None of the described FhuA variants has been crystalized and analyzed by X-ray diffraction analysis, yet. Nevertheless the functional analysis by indirectly monitoring substrate/product diffusion through liposome or polymersome reconstituted channel proteins, using a vesicle enclosed enzyme or by more accurate single channel conductance measurements or CD spectroscopy derived information on protein secondary structure can give good clues on the correct folding of the new variants.

Table 5.1 gives a summary of secondary structure information obtained by CD spectroscopy for FhuA WT and all variants with changed channel geometry. Each of the variants shows a rather high amount of β-structure indicating that the barrel structure is retained. A definite proof on the accurate folding and tertiary structure of a newly expressed protein variant however can be obtained by X-ray diffraction analysis of protein crystals, protein NMR or Cryo-EM (see also Chap. 3).

5.1.3 Modifications for Chemical/Physical Triggering and Specificity

Especially for drug-release or more general compound-release systems composed of a lipid or polymer membrane or vesicles and an OMP channel, a triggerable (irreversible) or

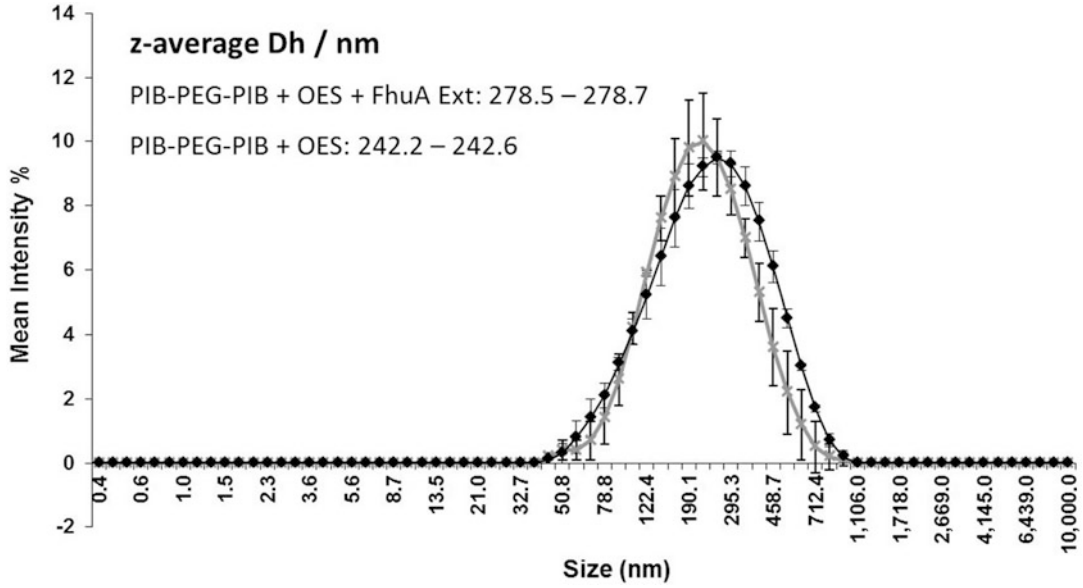

Fig. 5.8 DLS measurement results for PIB_{1000}-PEG_{6000}-PIB_{1000} vesicles harboring FhuA Ext (solubilized in OES detergent) and plain polymersomes in presence of OES (*grey crosses*). Z-average hydrodynamic diameters (Dh) are given in nm for both samples

Table 5.1 CD spectroscopy derived secondary structure amounts of FhuA WT and variants with changed channel geometry

FhuA variant	% β-structure	% α-helix	% random coil	Reference
FhuA WT	51	2	47	[42]
FhuA Δ1–159	49–65	3–13	23–37	[3, 4, 41]
FhuA Δ C/Δ 4L	60.3	3.7	37.2	[46]
FhuA Exp	63	7	30	[67]
FhuA Ext	75	20	5	[69]

switchable (reversible) channel opening/closing mechanism is desirable. Drugs or chemical substances contained in a nano-compartment can be transported to a particular target, at which a triggered channel opening leads to compound release, without the nano-compartment being destroyed (see also Sect. 6.2.1). As proteins are composed of amino acids of which some expose reactive side-chains the most obvious way to introduce an opening/closing mechanism to a channel protein is to selectively and chemically modify certain amino acid side chains with bulky substances that are able to efficiently seal the protein pore. The bond has to be cleavable upon an external stimulus, so to release the bulky label resulting in channel opening. Proteins as all biological molecules restrict chemical labeling reactions to certain conditions that ensure the maintenance of biomolecule integrity. Chemical reaction to modify proteins generally have to be performed in aqueous environment (OMPs often tolerate low concentrations of organic solvents), at ambient temperature and preferably at neutral pH.

For OMPs with a narrower inner channel diameter unlabeled amino acids can be used to introduce channel opening/closing switches, as amino acids, based on their pK_a values, answer with a charge change upon changes in pH values. This purpose however generally requires for protein modification by site-directed mutagenesis, *e.g.* the introduction of a certain amino acid at multiple inside facing positions, so to form a ring at the channel constriction site [77, 78].

Such amino acid substitutions by site-directed mutagenesis are not only useful in compound-release applications but furthermore are of importance for OMP based biosensor systems, too. The addition of charged rings to an OMP pore inside for instance can be used to detect charged molecules [78]. Single substitutions might change the specificity of OMPs that specifically transport one substance or substance class [79].

5.1.3.1 Chemical Modification (CM)

Several amino acid side chains are suitable to be chemically modified due to their reactive nature. The primary amine moiety of the amino acid lysine for instance can act as a nucleophile and forms amides with N-hydroxysuccinimide (NHS) esters, releasing the NHS. As the reaction occurs spontaneously at pH 8–9 and as no reagents such as bases are necessary it is well suited to label Lys residues in a protein. Sulfonated NHS ester derivatives reveal enhanced water solubility and are therefore preferable. Many protein labeling reagents based on NHS esters or NHS ester derivatives are commercially available. Isothiocyanides likewise react with primary amines under the formation of thiourea, at basic pH (9–9.5). Furthermore the Lys amino can be used for the reductive amination of aldehydes under reversible imine formation. The imine can then be reduced irreversibly to a secondary amine using hydrides such as $NaBH_3CN$. Furthermore Lys is able to form peptide-bonds with halogenated carboxylic acids by a nucleophilic substitution reaction.

Arginine residues can amongst others be modified by methyl glyoxal under formation of a pyrimidine derivative.

Carboxylic groups of glutamate or aspartate can be labeled by carbodiimide reagents under peptide-bond formation.

The cysteine thiol side residue is a stronger nucleophile than the amino group, therefore cysteine generally reacts faster than Lys. But since Cys residues in proteins form disulfide bonds, proteins have to be treated with reducing agents prior to performance of a labeling reaction. Free thiols can then be modified by a variety of different reactions including the reaction with maleimides, alkylation reactions or disulfide bond formation, to name a few.

Arg, His or Tyr are other targets for the introduction of chemical modifications [80]. The interested reader may find detailed information on the bioconjugate chemistry including step-wise description of reactions, commercially available reagents, and practical applications of labeled bio molecules in the book "Bioconjugate Techniques" [81].

In case of chemical amino acid modification to introduce an opening/closing trigger or switch to the inside of a protein channel, the decision for a chemical labeling agent and labeling position depends on several points:

1. The amino acid side chains to be modified have to be exposed within the channel interior. Therefore information on the target OMP structure is a pre-requisite for the selection of the type of amino acid to be labeled. It might be necessary to introduce one or several amino acid labeling targets to the inside of an OMP channel by site-specific mutagenesis.
2. It has to be ensured that labeling of certain positions will not destabilize a protein structure. Theoretical considerations prior to labeling experiments are advisable.
3. The labeling agent has to be bulky enough to block the OMP channel and is preferably commercially available.
4. It has to be possible to cleave off the bulky label to open the channel by an outside stimulus (*e.g.* reduction agents, light, pH or temperature changes).

A collection of literature examples for the OMP amino acid chemical modification to introduce trigger mechanisms or to change the conductance of nano-pore sensing elements is given below:

CM: Reduction Trigger

In case of the FhuAΔ1–159 the amino acid lysine has been successfully biotinylated using reagent 2-[biotinamido]ethylamido-3,3′-dithiodipropionic acid N-hydroxysuccinimide ester [3] or pyridylated using 3-(2-pyridyldithio) propionic-acid [3, 82] (see Fig. 5.9). Both labels contain a disulfide bond that allows removal by

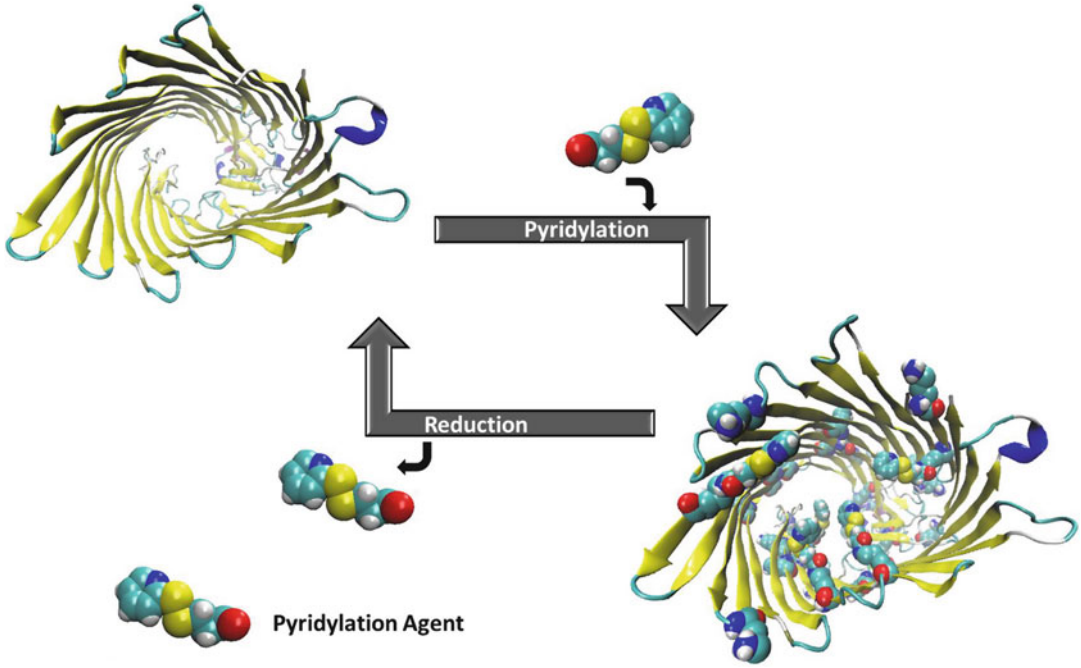

Fig. 5.9 Schematic representation of pyridylation of FhuAΔ1-159 (or variants) Lys residues. Labeling leads to effective channel blocking as determined by calcein release assay or indirect determination of TMB diffusion through the channel by measuring HRP TMB conversion kinetics. The label can be cleaved off by providing reducing conditions (*e.g.* addition of DTT), cleavage results in channel opening (Protein New Cartoon representations made with VMD)

addition of reductive agents, such as dithiothreitol (DTT) permitting reduction-triggered channel opening. Channel blocking properties of both agents were analyzed utilizing calcein release kinetics measurements or the TMB-HRP assay system upon reconstituting labeled proteins into PMOXA-PDMS-PMOXA polymersomes [3] or liposomes [82]. FhuAΔ1–159 contains 29 lysines of which 19 are located on the OMP surface, 6 are buried within the channel and 4 sit on the barrel rims [3]. As the surface exposed lysines are located in the proteins TM region they are covered by detergents after membrane extraction and therefore unlikely to be modifiable [82].

A mutational study on the six channel buried Lys came to the conclusion that pyridylation of Lys at amino acid position 556 is sufficient to sterically hinder compound flux. This finding has been in accordance with MD simulation based B-factor analysis identifying Lys556 as the most rigid of the investigated lysine residues [82].

The same bioconjugation chemistry had been used to biotinylate Lys residues in FhuA Ext and FhuA Exp. FhuA Ext contains 29 lysine residues of which 22 are facing the outside, four are buried within the channel and three are localized at both channel rims. Labeling of PIB_{1000}-PEG_{6000}-PIB_{1000} polymersome inserted FhuA Ext lysines led to a decrease in TMB conversion speed and thus TMB influx (about five-times more slow than for unblocked FhuA Ext), but did not lead to a complete channel blocking as it was found for liposome reconstituted FhuAΔ1-159 [69]. In FhuA Ext one finds 31 Lys residues of which 21 are located on the OMP surface, 6 face the channel inside and 4 are on both barrel ends. Biotinylation in this case again led to a decrease in TMB conversion speed as a measure for compound flux through the liposome inserted protein channel. However TMB conversion occurred only 3-times more slow through the blocked FhuA Exp as compared to the open barrel, most likely due

to the increased channel diameter [67]. These findings show that the label choice depends also on the newly designed protein features and an increase in length and especially in diameter necessitates the use of bulkier agents to guarantee efficient channel sealing.

CM: Light Trigger

Modification of lysine residues in FhuA furthermore allowed the introduction of a light triggerable channel opening system. For this purpose Lys amino acids were labeled using the well-known photo-cleavable compound 6-nitroveratryloxycarbonyl chloride (NVOC-Cl) [83]. NVOC-Cl reacts with the Lys primary amine by nucleophilic substitution. The Lys-NVOC complex is cleaved upon irradiation with light of a wavelength of 366 nm releasing the easily detectable yellow compound o-nitrosobenzaldehyde and CO_2 [4]. Channel blocking was monitored using the TMB/HRP assay system and Lys556 labeling was again found to be sufficient to completely block TMB diffusion through the protein pore [4].

CM: Further Examples

Cysteine amino acid labeling of the FhuA WT with biotin-maleimide [N-biotinoyl-N′-(6-maleimidohexanoyl)-hydrazide] had been carried out in order to obtain structural and functional information on cysteine residues in the surface exposed loop regions L4 and L11. As labeling was possible only with Cys in L4 after reduction, while C-terminal cysteines reacted only after replacement of one Cys and subsequent denaturation it was deduced that all four cysteine residues form disulfide bridges [84]. As these naturally occurring Cys in the FhuA protein do not easily react with thiol marking compounds single Cys may be introduced to the protein as labeling targets. In fact two Cys were newly added to L4 of which one proved to be more reactive than Cys in the WT [85]. However newly added Cys might lead to incorrect folding by formation of unwanted disulfide bridges and proteins should be expressed under reducing conditions.

The α-hemolysine heptamer had been modified by covalently attaching one 5,000 Da (or 3,000 Da) poly(ethylene glycol) (PEG) molecule to an introduced cysteine in the protein pore (by mixing modified and unmodified subunits in suitable ratios heptamers with a single PEG chain were obtained). As the PEG was attached via disulfide bond formation, it could be cleaved off by adding DTT. Single channel recordings showed that the modification reduced the pore conductance by 18 % [86]. This approach broadens the design possibilities for new biosensors by attaching polymer chains that respond to analytes. The attachment of responsive polymers to the interior of an OMP channel is furthermore useful to design new switches and triggers to open or close the channel.

Non-covalent modification of α-hemolysine by β-cyclodextrin as adapters for organic molecule analytes was reported. β-cyclodextrin blocked the WT channel by 64 %, it is thought to be retained in the channel by one of two restriction rings. As it is known that adamantane derivatives bind to cyclodextrines two adamantine based model molecules (2-adamantamine and 1-adamantanecarboxylic acid) were tested on their ability to further block the channel. Both molecules led to a further decrease in channel conductance, proving the adapter qualities of the channel retained cyclodextrin [87].

5.1.3.2 Site-Specific Mutagenesis (SSM)

Since easy to use site-directed mutagenesis kits for amino acid substitutions, point mutations and small deletions/insertions are available from various suppliers, mutagenesis became a lot more practical and is a useful tool to change a protein's functional features, provided that a protein's primary sequence is known and one has some information on its tertiary structure. As mentioned, site-specific mutagenesis to substitute amino acids can be used to introduce amino acid based switches. Furthermore it allows the introduction of protease recognition sites that permit channel opening upon cleavage of an introduced loop structure [88]. Some representative

examples of OMP site-specific mutagenesis resulting in property changes (*e.g.* substrate specificity) or the introduction of opening/closing switches are given in the following:

SSM: OmpF pH Switch

The *E. coli* porin OmpF has been used for the development of a pH based channel release switch. OmpF was chosen as it had previously been successfully reconstituted into PMOXA-PDMS-PMOX block copolymer vesicles [74]. As histidine has a pKa value of ~6, the introduction of a ring of histidines at a channel constriction site was thought to permit the release control of positively charged molecules by shifting the pH from 5 to 7. As target position was chosen a constriction site consisting of two amino acid half rings with three amino acids each (Arg42, Arg82 and Arg132; positively charged and Asp113, Glu117 and Asp121; negatively charged). All six amino acids were substituted by His. The pH dependent release was demonstrated by monitoring the acridine orange translocation by Fluorescence Correlation Spectroscopy at different pH values [77].

SSM: Zinc Binding α-Hemolysin

A Zn(II)-binding α-hemolysin subunit was designed by substituting amino acids Asn123, Thr125, Gly133 and Leu135 with histidine, while Thr292 was substituted by cysteine. The four histidine imidazole sidechains act as ligands to Zn(II), while the single cysteine was modified by 4-acetamido-4′-[(iodoacetyl)amino]stilbene-2,2′-disulfonate, leading to a change in SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) electrophoretic mobility allowing heteromer separation [89].

SSM: Nano-pore DNA-Sequencing MspA

Nano-pore DNA sequencing, using a membrane inserted protein pore has received much interest [90]. The method is based on single channel current measurements, exploiting that current changes when single-stranded DNA passes through the channel depending on the DNA sequence. The homo-octameric *Mycobacterium smegmatis* porin A (MspA) is one of the most stable proteins so far known [91]. Due to its geometrical features, *i.e.* pore constriction of ~1 nm length and ~1 nm width surrounded by sections with much larger diameter [92], MspA is rather suited as a nano-pore DNA sequencing device. The MspA pore has been engineered to allow electrophoretic passage of DNA by replacing negatively charged by neutral amino acids in the constriction site. Furthermore the addition of 24 positively charged residues at the channel entrance of another mutant led to an increase of DNA translocation through the protein pore of 20-times as compared to WT [93]. A combination with the latter mutant and the phage DNA polymerase phi29, allowed the control of the DNA translocation rate [94].

A combination of site-directed mutagenesis and subsequent chemical modification of the newly added amino acid(s) can be used to limit labeling to a defined position, adding a further element of control. The introduction of a single cysteine to the cysteine free α-hemolysine WT made possible the development of a light-activation of pore-formation by cysteine modification with 2-Bromo-2-(2-nitrophenyl)acetic acid (BNPA). The BNPA modified protein lost pore-forming abilities on rabbit erythrocyte membranes, while the cleavage of the label by irradiation with near UV light restored pore-forming properties [95].

All variants of β-barrel MPs have to be isolated and purified prior to characterization and to application. Apart from the multimeric pore forming toxins whose subunits are expressed in soluble form, OMPs are generally expressed into the bacterial outer membrane and their over-expression and retrieval can be rather challenging due to their localization and hydrophobic character as will be discussed in the following sections.

5.2 Outer Membrane Protein Production: Challenges and Solutions

When it comes to their over-expression, isolation and purification OMPs are, as all integral MPs, classified as challenging proteins. However when

working with bacterial MPs one has access to a number of rather developed standard procedures from which to start from, when a protocol for a new OMP variant has to be set-up. Protocols for eukaryotic MPs are less established, however certain general rules and guidelines apply to both pro- and eukaryotic expression and in some cases the heterologous expression of a eukaryotic OMP in bacteria such as *E. coli* might be possible.

This section and its sub-sections will deal mainly with the production of bacterial OMPs, as they are the main targets for the developments of new nano-materials.

In principle MPs can be obtained from their natural source the inner or outer membrane subsequent to their over-expression. This works however only, if the target protein is highly abundant in the respective membrane, as it is the case with the α-helical MP bacteriorhodopsin that can be obtained from the purple membrane of *Halobacterium halobium* [96] or for an outer membrane derived porin of *Rhodobacter capsulatus* [97]. In most cases though the low abundance of a MP in the corresponding membrane prohibits it's recovery from the natural source. Therefore most often the ORF coding for the target outer membrane protein will be cloned into a suitable over-expression vector allowing homologous or heterologous over-expression leading to higher expression and membrane insertion levels. Three of the first OMPs expressed from plasmid-DNA are *E. coli* malotoporin LamB [98], the *Salmonella typhimurium* sucrose porin ScrY produced in *E. coli* [99] and the *E. coli* FhuA [100]. The outer membrane of the expression host bacterium, rich in the desired OMP, will then be isolated and the protein has to be solubilized and purified prior to its structural or functional characterization or its use as nano-material.

As the membrane offers only a limited space, it restricts the number of correctly folded and inserted OMPs, lowering the obtainable yield. Therefore if the protein shows the ability to spontaneously refold from a fully or partially unfolded state, over-expression into inclusion bodies should be considered, as much higher yields can be reached. This method can be especially useful for engineered channel forming OMPs with changed geometry or with a higher than average hydrophobicity (when in folded state).

No matter if an OMP has been obtained from the outer membrane or from inclusion body material, the folded/refolded OMP has to be solubilized in detergent-, polymer-, or organic solvent solutions and extensive screening for the right solubilization agent (depending on protein and purpose) may become necessary. Furthermore the subsequent purification steps that are crucial for crystallization purposes and many applications are often selected following a trial-and-error approach, starting out from standard protocols, as each new variant might require the development of a new purification procedure. But since "*many roads lead to Rome*", good knowledge of standard OMP production protocols as well as more unconventional methods can be of advantage and sometimes, as it is so often in science, one has to be bold and try something entirely new.

Here a summary of the well-established OMP over-expression, solubilization and purification methods, as well as of newly developed/less conventional protocols, with a focus on adaptations suitable for the production of novel developed OMP variants, will be given. Of further consideration will be the various means of expression quality, protein functionality and yield control.

5.2.1 Conventional OMP Isolation from the Outer Membrane

OMPs are either derived from the outer membrane of an expression host or expressed into inclusion bodies from where they can be isolated and refolded [101]. Recently cell-free expression techniques have been recognized as promising tools for the production of MPs in general [102]. Both expression into inclusion bodies and by cell-free expression systems avoid the limited membrane surface bottle-neck.

Apart from the rare case in which a target MP can be obtained from its natural source without preceding over-expression, OMPs have to be over-expressed to reach sufficient concentrations for characterization purposes or for the

application as nano-sized building blocks. When working with a new OMP one will generally start by choosing a fitting expression host and in case of bacterial OMPs it is often possible to over-express the protein in the natural host (homologous) [101], while for eukaryotic, higher organism mitochondrial or plastidic OMPs heterologous expression into the outer membranes of Gram-negative bacteria or yeast mitochondrial outer membranes can be advantageous as unicellular organisms are more easy to handle and have much shorter generation times. However when expressed in heterologous systems, OMPs as all MPs can be toxic for the host or misfolded and expression levels are often low due to the limited space offered by the membrane [103]. Furthermore OMPs fold and insert solely into the Gram-negative bacterial or mitochondrial/chloroplasts outer membrane and can therefore not be membrane-expressed into Gram-positive bacteria or for instance the yeast plasma membrane. The most widely used host for the heterologous over-expression of Gram-negative bacterial OMPs is *E. coli*, as it is a Gram-negative organism itself and as it is a well-studied model organism that is widely used for heterologous protein expression. Numerous bacterial OMPs such as the *Chlamydia psittaci* major outer membrane protein MOMP [104], *Haemophilus ducreyi* OMP D15 [105], *Chlamydia trachomatis* MOMP [106], *Pasteurella multocida* OmpH [107], *Neisseria meningitidis* PorA [108] to name but a few, are examples for OMPs that have been over-expressed in *E. coli*. *E. coli* however is often not suitable for the recombinant expression of eukaryotic OMPs.

When a new variant of an already known OMP is to be expressed then often a good starting point is to choose the same expression host, conditions and protocols as used for the parent OMP. If necessary, procedures can then be adapted and optimized for the newly engineered protein.

The general work-flow for the conventional OMP production in bacterial host cells is shown in Fig. 5.10. As step one in Fig. 5.10 shows after selecting the host organism, the target OMP coding ORF is cloned into a suitable over-expression vector. These vectors are usually the same as used for the expression of soluble proteins therefore the same selection criteria are applied. Further considerations are that affinity tags useful for purification purposes, such as the Hexa-His-tag should be fused to the protein C-terminal end, so to not interfere with the N-terminal signal peptide sequence leading the protein to the outer membrane. The elongation of the His-tag from 6 to 8 or 10 histidines can increase purification success, while it might result in lower expression levels [109]. In case of OMPs however internal His-tags can be the better solution, as C- and N-terminus close the barrel by hydrogen bond formation and an affinity tag may inhibit the correct closing [100] (see also Sect. 5.2.4.1).

A standard combination of expression host and expression plasmid (not only for MPs) is *E. coli* BL21 (DE3) and the pET vector system. Both are readily available and generally present in any molecular biology laboratory. The *E. coli* BL21 (DE3) [F– ompT gal dcm lon $hsdS_B$ (r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])] is an *E. coli* B strain carrying a λ prophage carrying the T7 RNA polymerase gene and $lacI^q$, usable with plasmids containing the T7 promotor. Expression of the T7 RNA-polymerase is under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible lac promoter [110]. The pET expression vector system is one of the most widely used systems for the expression of recombinant proteins in *E. coli* since the T7 RNA polymerase is highly specific for the T7 promoter on the plasmid (host genes will not be expressed) and since it is highly active leading to high expression levels. The pET vectors have been derived from plasmid pBR322. Target cloned ORFs are under control of strong bacteriophage T7 transcription and translation signals. Upon induction with IPTG the T7 RNA polymerase is expressed in *E. coli* host cells, resulting in plasmid protein expression [110]. This element of control avoids "leaky" expression of the plasmid coded ORF (*i.e.* low expression in non-induced cells), this is of great importance since OMP over-expression can lead to toxic effects or reduced growth rates [111].

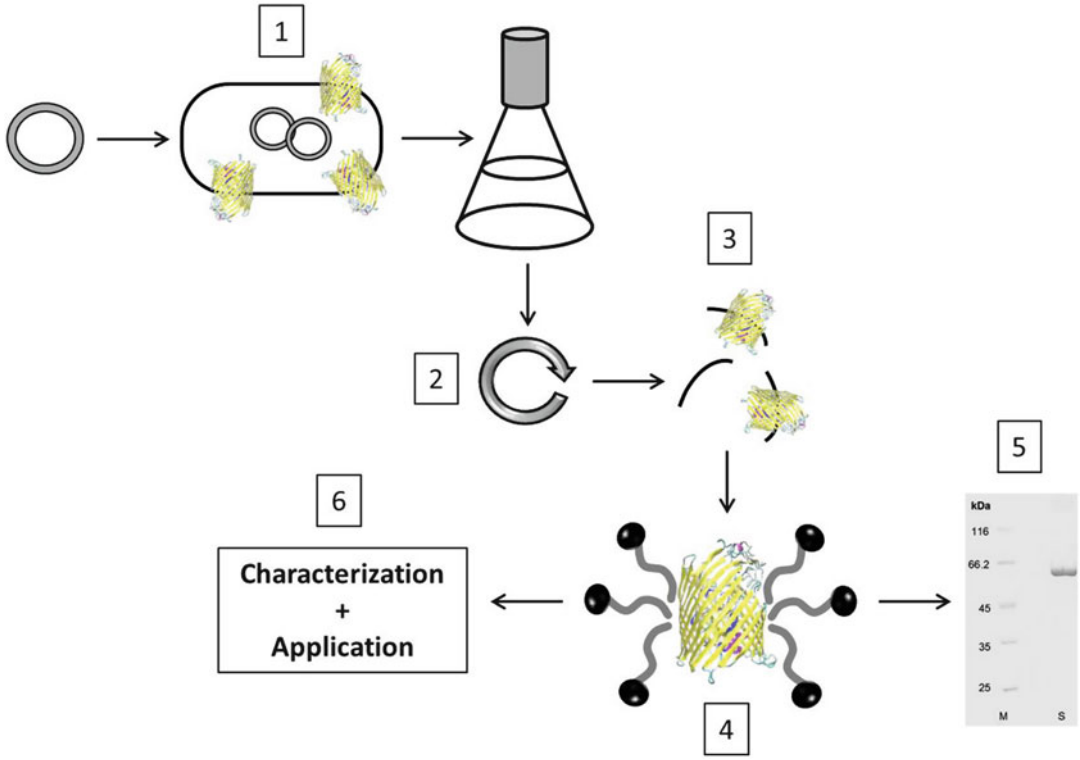

Fig. 5.10 Schematic representation of conventional membrane expressed OMP production process flow. 1 – ORF cloning into expression vector (*grey circle*), transformation of host cells (screening for best expressing cells and conditions not shown) and cell growth + protein expression in liquid culture; 2 – cell harvest (*e.g.* by centrifugation); 3 – cell disruption and outer membrane isolation (by differential centrifugation); 4 – protein solubilization (includes solubilization agent screening); 5 – expression yield and protein purity control (*e.g.* by SDS-PAGE); 6 – characterization and application of purified OMP variant (Protein New Cartoon representations made with VMD)

For the over-expression of OMPs *E. coli* BL21 (DE3) derivatives lacking genes for the major OMPs, such as *E. coli* BL21 (DE3) omp8, are available. *E. coli* BL21 (DE3) omp8 expresses only a small subset of naturally occurring *E. coli* porins assisting the purification of over-expressed barrel proteins [112]. Cloning, transformation, clone screening, small-scale expression and expression follow standard protocols [113] and will not be further discussed. Generally it can be said though, that decelerated expression rate and bacterial growth rate facilitates the functional OMP expression into the outer membrane of *E. coli* as it avoids unwanted inclusion body formation and toxic effects of OMP over-expression (but lower expression levels can lead to difficulties in OMP isolation and purification). Lower growth and expression rates can be achieved by lowering growth temperature after induction (*e.g.* to 30 °C), by low inducer concentration or by utilizing a vector system with a weaker promoter. Toxicity can be lowered furthermore by shortening the time of induction [99]. In case of OMPs that might lead to host-cell osmotic imbalances (*i.e.* caused by OMPs that have high conductivity) carefully considering the expression media can lead to better results. As described in Sect. 5.1.2.3. FhuA variant FhuA Exp, an OMP with a widened channel diameter [67], led to poor cell-densities and protein expression-rates when expressed under conventional conditions (*i.e.* TY or LB media and a growth temperature of 37 °C [68]). Lowering the expression temperature to 30 °C, combined with using the hypertonic NaPy

Table 5.2 Fatty acid composition of *E. coli* outer cell membranes applying different cultivation temperatures [115]

	% of overall fatty acid content			
Fatty acid	10 °C	20 °C	30 °C	40 °C
Myristic acid	4	4	4	8
Palmitic acid	18	25	29	48
Palmitoleic acid	26	24	23	9
Oleic acid	38	34	30	12
Hydroxy-myristic acid	13	10	10	8
Ratio unsaturated/saturated	2.9	2.0	1.6	0.38

medium, such as it is used for L-form *E. coli* [114] greatly enhanced the obtained cell density and allowed the new OMP variant to be isolated.

Another important effect of the cultivation temperature is that it influences the lipid composition of the bacterial outer membrane leading to differences in membrane fluidity [115]. Though *E. coli* can regulate to some extend the ratio of unsaturated/saturated membrane lipid fatty acids at different temperatures due to the FabF enzyme [116]. Nevertheless the *E. coli* outer membrane composition is temperature dependent and Table 5.2 lists the fatty acid composition of the *E. coli* outer membrane at different cultivation temperatures. It shows that at lower temperature the membrane contains more unsaturated fatty acids rendering the membrane less fluid (see also Chap. 2). These differences in membrane fluidity may affect the membrane isolation and OMP solubilization process. A comprehensive review on the bacterial membrane lipid homeostasis can be found in [117].

After expressing the target OMP in a liquid culture, protein expressing cells will be harvested (Fig. 5.10, step 2). In case of bacterial expression hosts harvest will generally be accomplished by low speed centrifugation at 4 °C and cell pellets are resuspended in suitable buffers (*e.g.* phosphate buffers pH 7.0–8.0) or unsuspended pellets can be stored at $\leq -20°$ prior to further processing. In order to be able to isolate the protein carrying outer membrane, harvested cells have to be disrupted and non-outer membrane cell components have to be separated and discarded. Many of the methods used to disrupt cells before the isolation of cytoplasmic proteins can be applied also prior to isolation of outer membrane proteins. Cell disruption procedures that cause destabilization or dissolving of the outer membrane by addition of high concentrations of detergents such as SDS, Sarkosyl or Triton X-100 are not advisable because they may lead to a premature OMP solubilization.

A rather complete review on cell disruption methods is given in [118] and Table 5.3 summarizes the most common methods leading to outer membrane fragments in vesicle shape that can be used prior to outer membrane isolation. The disruption method of choice depends often on the culture volume and cell mass, as some methods like the disruption by ultrasonic sound are only applicable for small volumes. Furthermore it depends on the target protein, as temperature sensitive proteins or proteins that do not tolerate strong shearing forces forbid the use of methods that lead to sample heating or that disrupt cells by shearing such as the high pressure homogenisator (HPH), French press or ultrasonic cavitations. The much milder enzymatic lysis of bacterial cells using lysozyme might be more suitable, shows however low disruption efficiency and is quite time consuming. Gram-negative cells have to be treated with lysozyme and ethylenediaminetetra-acetic acid (EDTA), as EDTA chelates divalent cations (*e.g.* Mg^{2+}) that stabilize the outer membrane. The outer membrane destabilization allows lysozyme to reach the inner membrane murein layer [118]. In all cases the addition of DNase is advisable to decrease the solutions high viscosity caused by released DNA. The addition of commercially available protease inhibitor cocktails is generally useful also, as cells contain proteases that are liberated upon disruption.

Disruption of cells is followed by the isolation of the outer membrane (Fig. 5.10, step 3) by differential centrifugation. In case of *E. coli* cell

Table 5.3 Classification of cell disruption methods applicable to disrupt Gram-negative bacterial cells prior to outer membrane isolation

Type	Method	Concept	Comments	Reference
Mechanical	HPH	Pressure of a cell suspension is raised to ~1,000 bar. Then high-velocity release of disrupted material through a valve leads to disruption.	Suitable for large volumes, but leads to sample heating (stringent cooling required).	[119]
	French Press	Pressure of up to ~1,400 bar applied to cell suspension. Abrupt pressure reduction by dropwise release of disrupted material.	Same as HPH.	[120]
Non-mechanical (physical)	Freeze/Thaw	Repeated freezing/thawing of cell paste; disruption through ice crystal formation. Often combined with grinding or enzymatic lysis.	Simple, inexpensive; but sensitive proteins can be damaged and efficiency is low.	[121]
	Osmotic shock	Rapid change of solution of high osmotic pressure to low osmotic pressure.	Simple, inexpensive; usable in combination with enzymatic lysis; but not for stationary phase cells.	[122]
	Ultrasonic cavitations	Ultrasonic sound leads to pressure changes in cell solution resulting in cavitations.	Simple, inexpensive; but only for small volumes. Heats sample and can destroy sensitive proteins.	[123]
Non-mechanical (biological)	Enzymatic	Lysozyme cleaves β-1,4 glycosidic bonds of peptidoglycan polysaccharide chains.	Mild, but slow and with low disruption efficiency. Has to be combined with EDTA to destabilize the outer membrane.	[124]

lysates can be cleared by low speed centrifugation to remove cell debris [125] (this step is not crucial though), cleared lysates are then treated with low concentrations of non-ionic detergents like Triton X-100 or N-lauroylsarcosine to solubilize the inner membrane [126, 127]. Centrifugation at speeds around 20,000–40,000 g allows the removal of solubilized inner membrane and other residual cell components with the resulting supernatant [40]. In a pre-solubilization step proteins loosely bound to the outer membrane can be removed by adding low concentration of solubilization detergent and subsequent ultracentrifugation [82].

If the fusion of a purification tag to the target OMP is possible, the outer membrane isolation can be skipped. To the uncleared lysate instead is added the solubilization detergent and the solution is loaded onto a suitable chromatographic column. This is of great advantage, as outer membrane preparation is a time consuming procedure due to lengthy centrifugation steps and incubation in pre-solubilization buffers. Furthermore the outer membrane pellets are very tough and sticky and resuspension in newly added buffer requires mechanical processing with homogenizer or by ultrasonication.

The most crucial step, the solubilization of the target protein from the outer membrane (see step 4 in Fig. 5.10), is achieved by the addition of a suitable detergent or other amphiphilic molecules (though detergents are often rather expensive due to a lack of effective alternatives they are still first choice for the MP solubilization) and a last high speed centrifugation step leading to separation of membrane lipids (pellet) and solubilized proteins (supernatant). The solubilization step is mainly dependent on the careful selection of the right detergent. In case of newly developed variants of an already known protein, one often can resort to protocols developed for the WT protein. While in case of new OMPs or variants with novel features a detergent screening is necessary.

Detergents first disintegrate the lipid bilayer, dissociating lipid-protein interactions then their hydrophobic tail-regions interact with hydrophobic surface areas of the released OMPs. Hydrophilic detergent and protein portions are

in contact with the aqueous environment (see Fig. 5.10, step 4) [128]. It is assumed that a solubilized membrane protein is properly folded when in contact with the detergent [129]. Small-angle x-ray scattering showed that this is true for β-barrel shaped proteins such as the *E. coli* OmpF, while detergent-solubilized helical MPs such as bacteriorhodopsin and Ste2p G-protein coupled receptor from *S. cerevisiae* were not properly folded [130]. Solubilized MPs are always complexes of protein, detergent and membrane lipids, the detergent and lipid content of these complexes lies between 10–50 %, depending on the utilized buffer.

Detergents are amphiphilic molecules that can be classified into four distinct groups [131]:

1. Ionic detergents (*e.g.* SDS)
2. Bile acid salts (*e.g.* Sodium cholate)
3. Non-ionic detergents (*e.g.* Triton X-100)
4. Zwitterionic detergents (*e.g.* 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate. – CHAPS)

Ionic detergents consist of a charged head group and a hydrophobic hydrocarbon chain tail. They effectively solubilize proteins from the outer membrane, but disadvantageously many ionic detergents have more or less strong denaturing effects [131]. They are useful to dissociate protein-protein interactions. Some proteins can be refolded from their SDS-solubilizes state into a lipid environment by renaturing detergents [131] or amphipathic diol solvents such as 2-methyl-2,4-pentanediol (MPD) [132]. SDS can be removed for instance by dialysis.

In an unpublished study the FhuA variant FhuAΔ1–159 has been solubilized from the *E. coli* outer membrane upon over-expression, using SDS [133]. Solubilization efficiencies were much higher than with the conventionally used non-ionic detergent n-octylpolyoxyethylene (octyl-POE), where most of the protein remains within the membrane fragments. (FhuAΔ1–159 expressed into the membrane and extracted using octyl-POE allowed to obtain ∼1.5 mg of correctly folded protein from 1L of culture [41]). Protein solubilized by SDS was present in an unfolded state and tended to precipitate. The protein was therefore refolded by SDS-removal via dialysis or during immobilized metal-affinity chromatography (IMAC), where the unfolded protein was bound via an internal His-tag to the column resin and the SDS was diluted by washing steps. In both cases the commercially available and relatively cheap diblock copolymer Polyethylene-Poly(ethyleneglycol) (PE-PEG) had been used as a refolding agent and only with the dialysis method could be obtained reasonable yields of ∼30 mg/L of culture (see Fig. 5.11), which is about 20-times higher than yields obtained by solubilization with octyl-POE and about 1.5-times more than when isolating and refolding the protein from inclusion bodies (where refolding is again facilitated by PE-PEG) (see also Sect. 5.2.2). Structural integrity of the so obtained protein was verified by CD-spectroscopy, revealing a β-structure content of ∼48 %. Obtained purities however were only up

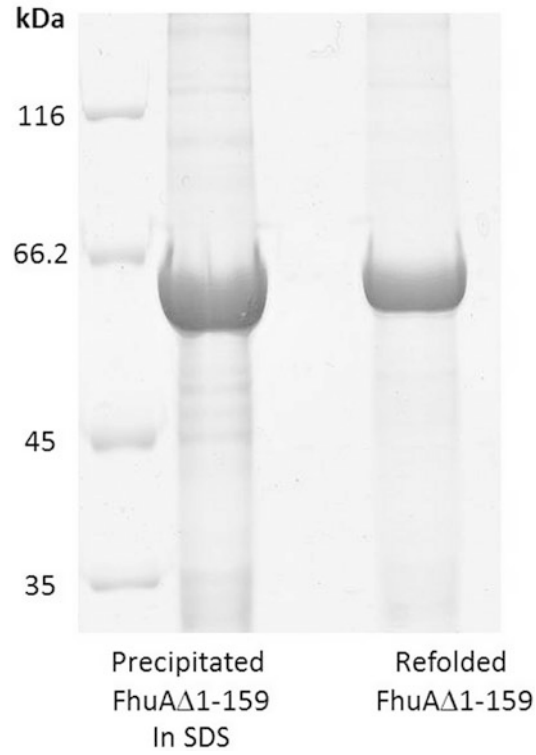

Fig. 5.11 SDS-PAGE result of outer membrane derived FhuAΔ1–159 solubilized in SDS and refolded by dialysis into PE-PEG containing buffer

to 85 %. However in this study only PE-PEG had been used to refold the protein without further optimization or screening for better refolding agents, and no further purification steps were employed. Therefore there might be noteworthy potential to improve yields and especially purity by further optimizing the protocol and by using other solubilization agents.

Bile acid salts (or saponin detergents) are mild ionic detergents with a rigid steroidal backbone. In contrast to linear chain detergents they do not form conventional micelles [128, 131].

Nonionic detergents have polyoxyethylene or glycosidic, hydrophilic head groups and they are mild and often not denaturing. They effectively break lipid-lipid and lipid-protein interactions, therefore they are commonly used to solubilize MPs in general. Disadvantages are however that they often lead to low yields [128] and as they do not break protein-protein interactions they do not prevent proteins from forming aggregates that may precipitate.

Zwitterionic detergents combine ionic and nonionic detergents properties. Though they are more denaturing than non-ionic detergents they are widely used, as they proved to be of advantage for protein structural study purposes. Examples are the BtuB protein that has been crystallized in complex with TonB using detergent lauryldimethyl amine oxide (LDAO) [134] while FhuA and OmpF were crystallized in dimethyldecylamine-N-oxide (DDAO) [56, 135]. Furthermore many NMR-based structural studies have been carried out using zwitterionic detergents such as dodecylphosphocholine [136, 137].

In order to find the best detergent for a certain outer membrane protein (it should solubilize the protein without denaturing it) several different detergents and solubilization conditions have to be screened.

The construction of a new OMP variant from an existing OMP often necessitates to newly screen for optimal solubilization conditions, especially if the new variant shows changes in hydrophobicity. One example is FhuA variant FhuA Ext with extended hydrophobic, TM region (see Sect. 5.1.2.4). In contrast to the parent FhuAΔ1–159 that can be solubilized from the membrane using detergent octyl-POE, FhuA Ext could be obtained from the membrane only by a serial extraction with organic solvent mixtures of chloroform:methanol and TFE (2,2,2-trifluoroethanol):Chloroform prior to solubilization in detergent OES [69].

Solubilization yield can be monitored for instance by SDS-PAGE of supernatant and pellet after ultra-centrifugation (Fig. 5.10, step 5). The target protein should be mainly in the supernatant. As a rule of thumb a membrane protein is properly solubilized if it stays in solution after 1 h centrifugation at 100,000 g. A more accurate method to determine whether a MP is solubilized is the gelfiltration of protein in detergent solution. When using Sepharose 6B for instance the protein elutes in the columns inclusion volume in case it is solubilized. Recently an ultracentrifugation dispersity sedimentation assay (combining small volume ultracentrifugation and subsequent SDS-PAGE) has been developed that allows determining rapidly whether a MP is monodispersedly solubilized, thus speeding up screening processes [138]. Nowadays commercially obtainable detergent screening kits containing sets of commonly used detergents may speed up the selection of a detergent for solubilization or crystallization.

Above a certain detergent concentration, the critical micellar concentration (cmc), in an aqueous environment, the detergent molecules associate and form multimolecular complexes, so called micelles. Micelles are aggregates with hydrophobic interior and hydrophilic exterior surfaces. The cmc depends on the detergent type, the solution pH, temperature and ionic strength [139]. In general the cmc decreases with alkyl chain length and increases with the introduction of double bonds. Additives that break up the water structure (*e.g.* urea) increase the cmc in all detergent types, while increased concentrations of counter ions result in cmc-reduction for ionic detergents. When using detergents for protein solubilization purposes, buffers and solutions should have a detergent concentration above the cmc, especially when a detergent might have to be removed/exchanged by dialysis [131].

Arachea et al. extensively studied the extraction profiles of different detergent types for membranes isolated from bacteria and yeast, on a set of recombinant target proteins. Some general trends were found [140]:

1. The extraction efficiencies of the analyzed detergents increased at higher concentrations. At concentrations below a detergent's cmc extraction efficiency dropped significantly. The optimal concentration is detergent-dependent.
2. SDS, two alkyl sugar detergents, octyl-b-D-glucoside (OG) and 5-cyclohexyl-1-pentyl-b-D-maltoside (Cymal-5), and a zwitter-ionic detergent, N-decylphosphocholine (Fos-choline-10), were effective in the extraction of a broad range of MPs.
3. In case of *E. coli*, SDS was the most efficient at extracting proteins from the inner membrane as well as from the outer membrane, while OG was among the most effective non-ionic detergents. However membrane protein extraction efficiencies of detergents vary between different *E. coli* strains.
4. Fos-choline is not very effective for the extraction of outer membrane proteins.
5. Protein extraction from yeast membranes was reported to be in general more difficult.

These findings underline the necessity for thorough detergent screening in order to obtain the highest possible concentration of a correctly folded and functional target MP.

Some proteins can be solubilized in detergents but need a lipid environment to be functional, here the use of lipid/detergent systems for solubilization can be an alternative [131]. As detergents that allow functional OMP solubilization are often rather expensive, commercially available, amphiphilic block copolymers [41] or if the protein target is stable enough organic solvents [141] may be considered. A recent approach uses nanoscale phospholipid bilayers stabilized by an encircling membrane scaffold lipo-protein, so called nano-discs [142]. Nano-discs are getting more and more attention as they are now commercially available and have been found to be especially useful for protein biophysical characterization [143].

5.2.1.1 Yield, Purity and Quality Control (OMP Structure and Functionality)

Supernatants obtained from centrifugation after solubilization have to be analyzed on target OMP presence, amount and purity (see step 5, Fig. 5.10). Most often this will be accomplished by carrying out SDS-PAGE following the "Laemmli" method [144]. However the sample boiling with SDS and reducing agent containing buffer to fully denature the protein can lead to protein aggregate formation, when working with membrane proteins in general. If this is the case, samples can be incubated at 60 °C for 30 min or at 37 °C for 60 min instead. The experimentalist has to be aware of the fact that MPs often do not move according to their molecular weight in SDS-PAGE. They may move faster and will thus appear smaller, which might be due to differences in bound SDS amounts as compared to soluble proteins [145]. For instance for the 37.2 kDa OmpF it was reported that its band position on stained SDS gels varied significantly depending on the detergent used for solubilization. While in SDS it migrated as a protein with lower molecular mass in OG it migrated more slowly, showing a false higher molecular mass [140]. The commonly observed "smearing or tailing" of MPs on SDS gels is due to lipids sticking to the protein or to protein precipitation resulting from too low detergent concentrations (as the protein is concentrated during the gel run), often a higher SDS concentration in sample and running buffer avoids "smearing". After electrophoresis proteins can be visualized by staining the gel with Coomassie brilliant blue or by a silver staining (for lower protein concentrations).

In order to check the target OMP purity using the SDS-PAGE method it is suggestible to further analyze the stained gel by image processing programs. A good freeware program to analyze SDS-gels is the Java based "ImageJ" (available at http://rsb.info.nih.gov/ij; developed by Wayne Rasband, National Institutes of Health, Bethesda, MD).

Western-blot analysis, using target protein (or purification tag) specific antibodies can be used

to detect poorly expressing proteins after protein transfer from an SDS gel to a blotting membrane. Due to the OMPs hydrophobicity however their transfer to a membrane can be difficult [146].

The protein concentration can be determined using standard assays, such as the bicinchoninic acid (BCA) [147] or Lowry assay [148], while the commonly used Bradford method [149] is incompatible with detergents.

An outer membrane protein's structural integrity and overall folding can be verified by secondary structure analysis via CD spectroscopy or X-ray diffraction measurements of crystallized OMPs and NMR studies on the tertiary structure (see Chap. 3).

OMP-functionality is generally studied after reconstituting/inserting the protein into lipid or polymer membranes and analyzing transport activity or conductance either directly by patch-clamp methods or indirectly by enzymatic assays that allow the catalytic transformation of a substrate that passes the OMP to reach the enzyme (unable to pass the OMP channel) and by subsequently detecting the product that again passes the OMP (see Chap. 6, Sect. 6.1.3).

While the purity of solubilized OMPs is generally sufficient for functional tests, it often does not suffice for structural analysis. Therefore further purification steps might become necessary. A general discussion on the purification of outer membrane proteins with and without purification tag will be given in Sect. 5.2.4.

5.2.2 OMP Isolation from Inclusion Bodies for Improved Yields and for the Expression of Toxic OMPs and Their Variants

The described OMP over-expression and subsequent targeting to the outer membrane often has toxic or lethal effects on the expression-host (see also Sects. 5.1.2.2 and 5.1.2.3) limiting the reachable expression-levels and range of obtainable proteins [101, 150]. A valid alternative to obtain OMPs is their non-functional expression into inclusion bodies with subsequent solubilization of inclusion body material in denaturing buffer (*e.g.* urea) and refolding by exchanging denaturing agents with amphiphilic solubilization agents (*e.g.* detergents) (see Fig. 5.12).

In *E. coli* over-expressed proteins that accumulate in the cell are deposited in the form of inclusion bodies. Inclusion bodies are insoluble aggregates of misfolded and inactive protein. These aggregates can be observed by phase contrast microscopy as dark intracellular particles (refractile bodies) with a size of 0.2–0.6 μm [151]. As proteins from inclusion bodies can be refolded efficiently by well-established protocols, the expression of proteins in inclusion bodies is a suitable method for the high-level production of recombinant proteins [152]. In general the protein production in inclusion bodies holds several advantages over their functional expression:

1. The over-expression of proteins into inclusion bodies is less stressful for host cells.
2. Proteins that are toxic to the expression host when expressed in their functional state are non-toxic in inclusion bodies.
3. Higher expression levels can be reached.
4. Proteins in inclusion bodies are less prone to proteases.
5. The target protein in inclusion bodies is often already relatively pure.

A disadvantage is however that proteins obtained from inclusion bodies have to be refolded back to their native states, which can be quite difficult, especially if the target protein contains disulfide bridges. Generally one has to screen for the optimal conditions to refold a certain protein.

The inclusion body expression is a promising way to express bacterial OMPs and their variants that when expressed functionally have toxic effects on the host cells [31–34, 62, 101, 150] or that can be expressed in low concentration only. Especially so since the OMP expression into inclusion bodies followed by solubilization and refolding works rather well in contrast to the production of α-helical plasma membrane proteins by inclusion body expression. Furthermore the isolation of inclusion body material is rather less time consuming the isolation of outer membrane vesicles (however protein refolding steps may take considerable time, especially when carried out by dialysis). One of the first OMPs ob-

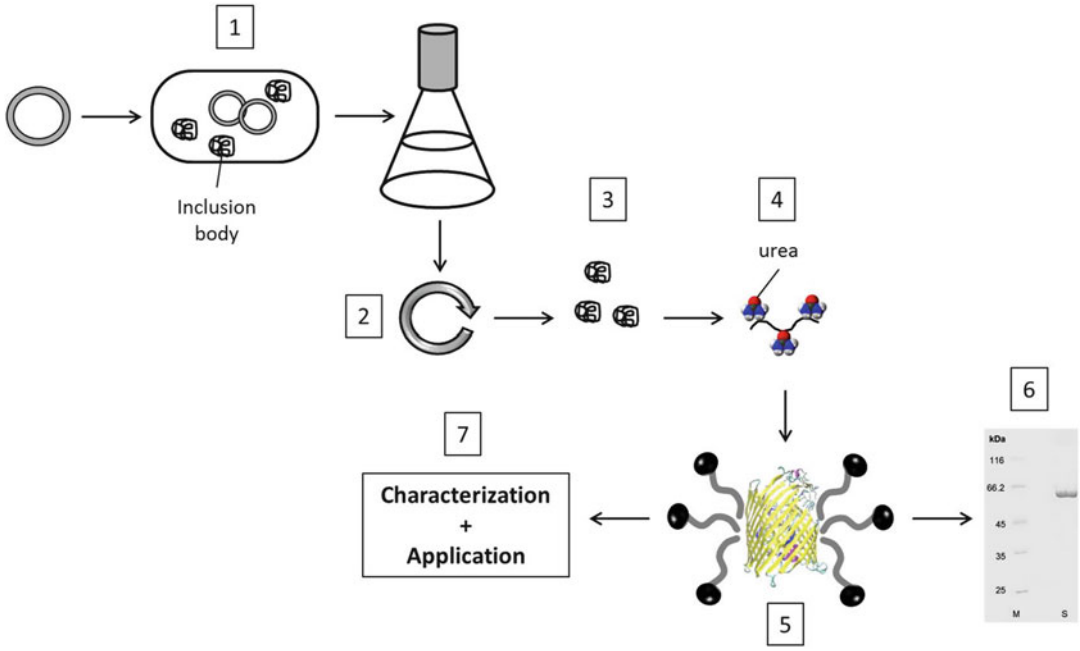

Fig. 5.12 Schematic representation of conventional inclusion body expressed OMP production process flow. 1 – ORF cloning into expression vector (*grey circle*), transformation of host cells and cell growth + protein expression in liquid culture; 2 – cell harvest (*e.g.* by centrifugation); 3 – cell disruption and isolation of inclusion body material; 4 – solubilization of inclusion body material using high concentrations of *e.g.* urea; 5 – urea removal and addition of OMP solubilization agent, for instance by step-wise dialyzing the solubilized inclusion body material; 6 – yield and protein purity control (*e.g.* by SDS-PAGE); 7 – characterization and application of obtained OMP variant (Protein New Cartoon representations made with VMD)

tained in this way and successfully crystalized afterwards, was the *E. coli* outer membrane phospholipase A (OmpLA) [153].

Inclusion bodies are formed when the transcription and translation rates are high. Generally it can be said that if expression of a plasmid-coded, over-expressed protein is higher than 2 % of cellular protein expression, inclusion bodies are likely to form [154]. Therefore in order to express a target protein into inclusion bodies a plasmid system with a strong promoter, such as the T7 promoter in combination with growth temperatures that facilitate fast growth are suggestible. Furthermore the use of an *E. coli* strain optimized to better endure stress caused by the formation of inclusion bodies should be considered. Examples for such strains are the *E. coli* BL21(DE3) derivatives C41(DE3) and C43(DE3) [103]. Moreover the outer membrane protein N-terminal signal sequence that leads to the transport to the Gram-negative bacteria outer membrane should be deleted, as over-expressed OMPs that lack their signal sequence accumulate in inclusion bodies [101, 155].

The production of an OMP in its unfolded state in inclusion bodies takes course similar to the inclusion body production of soluble proteins and starts after deleting the N-terminal signal sequence of the outer membrane protein ORF. The ORF is then cloned into an adequate vector (with a strong promoter). The vector is then inserted into the host strain by transformation, cells are allowed to grow and induction is carried out at temperatures that facilitate high growth rates (for *E. coli* 37 °C) (see Fig. 5.12, step 1). Cells are harvested and disrupted (see Fig. 5.12, step 2), following the same considerations as when isolating OMPs from the outer membrane. But cells have to be disrupted as completely as possible to avoid contaminations with other cellular components. Best results can be obtained by a combination of ultra-sonication or high pressure

homogenization in presence of EDTA (to destabilize the outer membrane) and detergents, such as Triton X-100 followed by lysozyme and DNase treatment. Here it is important to add high concentrations of Mg in order to chelate and inactivate EDTA, as DNase requires Mg as a cofactor [156].

Since inclusion bodies have a rather high specific density, they can be obtained after cell disruption by centrifugation at moderate speeds [152, 156, 157] (Fig. 5.12, step 3). Lysozyme and DNase treatment can be repeated for enhanced purity. Further washing of inclusion body material with Triton X-100 containing buffer removes residual membrane fractions and purities of the target protein of ∼95 % can be reached [158].

After washing, the purified inclusion bodies have to be solubilized (Fig. 5.12, step 4). Solubilization is generally achieved by the addition of highly concentrated denaturants (*i.e.* 6 M guanidinium chloride – GdmCl or 6–8 M urea.) and reducing agents (to keep disulfide bridges from forming). Since GdmCl is a strong chaotroph it allows the solubilization of inclusion bodies that are resistant to solubilization by urea [156, 159] but in certain proteins it might inhibit ionic interactions necessary for correct refolding [160]. Alternatives include the use of SDS [161], sodium N-lauroyl sarcosine [162] or N-cethyl trimethyl ammonium chloride [163].

Protein contained in solubilized inclusion bodies has to be folded back to its native state (see step 5 in Fig. 5.12). Prior to refolding the target protein can be purified further, in case it carries a His-tag, purification can be carried out by IMAC in presence of a denaturant allowing on-column refolding (see Sect. 5.2.4) [164]. Refolding is generally initiated by dilution of the denaturing agent into a suitable buffer. In case of membrane proteins this buffer has to contain lipids, detergent micelles, mixed lipid-detergent micelles or other amphiphilic substances. In case of FhuA variant FhuAΔ1–159 expressed into inclusion bodies after signal sequence deletion, the protein could be refolded into a solution containing the commercially available diblock copolymer PE-PEG as an alternative to costly detergents. By this

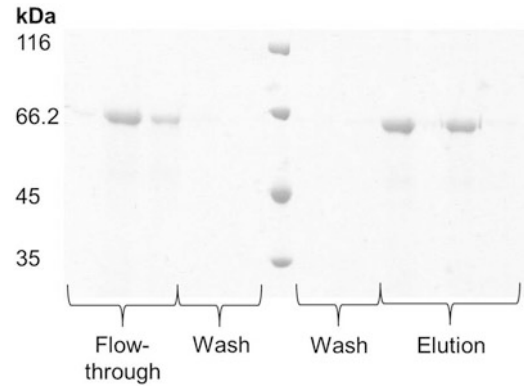

Fig. 5.13 SDS-PAGE result of inclusion body derived FhuAΔ1–159 after IMAC on-column refolding, using PE-PEG diblock copolymer

method ∼19.5 mg of protein (purity of ≥92 %) per L of culture could be obtained. Furthermore FhuAΔ1–159 was successfully reconstituted into liposomes from the PE-PEG solubilized state [41] showing that the diblock copolymer is suitable for protein reconstitution purposes.

Dilution of the denaturing agent can be carried out either rapidly by drop-wise addition of solubilized inclusion bodies to a refolding buffer, stirring the solution vigorously or slowly and step-wise by dialysis, or by on-column refolding during chromatography, decreasing the denaturant concentration with each new dialysis or chromatography washing step, in presence of the refolding buffer [152]. As slow folding might lead to the formation of folding-intermediates that precipitate and as the rapid dilution might equally lead to protein precipitation, both depending on the respective protein, the method of choice has to be selected by testing. Also a screening for the best refolding agent is necessary. The on-column refolding method can facilitate protein refolding especially when combined with the use of an ionic detergent, as it inhibits protein aggregation [101, 165].

In an unpublished study the diblock copolymer PE-PEG has been used to on-column refold FhuAΔ1–159 (that includes an internal His-tag; see Sect. 5.2.4) during IMAC. As the SDS-PAGE result depicted in Fig. 5.13 shows, the highly pure protein (92 %) that was successfully bound and eluted from the column, though a relatively high

Table 5.4 Examples for bacterial OMPs expressed in *E. coli* inclusion bodies, listing protein origin and applied refolding method

OMP	Origin	Refolding method	Reference
OprM	*Pseudomonas aeruginosa*	Dilution with detergent solution	[167]
Major porin	*Rhodopseudomonas blastica*	Dilution with detergent solution	[62]
Omp21	*Comamonas acidovorans*	On column refolding using detergents	[168]
Opc	*Neisseria meningitides*	Dilution with detergent solution	[169]
OmpC	*Salmonella typhi*	Dilution with detergent solution	[170]
OmpA, OmpX	*E. coli*	Dialysis into detergent solution	[171]
OmpT	*E. coli*	Dilution with detergent solution	[172]
FepA	*E. coli*	Dialysis into detergent solution including SDS	[173]
OmpG	*E. coli*	Dilution with detergent solution	[174]
FhuAΔC/Δ4L	Variant of *E. coli* FhuA	On column refolding using detergents	[46]
FhuAΔ1–159	Variant of *E. coli* FhuA	Dialysis into PE-PEG diblock copolymer solution	[41]
FhuAΔ1–159	Variant of *E. coli* FhuA	On column refolding using PE-PEG diblock copolymer	Unpublished results [166]

amount of protein did not bind to the affinity resin and yields were rather poor and in the same range as when isolating the protein from the membrane using octyl-POE for protein solubilization (See Sect. 5.2.1 and see below). Structural integrity of the refolded protein had been verified by CD spectroscopy, with an ascertained β-structure content of 48 % [166].

Refolding is influenced by a set of parameters, such as protein concentration, ionic strength (high ionic strength might increase the yield), residual low concentrations of denaturing agents (in some cases the retention of denaturing agents in low concentrations can lead to yield-increase) and temperature (often room temperature is preferable to lower temperature, such as 4 °C) [160].

The refolding methods employed for the refolding of several *E. coli* inclusion body derived bacterial OMPs and OMP variants are listed in Table 5.4. The yields of correctly folded protein are often higher than when the protein is isolated from the outer membrane. In case of the FhuAΔ1–159 ∼19.5 mg of refolded protein were obtained from 1L of culture, while the same culture volume led to only ∼1.5 mg of outer membrane derived protein [41].

However as the aggregation of unfolded protein and the precipitation of folding-intermediates occur faster than the correct refolding, the yield of correctly folded protein is often limited to 2–5 % (in rare cases up to 20 %) of the total expressed protein. Recent developments in OMP expression using cell-free expression systems, with attempts to directly express and fold the OMP into a suitable folding-buffer containing detergents might lead to further increased yields in the future and will be discussed below. Yields are again checked by SDS-PAGE and protein concentration measurements (see Sect. 5.2.1). The refolding itself can be monitored by CD spectroscopy or by measuring protein tryptophane fluorescence, providing the target protein contains tryptophane residues [175]. Spectra of folded and unfolded proteins as well as folding intermediates differ from each other, as a blue-shift of the excitation wavelength is observed upon folding to the native state [175]. As tryptophane fluorescence measurements can also give clues on the stability of membrane proteins in different environments, such as when reconstituted into liposomes or polymersomes [176], this method will be discussed further in Sect. 6.1.4.

Useful information on the inclusion body expression and protein refolding methods for many different proteins can be found in the online database REFOLD (http://refold.med.monash.edu.au/) that currently (2013) holds 1,165 entries, of which about 40 concern OMPs [177].

A recent attempt to develop a universally applicable refolding method is based on the finding that the anionic detergent SDS that is very effective to solubilize proteins (also from the membrane; see Sect. 5.2.1), but has denaturing effects, when in presence of the amphipathic diol solvent (2-methyl-2,4-pentanediol) is transformed into a non-denaturing solubilization agent. Using this solvent mix the inclusion body derived eight-stranded bacterial OMP PagP could be refolded after inclusion body solubilization in SDS by the addition of 2-methyl-2,4-pentanediol [178]. The same solvent system has been used to refold the Omp2a from *Brucella melitensis* heterologous expressed into *E. coli* inclusion bodies [132].

5.2.3 A New Alternative: OMP Production Using Cell-Free Expression Systems

Though first attempts in cell-free protein expression have been made as early as in the 1960s [179] and were carried out chiefly to understand the involved cellular processes, the developments made during the last decade, render cell-free expression systems a powerful, rapid and efficient alternative for the production of proteins in general (allowing product yields higher than the mg/ml scale) and are used since around 2004 also for integral membrane proteins [180]. Especially for α-helical MPs the cell-free expression approach is a valid alternative to the membrane expression, as their refolding from inclusion body material works only in rare cases. However also in case of OMPs it might be a further way to overcome the limitations implied by the expression into and isolation from the outer membrane, namely low yields due to limited (membrane) space for correctly folded protein, toxic effects of OMP over-expression (especially when working with wide and open passive diffusion channel variants that lead to osmotic imbalances; see Sect. 5.1.2.2) or negative effects on the expression host cells caused by blocked OMP trafficking. Cell-free expression systems contain either the coupled transcription/translation mechanism or translation mechanism only of bacteria (generally of *E. coli* [181]) or of eukaryotic organisms such as yeast [182], insect cells [183], wheat germ cells [184] or rabbit reticulocytes [185]. Coupled systems allow to start from a DNA template, while translation mechanism systems require less facile mRNA templates [186]. A cell-free expression system can either be just a crude cell extract supplemented with essential amino acids, nucleotides, salts and ATP or guanosine triphosphate (GTP) replenishing factors, such as creatine kinase or creatine phosphatase [187, 188] or can consist of purified components (with the same supplements). The *E. coli* translational machinery (consisting of more than 100 proteins) for instance had been successfully purified and functionally reconstituted in 2001 [189] and has been further advanced into the commercially available PURE system 2010 [190]. While crude extracts are much easier to obtain and thus much less expensive, they are useful for the expression of proteins harboring a purification tag only, as the extracts contain a multitude of other proteins. Purified protein cell-free systems instead are difficult to obtain, as each protein has to be purified individually, they have the advantage though that the reaction conditions are better controllable and that target proteins without purification tags can be expressed and isolated. Various systems can be commercially obtained or prepared by standard protocols.

As the expression of bacterial OMPs does not necessitate the introduction of post-translational modifications, as is the case for many eukaryotic proteins, *E. coli* based cell-free systems should generally be suitable and the following discussion will therefore focus on such systems. Comprehensible, recent reviews on cell-free protein expression and their use for the biotechnological protein production can be found in [191–194].

E. coli cell-free expression systems are often based on the optimized coupled transcription/translation *E. coli* extracts [195]. All coupled cell-free systems contain the following major components:

1. For transcription: recombinant T7 RNA polymerase, nucleotides;

2. For translation: tRNAs, initiation, elongation, release factors, ribosomes, amino acyl-tRNA, essential amino acids synthetase;
3. General: ATP, co-factors, adenosine triphosphate (ATP)-replenishing factors.

Template DNA can be either plasmid DNA or polymerase chain reaction (PCR) derived DNA carrying a T7 promoter (most commonly used) as well a Shine-Dalgarno box as translation initiation signal.

The simplest cell-free protein expression format is the batch, where all components and template-DNA are mixed in a closed reaction container, without further substance addition during the reaction. Though easy to handle the batch format limits obtainable yields significantly, as it has rather short lifetimes below 1 h, due to the fast energy-delivering phosphate pool consumption and due to accumulating free phosphates that complex essential enzyme co-factors such as Mg [196]. This problem can be solved by either employing formats such as the complex continuous-flow cell-free expression with a continuous supply of substrates and energy [197], the simpler continuous-exchange cell-free expression with passive substrate and by-product exchange [198] and the efficient bilayer diffusion system [199] or by finding better alternatives to traditional ATP and/or GTP replenishing systems, while using the user-friendly batch format, such as the "Cytomim" system that uses the energy source pyruvate to obtain high yields of expressed protein [200].

As already mentioned cell-free OMP expression can be a good alternative especially to the expression into the outer membrane as it decouples protein production from cell growth and viability, as well as it avoids protein insertion into the outer membrane, thus avoiding problems caused by toxic effects an over-expressed OMP or OMP variant might cause and allowing higher yields due to the omitted membrane-caused space limitation. However until recently cell-free expression had been utilized to produce α-helical MPs exclusively, due to the lacking possibility to refold these proteins from inclusion body material. In 2005 the first OMP (*E. coli* nucleoside transporter Tsx) had been expressed using an *E. coli* derived cell-free expression system [201]. One can distinguish between two different modes of cell-free MP expression:

1. The expression mix does not contain detergents or other solubilization agents resulting in the precipitation of unsolubilized protein (somewhat similar to the expression into inclusion bodies, though mild detergents are sufficient for solubilization). Proteins are subsequently solubilized by the addition of detergents or other amphiphilic substances [180] (see Fig. 5.14a, step 1 and 2). Unsolubilized precipitate can be removed by centrifugation. The detergent 1-myristoyl-2-hydroxy-sn-glycero-3-[phosphorac-(1-glycerol)] (LMPG) has been proven to efficiently solubilize obtained protein precipitates [201]. In general the best suitable detergent has again to be decided on after carrying out a screening. Apart from lipids also lipid-protein nano-discs have been used to solubilize MPs from precipitates obtained by cell-free expression [202, 203]. The use of nano-discs often assists NMR-studies of the reconstituted protein (see also Chap. 3).
2. The expression mix is provided with detergents or other solubilization agents, resulting in the production of already solubilized protein [201, 204] (see Fig. 5.14b, step 1). The detergent class of polyoxyethylene-alkyl-ethers, with higher polymerization number of the polyoxyethylene moieties, has been shown to effectively solubilize membrane proteins during cell-free expression [201]. Several α-helical MPs have furthermore been successfully cell-free expressed in the presence of liposomes, leading to protein insertion into the lipid membrane [203, 205–207]. Recently the α-helical MP claudin-2, has been expressed in a wheat germ extract based cell-free system containing block copolymer vesicles of the polymer polybutadiene-polyethyleneoxide (PBD-PEO) [208].

Protein derived from both of the described cell-free expression modes has then to be analyzed by SDS-PAGE, to obtain information on

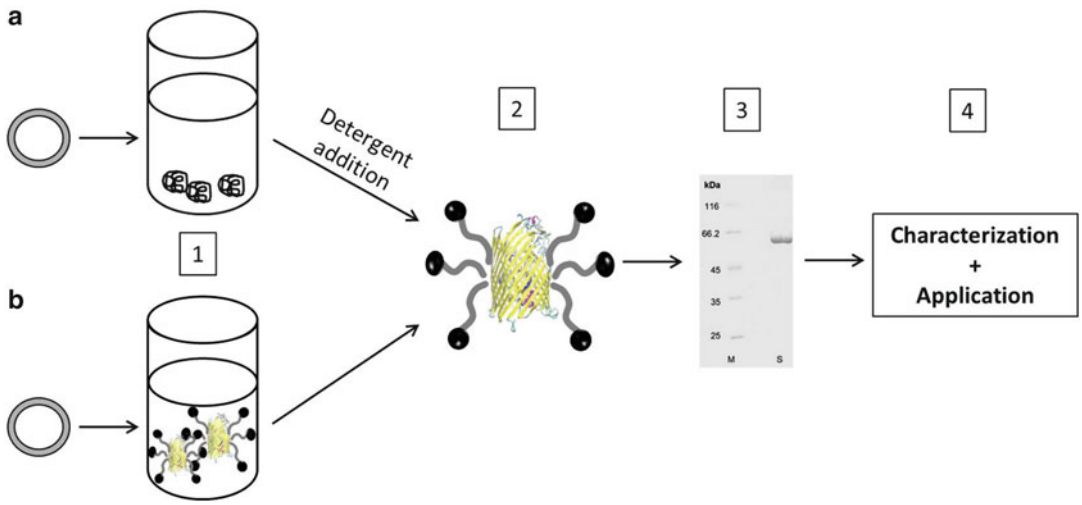

Fig. 5.14 Schematic representation of cell-free OMP expression. a – expression mix lacking detergents (or other solubilization agents), leading to protein precipitates that can be refolded by detergent addition; b – expression mix provided with detergents (or other solubilization agents); 1 – addition of template-DNA and cell-free expression; 2 – purification of solubilized OMP; 3 – yield and protein purity control (*e.g.* by SDS-PAGE); 4 – characterization and application of obtained OMP variant (Protein New Cartoon representations made with VMD)

yield and purity, before further characterization studies can be carried out or the protein can be used for the desired application (Fig. 5.14, step 3 and 4).

Up to now only few examples of cell-free expressed β-barrel shaped MPs including *E. coli* Tsx [201] and the mitochondrial, voltage-dependent anion channel (VDAC1) [209] and its chloroplast homologue OEP24 [210] are reported. Nevertheless together with the inclusion body expression/refolding approach the cell-free expression might become even more important to produce OMP variants for nano-technological applications, as both systems allow the expression of variants with "extreme" geometry, thus going "beyond natural limits". One thinkable example would be a β-barrel protein with hydrophobic region extended above the limit the biological lipid-bilayer can accommodate. In order to obtain such a protein it has to either be produced as insoluble precipitate (*i.e.* inclusion body or cell-free expression) with subsequent reconstitution using detergents or thick enough lipid/polymer membranes or solubly expressed by use of a cell-free system containing suitable detergents, lipids or polymers. In conclusion both introduced alternatives to the conventional OMP membrane expression allow higher protein yields, avoid toxic effects on an expression host and open the possibility to produce OMP variants with novel characteristics or altered geometries that cannot be found in nature, allowing to adjust the protein nano-channel's features to the desired application. Furthermore both inclusion body expression and cell-free expression can in the future probably be coupled with the OMP channel insertion into lipid or polymer membranes, which is a step towards functional nano-systems for various applications.

5.2.4 OMP Purification and Concentration Methods

Solubilized outer membrane proteins often have to be further purified to allow functional and structural characterization, especially for crystallization purposes or to perform CD spectroscopy samples have to be exceptionally pure. This is the case particularly for membrane-derived and cell-free expressed OMPs, but can be necessary

also for OMPs refolded from inclusion bodies (though inclusion body derived proteins are often already rather pure). Purification by non-specific chromatographic (*e.g.* gel filtration) methods often leads to a dilution of the purified target, several concentration techniques may be used to obtain highly concentrated samples, however special considerations have to be made when concentrating membrane proteins (see below).

In general the same purification techniques as used for water-soluble proteins can be used to purify MPs also. MPs however require the presence of detergents (or other amphiphiles) to remain in solution. Therefore they are usually purified in complex with detergents or detergents and lipids. Detergent concentrations during purification steps should be above cmc, but may be lower than when used for solubilization.

In the following will be given an overview on the common purification and protein concentration methods applied to bacterial OMPs. A complete work on protein purification methods at large can be found in [211] while [212] focusses on the purification of membrane proteins.

5.2.4.1 Immobilized Metal Affinity Chromatography

In the easiest and most universally applicable case the target outer membrane protein contains a His-tag allowing the specific purification and concentration by IMAC. Proteins containing a sequence of at least six histidine residues (His-tag) bind to Ni-ions that are immobilized to a chromatography resin, allowing the specific purification of tagged proteins [213]. IMAC is one of the preferred OMP purification methods, especially since the presence of a His-tag allows skipping the time consuming membrane isolation and OMP extraction, by applying uncleared bacterial lysate to the IMAC column and providing a suitable solubilization agent during chromatography (though in this case soluble metal-binding proteins present in the lysate might contaminate the purified fraction). Furthermore His-tagged OMPs derived from purified inclusion bodies can be loaded to an IMAC column after solubilizing the inclusion body material. The protein binds to the Ni-affinity resin, the column can be washed and by adding detergents (or other amphiphiles) during elution the protein can be refolded on-column (see also Table 5.4). The IMAC method can furthermore be used to change the detergent solubilizing the protein, as detergents that are used to extract a protein from the outer membrane are not always suitable for characterization purposes or applications involving the OMP reconstitution into polymer or lipid membranes.

Regarding the tag position one has to consider that the N-terminus is often less suitable for OMPs that are functionally expressed into the membrane, since the N-terminal signal sequence is processed and cleaved by a signal peptidase. Furthermore C- and N-terminus close to form the barrel with which the His-tag might interfere. For OMPs that do not tolerate a C-terminal His-tag the tag can be placed within an external loop that is not covered by detergents. Ferguson et al. for instance introduced an internal hexa-His-tag to the FhuA protein at the surface exposed amino acid position 405 [100].

The presence of detergents on hydrophobic protein portions can weaken protein binding to the column matrix, this problem can be addressed by increasing the tag-length (generally up to 10 His), and by increasing binding time. Binding time can be elongated either by batch-wise binding, *i.e.* mixing the protein sample and affinity matrix prior to column packing (the mixture can be incubated overnight) or by lowering the chromatography buffer flow rate during protein binding. Though batch-wise binding can further improve the overall yield, the long incubation time increases the probability of protein degradation.

Several bacterial OMPs have been provided with a His-tag and purified by IMAC either after being expressed into inclusion bodies, including on-column refolding (Table 5.4) or after functional expression into the bacterial outer membrane. Table 5.5 sums few examples of bacterial OMPs (with different tag locations) that were expressed into the *E. coli* outer membrane and were purified by IMAC after protein solubilization from the membrane fraction.

IMAC derived protein samples are often purified further by gel filtration to remove remaining

Table 5.5 Examples for bacterial OMPs expressed functionally into the *E. coli* outer membrane that were purified by Ni-IMAC; listing protein origin and tag position

OMP	Origin	Tag position	Reference
FhuA	*E. coli*	Internal	[101]
FhuAΔC/Δ4L	Variant of *E. coli* FhuA	Internal	[46]
PhoE	*E. coli*	N-terminal	[214]
Omp85	*Neisseria meningitidis*	N-terminal	[215]
OprM	*Pseudomonas aeruginosa*	C-terminal	[216]
OprN	*Pseudomonas aeruginosa*	C-terminal	[216]

aggregates or to exchange the buffer, as many reconstitution experiments or structural and functional studies in general require special buffer conditions and the imidazole used to elute proteins from the IMAC column may interfere with certain analytical methods, such as CD spectroscopy, as imidazole is a chiral molecule. As the His-tag might inhibit protein crystallization the insertion of a tag removal protease site should be considered prior to crystallization experiments. The TEV (tobacco etch virus) protease site is commonly used, however one has to be aware that the TEV protease is partly inactive in several detergents that are often used to solubilize OMPs after purification (*e.g.* OG, LDAO) [217].

5.2.4.2 Ion Exchange Chromatography

Ion exchange chromatography (IEXC) involves the use of a charged column matrix. The separation is based on competition between proteins carrying different surface charges for oppositely charged groups present on the matrix. IEXC is a good method to purify un-tagged OMPs or can be used in combination with IMAC (or other affinity chromatography methods) whenever especially high purity is necessary. Elution can be accomplished either by changing buffer pH and thus changing the protein net surface charge or more commonly by changing the ionic strength of the buffer.

In order not to shield the matrix surface charge, ionic detergents have to be avoided and non-ionic detergents (such as octyl-POE) or zwitterionic detergents have to be used.

Ion exchange chromatography can be used to separate folded OMPs from unfolded or partially folded proteins as the unfolded counterparts often elute at different ionic strengths than the correctly folded OMPs. *E. coli* FepA for instance elutes at lower ionic strength when in the unfolded state than when in its native form [101, 173]. Thus IEXC is a feasible way to further condition proteins refolded from inclusion body material. Since several OMPs bind to ion exchange resins when in high concentrations of urea, IEXC can be used to on-column refold inclusion body derived proteins (similar to the on-column refolding by IMAC) [101]. This technique has for instance been successfully deployed to refold the *Rhodopseudomonas blasticus* porin [62].

5.2.4.3 Gel Filtration (Size Exclusion Chromatography)

In gel filtration (GF) or size exclusion chromatography proteins in a sample are fractionated due to their relative size and GF in contrast to other chromatographic techniques does not involve any chemical interaction between protein and matrix. The GF matrix consists of porous beads and while molecules with hydrodynamic diameter above pore size are not able to enter the bead pores and will elute immediately, molecules whose hydrodynamic radii allow pore entering will be fractionated according to their radii (big radius > small radius) [218].

GF is generally the last step of a multi-step OMP purification protocol. Especially for crystallization purposes, when a completely homogeneous sample is necessary, gel filtration can be used to remove any remaining folding intermediates or protein aggregates, as they show different hydrodynamic diameters as compared to the hydrodynamic diameter of the correctly folded protein [101] and elute at different chromatography stages. GF allows a

buffer exchange and the removal of imidazole used to elute His-tagged proteins from a Ni-column.

As automated protein chromatography methods often rely on the detection of aromatic amino acid residues present in the proteins within a sample by UV absorption measurement at 280 nm, it has to be considered that detergents containing aromatic rings, such as Triton X-100 absorb in the UV region and lead to false positive signals or false high signals.

5.2.4.4 Concentration Methods

Especially for crystallization purposes but also for many applications that involve the OMP reconstitution into polymer or lipid membranes (*e.g.* for polymersome or liposome based drug-delivery systems) considerably high protein concentrations are required and further concentration of purified protein samples might be necessary. In case of IEXC and especially IMAC purification and concentration can be achieved at the same time, as the target-protein interacts with the column resin and elutes when a certain buffer is added, by applying an as much as possible small elution buffer volume the protein concentration can be increased. If the employed purification protocol ends however with a gel filtration step the sample will be diluted (often by a factor of 3). IMAC or IEXC can again be applied or the protein sample can be concentrated by for example:

1. Ultrafiltration: Ultrafiltration-based concentration techniques can be carried out either by commercially available stirred ultrafiltration cells (for volumes up to 500 ml) that can be pressurized to allow water and small solutes to leave through a narrow pore membrane or by centrifugation filter units (for smaller volumes) that remove liquid by centrifugation. Though commonly used for membrane proteins these devices co-concentrate the present detergent micelles, resulting in a change in detergent:protein ratio. This ratio change often leads to detergent or detergent-protein complex precipitation. Furthermore ultrafiltration membranes can be blocked by the protein or by detergent micelles
2. Dialysis: By dialysis of an already pure OMP sample against high molecular weight solutions of polyethylene glycol (PEG; most often PEG 20,000) combined with the use of dialysis membranes with very small average pore size the protein concentration can be increased, since PEG will deprive the sample of water [212]. The PEG can be dissolved to form a very high concentrated solution ($\geq$20 %), or can even be used as a solid. However this method will increase the detergent concentration as well, leading to the mentioned negative effects.
3. Lyophilization: Since bacterial porins like OmpF and OmpC are known to be stable even after lyophilization [219], this technique might be a very good way to concentrate a target OMP providing the protein resists such treatment. The lyophilizate in powder form can be resolubilized in a suitable amount of buffer. In this way a co-concentration of solubilization agents (*e.g.* detergent) can be avoided.

5.3 OMP Scale-Up Production

Up to now and to the best of the author's knowledge the only α-helical membrane protein produced on industrial scale is the aquaporin protein class. Since aquaporins can be defined as nano-sized water-treatment filters, as they allow the specific passage of water molecules at rates near to the diffusion limit [220] they have been used for the development of water purifying systems by the Danish company Aquaporin A/S (http://www.aquaporin.dk/) (see also Sect. 2.2.3). Scale-up procedures however mostly rely still on the conventional expression, extraction and purification methods, though a cell-free expression system in synthetic liposomes allowing the production of aquaporin Z amounts on the milligram scale has been demonstrated [221] and the cell-free expression approach has been suggested to have great potentials as a platform for the industrial scale protein production [222].

As no other membrane proteins and therefore OMPs are used for industrial applications, yet, no

high-end production procedures have been developed. However some general considerations can be made based on the pre-requisite that protein yield and purity have to be maximized, while time-loss and costs need to be minimized in order to make a production procedure industrially interesting:

1. The β-barrel OMP *expression into the outer membrane* and subsequent extraction/solubilization by *mild detergents* cannot feasibly be used to develop high level production procedures as the membrane offers rather limited space and especially channel forming OMPs can have toxic effects for the expression host cells. Furthermore mild detergents often show low extraction efficiency and the necessary membrane isolation is a time consuming affair (see Sect. 5.2.1). Expression systems with large internal membrane systems, such as *Rhodobacter* that are used to over-produce α-helical MPs [223] are not suitable for OMPs due to the nature of the outer membrane.
2. The β-barrel OMP *expression into the outer membrane* and subsequent extraction/solubilization using *strong detergents* such as SDS lead to much higher extraction efficiencies but the protein becomes unfolded due to the denaturing effects of SDS. The protein can however be refolded by procedures used also to refold inclusion body derived proteins resulting in considerably high yields (see Sect. 5.2.1, Figs. 5.11 and 5.15). A recent report of the successful transformation of SDS from a denaturing detergent to a non-denaturing one with high extraction/solubilization efficiency [132, 178] shows that SDS surely is a solubilization agent that should be considered when OMPs need to be produced in high concentrations. Especially as SDS also solubilizes inclusion body material.
3. The β-barrel OMP *expression into inclusion bodies* followed by inclusion body solubilization and protein refolding avoids the space limit posed by the outer membrane and avoids toxic effects (such as osmotic imbalances caused by the over-expression of a passive diffusion channel OMP) on the host cells. Obtained yields are high when compared to membrane expression and extraction by mild detergents (Fig. 5.15). The expression into inclusion bodies is a valid outer membrane protein over production platform especially for engineered OMP variants that affect host cells in a negative way when expressed into the membrane or that cannot be accommodated by the membrane due to newly introduced features (*i.e.* elongated hydrophobic portion) (see Sect. 5.2.2). A drawback are the often time consuming refolding procedures, however this problem can be avoided by the use of SDS in combination with the amphipathic diol solvent 2-methyl-2,4-pentanediol that upon addition leads to protein folding [132, 178].
4. Cell-free expression systems equally avoid the membrane space limit as well as OMP over-expression related toxic effects on the expression host, as it even is completely independent from the cellular system of a living organism. As optimized cell-free expression systems already report good overall yields (see Sect. 5.2.3) it is surely a powerful protein over-expression method that once efficient ways of direct solubilization or protein reconstitution are found and optimized can be used as a basis for an OMP production scale-up.

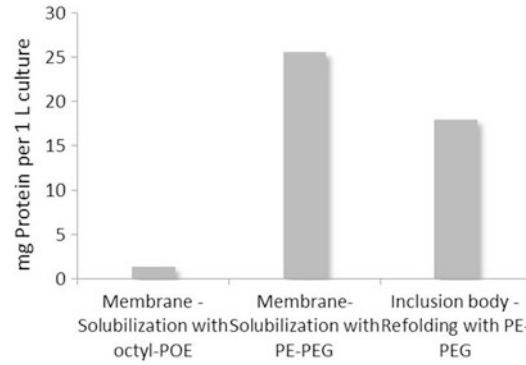

Fig. 5.15 FhuAΔ1–159 obtainable yields (in mg per 1 L of culture) by membrane extraction with the mild detergent octyl-POE, membrane extraction by SDS followed by refolding using PE-PEG and inclusion body expression and refolding again using PE-PEG

5. The use of cost-effective solubilization agents (*e.g.* PE-PEG diblock copolymer) is preferable over often costly detergents.

Figure 5.15 shows obtainable yields (in mg protein per L of initial culture) by expression into the membrane and mild detergent extraction [40], expression into the membrane, solubilization by SDS and subsequent dialysis based refolding using diblock copolymer PE-PEG [133] and expression into inclusion bodies, followed by dialysis based refolding again using diblock copolymer PE-PEG [41] on the example of FhuA variant FhuAΔ1–159. In all cases *E. coli* had been used as expression host. For better comparison values have been normalized assuming a purity of 100 %. While the conventional membrane expression method shows low yields of 1–2 mg/L of culture rendering it unfeasible for any industrial scale production, both membrane expression and solubilization by SDS and inclusion body expression lead to much higher yield, permitting 10–20 times more obtained protein from the same culture volume.

Though SDS solubilization of proteins from the outer membrane led to highest yields it has to be mentioned that the involved procedure is the most time consuming one, as it includes the isolation of the outer membrane, a protein extraction step and protein refolding by step-wise dialysis.

The step-wise dialysis necessary to refold inclusion body derived protein renders the respective protocol almost equally time ineffective, even though the isolation of inclusion bodies and inclusion body solubilization can be achieved rather fast. The combination of inclusion body expression with SDS solubilization and refolding by simple addition of 2-methyl-2,4-pentanediol or similar substances seems therefore rather promising and should be tried in the future.

In conclusion however it has to be noted that the nature of membrane proteins in general and the resulting special demands they make in terms of hydrophobicity of their environment and presence of a boundary surface (hydrophobic/hydrophilic) an industrial large-scale production in the range of bulk chemicals is very unlikely and scale-up intentions will most likely reach to the scale of fine chemical or pharmaceutics production rates, as required for nano-channel applications such as drug-delivery from nano-containers or OMP use as stochastic nano-sensors (Details on the nano-technological applications of OMPs can be found in Chap. 6).

5.4 Artificial β-Barrel Structures

In Parallel to the efforts to produce nano-channels based on engineered bacterial β-barrel OMPs a different scientific approach instead considered the development of synthetic nano-channels that are based on artificial β-barrel structures able to insert into the lipid bilayer [224–227]. These artificial and self-assembling β-barrels are based on rigid rod molecules. Rigid rod molecules are extremely rigid, synthetic rod-shaped molecules with great potential in material sciences as they have one huge advantage; they do not fold, avoiding folding problems one might encounter during production of any protein derived materials [224].

The repeating rigid rod unit of an artificial β-barrel can be the p-octiphenyl rod. To each of the rods phenyl rings short peptide sequences that show β-strand shape are attached, leading to the barrel monomer. These monomers spontaneously assemble to form antiparallel β-sheets that are rolled up into a cylindrical structure promoted by the amphiphilicity of the individual sheets. In the final barrel adjacent amino acid residues face alternately the inner and outer barrel surface, similar to the geometry of β-barrel protein (see also Sect. 2.3.1) [224, 228].

By introducing hydrophobic and hydrophilic amino acids of which the first face the barrel outside, while the latter face the barrel inside, resulting barrel structures are soluble in organic solvents and can form pores in bilayer membranes [224, 229].

Triggering of opening and closing upon an external stimulus (*e.g.* pH, ionic strength, voltage) of artificial barrels with further chemically modified inside amino acid residues had been achieved and other chemically modified inside facing residues allowed the recognition of for example magnesium cations, and phosphate anions,

nucleotides, carbohydrates, inositol phosphates, pyrenes, calixarenes, and fullerenes, oligosaccharides, RNA, DNA, peptides, and synthetic polymers [224], rendering such pore structures useable as stochastic nano-sensing elements.

Future further development of this technology can be considered as an alternative to biologically derived nano-channel materials especially considering the possibility to tune length and diameter of the pores. However the synthesis of these molecules is not trivial and large scale production might be difficult [225] and as described in Sects. 5.1.2.3 and 5.1.2.4 the *E. coli* FhuA protein structure has been shown to allow the engineering of protein nano-channels with altered geometrical features (changing length and diameter) reaching a similar *tune-ability*.

In the following Chap. 6 protein nano-channel reconstitution methods will be briefly introduced, the main applications of these reconstituted nano-channels will be explored and methods of system characterization will be explained.

References

1. Howorka S, Siwy Z (2009) Nanopore analytics: sensing of single molecules. Chem Soc Rev 38:2360–2384
2. Onoda A, Fukumoto K, Arlt M, Bocola M, Schwaneberg U, Hayashi T (2012) A rhodium complex-linked β-barrel protein as a hybrid biocatalyst for phenylacetylene polymerization. Chem Commun 48:9756–9758
3. Onaca O, Sarkar P, Roccatano D, Friedrich T, Hauer B, Grzelakowski M, Güven A, Fioroni M, Schwaneberg U (2008) Functionalized nanocompartments (Synthosomes) with a reduction-triggered release system. Angew Chem Int Ed 47:7029–7031
4. Güven A, Dworeck T, Fioroni M, Schwaneberg U (2011) Residue K556-A light triggerable gatekeeper to sterically control translocation in FhuA. Adv Eng Mater 13:B324–B329
5. Lolicato M, Reina S, Messina A, Guarino F, Winterhalter M, Benz R, De Pinto V (2011) Generation of artificial channels by multimerization of β-strands from natural porin. Biol Chem 392:617–624
6. Richardson SM, Wheelan SJ, Yearrington RM, Boeke JD (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Res 16:550–556
7. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of Escherichia coli K-12. Science 277:1453–1474
8. Hershberg R, Petrov DA (2008) Selection on codon bias. Annu Rev Genet 42:287–299
9. Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in Escherichia coli. BMC Mol Biol 8:100–113
10. Zhang W, Xiao W, Wei H, Zhang J, Tian Z (2006) mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli. Biochem Biophys Res Commun 349:69–78
11. Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Nat Biotechnol 6:1321–1325
12. Smith DB, Johnson K (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Genetics 67:31–40
13. Schmidt TG, Skerra A (1993) The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. Protein Eng 6:109–122
14. Voss S, Skerra A (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein Eng 10:975–982
15. 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
16. 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
17. Davis GD, Elisee C, Newham DM, Harrison RG (1999) New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 65:382–388
18. Waugh DS (2005) Making the most of affinity tags. Trends Biotechnol 23:316–320
19. Raghava G, Sahni G (1994) GMAP: a multi-purpose computer program to aid synthetic gene design, cassette mutagenesis and the introduction of potential restriction sites into DNA sequences. Biotechniques 16:1116–1123
20. Hoover DM, Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR- based gene synthesis. Nucleic Acids Res 30:e43
21. Fuglsang A (2003) Codon optimizer: a freeware tool for codon optimization. Protein Expr Purif 31:247–249
22. Gao W, Rzewski A, Sun H, Robbins P, Gambotto A (2004) UpGene: application of a web-based DNA codon optimization algorithm. Biotechnol Prog 20:443–448

23. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 1:W526–W531
24. Jayaraj S, Reid R, Santi DV (2005) GeMS: an advanced software package for designing synthetic genes. Nucleic Acids Res 33:3011–3016
25. Lorimer D, Raymond A, Walchli J, Mixon M, Barrow A, Wallace E, Grice R, Burgin A, Stewart L (2009) Gene Composer: database software for protein construct design, codon engineering, and gene synthesis. BMC Biotechnol 9:36
26. Raab D, Graf M, Notka F, Schödl T, Wagner R (2010) The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Syst Synth Biol 4:215–225
27. Wu G, Bashir-Bello N, Freeland SJ (2006) The synthetic gene designer: a flexible web platform to explore sequence manipulation for heterologous expression. Protein Expr Purif 47:441–445
28. Richardson SM, Nunley PW, Yarrington RM, Boeke JD, Bader JS (2010) GeneDesign 3.0 is an updated synthetic biology toolkit. Nucleic Acids Res 38:2603–2606
29. Jung S, McDonald K (2011) Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics 12:340
30. Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7:285–293
31. Freigassner M, Pichler H, Glieder A (2009) Tuning microbial hosts for membrane protein production. Microb Cell Fact 8:69
32. Steffensen L, Pedersen PA (2006) Heterologous expression of membrane and soluble proteins derepresses GCN4 mRNA translation in the yeast Saccharomyces cerevisiae. Eukaryot Cell 5:248–261
33. Wagner S, Bader ML, Drew D, de Gier JW (2006) Rationalizing membrane protein overexpression. Trends Biotechnol 24:364–371
34. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier JW (2008) Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci USA 105:14371–14376
35. Braun M, Killmann H, Braun V (1999) The beta-barrel domain of FhuADelta5–160 is sufficient for TonB-dependent FhuA activities of Escherichia coli. Mol Microbiol 33:1037–1049
36. Killmann H, Braun M, Herrmann C, Braun V (2001) FhuA barrel-cork hybrids are active transporters and receptors. J Bacteriol 183:3476–3487
37. Scott DC, Cao Z, Qi Z, Bauler M, Igo JD, Newton SM, Klebba PE (2001) Exchangeability of N termini in the ligand-gated porins of Escherichia coli. J Biol Chem 276:13025–13033
38. Braun M, Killmann H, Maier E, Benz R, Braun V (2002) Diffusion through channel derivatives of the Escherichia coli FhuA transport protein. Eur J Biochem 269:4948–4959
39. Nallani M, Onaca O, Gera N, Hildenbrand K, Hoheisel W, Schwaneberg U (2006) A nanophosphor-based method for selective DNA recovery in Synthosomes. Biotechnol J 1:828–834
40. Nallani M, Benito S, Onaca O, Graff A, Lindemann M, Winterhalter M, Meier W, Schwaneberg U (2006) A nanocompartment system (Synthosome) designed for biotechnological applications. J Biotechnol 123:50–59
41. Dworeck T, Petri AK, Muhammad N, Fioroni M, Schwaneberg U (2011) FhuA deletion variant Δ1–159 overexpression in inclusion bodies and refolding with Polyethylene-Poly(ethylene glycol) diblock copolymer. Protein Expr Purif 77:75–79
42. Boulanger P, Le Maire M, Bonhivers M, Dubois S, Desmadril M, Letellier L (1996) Purification and structural and functional characterization of FhuA, a transporter of the Escherichia coli outer membrane. Biochemistry 35:14216–14224
43. Rodríguez-Ropero F, Fioroni M (2012) Structural and dynamical analysis of an engineered FhuA channel protein embedded into a lipid bilayer or a detergent belt. J Struct Biol 177:291–301
44. Locher KP, Rees B, Koebnik R, Mitschler A, Moulinier L, Rosenbusch JP, Moras D (1998) Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. Cell 95:771–778
45. Endriß F, Braun V (2004) Loop deletions indicate regions important for FhuA transport and receptor functions in Escherichia coli. J Bacteriol 186:4818–4823
46. Mohammad M, Howard KR, Movileanu L (2011) Redesign of a plugged β-barrel membrane protein. J Biol Chem 286:8000–8013
47. Chen M, Khalid S, Sansom MSP, Bayley H (2008) Outer membrane protein G: engineering a quiet pore for biosensing. Proc Natl Acad Sci 105:6272–6277
48. Nestorovich EM, Danelon C, Winterhalter M, Bezrukov SM (2002) Designed to penetrate: time-resolved interaction of single antibiotic molecules with bacterial pores. Proc Natl Acad Sci 99:9789–9794
49. Jung Y, Bayley H, Movileanu L (2006) Temperature-responsive protein pores. J Am Chem Soc 28:15332–15340
50. Naveed H, Jimenez-Morales D, Tian J, Pasupuleti V, Kenney LJ, Liang J (2012) Engineered oligomerization state of OmpF protein through computational design decouples oligomer dissociation from unfolding. J Mol Biol 419(1–2):89–101
51. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science 274:1859–1866

52. Subbarao GV, van den Berg B (2006) Crystal structure of the monomeric porin OmpG. J Mol Biol 360:750–759
53. Cowan SW, Garavito RM, Jansonius JN, Jenkins JA, Karlsson R, Konig N, Pai EF, Pauptit RA, Rizkallah PJ, Rosenbach JP (1995) The structure of OmpF porin in a tetragonal crystal form. Structure 3:1041–1050
54. Tieleman DP, Berendsen HJ (1998) A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. Biophys J 74:2786–2801
55. Schulz GE (2002) The structure of bacterial outer membrane proteins. Biochim Biophys Acta 1565:308–317
56. Ferguson AD, Hofmann E, Coulton JW, Diederichs K, Welte W (1998) Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science 282:2215–2220
57. Mohammad M, Iyer R, Howard KR, McPike M, Borer PN, Movileanu L (2012) Engineering a rigid protein tunnel for biomolecular detection. J Am Chem Soc 134:9521–9531
58. Tenne SJ, Dworeck T, Fioroni M (2010) Unpublished raw data. Department for Biotechnology, RWTH Aachen University, Aachen
59. Bendezú FO, de Boer PAJ (2008) Conditional lethality, division defects, membrane involution, and endocytosis in mre and mrd shape mutants of Escherichia coli. J Bacteriol 190:1792–1811
60. Liu P, Duan W, Wang Q, Li X (2010) The damage of outer membrane of Escherichia coli in the presence of TiO2 combined with UV light. Colloids Surf B 78:171–176
61. Glover WA, Yang Y, Zhang Y (2009) Insights into the molecular basis of L-form formation and survival in Escherichia coli. PLoS One 4:e7316
62. Schmid B, Krömer M, Schulz GE (1996) Expression of porin from Rhodopseudomonas blastica in Escherichia coli inclusion bodies and folding into exact native structure. FEBS Lett 381:111–114
63. Sidorova OV, Isaeva MP, Khomenko VA, Portniagina O, Likhatskaia GN, Kim N, Novikova OD, Chistiulin DK, Solov'eva TF (2012) Yersinia pseudotuberculosis mutant OmpF porins with deletions of the external loops: genetic constructions design, expression, isolation and refolding. Bioorg Khim 38:156–165
64. Koide S, Huang X, Link K, Koide A, Bu Z, Engelman DM (2000) Design of single-layer b-sheets without a hydrophobic core. Nature 403:456–460
65. Schulz GE (2000) β-barrel membrane proteins. Curr Opin Struct Biol 10:443–447
66. Huang Y, Smith BS, Chen LX, Baxter RH, Deisenhofer J (2009) Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. Proc Natl Acad Sci USA 106:7403–7407
67. Krewinkel M, Dworeck T, Fioroni M (2012) Engineering of an E. coli outer membrane protein FhuA with increased channel diameter. J Nanobiotechnol 9:33
68. Krewinkel M, Dworeck T, Fioroni M (2011) Unpublished data. RWTH Aachen University, Aachen
69. Muhammad N, Dworeck T, Fioroni M, Schwaneberg U (2011) Engineering of the E. coli outer membrane protein FhuA to overcome the hydrophobic mismatch in thick polymeric membranes. J Nanobiotechnol 9:8
70. Ahmed F, Pakunlu R, Brannan A, Bates F, Minko T, Discher D (2006) Biodegradable polymersomes loaded with both paclitaxel and doxorubicin permeate and shrink tumors, inducing apoptosis in proportion to accumulated drug. J Control Release 116:150–158
71. Lee J, Bermudez H, Discher B, Sheehan M, Won Y, Bates F, Discher D (2001) Preparation, stability, and in vitro performance of vesicles made with diblock copolymers. Biotechnol Bioeng 73:135–145
72. Nardin C, Hirt T, Leukel J, Meier W (2000) Polymerized ABA triblock copolymer vesicles. Langmuir 16:1035–1041
73. Nardin C, Thoeni S, Widmer J, Winterhalter M, Meier W (2000) Nanoreactors based on (polymerized) ABA-triblock copolymer vesicles. Chem Commun 15:1433–1434
74. Nardin C, Widmer J, Winterhalter M, Meier W (2001) Amphiphilic block copolymer nanocontainers as bioreactors. Eur Phys J E 4:403–410
75. Mouritsen O, Bloom M (1984) Mattress model of lipid-protein interactions in membranes. Biophys J 46:141–153
76. Muhammad N, Dworeck T, Fioroni M (2011) Unpublished data. RWTH Aachen University, Aachen
77. Ihle S, Onaca O, Rigler P, Hauer B, Rodrıguez-Ropero F, Fioroni M, Schwaneberg U (2011) Nanocompartments with a pH release system based on an engineered OmpF channel protein. Soft Matter 7:532–539
78. Cheley S, Gu LQ, Bayley H (2002) Stochastic sensing of nanomolar inositol 1,4,5-trisphosphate with an engineered pore. Chem Biol 9:829–838
79. Fsihi H, Kottwitz B, Bremer E (1993) Single amino acid substitutions affecting the substrate specificity of the Escherichia coli K-12 nucleoside-specific Tsx channel. J Biol Chem 268:17495–17503
80. Basle E, Joubert N, Pucheaul M (2010) Protein chemical modification on endogenous amino acids. Chem Biol 17:213–227
81. Hermanson GT (2008) Bioconjugate techniques. Academic, Amsterdam
82. Güven A, Fioroni M, Hauer B, Schwaneberg U (2010) Molecular understanding of sterically controlled compound release through an engineered channel protein (FhuA). J Nanobiotechnol 8:14
83. Patchornik A, Amit B, Woodwars RB (1970) Photosensitive protecting groups. J Am Chem Soc 92:6333–6335

84. Bös C, Braun V (1997) Specific in vivo thiol-labeling of the FhuA outer membrane ferrichrome transport protein of Escherichia coli K-12: evidence for a disulfide bridge in the predicted gating loop. FEMS Microbiol Lett 153:311–319
85. Bös C, Lorenzen D, Braun V (1998) Specific in vivo labeling of cell surface-exposed protein loops: reactive cysteines in the predicted gating loop mark a ferrichrome binding site and a ligand-induced conformational change of the Escherichia coli FhuA protein. J Bacteriol 180:605–613
86. Howorka S, Movileanu L, Lu X, Magnon M, Cheley S, Braha O, Bayley H (2000) A protein pore with a single polymer chain tethered within the lumen. J Am Chem Soc 122:2411–2416
87. Gu LQ, Braha O, Conlan S, Cheley S, Bayley H (1999) Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. Nature 398:686–690
88. Walker BJ, Bayley H (1994) A pore-forming protein with a protease-activated trigger. Protein Eng 7:91–97
89. Braha O, Walker BJ, Cheley S, Kasianowicz JJ, Song L, Gouaux E, Bayley H (1997) Designed protein pores as components for biosensors. Chem Biol 4:497–505
90. Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci USA 93:13770–13773
91. Heinz C, Engelhardt H, Niederweis M (2003) The core of the tetrameric mycobacterial porin MspA is an extremely stable β-sheet domain. J Biol Chem 278:8678–8685
92. Faller M, Niederweis M, Schulz GE (2004) The structure of a mycobacterial outer-membrane channel. Science 303:1189–1192
93. Butler T, Pavlenok M, Derrington IM, Niederweis M, Gundlach JH (2008) Single-molecule DNA detection with an engineered MspA protein nanopore. Proc Natl Acad Sci 105:20647–20652
94. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH (2012) Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat Biotechnol 30:349–354
95. Chang CY, Niblack B, Walker BJ, Bayley H (1995) A photogenerated pore-forming protein. Chem Biol 2:391–400
96. Miercke L, Ross PE, Strout RM, Dratz EA (1989) Purification of bacteriorhodopsin and characterization of mature and partially processed forms. J Biol Chem 264:7531–7535
97. Nestel U, Wacker T, Woitzik D, Weckesser J, Kreutz W, Welte W (1989) Crystallization and preliminary X-ray analysis of porin from Rhodobacter capsulatus. FEBS Lett 242:405–408
98. Stauffer KA, Page MGP, Hardmeyer A, Keller TA, Pauptit RA (1989) Crystallization and preliminary X-ray characterization of maltoporin from Escherichia coli. J Mol Biol 211:297–299
99. Forst D, Schülein K, Wacker T, Diederichs K, Kreutz W, Benz R, Welte W (1993) Crystallization and preliminary X-ray diffraction analysis of ScrY, a specific bacterial outer membrane porin. J Mol Biol 229:258–262
100. Ferguson AD, Breed J, Diederichs K, Welte W, Coulton JW (1998) An internal affinity-tag for purification and crystallization of the siderophore receptor FhuA, integral outer membrane protein from Escherichia coli K-12. Protein Sci 7:1636–1638
101. Bannwarth M, Schulz GE (2003) The expression of outer membrane proteins for crystallization. Biochim Biophys Acta 1610:37–45
102. Schwarz D, Klammt C, Koglin A, Löhr F, Schneider B, Dötsch V, Bernhard F (2007) Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. Methods 41:355–369
103. 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
104. Dascher C, Roll D, Bavoil PM (1993) Expression and translocation of the chlamydial major outer membrane protein in Escherichia coli. Microb Pathog 15:455–467
105. Thomas KL, Leduc I, Olsen B, Thomas CE, Cameron DW, Elkins C (2001) Cloning, overexpression, purification, and immunobiology of an 85-kilodalton outer membrane protein from Haemophilus ducreyi. Infect Immun 69:4438–4446
106. Findlay HE, McClafferty H, Ashley RH (2005) Surface expression, single-channel analysis and membrane topology of recombinant Chlamydia trachomatis major outer membrane protein. BMC Microbiol 5:5
107. Singh R, Gupta PK, Durga V, Rao P (2009) Expression and purification of the major outer membrane protein (OmpH) of Pasteurella multocida P52 from Escherichia coli. Vet Arhiv 79:591–600
108. Haghi F, Peerayeh SN, Siadat SD, Montajabiniat M (2011) Cloning, expression and purification of outer membrane protein PorA of Neisseria meningitidis serogroup B. J Infect Dev Ctries 5:856–862
109. Lee J, Kim S-H (2009) High-throughput T7 LIC vector for introducing C-terminus poly-histidine tags with variable lengths without extra sequences. Protein Expr Purif 63:58–61
110. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189:113–130
111. Gosh R, Steiert M, Hardmeyer A, Wang YF, Rosenbusch JP (1998) Overexpression of outer membrane porins in E. coli using pBluescript-derived vectors. Gene Expr 7:149–161

112. Prilipov A, Phale PS, Van Gelder P, Rosenbusch JP, Koebnik R (1998) Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli. FEMS Microbiol Lett 163:65–72
113. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, New York
114. Onoda T, Enokizono J, Kaya H, Oshima A, Freestone P, Norris V (2000) Effects of calcium and calcium chelators on growth and morphology of Escherichia coli L-form NC-7. J Bacteriol 182:1419–1422
115. Marr AG, Ingraham JL (1962) Effect of temperature on the composition of fatty acids in Escherichia coli. J Bacteriol 84:1260–1267
116. Garwin JL, Klages AL, Cronan JE (1980) β-ketoacyl-acyl carrier protein synthase II of Escherichia coli. evidence for function in the thermal regulation of fatty acid synthesis. J Biol Chem 255:3263–3265
117. Zhang YM, Rock CO (2008) Membrane lipid homeostasis in bacteria. Nat Rev Microbiol 6:222–233
118. Harrison STL (1991) Bacterial cell disruption: a key unit operation in the recovery of intracellular products. Biotechnol Adv 9:217–240
119. Kleinig AR, Middelberg APJ (1998) On the mechanism of microbial cell disruption in high-pressure homogenisation. Chem Eng Sci 53:891–898
120. French CS, Milner HW (1955) Disintegration of bacteria and small particles by high-pressure extrusion, vol 1, Methods in enzymology. Academic, New York
121. Johnson BH, Hecht MH (1994) Recombinant proteins can be isolated from E. coli cells by repeated cycles of freezing and thawing. Biotechnology 12:1357–1360
122. Felix H (1982) Permeabilised cells. Anal Biochem 120:211–234
123. Wase DAJ, Patel YR (1985) Effect of cell volume on disintegration by ultrasonics. J Chem Technol Biotechnol 35B:165–173
124. Andrews BA, Asenjo JA (1987) Enzymatic lysis and disruption of microbial cells. Biotechnol Lett 5:273–277
125. Lutkenhaus JF (1977) Role of a major outer membrane protein in E. coli. J Bacteriol 131:631–637
126. Hantke K (1981) Regulation of ferric iron transport in Escherichia coli K-12: isolation of a constitutive mutant. Mol Gen Genet 191:288–292
127. Beis K, Nesper J, Whitfield C, Naismith JH (2004) Crystallization and preliminary X-ray diffraction analysis of Wza outermembrane lipoprotein from Escherichia coli serotype O9a:K30. Acta Crystallogr D60:558–560
128. Arnold T, Linke D (2008) The use of detergents to purify membrane proteins. Curr Protoc Protein Sci Chapter 4:Unit 4.8.1–4.8.30. doi: 10.1002/0471140864.ps0408s53
129. Gohon Y, Popot JL (2003) Membrane protein-surfactant complexes. Curr Opin Colloid Interface Sci 8:15–22
130. Mo Y, Lee B-K, Ankner JF, Becker JM, Heller WT (2008) Detergent-associated solution conformations of helical and beta-barrel membrane proteins. J Phys Chem B 112:13349–13354
131. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta 1666:105–117
132. Roussel G, Perpete EA, Matagne A, Tinti E, Michaux C (2013) Towards a universal method for protein refolding: the trimeric beta barrel membrane Omp2a as a test case. Biotechnol Bioeng 110:417–423
133. Petri AK, Hariskos I, Dworeck T (2010) Unpublished data. RWTH Aachen University, Aachen
134. Shultis DD, Purdy MD, Banchs CN, Wiener MC (2006) Outer membrane active transport: structure of the BtuB:TonB complex. Science 312:1396–1399
135. Peybay-Peyroula E, Garavito RM, Rosenbusch JP, Zulauf M, Timmins PA (1995) Detergents structure in tetragonal crystals of OmpF porin. Structure 3:1051–1059
136. Evanics F, Hwang PM, Cheng Y, Kay LE, Prosser RS (2006) Topology of an outer-membrane enzyme: measuring oxygen and water contacts in solution NMR studies of PagP. J Am Chem Soc 128:8256–8264
137. Hwang PM, Choy WY, Lo EI, Chen L, Forman-Kay JD, Raetz CR, Privé GG, Bishop RE, Kay LE (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR. Proc Natl Acad Sci 99:13560–13565
138. Gutmann DAP, Mizohata E, Newstead S, Ferrandon S, Henderson PJF, van Veen HW, Byrne B (2007) A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay. Protein Sci 16:1422–1428
139. le Maire M, Champeil P, Möller JV (2000) Interaction of membrane proteins and lipids with solubilizing detergents. Biochim Biophys Acta 1508:86–111
140. Arachea BT, Sun Z, Potente N, Malik R, Isailovic D, Viola RE (2012) Detergent selection for enhanced extraction of membrane proteins. Protein Expr Purif 86:12–20
141. Winstone T, Duncalf KA, Turner RJ (2002) Optimization of expression and the purification by organic extraction of the integral membrane protein EmrE. Protein Expr Purif 26:111–121
142. Civjan NR, Bayburt TH, Schuler MA, Sligar SG (2003) Direct solubilization of heterologously expressed membrane proteins by incorporation into nanoscale lipid bilayers. Biotechniques 35:556–563
143. Inagaki S, Ghirlando R, Grisshammer R (2013) Biophysical characterization of membrane proteins in nanodiscs. Methods 59:287–300
144. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

145. Ratha A, Glibowickaa M, Nadeaua VG, Chena G, Debera CM (2008) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci 106:1760–1765
146. Schlegel S, Klepsch MM, Gialama D, Wickström D, Slotboom DJ, de Gier JW (2010) Revolutionizing membrane protein overexpression in bacteria. Microb Biotechnol 3:403–411
147. Walker J (1994) The bicinchoninic acid (BCA) assay for protein quantitation. Methods Mol Biol 32:5–8
148. Waterborg J, Matthews H (1984) The Lowry method for protein quantitation. Methods Mol Biol 1:1–3
149. Kruger N (1994) The Bradford method for protein quantitation. Methods Mol Biol 32:9–15
150. Laage R, Langosch D (2001) Strategies for prokaryotic expression of eukaryotic membrane proteins. Traffic 2:99–104
151. Middelberg A (2002) Preparative protein refolding. Trends Biotechnol 20:437–443
152. Rudolph R (1995) Successful protein folding on an industrial scale. In: Cleland JL, Craik CS (eds) Protein engineering: principles and practices. Wiley, New York, pp 283–298
153. Dekker N, Merck K, Tommassen J, Verheij HM (1995) In vitro folding of Escherichia coli outer-membrane phospholipase A. Eur J Biochem 232:214–219
154. Mitraki A, Fane B, Haase-Pettingell C, Sturtevant J, King J (1991) Global suppression of protein folding defects and inclusion body formation. Science 253:54–58
155. Tsumoto K, Ejima D, Kumagai I, Arakawa T (2003) Practical considerations in refolding proteins from inclusion bodies. Protein Expr Purif 28:1–8
156. Steinle A, Li P, Morris DL, Groh V, Lanier LL, Strong RK, Spies T (2001) Interactions of human NKG2D with its ligands MICA, MICB, and homologs of the mouse RAE-1 protein family. Immunogenetics 53:279–287
157. O'Callaghan CA, Tormo J, Willcox BE, Blundell CD, Jakobsen BK, Stuart DI, McMichael AJ, Bell JI, Jones EY (1998) Production, crystallization, and preliminary X-ray analysis of the human MHC class Ib molecule HLA-E. Protein Sci 7:1264–1266
158. Khan RH, AppaRao KBC, Eshwari ANS, Totey SM, Panda AK (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
159. Roudolph R, Lilie H (1996) In vitro folding of inclusion body proteins. FASEB J 10:49–56
160. Buchanan SK (1999) Beta-barrel proteins from bacterial outer membranes: structure, function and refolding. Curr Opin Struct Biol 9:455–461
161. Stockel J, Doring K, Malotka J, Jahnig F, Dornmair K (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
162. Burgess RR (1996) Purification of overproduced Escherichia coli RNA polymerase sigma factor by solubilizing inclusion bodies and refolding from sarkosyl. Methods Enzymol 273:145–149
163. Cardamone M, Puri NK, Brandon MR (1995) Comparing the refolding and reoxidation of recombinant porcine growth hormone from a urea denatured state and from Escherichia coli inclusion bodies. Biochemistry 34:5773–5794
164. Hochuli E, Döbeli H, Schacher A (1987) New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine residues. J Chromatogr 411:177–184
165. De Bernadez-Clark E, Schwarz E, Rudolph R (1999) Inhibition of aggregation side reactions during in vitro protein folding. Methods Enzymol 309:217–236
166. Petri AK, Dworeck T (2010) Unpublished data. RWTH Aachen University, Aachen
167. Charbonnier F, Köhler T, Pechère JC, Ducruix A (2001) Overexpression, refolding, and purification of the histidine-tagged outer membrane efflux protein OprM of Pseudomonas aeruginosa. Protein Expr Purif 23:121–127
168. Baldermann C, Engelhardt H (2000) Expression, two-dimensional crystallization, and three-dimensional reconstruction of the beta8 outer membrane protein Omp21 from Comamonas acidovorans. J Struct Biol 131:96–107
169. Prince SM, Achtman M, Derrick JP (2002) Crystal structure of the OpcA integral membrane adhesin from Neisseria meningitidis. Proc Natl Acad Sci USA 99:3417–3421
170. Kumar PD, Krishnaswamy S (2005) Overexpression, refolding, and purification of the major immunodominant outer membrane porin OmpC from Salmonella typhi: characterization of refolded OmpC. Protein Expr Purif 40:126–133
171. Pautsch A, Vogt J, Model K, Siebold C, Schulz GE (1999) Strategy for membrane protein crystallization exemplified with OmpA and OmpF. Proteins 34:167–172
172. Kramer RA, Zandwijken D, Egmond MR, Dekker N (2000) In vitro folding, purification and characterization of Escherichia coli outer membrane protease OmpT. Eur J Biochem 267:885–893
173. Buchanan SK (1999) Overexpression and refolding of an 80-kDa iron transporter from the outer membrane of Escherichia coli. Biochem Soc Trans 27:903–908
174. Anbazhagan V, Qu J, Kleinschmidt JH, Marsh D (2008) Incorporation of outer membrane protein OmpG in lipid membranes: protein-lipid interactions and beta-barrel orientation. Biochemistry 47:6189–6198
175. Royer CA (1995) Fluorescence spectroscopy, Methods in molecular biology. Humana Press, Totowa
176. Moon CP, Fleming KG (2011) Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes, vol 492, Methods in enzymology. Elsevier, Baltimore

177. Chow MK, Amin AA, Fulton KF, Fernando T, Kamau L, Batty C, Louca M, Ho S, Whisstock JC, Bottomley SP, Buckle AM (2006) The RE-FOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res 34:D207–D212
178. Michaux C, Pomroy NC, Prive GG (2008) Refolding SDS-denatured proteins by the addition of amphipathic cosolvents. J Mol Biol 375:1477–1488
179. Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci USA 47:1588–1602
180. Klammt C, Löhr F, Schäfer B, Haase W, Dötsch V, Rüterjans H, Glaubitz C, Bernhard F (2004) High level cell-free expression and specific labeling of integral membrane proteins. Eur J Biochem 271:568–580
181. Zubay G (1973) In vitro synthesis of protein in microbial systems. Annu Rev Genet 7:267–287
182. Gaisor E, Herrera F, Sadnik I, McLaughlin CS, Moldave K (1979) The preparation and characterization of a cell-free system from Saccharomyces cerevisiae that translates natural messenger ribonucleic acid. J Biol Chem 254:3965–3969
183. Swerdel MR, Fallon AM (1989) Cell-free translation in lysates from Spodoptera frugiperda (Lepidoptera:Noctuidae) cells. Comp Biochem Physiol 93B:803–806
184. Roberts BE, Patterson BM (1973) Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc Natl Acad Sci 70:2330–2334
185. Mosca JD, Wu JM, Suhadolnik PJ (1983) Restoration of protein synthesis in lysed rabbit reticulocytes by the enzymatic removal of adenosine 5′-monophosphate with either AMP deaminase or AMP nucleosidase. Biochemistry 22:346–354
186. Jermutus L, Ryabova LA, Plückthun A (1998) Recent advances in producing and selecting functional proteins by using cell-free translation. Curr Opin Biotechnol 9:534–548
187. Jackson AM, Boutell J, Cooley N, He M (2004) Cell-free protein synthesis for proteomics. Brief Funct Genomic Proteomic 2:308–319
188. Kigawa T, Yabuki T, Yoshida Y, Tsitsui M, Ito Y, Shibata T, Yokoyama S (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett 442:15–19
189. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755
190. Ohashi H, Kanamori T, Shimizu Y, Ueda T (2010) A highly controllable reconstituted cell-free system-a breakthrough in protein synthesis research. Curr Pharm Biotechnol 11:267–271
191. Katzen F, Chang G, Kudlicki W (2005) The past, present and future of cell-free protein synthesis. Trends Biotechnol 23:150–156
192. Carlson ED, Gan R, Hodgman CE, Jewett MC (2012) Cell-free protein synthesis: applications come of age. Biotechnol Adv 30:1185–1194
193. He M (2008) Cell-free protein synthesis: applications in proteomics and biotechnology. Nat Biotechnol 25:126–132
194. Gite S, Lim M, Rothschild KJ (2006) Cell-free protein synthesis systems: biotechnological applications. Biotechnol Genet Eng Rev 22:151–169
195. Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4:220
196. Kim DM, Swartz JR (1999) Prolonging cell-free protein synthesis with a novel ATP regeneration system. Biotechnol Bioeng 66:180–188
197. Spirin AS, Baranov VI, Ryabova LA, Ovodov SY, Alakhov YB (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. Science 242:1162–1164
198. Kim DM, Choi CY (1996) A semi-continuous prokaryotic coupled transcription/translation system using a dialysis membrane. Biotechnol Prog 12:645–649
199. Sawasaki T, Hasegawa Y, Tsuchimochi M, Kamura N, Ogasawara T, Kuroita T, Endo Y (2002) A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett 514:102–105
200. Jewett MC, Swartz JR (2004) Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng 86:19–26
201. Klammt C, Schwarz D, Fendler K, Haase W, Dötsch V, Bernhard F (2005) Evaluation of detergents for the soluble expression of a-helical and b-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. FEBS J 272:6024–6038
202. Shenkarev ZO, Lyukmanova EN, Butenko IO, Petrovskaya LE, Paramonov AS, Shulepko MA, Nekrasova OV, Kirpichnikov MP, Arseniev AS (2013) Lipid-protein nanodiscs promote in vitro folding of transmembrane domains of multi-helical and multimeric membrane proteins. Biochim Biophys Acta 1828:776–784
203. Periasamy A, Shadiac N, Amalraj A, Garajová S, Nagarajan Y, Waters S, Mertens HD, Hrmova M (2013) Cell-free protein synthesis of membrane (1,3)-β-d-glucan (curdlan) synthase: co-translational insertion in liposomes and reconstitution in nanodiscs. Biochim Biophys Acta 1828:743–757
204. Ishihara G, Goto M, Saeki M, Ito K, Hori T, Kigawa T, Shirouzu M, Yokoyama S (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
205. Kalmbach R, Chizhov I, Schumacher MC, Friedrich T, Bamberg E, Engelhard M (2007) Functional

cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes. J Mol Biol 371:639–648
206. Goren MA, Fox BA (2008) Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. Protein Expr Purif 62:171–178
207. Jarecki BW, Makino S, Beebe ET, Fox BG, Chanda B (2013) Function of Shaker potassium channels produced by cell-free translation upon injection into Xenopus oocytes. Sci Rep 3:1040
208. Nallani M, Andreasson-Ochsner M, Tan CW, Sinner EK, Wisantoso Y, Geifman-Shochat S, Hunziker W (2011) Proteopolymersomes: in vitro production of a membrane protein in polymersome membranes. Biointerphases 6:153–157
209. Nguyen TA, Lieu SS, Chang G (2010) An Escherichia coli-based cell-free system for large-scale production of functional mammalian membrane proteins suitable for X-ray crystallography. J Mol Microbiol Biotechnol 18:85–91
210. Liguori L, Marques B, Villegas-Méndez A, Rothe R, Lenormand JL (2007) Production of membrane proteins using cell-free expression systems. Expert Rev Proteomics 4:79–90
211. Burgess RR, Deutscher MP (2009) Guide to protein purification, vol 463, 2nd edn, Methods in enzymology. Academic, Waltham
212. von Jagow G, Schägger H (1994) A practical guide to membrane protein purification, vol 2, Separation, detection, and characterization of biological macromolecules. Academic, San Diego
213. Arnold FH, Haymore BL (1991) Engineering metal-binding proteins: purification to protein folding. Science 252:1796–1797
214. Van Gelder P, Steiert M, El Khattabi M, Rosenbusch JP, Tommassen J (1996) Structural and functional characterization of a His-tagged PhoE pore protein of Escherichia coli. Biochem Biophys Res Commun 229:869–875
215. Volokhina EB, Beckers F, Tommassen J, Bos MP (2009) The β-barrel outer membrane protein assembly complex of Neisseria meningitidis. J Bacteriol 191:7074–7085
216. Broutin I, Benabdelhak H, Moreel X, Lascombe MB, Lerouge D, Ducruix A (2005) Expression, purification, crystallizationand preliminary X-ray studies of the outer membrane efflux proteins OprM and OprN from Pseudomonas aeruginosa. Acta Crystallogr Sect F Struct Biol Cryst Commun 61:315–318
217. Mohanty AK, Simmons CR, Wiener MC (2003) Inhibition of tobacco etch virus protease activity by detergents. Protein Expr Purif 27:109–114
218. Stellwagen E (2009) Gel filtration. In: Burgess RR, Deutscher MP (eds) Guide to protein purification, vol 463, 2nd edn, Methods in enzymology. Academic, Waltham
219. Luckey M (2008) Membrane structural biology – with biochemical and biophysical foundations. Cambridge University Press, New York
220. Savage DF, Stroud RM (2007) Structural basis of aquaporin inhibition by mercury. J Mol Biol 368:607–617
221. Hovijitra NT, Wuu JJ, Peaker B, Swartz JR (2009) Cell-free synthesis of functional aquaporin Z in synthetic liposomes. Biotechnol Bioeng 104:40–49
222. Tang CY, Zhao Y, Wang R, Helix-Nielsen C, Fane AG (2013) Desalination by biomimetic aquaporin membranes: review of status and prospects. Desalination 308:34–40
223. Laible PD, Mielke DL, Hanson DK (2005) Membrane protein production: a bacterial "factory" in Rhodobacter. Screening 2:30–32
224. Sakai N, Mareda J, Matile S (2008) Artificial β-barrels. Acc Chem Res 41:1354–1365
225. Baumeister B, Matile S (2000) Rigid-rod beta-barrels as lipocalin models: probing confined space by carotenoid encapsulation. Chemistry 6:1739–1749
226. Sakai N, Matile S (2003) Synthetic multifunctional pores: lessons from rigid-rod beta-barrels. Chem Commun 21:2514–2523
227. Schwab PFH, Levin MD, Michl J (1999) Molecular rods. 1. Simple axial rods. Chem Rev 99:1863–1934
228. Sakai N, Mareda J, Matile S (2005) Rigid-rod molecules in biomembrane models: from hydrogen-bonded chains to synthetic multifunctional pores. Acc Chem Res 38:79–87
229. Clark TD, Buehler LK, Ghadiri MR (1998) Self-assembling cyclic b-peptide nanotubes as artificial transmembrane ion channels. J Am Chem Soc 120:651–656

Technological Applications

6

The chapter describes the reconstitution of a series of non-bacterial β-barrel proteins as well as bacterial outer membrane β-barrel proteins (OMPs) into lipid or polymer vesicle membranes or flat membranes, to function as nano-channels. The microscopic reconstitution process is described based on the actual knowledge, introducing the concepts of hydrophobic mismatch and polydispersity of a polymer sample, both being variables affecting protein reconstitution. The main experimental reconstitution methods are reported emphasizing their limits to obtain a functional protein reconstitution. Furthermore an ensemble of experimental methods to characterize the assembled systems is reported, *i.e.* dynamic light scattering (DLS), spectroscopic flux assay, patch clamp, electron microscopy (EM) and tryptophan fluorescence. At the end of the chapter a description of the technological applications based on the reconstituted protein nano-channels is reported, including drug delivery, stochastic nanosensors and bionano-electronics.

6.1 Reconstitution into Lipid/Polymer Vesicles or Membranes and Characterization of the New Systems

As in nature β-barrel membrane proteins reside in the lipid bilayer of the outer membrane, with the lipid composition of the membrane influencing the protein structure and function, any nano-technological use of these channel proteins necessitates their reconstitution or the insertion into artificial lipid or polymer membranes that can mimic to some extent the proteins natural environment.

The processes involved in protein reconstitution are therefore rather important and need to be understood in order to develop new β-barrel membrane protein based nano-materials and will be discussed in the following section.

6.1.1 Membrane Protein Reconstitution

Bacterial β-barrel outer membrane proteins (OMPs) and β-barrel membrane proteins in general have been selected by evolution to fit within a lipid bilayer. The thickness of a lipid bilayer is due to two main components: the hydrophilic head group spanning approximately 1.0–1.3 nm function of the hydration condition and the hydrophobic central part, showing a higher variability depending on the length of the acylic tails, comprised between 2.5 and 3.5 nm [1, 2]. For a better description of the lipid bilayer check Sect. 2.1, while for a complete and exhaustive lipid chemical and physical properties summary with various important links to databases check the Cyberlipid Center at http://www.cyberlipid.org/index.html.

Though cell membranes are complex systems characterized by the coexistence of a series of lipids with different chemical and physical

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_6,

properties [3], the hydrophobic core of natural lipid membranes is mainly constituted by saturated fatty acids that can be well defined as constituted by short or low molecular weight polyethylene chains ($-[CH_2]_n-$) with a degree of polymerization n strongly dependent from the organism and cell ranging from $n = 4$ (butanoic acid) to $n = 35$ (pentatriacontanoic acid).

The modulation of the membranes physical-chemical properties is function of the main components *i.e.* derivatives of di- and tri-glycerides, "doped" by supplementary constituents like unsaturated lipid chains with one or more unsaturated double carbon bonds, sterols, sphingolipids and glycolipids [4].

The resulting "two dimensional liquid membrane" possesses peculiar properties to satisfy complex cellular functions as requested by the cell's inner physiology as well as functions involving communication with the outer world *i.e.* for the flux of matter [4].

The complexity of the system can be described by considering the temperature dependent phase transitions for a "pure" lipid bilayer which is acyl chain length dependent. By increasing the temperature the L_β phase or subgel phase is followed by the $L_{\beta'}$ phase or gel phase (in both phases the hydrocarbon tails are tilted toward the bilayer surface normal but, in the $L_{\beta'}$ phase the head group shows higher hydration). Increasing further the temperature leads to the rippled ($P_{\beta'}$) phase (the lipid bilayer shows a corrugated surface) to finally end up in the liquid crystalline or fluid L_α phase (where the lipid chains show an order/disorder transition, defining the bilayer as a two-dimensional fluid) [5].

The acyl chains length mainly characterizes the lipid bilayer transition temperature while the head group and the presence of small molecules interacting with the head of the lipid molecules influence the structure of the bilayer low temperature phases. When the head group is small, the L_β gel phase is stable (the tails do not show a tilt with respect to the bilayer normal like in the $L_{\beta'}$) while if small amphiphilic molecules, for example ethanol, are added to the bilayer, the low temperature interdigitated $L_\beta I$ phase is the most stable (where the terminal methyl groups of the lipid chains are located near the head group region of the opposite layer) [6].

As a consequence due to the different chemical-physical properties of the lipid constituents, the lipid bilayer itself is not a homogeneous system but shows inhomogeneity where local phase separation is characterized by the presence of "rafts" or islands of chemically distinct lipids immersed in the lipid membrane [7]. It must be also underlined that such rafts do not exist only in "pure" bilayers or monolayers [7, 8] but are also strongly influenced by the interactions with proteins, where for example the sphingolipid-cholesterol self-assembly shows some protein specificity strongly affecting the membrane bio-activity [4, 9].

When considering β-barrel outer membrane proteins (OMPs) specifically, the evolution into a lipidic environment length of the hydrophobic patch has been optimized to interact with the central hydrophobic core of the bilayer.

In fact an OMP can be well represented by a cylinder where a central hydrophobic section is embedded into the hydrophobic bulk of the lipid bilayer, topped by two hydrophilic rings interacting with the outer hydrophilic membrane portion and the inner and outer cell aqueous environments (see Fig. 6.1).

However the length of the hydrophobic patch of the β-barrel OMP and the lipid bilayer is not always perfectly overlapping. This difference defines the so-called *hydrophobic mismatch*, *i.e.* the difference in length between the hydrophobic core of the lipid bilayer and the hydrophobic tertiary domain of the protein.

As a consequence there is an important reciprocal influence/perturbation between the bilayer and the protein itself, where the bilayer regulates the membrane protein function by inducing protein conformational changes which involve a perturbation on the protein/bilayer boundary affecting the adjacent bilayer characteristics [10]. The *hydrophobic mismatch* between a β-barrel membrane protein and a BAB block copolymer membrane (explanation on polymer blocks, see below) is shown schematically in Fig. 6.2.

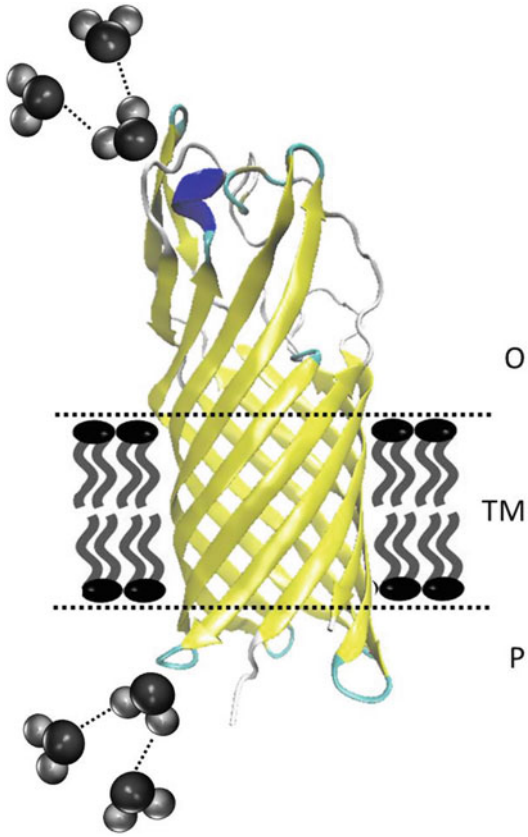

Fig. 6.1 Schematic representation of a bacterial β-barrel outer membrane protein embedded in a lipid bilayer, with the O – outer aqueous environment, P – periplasm and TM – trans-membrane part. Water molecules surrounding the external protein parts (loop and turn regions) are indicated

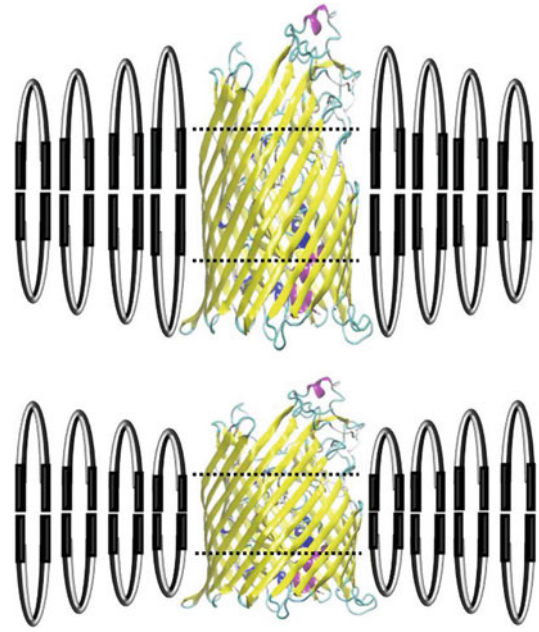

Fig. 6.2 Schematic representation of hydrophobic mismatch between β-barrel membrane protein and a BAB block copolymer membrane. *Top* – hydrophobic portion (indicated by *broken line*) of the protein is longer than the unperturbed hydrophobic membrane thickness and *bottom* – hydrophobic portion of the protein is shorter than the unperturbed hydrophobic membrane thickness

The mentioned bilayer perturbations/deformations involve an energetic cost, function of the conformational change varying with the bilayer physical-chemical properties, *i.e.* bilayer thickness, intrinsic lipid curvature, elastic compression and bending moduli. Therefore the protein function is regulated by the bilayer properties determining its free-energy changes caused by the protein-induced bilayer deformation, defining the lipid bilayer as a possible allosteric regulator of the membrane function [10].

Such analysis is not only important for the membrane protein regulation but also when a reconstitution of a membrane protein into a lipid membrane or other specific systems takes place. These reconstituted protein systems are useful to set up a simplified environment, as compared to the original complex cell lipid membrane, so to obtain a model system for studies to be conducted on the membrane protein itself.

In general the reconstitution of membrane proteins into model membranes is a complex experimental procedure performed to retain the original protein activity by the use of different structures such as monolayers, supported lipid bilayers, liposomes and nano-discs [11].

Specifically interesting for biotechnological applications are so-called liposomes [12]; lipid based supra-molecular assemblies and artificially-prepared vesicles composed by a lipid bilayer.

Their use and associated techniques have been developed within the last 30 years being relatively easy to construct by procedures

based on reverse-phase evaporation followed by extrusion methods or ultrasonication. Furthermore, under optical microscope both unilamellar or multilamellar giant vesicles can be easily micro-manipulated. The reconstitution of membrane proteins into liposomes is generally conducted within the presence of detergents; the latter are necessary to solubilize membrane proteins during protein extraction from the original cell membrane; then mixed with the desired phospholipid vesicles forming an isotropic solution of mixed phospholipid-protein-detergent micelles.

The subsequent detergent removal can be performed by dialysis, gel filtration, affinity chromatography or Biobead adsorption (for details see Sect. 5.2). As a consequence when the decreasing detergent concentration reaches a critical level, the protein spontaneously associates with the phospholipid membrane to form, possibly, biologically active, protein-functionalized liposomes defined as "proteoliposomes".

Although the mentioned techniques are able to successfully embed membrane proteins into a liposome membrane, seldom there is a control on the protein number and orientation, especially important if ion-channel proteins are to be used; though some success has been obtained [13].

The reconstitution of a membrane protein into a liposome membrane has been proposed to follow a similar energetic scheme as proposed for an inserting peptide [14] where four steps must be considered: partitioning between the aqueous and membrane environment; folding/unfolding of the peptide/protein within the two different environments; insertion into the lipid bilayer and association, with other peptides or proteins.

Such a scheme though based on four thermodynamic separated steps, presents many cross-couplings such as, for example, the energetics of partitioning-folding for an α-helix formation by the bee-venom melittin peptide or β-sheet formation by the hexapeptide AcWL5 [14].

The aforementioned energetic scheme is quite well known for helical transmembrane proteins characterized by their ability to dissolve and fold into lipid bilayers [15].

Furthermore the helical transmembrane proteins have been proposed to interact with only one of the α-helices with the lipid bilayer while the association of the following helices can occur after insertion by hydrogen bonds [16] (for details see Fig. 2.5).

In case of the β-barrel membrane proteins, instead, the folding and insertion have been proposed as synchronous events [17] and their assembly follows an "all or none" mechanism [18].

It should be underlined that while the folding and insertion of a β-barrel protein are quite well understood in "*in vivo*" conditions, when dealing with "*in vitro*" conditions the knowledge is scarce (for details check Sect. 2.3.2), where the "*in vitro*" term includes the interaction, insertion and folding within a liposome membrane.

Though the "*in vivo*" folding conditions are not fully understood, they are still better characterized as compared to the "*in vitro*" conditions due to the "*in-vitro*" absence of chaperones and the molecular machinery found in "*in vivo*" defined as a "*translocon or translocation channel*" constituted by protein complexes able to translocate proteins and polypeptides across the inner membrane [19] or the outer membrane [20] and to integrate nascent proteins into the membrane itself (see also Fig. 2.10).

Some steps toward the understanding of the two step model and "*in vivo*" behavior have been recently obtained. Measurements referred to the free energies of folding for the transmembrane proteins PhoP/PhoQ-activated gene product (PagP), outer membrane protein W (OmpW) and *Escherichia coli* outer membrane phospholipase A (OmpLA), suggest that the stability of these proteins are strongly correlated to the water-to-bilayer transfer free energy of the lipid-facing residues in their transmembrane regions showing differences in solvent exposure between their folded and unfolded states. From a biological point of view, these findings suggest that the folding free energies for these membrane proteins may be the thermodynamic driving force establishing an energy gradient across the periplasm, thus driving their sorting by chaperones to the outer membranes in living bacteria [21].

Table 6.1 Hydrophobic region length of the FhuAΔ 1–159 protein; h and the α angle defined as the tilt of the strands toward the barrel axis simulated in DNPC and DOPC lipid bilayers (310 K)

	FhuA Δ1–159 (DOPC)	FhuA Δ1–159 (DNPC)
h (nm)	2.03 ± 0.14	2.24 ± 0.13
α (°)	37.07 ± 4.13	37.66 ± 3.94

Regarding the hydrophobic mismatch effect [10] and its influence on the secondary/tertiary structure of a β-barrel membrane like *E. coli* FhuA (ferric hydroxamate uptake component A) and a related engineered variant FhuAΔ1–159, where the first 159 amino acids corresponding to the internal cork domain have been deleted (for details check Sect. 5.1.2.1) a Molecular Dynamics simulation study was performed [22]. The structural changes on the aforementioned proteins where analyzed when embedded into two different environments based on a DNPC (1,2-dinervonyl-sn-glycero-3-phosphocholine, a 24:1 lipid represented by the nervonyl acyl radical, $-[CH_2]_7CH{=}{=}CH[CH_2]_{13}$-cis-15-Tetracoseno) lipid bilayer and an OES detergent cell (N-octyl-2-hydroxyethyl sulfoxide) showing no substantial length differences or hydrophobic stress. However in a subsequent simulation [23], the FhuAΔ1–159 variant was embedded into two different lipid bilayers, the aforementioned long chain DNPC and the shorter DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine, a 18:1 lipid represented by the oleyl radical $-[CH_2]_7CH{=}{=}CH[CH_2]_7$-cis-9-). The difference in the length of the hydrophobic core between the two lipids was reported to be $\Delta_{DNPC-DOPC} = 1$ nm and as a consequence the hydrophobic mismatch of 1 nm is expected to induce some distortion or "stress" on the protein secondary/tertiary structure.

As reported in Table 6.1 two parameters have been defined to analyze the effect of the hydrophobic mismatch on the FhuAΔ1–159 engineered protein, defining the length of the hydrophobic protein region and the α angle, *i.e.* the tilt of the strands toward the barrel axis [22]. The results do show very little differences, though reporting a protein elongation in the DNPC bilayer compared to the DOPC.

The robustness of OMPs even when engineered (FhuAΔ1–159), adds to their value for the development of new nano-materials for nanotechnological applications. Details on the main applications will be described later in Sect. 6.2.

In principle a difference must be set when using the terms "insertion" or "reconstitution". The term insertion is well represented by the insertion of a α-helical transmembrane protein or polypeptide, like bee-venom Melittin or *Hyalophora cecropia* (Cecropia moth) derived Cecropin A, into a lipid or polymer membrane following the aforementioned four step model, while reconstitution can be generally intended as the embedding of a membrane protein into a lipid membrane during the forming process of a liposome.

In fact, because a β-barrel protein has been suggested to follow a two-step simultaneous mechanism [17] considering the cross section of a polypeptide compared to a classical β-barrel protein (∼3–5 nm) the energy penalty to produce a cavity within a lipid or polymer membrane in case of a β-barrel protein is much higher.

Such a simple model fits well to many of the experimental procedures where a β-barrel membrane protein is "reconstituted" into a lipid bilayer adding slowly the lipidic phase in a solution already containing purified and detergent-solubilized β-barrel proteins such as a FhuA variant [24] or the BamA transmembrane protein of the *E. coli* Bam complex necessary for the outer membrane protein translocation and insertion [25]. In both cases the detergent used to solubilize the membrane proteins is slowly diluted and at the same time lipid extract is added.

Though the reconstitution of β-barrel proteins into liposomes is an interesting method to study the embedded proteins itself, the reconstitution of β-barrel proteins into polymersomes has attracted considerable attention during the last decades, also.

Polymersomes are artificially made vesicles where the membrane is constituted by amphiphilic synthetic block copolymers instead of lipids as in the liposome case. The term "amphiphilic (or amphipatic)" defines a chemical compound possessing both hydrophilic (*water-loving*, polar) and lipophilic (*fat-loving*)

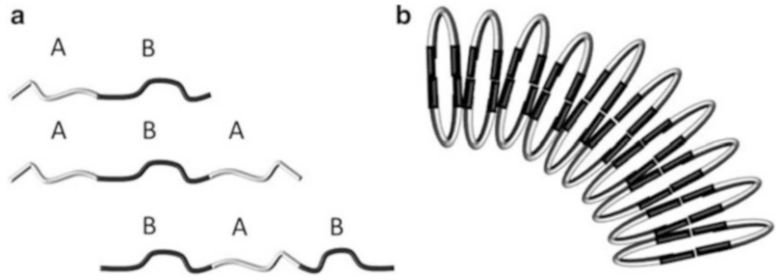

Fig. 6.3 Schematic representation of (**a**) different examples for block copolymers with A – hydrophilic block; B – hydrophobic block; *top* – AB diblock; *middle* – ABA triblock and *bottom* – BAB triblock and (**b**) BAB triblock copolymers assembled to for a bi-layer membrane

characteristics. For example lipids constituting a cellular membrane or a liposome are amphiphilic, with the classical polar head and a hydrophobic tail. However in the synthetic block co-polymers two or more hydrophilic or hydrophobic polymers are chemically linked aggregating in solution to produce vesicular structures [26] or to microphase into a more complex variety of morphologies, *i.e.* spheres, lamellae, inverse spheres and several more complex shapes depending from the surrounding physical-chemical conditions like solvent and temperature or depending from the chemical characteristics of the polymers used and their volume fractions [27]. The hydrophilic or hydrophobic blocks of an amphiphilic block copolymer are generally termed A (hydrophilic block) and B (hydrophobic block). They form then for instance AB diblock or ABA triblock copolymers (Fig. 6.3a) that self-assemble to form membranes (Fig. 6.3b).

The use of synthetic polymers increases dramatically the number and combinations of possible physical-chemical characteristics of the single block copolymers, enabling chemists to have some control on the permeability and stability of the polymersome itself. Although the stability of liposomes and polymersomes is different due to their different chemical composition, the origin of their formation is practically the same being both held together by non-covalent interactions.

From the mechanistic point of view, block copolymers self-assemble by a thermodynamically driven process where the chemical disaffinity between the blocks results in an unfavorable mixing enthalpy and a small mixing entropy while covalent bonds between the blocks prevent macrophase separation.

Experimentally the formation of such nanostructures like micelles and vesicles can be achieved by two main methods: by molecular aggregation or by mechanical disintegration of a macroscopic phase of matter [28].

When considering the insertion of a membrane protein into a polymer membrane further points must be considered in order to understand whether the reconstitution will be successful or not.

Specifically there is a lack of general and reproducible protocols regarding the functional reconstitution of channel-forming membrane proteins, quite often strongly protein dependent. A first problem derives from the use of organic solvents necessary to solve the synthetic polymers. One of the classical techniques is the drop-wise addition of a polymer–protein–organic solvent mixture, like ethanol or tetrahydrofurane (THF), to an aqueous solution leading to the formation of proteo-polymersomes.

Another approach, again involves the drop-wise addition of a solution of organic solvent + polymer into an aqueous solution of the membrane protein stabilized by detergents. However, in both cases, when membrane proteins are mixed with an organic solvent, they get mostly denatured.

Furthermore organic solvents remaining within the solution, especially if totally or partially soluble with water, may keep the reconstituted proteins in an unfavorable environment

and destabilize the vesicular structures by fluidifying the polymersome membrane made by amphiphiles. As a consequence the aforementioned methods limit the use of membrane proteins due to their de-functionalization by organic solvents.

A solution toward the functional reconstitution may derive from two approaches:

1. the use of membrane proteins that show an intrinsic robustness to the presence of organic solvents or;
2. the development of methods avoiding the use of organic solvent.

Following the first approach, an engineered FhuA channel protein with a specifically elongated hydrophobic section to fit within the hydrophobic thick layer of the PIB_{1000}-PEG_{6000}-PIB_{1000} (PIB = polyisobutylene, PEG = polyethylene glycol) tri-block copolymer (see below for further explanation) shows an extremely strong resistance toward organic solvents, reaching 10 % v/v of THF still obtaining a valid functional reconstitution [29].

While the second technique was used to functionally reconstitute a BR/F_0F_1-ATP synthase into a PMOXA-PDMS-PMOXA (PMOXA: poly-2-methyl-2-oxazoline, PDMS: poly-dimethyl-siloxane) polymersome membrane [30].

Shifting the attention to the microscopic origin of the problems related to membrane protein reconstitution into a polymer bilayer, the main variable has been detected in the strong hydrophobic mismatch between the hydrophobic domain of the protein and the hydrophobic thickness of the polymer itself.

In fact membranes formed by block copolymers are often thicker (5–22 nm) than those formed by "natural" phospholipids (3–4 nm) leading to better mechanical strength [31] resulting in to a drop in efficiency of channel insertion (when comparing polymersomes to liposomes). The strategic solution for the functional reconstitution of membrane proteins into polymeric membranes requires to design polymer membranes as thin and fluid as possible, in order to minimize the energetic penalty associated with the exposure of the nonpolar/polar interface.

For example, simulation studies conducted on the outer membrane proteins OmpF insertion into a di-block copolymer $EO_{29}EE_{28}$ (EO = Ethyleneoxide, EE = Ethylethylene) membrane show a considerable symmetric deformation of the hydrophobic region of the polymer. The hydrophobic mismatch upon insertion is 1.32 nm, corresponding to 22 % of the polymer thickness [32]. As a consequence, if copolymer bilayers cannot withstand the hydrophobic mismatch, channel protein insertion is prevented.

The polymers till now used to reconstitute membranes proteins reproduce a bilayer lipid like organization, though the polymer itself is chemically different from the original lipids. For example different di- or tri-block copolymers have been used to reconstitute several membrane proteins, like the tri-block copolymer PMOXA-PDMS-PMOXA (PMOXA: poly-2-methyl-2-oxazoline, PDMS: poly-dimethyl-siloxane) for the FhuA, OmpF and Tsx [33–35], the di-block copolymer PBD-PEO (PB: poly-butadiene, PEO: poly-ethylene oxide) to reconstitute the dopamine receptor D2 (DRD2) [36] or with both of the aforementioned polymers the reconstitution at high-densities of water channels into vesicular and planar membranes [37]. Another tri-block copolymer PIB_{1000}–PEG_{6000}–PIB_{1000} (PIB = polyisobutylene, PEG = polyethylene glycol) was used to embed an engineered FhuA variant with extended hydrophobic portion (FhuA Ext) [29] while the tri-block PetOz-PDMS-PEtOz (PEtOz = poly(2-ethyl-2-oxazoline)) was successfully used for the reconstitution of the F_0F_1-ATP synthase and bacteriorhodopsin [38].

A solution of the hydrophobic mismatch problem can be obtained by increasing the hydrophobic length of a membrane protein by changing the point of view: matching the protein to the polymer instead of matching the polymer to the protein. For example the engineered FhuAΔ1–159 [39] often cannot insert into thick polymeric membranes.

Using a simple amino acid "copy-paste" approach the protein hydrophobic transmembrane section was increased by 1 nm, leading to a predicted lower hydrophobic mismatch between

the protein and the PIB_{1000}-PEG_{6000}-PIB_{1000} (PIB = polyisobutylene, PEG = polyethylene glycol) tri-block copolymer polymersome membrane, minimizing the insertion energy penalty [29]. The strategy of adding amino acids to the FhuAΔ1–159 hydrophobic part can be further used to build even more hydrophobic proteins, promoting the efficient embedding into thicker or more hydrophobic block copolymer membranes (for details on the FhuA variant FhuA Ext check Sect. 5.1.2.4).

In general the mechanical properties of a polymersome membrane result in a higher stiffness as compared to the liposome membrane; a general problem especially when considering the embedded protein stability or reconstitution. However in some specific polymer cases such as PDMS-PMOXA, the elastic response of the PDMS-PMOXA polymersomes to external stimuli is much closer to that of lipid vesicles compared to other types of polymersomes, such as polystyrene-block-poly(acrylic acid) (PS-b-PAA) [40]. This is a rather interesting finding explaining the success of the PDMS-PMOXA polymer to embed natural membrane proteins.

As previously reported the hydrophobic central part of a lipid bilayer is mainly based on low molecular weight -CH_2- chains where the hydrophobic mismatch can be thought mainly driven by enthalpic energy penalties, *i.e.* the lengthening of the -CH_2- angles. In a polymer bilayer, instead, the higher molecular weight of the central hydrophobic part is mainly entropically driven, *i.e.* folding and/or unfolding of the polymer chain.

Due to the length of the polymer chains the two independent layers are expected to be entangled in the hydrophobic region [41], a situation less common though existing in the lipid bilayers ($L_\beta I$ phase). Such a behavior in case of block copolymers originates from the shielding of the hydrophobic membrane from water by a polymeric brush due to the partial coiling [42] of the hydrophile. As a consequence the higher molecular weight hydrophobic polymer coils are more likely to interdigitate and become entangled and the reconstitution of a β-barrel protein in a polymer bilayer can be hindered by an entropic factor with the polymer chain interacting with the hydrophobic "cylinder" of the membrane protein.

Another problem can be found considering the miscibility of the detergents used to solubilize/stabilize the membrane protein in aqueous solutions. In fact membrane proteins are strongly hydrophobic and insoluble in water and a series of detergents are employed to extract and stabilize the membrane proteins. The protein itself when in a water solution will be surrounded by a detergent belt or small micelles, depending on the detergent concentration. The chemical characteristic of such an environment can be very important when protein reconstitution in a polymer membrane is considered. In fact polymers can be immiscible [43] and if a membrane protein is solubilized with a detergent based on a polymer chain immiscible with a determined polymersome membrane, a reconstitution is difficult or impossible.

A last variable when considering polymeric membranes must be analyzed: *polydispersity*. The polydispersity index (PDI) of a polymer refers to the distribution of the single polymer chains as function of the molecular weight and defined by the relation:

$$PDI = \frac{M_W}{M_n} \tag{6.1}$$

where M_w is the weight average molar mass defined by:

$$M_W = \frac{\sum M_i{}^2 N_i}{\sum M_i N_i} \tag{6.2}$$

and M_n is the number average molar mass, defined by:

$$M_n = \frac{\sum M_i N_i}{\sum N_i} \tag{6.3}$$

with i representing the i-monomer within the chain. M_n is more sensitive to molecules of low molecular mass, while M_w is more sensitive to molecules of high molecular mass.

All biological polymers have a PDI of 1, defining all molecules having the same molecular weight while industrial polymers very seldom show a PDI near to 1. Until now when relating to the difficulties to include a membrane protein into a polymer membrane however the polymer is considered to have a PDI = 1.

When a protein reconstitutes into polymer membranes, the protein will be most probably in contact with a "gradient" of shorter polymer chains gradually lengthening toward the farthest point from the protein. In fact as reported in the theoretical analysis of Pata et al. [44] entropy should mix the different chains uniformly, but the perturbation induced by embedded proteins can lead to a local segregation, where shorter chains that match the protein dimensions more closely would concentrate in the region adjacent to the protein boundary, inducing a kind of "polymer raft" around the protein itself. Such a phenomenon can introduce an easier way for the membrane protein to be reconstituted into a polymer layer.

In conclusion polymersome research as well as the development of biomimetic membranes for sensor and separation technology is a vibrant field with ample possibilities of improvement [45] where the microscopic understanding of a "simple" membrane protein embedded into a polymer membrane would boost a rational approach.

6.1.2 Dynamic Light Scattering

Light scattering is a phenomenon resulting from the interaction of an electromagnetic wave with a portion of matter. When an incident light ray or electromagnetic (em) wave encounters a body or an inhomogeneity light itself is re-diffused or redirected in all directions.

The phenomenon itself is a complex interaction between the incident em wave and the molecular/atomic structure of the scattering object and cannot be simply considered as a reflecting or bouncing of light rays from the surface of the scattering object [46].

The interaction between the em wave and matter results in an induced dipole moment within the object due to the electrons that are periodically perturbed with the same phase as the electric field of the incident light. As a consequence molecules themselves constituting the analyzed object are a secondary source of light homogeneously emitted in space (scattering). The frequency shifts compared to the original source, the angular distribution, the polarization, and the intensity of the scattered light are function of the size, shape and molecular interactions in the scattering material [46].

The two main theories of scattering are based on the Rayleigh and Mie scattering. The former is mainly applicable to small (compared to the incident wavelength), dielectric (non-absorbing), spherical particles while the second one deals with the general problem of scattering from a solution of spheres, absorbing or non-absorbing, not anymore depending on the particle size.

The Mie scattering theory does not show particle size limitations, converges to the limit of geometric optics for large particles and can include the Rayleigh scattering as subset. In general Rayleigh scattering theory is preferred when possible to apply, due to the complexity of the Mie scattering formulation.

Some other types of scattering exist for example when analyzing colloidal mixtures or suspensions (Tyndall scattering), while inelastic forms of scattering like the Brillouin; generated by the interaction of photons with phonons in solids; or the Raman inelastic scattering; where photons interact with optical photons in solids; will not be considered for further analysis in the liposome or polymersome characterization [47].

From the experimental point of view a differentiation between static and dynamic light scattering is necessary. In the static light scattering (SLS) case, the considered experimental parameter is a function of the time-averaged intensity of the scattered light, while in the dynamic light scattering (DLS) case the fluctuations in light intensity by itself are considered so to extract the information from the desired experimental variable. The DLS technique is also quite often referred to as photon correlation spectroscopy (PCS) and quasi-elastic light scattering (QELS).

With a combination of the aforementioned theoretical apparatus for the interpretation of the light scattering and some fluctuation theory derived by statistical mechanics, the information on the structure and dynamics of the scattering medium can be obtained.

By introducing a short physical description (for a better and complete overview of DLS theory and some examples of application please check the following literature [48, 49]) on how DLS scattering is related to macroscopic variables like diffusion, a direct link between the Brownian motion, arising from collisions between the suspended particles and the solvent molecules, and the resulting scattered light is obtained by considering the "correlation coefficients" described by a relation like Eq. 6.4:

$$G(\tau) = \int_0^\infty I(t)I(t+\tau) = B + Ae^{-2q^2D\tau} \tag{6.4}$$

where G(τ) is the correlation coefficient, I is the intensity at a given time t and t + τ, B is the baseline, A the amplitude and D the diffusion coefficient. In fact as a consequence of the particle motion, the light scattered from the particle ensemble itself will randomly fluctuate in time. At this point the necessary information regarding the motion or diffusion of the particles in the solution is contained within the measured correlation curve. By measuring the fluctuations in very short time intervals a correlation curve is built by which it is possible to extract the diffusion coefficient and subsequently the hydrodynamic particle size or hydrodynamic radius. It should be underlined that the hydrodynamic radius differs from the "real" particle size, due to the fact the particle itself is hydrated/solvated and the diffusion coefficient of the particle derived by DLS gives the total radius *i.e.* particle + solvation/hydration sphere.

Finally by using the Stokes-Einstein equation given in Eq. 6.5:

$$R(H) = \frac{kT}{6\pi\eta D} \tag{6.5}$$

the hydrodynamic radius R(H) is derived where k is the Boltzmann constant, T is the absolute temperature, and η is the solvent viscosity.

Applications of DLS within the biophysical-chemistry realm are devoted to the characterization of particle sizes and distribution or diffusion coefficients including proteins, polymer solutions, particles in solution, polymersomes and liposomes [49]. Stability or aggregation studies can be also performed by DLS analysis checking whether the hydrodynamic radius changes in time or during temperature dependence experiments.

Operationally DLS is a powerful way to perform short time, non-invasive measurements on samples including a quite user-friendly and simple preparation of the same samples. Modern instrumentation can detect and resolve particle sizes of diameters ranging between 0.6 nm and 6 μm converting a wide range of sample concentrations. If the considered system is monodispersed or polydispersed a fitting procedure based on a CONTIN analysis must be performed and in case of polydispersed systems various scattering angles should be considered.

It can be well stated that DLS is a commonly used and affirmed technique for biomolecular studies and interaction studies [50]. Modern application of DLS include pre-screenings for protein crystallization to test the homogeneity of the sample [51] or to obtain molecular parameters such as size, molar mass and intermolecular interactions to identify and characterize intrinsically disordered proteins (IDPs) [52].

The literature on and application of SLS and DLS regarding the studies of liposomes and polymersomes is vast and, practically, quite often each new characterization and study of such systems includes a high number of techniques where scattering methods are routinely used to understand the hydrodynamic radius and polydispersity of the samples [53–56].

An interesting non published study on polymersomes with the embedded engineered FhuA membrane protein (FhuA Ext) noticed an increase in polymersome diameter in presence of the protein compared to the pure polymersomes hinting towards a correlation between diameter increase and protein reconstitution (for details

Fig. 6.4 Schematic representation of the two step oxidation products of TMB (3,3′-5,5′-tetramethylbenzidine)

see Sect. 5.1.2.4). From this finding might be deduced a first clue on the number of proteins embedded in the polymersome membrane with the pre-requisite of a constant membrane thickness, polymer membrane total volume and maintained spherical symmetry of polymersomes.

It would be interesting to study in greater detail the aforementioned system by a multi-angle DLS method to understand if these results are due to a change in the geometric symmetry of the polymersome or simply due to the number of proteins inserted in the polymersome membrane.

In conclusion there is still a wide unexplored range of possible sophisticated applications of DLS beyond the simple characterization methods explained above.

6.1.3 Spectroscopic Flux Assay and Patch-Clamp

The insertion of a channel protein into a liposome or polymersome membrane establishes a communication between the external environment and the internal vesicle. The functionality of the channel can be tested by using a simple conceptual experiment by determining the flux of a detectable chemical compound through the channel itself. The detection of the compound flux through the channel can be measured by two possible main strategies:

1. the detection of a "macroscopic" flux performed on a statistical ensemble of liposomes/polymersomes or
2. by detecting the "microscopic" flux on a single protein channel.

These two approaches rely on different techniques whose application and selection is a function of the information to be extracted. In general a spectroscopic flux assay, as will be shown in the following, applied to an ensemble of channel proteins can only certify if the channel is inserted into the membrane while it generally does not give detailed information on the conditions of the channel itself, *i.e.* if the channel has changed its ability to work more or less efficiently as a diffusion channel depending on the membrane characteristics.

As reported in a study where the engineered FhuA Ext was reconstituted into PIB_{1000}–PEG_{6000}–PIB_{1000} polymersome membranes, the influx kinetics could be followed based on the TMB/HRP detection system (HRP = Horse Radish Peroxidase, TMB = 3,3′,5,5′-tetramethylbenzidine). The HRP/TMB assay system is widely used in enzyme immunoassays (EIA), due to its robustness. The TMB/HRP detection system is based on a two-step irreversible consecutive reaction $A \rightarrow B \rightarrow C$ (A = TMB; B and C = first and second TMB oxidation products, Fig. 6.4) catalyzed by HRP in presence of H_2O_2.

Since the final TMB oxidation product is only stable under very acidic conditions [57], the intermediate product is used as a reporter with a characteristic absorbance maximum at 370 and 652 nm [29].

The insertion of the modified FhuA was tested by detecting the flux of the second TMB oxidation product catalyzed by the entrapped HRP enzymes after the biotin label-closed channel (further details on the channel labeling can be

found in Sect. 5.1.3.1) was opened to permit flux of hydrogen peroxide and TMB internally to the polymersome. The same assay system had been used to analyze the reconstitution ability of the cork-less FhuAΔ1–159 [58–60] and of a FhuA variant with enlarged channel diameter (FhuA Exp) [24] both reconstituted into liposomes.

In another study Fluorescence Correlation Spectroscopy (FCS) was used to determine the number, diffusion properties and brightness of the fluorophore acridine orange dye between $5 \leq \text{pH} \leq 7$ being FCS able to discriminate between fluorescent species inside and outside polymer vesicles [61]. The FCS method is based on a temporal auto-correlation analysis of fluorescence intensity fluctuations collected in time from a tiny focal volume defined by the microscope focus of the excitation laser beam within a sample [62, 63].

Though interesting such techniques are unable to give specific answers regarding the single channel and how the channel itself is reacting under the new environment. To reach such a goal the patch-clamp technique is used though till now only on cells or liposomes with reconstituted channel proteins. The application of this technique to membrane protein functionalizing polymersomes would be a very welcome analysis giving important clues on how the polymer itself changes the channel protein behavior (stochastic diffusion) compared to a liposome reconstituted one.

Historically the patch clamp technique was developed by Neher and Sakmann [64, 65] to detect ion currents passing through single acetylcholine-activated channels in cell-attached patches of frog skeletal muscle membrane. Further developments and refinements [66, 67] resulted in techniques for current recording at high resolution in excised membrane patches in addition to those that remain cell-attached.

The importance of this technique results from its ability to record single channel conductance and kinetic behavior of ion channels partly investigated by classical voltage clamping and by noise analysis. However the sensitivity on the technique permits investigations on the physiological role of ion channels in cells inaccessible to voltage clamp and to cells that are not electrically excitable.

Methodologically a membrane patch is electrically isolated from the external solution to record the current flow within the patch itself. This is achieved by pressing a fire-polished glass pipette filled with an electrolyte solution against the cell or liposome surface applying light suction. Under optimal experimental conditions, *i.e.* when both the glass pipette and the cell membrane are clean and the distance between the pipette and the membrane is $\sim$1 nm, a seal with a corresponding resistance of 10 GΩ is formed (gigaseal) [68]. The high resistance of the seal gives the possibility to specifically isolate electronically the currents measured across the membrane patch with little competing noise, as well as providing some mechanical stability to the recording.

Often patch-clamp is applied to the analysis of the reconstituted channel proteins using unilamellar vesicles (ULV) that can be sub-divided in small unilamellar vesicles (SUV $\leq$ 100 nm), large unilamellar vesicles (LUV: 100 nm to 1 μm) and giant unilamellar vesicles (GUV $\geq$ 1 μm).

For example, recently a characterization of the mechanosensitive MscSP channel was performed in giant spheroplasts and in azolectin liposomes [69] while GUV where used to test the functional reconstitution of the voltage-gated potassium channel [70].

Apart from liposomes, ion channels and protein channels are reconstituted also into planar lipid bilayers [71] applying single-ion channel current measurements at a constant applied voltage.

Recently a high-throughput screening of ion channels has been developed on the advent of the planar patch clamp system [72]. This new technology based on patch clamp will drastically increase the number of ion channels that can be studied, as multiple ion channel experiments can now be conducted in parallel. On the same research line the development of the planar patch clamp [73] was undertaken to conduct measurements including artificial lipid bilayers.

While single channel conductance measurements are especially suited for stochastic

nanochannel sensing applications as sensing is based on the changes in electrical current caused by analytes that interact transiently or permanently with a nano-pore, the flux enzymatic assay instead can be applied to get more general information on channel functionality.

6.1.4 Further Characterization Methods (Electron Microscopy, Tryptophan Fluorescence)

In the previous section, a series of flux based detection systems have been explained that can be used to check whether a channel protein is functionally reconstituted and if the new environment (lipid or polymer based) affects the channel properties of the embedded protein.

However if the aforementioned measurements miss to detect a molecular flux out of the liposome/polymersome lumen, the lack of detection can be originated by the failed reconstitution of the channel protein or it might be due to the partial/total unfolding of the protein within membrane, implying a non-functional reconstitution. To discern such experimental cases, further techniques must be considered by which the detection of the embedded protein is realizable. Examples for such complementing techniques are Electron Microscopy (EM) and Trp fluorescence measurements.

6.1.4.1 Electron Microscopy

Electron microscopy is a microscopic technique using electrons instead of light (light microscopy) to resolve structures <1 μm. After a first demonstration by Ernst Ruska and Maximillian Knoll in 1931, the first electron microscope was built in 1933 by Ernst Ruska, it was able to resolve structures down to 50 nm [74].

Today under the realm of EM a plethora of techniques exists on how to study and prepare the samples to be analyzed [75].

However considering the EM application to the liposome/polymersome research field, a comparative study between atomic force microscopy (AFM), environmental scanning electron microscopy (ESEM), transmission electron microscopy (TEM) and confocal laser scanning microscopy (CLSM; labeling using a fluorochrome marker) on liposomal characterization was undertaken [76].

Each of the techniques gives a characterization of the nano-scale structures of liposomes, with AFM, TEM giving the best information on the shape and morphology, AFM, ESEM, TEM, and CLSM on dimensions, AFM on surface properties, and CLSM on the internal structure.

It should be underlined that while EM is well known and routinely used to check shape and geometry of liposomes and polymersomes, not many studies have been developed to consider the organization of membrane proteins on the membrane of liposomes [77–79] and in particular of polymersomes, apart from the routinely used application to check the geometry and polymer membrane thickness.

An EM technique getting popular in the structural biology field is the cryo-electron microscopy (Cryo-EM). Cryo-EM is a transmission electron microscopy (TEM) technique where samples are studied under cryogenic temperatures (liquid N_2 temperatures). The wide use of cryo-electron microscopy originates from the possibility to observe samples that have not been stained or fixed and are as a consequence near to their native structure.

For example, Cryo-EM of vitrified samples was applied as a quality control of the proteoliposome reconstitution while freeze fracture TEM was used to test the incorporation of RhCG-HA (Rh glycoproteins) into the lipid bilayer and to estimate the number of protein particles per proteoliposome [80].

In another study Cryo-EM was applied to study the BK (*Big potassium*) potassium channel in a lipid membrane [81] or to understand the structure of reconstituted bacterial membrane efflux pumps (by Cryo-electron tomography) [82].

Future use and applications of the EM technique will surely improve and increase the details on how and if membrane proteins interact with the polymersome membrane.

In general though the functionality can reasonably suggest the proper folding of the protein channel, such an inductive step is not always granted.

6.1.4.2 Tryptophan Fluorescence

It is well known that the spectral parameters of tryptophan fluorescence emission (Trp absorbs light at a wavelength of 280 nm) are sensitive to the environment and a protein's tryptophans can report on the environmental characteristics and changes during events such as folding or unfolding. For example, the tryptophan λ_{max} of many protein emission scans will red-shift upon unfolding as the tryptophans become more exposed to polar solvent. Based on this phenomenon has been shown how the amphiphilic peptide Cecropin A characterized by the presence of one Trp in its primary sequence was experiencing a hydrophobic environment when interacting with vesicles made by the artificial tri-block copolymer PIB_{1000}–PEG_{6000}–PIB_{1000} (PIB = polyisobutylene, PEG = polyethylene glycol) by showing a 30 nm blue-shifted emission when compared to water [83].

Tryptophan fluorescence is also often used to study peptide insertion/interaction with liposomes [84, 85].

Regarding the proteoliposomes, the mechanosensitive channel MscL and its interactions with the lipid membrane was studied by measuring Trp fluorescence and analyzing the obtained signal [86] showing specific preference for a particular class of anionic lipids, stressing the dependence of the mechanosensitive channel protein function on the lipid-protein interactions [87].

A series of recommendations on the technical use of tryptophan fluorescence measurements to analyze the stability of membrane proteins folded in liposomes has been recently published [88].

All these studies point out the validity of Trp fluorescence measurements as complementing technique to study the reconstitution of membrane proteins.

6.2 Nano-channel Applications

The use of membrane protein channels and especially of β-barrel outer membrane protein channels for nano-technological applications obtained considerable attention since the early 2000s. For instance the field of drug delivery using protein functionalized vesicular nano-compartments, the field of protein-based stochastic nano-sensing elements or the bio-nanoelectronics field more and more often report the successful use of OMPs. This development can be seen as a co-evolution to the likewise successful area of artificial β-barrels (as discussed also in Sect. 5.4).

The main applications of biological β-barrel outer membrane proteins will be introduced in the following sub-section.

6.2.1 Drug Delivery

The term "drug delivery" defines a series of technologies for the transport of therapeutic molecules within the body towards a certain target in order to achieve a desired therapeutic effect [89]. Classically drug delivery is actuated by the administration of a particular chemical compound by using a non-invasive supply. The ideal molecule is showing a selective activity toward the targeted virus, microbe or cell only, not affecting the main function/physiology of the patient body *i.e.* showing no deleterious or dangerous side effects.

Nowadays the side effects problem is addressed by the development of targeted delivery in which the drug is only active in the specific area (the target) of the body (for example, in cancerous tissues) where the active compound is specifically released following a controlled and designed release kinetic.

The technology engaged with achieving an efficient targeted delivery is quite complex, as for example, it must consider the host's defense mechanisms avoiding them so to reach its specific site of action. However the drug release methods have been over the years always more and more established and optimized and a step towards the development of an efficient drug delivery technology can be assisted by the use of (functionalized) liposomes and polymersomes.

Liposomes have been already used successfully as pharmaceutical carriers for drugs and DNA, in particular for cancer treatment. To

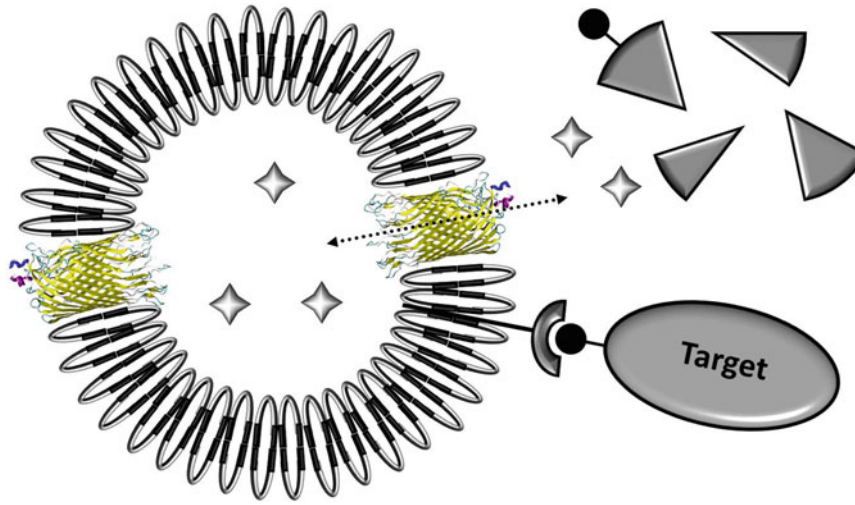

Fig. 6.5 A polymersome with a targeting device decorated membrane including channel proteins embedded into the polymer membrane releasing the compound contained within the polymersome lumen upon specific stimuli

increase the liposomal pharmaceutical efficiency towards cancer cells, two strategies are applied: passive targeting and active targeting. The former shows a balanced release kinetic of the active compound prolonging the liposomes half-life in blood circulation so to accumulate in the pathological sites, while the latter is based on the attachment of specific ligand molecules to the liposomal surface to actively and selectively target specific antigens on the target cells. At the state of the art, antibody-targeted liposomes loaded with anticancer drugs demonstrate high potential for clinical applications [90, 91].

As reported in Sect. 6.1.1, while liposomes are constrained by the limited chemical characteristics of the lipids, the use of synthetic polymers substantially increases the number of possible physical-chemical characteristics of the single block copolymers, enabling researchers to have some control on the permeability and stability of the polymersome itself. As a consequence the permeability and stability of the system can be tuned in order to satisfy the conditions needed for the passive or active drug delivery. For example a biodegradable polymersome characterized by a pH dependent permeable membrane for drug delivery [92] without the inclusion of channel proteins was developed.

Though recent research has focused on the development of multi-functional polymersomes as targeted drug delivery systems for combined therapeutic applications, there are some efforts to develop "*theranostic*" applications where the therapeutic delivery via passive or active targeting is combined with diagnostic capabilities of the system [93].

However a further step toward the controlled release deals with the inclusion of engineered channel proteins to specifically and actively control the in and out compound fluxes under a determined external perturbation of physical origin like pH, temperature, light or chemical origin such as a reducing agent [39].

It should be underlined that the class of polymers that must be considered under the active controlled release should be not permeable or at least showing an extremely low release kinetic to maximize the effect once arrived at their destination.

The general concept of an active controlled release polymersome is depicted in Fig. 6.5, where channel proteins are embedded into a polymersome membrane opening at a precise external perturbation while to the polymersome surface are anchored targeting devices (*e.g.* antibodies) specifically designed to bind to the desired target (*e.g.* cells, bacteria, viruses).

Within this branch of research a series of mutated and engineered *E. coli* FhuA and OmpF channel proteins have been used to develop liposome or polymersome based release systems. As the cork-lacking FhuA variant FhuAΔ1–159 can be reconstituted only into either liposomes [58–60] or polymersomes of the PMOXA-PDMS-PMOXA type [39], variant FhuA Ext was developed to overcome the hydrophobic mismatch of insertion into thick polymersome membranes of the PIB_{1000}–PEG_{6000}–PIB_{1000} type (hydrophobic mismatch problem, see Sect. 6.1.1) [29]. The PIB_{1000}–PEG_{6000}–PIB_{1000} block copolymer has certain advantages such as the biocompatibility of both building blocks (PIB/PEG) [94, 95], the impermeability toward gases and other compounds of the PIB unit [96]. Moreover PIB_{1000}–PEG_{6000}–PIB_{1000} is commercially available and comparably cost effective.

A further FhuA variant (FhuA Exp) was engineered to increase the channel cross section to possibly tune the molecular cut-off of the molecules to be externally diffused, *i.e.* from the *lumen* of the liposome/polymersome to the external environment [24] (see also Sects. 5.1.2.3 and 5.1.2.4).

The controlled release was obtained by including within the free channel of the engineered plug-less FhuAΔ1–159 protein an irreversible light triggered release system by employing the photo-cleavable lysine label 6-nitroveratryloxycarbonyl chloride (NVOC-Cl) [59] or an irreversible chemically triggered system where the lysine were labeled with 3-(2-pyridyldithio)propionic-acid-N-hydroxysuccinimide-ester or 2-[biotinamido] ethylamido-3,3′-dithiodipropionic acid N-hydroxysuccinimide, ester that are sensitive to the S–S bond cleavage under reducing conditions [39] (see Sect. 5.1.3.1).

Employing a different channel protein to functionalize nano-compartments for the controlled release, in an engineered OmpF protein, six hystidines were inserted within the constriction site of the protein obtaining a reversible pH sensitive nano-channel working between $5 \leq pH \leq 7$ [61] (see Sect. 5.1.3.2).

On the same protein a different approach using a non-triggered blocking of the OmpF channel with a crown compound to understand the structure-function relationship of synthetically modified porins was also undertaken [97].

The potential for further advancements and optimizations of the liposome and especially the polymersome technology is vast. Future research will attack the problem from two sides: on one side developing and engineering a new series of channel proteins that will stably reconstitute into the synthetic polymers while, in parallel new modified classes of polymers will be synthesized to accomplish the strict conditions requested especially when considering an active delivery.

6.2.2 Stochastic Nano-sensors

In the last two decades nano-pore stochastic sensors have received a great research impulse due to their multi-variate applications such as multi-analyte detection systems, DNA sequencing purposes [98–100], the study of covalent and non-covalent bonding interactions [101, 102], biomolecular folding and unfolding studies [103, 104], differentiation of chiral molecules, enzyme kinetics, determination of sample purity and composition [105] to mention only few.

Historically the first nano-sensor or nano-counter was patented by Wallace H. Coulter. Nowadays a Coulter counter [106] is a system for counting and sizing particles suspended in electrolytes, mainly used as a routine biomedical application for the counting of cells and virus particles [107].

The physical principle defining a nano-pore detection system is simple: the interesting molecules passes through or interacts with the interior of a pore causing detectable changes in ionic pore current [108, 109].

The actual research is following two different systems based on inorganic/organic solid state pores [110] or protein based pores.

Solid-state nano-pores are mainly obtained by using silicon nitride, glass or organic polymers (polyesters, polyimides). Compared to protein

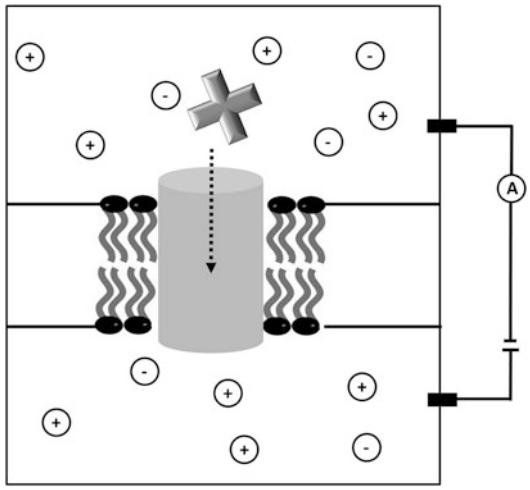

Fig. 6.6 Schematic representation of a nano-pore channel embedded into a lipid bi-layer. The analyte interacting with the pore is indicated by a *cross*. The current over the membrane is measured and changes upon interaction between analyte and pore-interior

pores they show some advantages, because of their tunable pore size and shape, physical robustness allowing to support high voltages and harsher chemical conditions [110].

Protein based nano-pores involve the use of a series of proteins, like β-barrel based bacterial outer membrane proteins such as OmpF [111, 112], OmpG [113], FhuA [114], β-barrel pore forming bacterial toxins as α-hemolysine [108, 115, 116] or α-helical membrane proteins as the K^+-channels (KcsA) [117], though very first bio-organic nano-pores were based on natural channel-forming peptides incorporated into a bilayer lipid membrane able to detect the passage of single molecules with gyration radii of 0.5–1.5 nm [118].

Nanosensors are assembled by placing a nanometre-sized pore into an insulated membrane so to measure the ionic transport through the pore as a measure for the interaction with the molecules of interest. The general principle of a nano-pore sensing element is given in Fig. 6.6.

When the molecules pass through or interact with the nano-channel, the nano-channel itself is obstructed and an interruption in the current flow is registered. Current blockades magnitude, duration, and rates of occurrence allow a fast determination of analyte concentrations and discrimination between similar molecular species [109].

A major issue in nano-pore sensor technology is the fast transport of analyte molecules through the nano-pore. Limitations are due to the current recording techniques that do not possess the right time resolution and sensitivity for the detection of such rapid events. One solution is to slow down the molecular and ion transport through the channel changing and actuating a series of different experimental strategies such as using a host compound or by modifying the analyte molecule and nano-pore sensor.

Other solutions to control molecular transport and, as a consequence, to improve the resolution and sensitivity of nano-pore stochastic sensors, are based on the introduction of functional groups to the nano-pore interior such as hydrophobic, aromatic, positively and negatively charged residues or by increasing the ionic strength of the electrolyte solution and the use of ionic liquids as electrolytes instead of inorganic salts [105].

As a prototype example of a protein channel based stochastic sensor the pore-forming toxin α-hemolysin secreted by *Staphylococcus aureus* can be considered. α-hemolysin is a mushroom-shaped pore, consisting of seven identical subunits arranged around a central axis (see Fig. 6.7).

The two β-strands of each subunit forming one β-hairpin, assemble in the hydrophobic portion of a lipid bilayer forming a β-barrel pore of ∼2 nm constituted by 14 β-strands. Interestingly each of the subunits is expressed independently in a soluble cytoplasmic form and only after assembly the protein folds in its quaternary insoluble structure.

The total length of the channel is ∼10 nm. The external non-barrel portion has a pore opening of 2.9 nm broadening into a cavity of ∼4.0 nm linked to the β-barrel pore.

Its vast use as a stochastic nano-sensor is due to a series of properties such as knowledge of its 3D structure (as solved by X-ray diffraction), robustness toward mutations without losing functionality, relatively large single-channel conductance and a rather open channel showing no transient background current modulations [108].

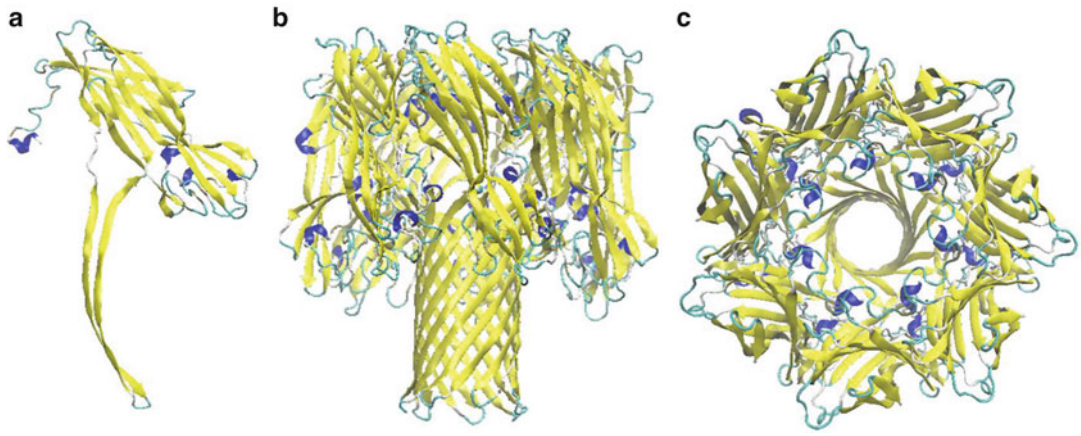

Fig. 6.7 α-hemolysine of *Staphylococcus aureus*, (**a**) one subunit, (**b**) lateral view, (**c**) top view

Due to the cytoplasmic soluble expression of α-hemolysine monomers the production of the assembled protein to be used as nano-pore sensing element is considerably easier than the production of a conventional outer membrane protein. However a disadvantage can be found on the genetic level, when genetic modifications should be introduced, due to the fact that the seven distinct subunits are expressed by the same gene obtaining a mutated protein with all seven subunits modified upon expression. However if only one subunit should be altered, mutated and wild type proteins have to be co-expressed and independently purified (see Sect. 5.1).

Several factors can be used to tune the molecular or ion translocation through the channel, affecting the performance (*i.e.* resolution and sensitivity) of nano-pore stochastic sensing elements, including physical conditions (*i.e.*, pH, voltage, temperature) and structural characteristics of the nano-pore [119–123].

One of the drawbacks of the use of protein based nano-pores is their limited flexibility with regard to the tuning of the pore dimensions, a target easily obtainable by using solid-state nano-pores [110].

To increase protein flexibility toward a tunable pore diameter or length, the FhuA β-barrel protein can be used as nano-channel as demonstrated by Mohammad et al. [114]. The previously mentioned FhuA variants with enlarged inner channel diameter (∼0.4 nm increase) [24] or elongated hydrophobic channel transmembrane portion (∼1 nm increase) [29] that have previously been shown to be applicable in the drug release system development, could potentially be used to design stochastic nano-pore sensors likewise.

However in conclusion and as reported in the article of Guan et al. [105] though the scientific development of nano-pore sensors has reached a high level of sophistication, technological applications of such nano-devices as in medical diagnosis, homeland security, pharmaceutical screening, environmental monitoring is still not feasible and many technological challenges remain to be solved.

6.2.3 Bio-nanoelectronics

A further interesting research field that employs membrane proteins, is the field of bio-nanoelectronic interfaces, that relies on the functional integration of nano-materials and membrane proteins.

By definition bio-nanoelectronic is the interface between bio-based materials and silicon based electronic components, fundamentally developing a bio-inorganic inter-phase. The integration of biological systems with micro-electronics started with the works on capacitive stimulation of cells [124] to monitor neuronal activity with field-effect transistors (FETs) [125]. More recently this field has been expanded toward the use of nanowire (NW) transistor arrays to follow neuronal signal propagation [126].

Recently a very elegant system using lipid membranes and transmembrane peptides, with Silicon nanowires (SiNW) was obtained. A lipid bilayer membrane was incorporated into SiNW transistors by covering the NW with a continuous lipid bilayer shell forming a barrier between the NW surface and solution species. When the "shielded wire" structure incorporates transmembrane peptide pores it enables ionic to electronic signal transduction by using voltage-gated and chemically gated ion transport through the membrane pores, opening new possible applications in biosensing, bioelectronics, neuroscience, and medicine [127].

As a further sophistication regarding the membrane proteins use in nano-electronics, a technology has been developed integrating ion channels and pumps into single-walled carbon nano-tubes and SiNWs in which membrane proteins are embedded in a lipid bilayer shell covering the nanotube or nano-wire component coupling biological transport to electronic signaling [128].

This considerable progress obtained during roughly the last 10 years is mainly due to the miniaturization of nano-materials reaching dimensions that are comparable to the size of biological molecules [129].

References

1. Mashaghi A, Partovi-Azar P, Jadidi T, Nafari N, Maass P, Tabar MRR, Bonn M, Bakker HJ (2012) Hydration strongly affects the molecular and electronic structure of membrane phospholipids. J Chem Phys 136:114709
2. Lewis BA, Engelman DM (1983) Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. J Mol Biol 166:211–217
3. van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9:112–124
4. Simons K, Sampaio JL (2011) Membrane organization and lipid rafts. Cold Spring Harb Perspect Biol 3:a004697
5. Heimburg T (2007) Thermal biophysics of membranes. Wiley VCH, Berlin
6. Vanegas JM, Contreras MF, Faller R, Longo ML (2012) Role of unsaturated lipid and ergosterol in ethanol tolerance of model yeast biomembranes. Biophys J 102:507–516
7. Heberle FA, Petruzielo RS, Pan J, Drazba P, Kučerka N, Standaert RF, Feigenson GW, Katsaras J (2013) Bilayer thickness mismatch controls domain size in model membranes. J Am Chem Soc 135(18): 6853–6859
8. Baoukina S, Mendez-Villuendas E, Tieleman DP (2012) Molecular view of phase coexistence in lipid monolayers. J Am Chem Soc 134:17543–17553
9. Scheve CS, Gonzales PA, Momin N, Stachowiak JC (2013) Steric pressure between membrane-bound proteins opposes lipid phase separation. J Am Chem Soc 135:1185–1188
10. Andersen OS, Koeppe RE (2007) Bilayer thickness and membrane protein function: an energetic perspective. Annu Rev Biophys Biomol Struct 36: 107–130
11. Shen H-H, Lithgow T, Martin LL (2013) Reconstitution of membrane proteins into model membranes: seeking better ways to retain protein activities. Int J Mol Sci 14:1589–1607
12. Jesorka A, Orwar O (2008) Liposomes: technologies and analytical applications. Annu Rev Anal Chem 1:801–832
13. Yanagisawa M, Iwamoto M, Kato A, Yoshikawa K, Oiki S (2011) Oriented reconstitution of a membrane protein in a giant unilamellar vesicle: experimental verification with the potassium channel KcsA. J Am Chem Soc 133:11774–11779
14. White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 28:319–365
15. Rath A, Deber CM (2012) Protein structure in membrane domains. Annu Rev Biophys 41:135–155
16. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29:4031–4037
17. Tamm LK, Arora A, Kleinschmidt JH (2001) Structure and assembly of beta-barrel membrane proteins. J Biol Chem 276:32399–32402
18. Wimley WC (2003) The versatile β-barrel membrane protein. Curr Opin Struct Biol 13:404–411
19. Gold VA, Duong F, Collinson I (2007) Structure and function of the bacterial Sec translocon. Mol Membr Biol 24:387–394
20. Kim KH, Kang HS, Okon M, Escobar-Cabrera E, McIntosh LP, Paetzel M (2011) Structural characterization of Escherichia coli BamE, a lipoprotein component of the β-barrel assembly machinery complex. Biochemistry 50:1081–1090
21. Moon CP, Zaccai NR, Fleming PJ, Gessmann D, Fleming KG (2013) Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. Proc Natl Acad Sci USA 110:4285–4290
22. Rodríguez-Ropero F, Fioroni M (2012) Structural and dynamical analysis of an engineered FhuA channel protein embedded into a lipid bilayer or a detergent belt. J Struct Biol 177:291–301
23. Fioroni M (2011) Unpublished study. RWTH Aachen University, Aachen

24. Krewinkel M, Dworeck T, Fioroni M (2012) Engineering of an E. coli outer membrane protein FhuA with increased channel diameter. J Nanobiotechnol 9:33
25. Hagan CL, Kim S, Kahne D (2010) Reconstitution of outer membrane protein assembly from purified components. Science 14:890–892
26. Meier W (2000) Polymer nanocapsules. Chem Soc Rev 29:295–303
27. Matyjaszewski K, Möller M (2012) Polymer science: a comprehensive reference, vol 1. Elsevier, Oxford
28. Rusanov AI (2002) Striking world of nanostructures. Russ J Gen Chem 72:495–511
29. Muhammad N, Dworeck T, Fioroni M, Schwaneberg U (2011) Engineering of the E. Coli outer membrane protein FhuA to overcome the hydrophobic mismatch in thick polymeric membranes. J Nanobiotechnol 9:8
30. Choi H-J, Germain J, Montemagno CD (2006) Effects of different reconstitution procedures on membrane protein activities in proteopolymersome. Nanotechnology 17:1825–1830
31. Discher B, Won Y, Ege D, Lee J, Bates F, Discher D, Hammer D (1999) Polymersomes: tough vesicles made from di-block copolymers. Science 284: 1143–1146
32. Srinivas G, Discher D, Klein ML (2005) Key roles for chain flexibility in block copolymer membranes that contain pores or make tubes. Nano Lett 5: 2343–2349
33. Ranquin A, Versees W, Meier W, Steyaert J, Van Gelder P (2005) Therapeutic nanoreactors: combining chemistry and biology in a novel triblock copolymer drug delivery system. Nano Lett 5: 2220–2224
34. Kumar M, Grzelakowski M, Zilles J, Clark M, Meier W (2007) Highly permeable polymeric membranes based on the incorporation of the functional water channel protein aquaporin Z. Proc Natl Acad Sci USA 104:20719–20724
35. Nardin C, Thoeni S, Widmer J, Winterhalter M, Meier W (2000) Nanoreactors based on (polymerized) ABA-triblock copolymer vesicles. Chem Commun 15:1433–1434
36. May S, Andreasson-Ochsner M, Fu Z, Low YX, Tan D, de Hoog HP, Ritz S, Nallani M, Sinner EK (2013) In vitro expressed GPCR inserted in polymersome membranes for ligand-binding studies. Angew Chem Int Ed 52:749–753
37. Kumar M, Habel JEO, Shen Y-X, Meier W, Walz T (2012) High-density reconstitution of functional water channels into vesicular and planar block copolymer membranes. J Am Chem Soc 134:18631–18637
38. Choi HJ, Montemagno CD (2005) Artificial organelle: ATP synthesis from cellular mimetic polymersomes. Nano Lett 5:2538–2542
39. Onaca O, Sarkar P, Roccatano D, Friedrich T, Hauer B, Grzelakowski M, Güven A, Fioroni M, Schwaneberg U (2008) Functionalized nanocompartments (synthosomes) with a reduction-triggered release system. Angew Chem Int Ed 47:7029–7031
40. Jaskiewicz K, Makowski M, Kappl M, Landfester K, Kroeger A (2012) Mechanical properties of poly(dimethylsiloxane)-block-poly(2-methyloxazoline) polymersomes probed by atomic force microscopy. Langmuir 28:12629–12636
41. Battaglia G, Ryan AJ (2005) Bilayers and interdigitation in block copolymer vesicles. J Am Chem Soc 127:8757–8764
42. Zheng Y, Won Y-Y, Bates FS, Davis HT, Scriven LE, Talmon Y (1999) Directly resolved core-corona structure of block copolymer micelles by cryo-transmission electron microscopy. J Phys Chem B 103:10331–10334
43. Robeson LM (2007) Polymer blends: a comprehensive review. Carl Hanser, Munich
44. Pata V, Dan N (2003) The effect of chain length on protein solubilization in polymer-based vesicles (polymersomes). Biophys J 85:2111–2118
45. Hélix-Nielsen C (2012) Biomimetic membranes for sensor and separation applications, Biological and medical physics, biomedical engineering. Springer, Dordrecht
46. Jackson JD (1998) Classical electrodynamics. Wiley, New York
47. Roke S, Gonella G (2012) Nonlinear light scattering and spectroscopy of particles and droplets in liquids. Annu Rev Phys Chem 63:353–378
48. Santos NC, Castanho MARB (1996) Teaching light scattering spectroscopy: the dimension and shape of the tobacco mosaic virus. Biophys J 71:1641–1646
49. Bernes BJ, Pecora R (1976) Dynamic light scattering with applications to chemistry, biology and physics. Courier Dover, New York
50. Some D (2013) Light-scattering-based analysis of biomolecular interactions. Biophys Rev 5:147–158
51. Proteau A, Shi R, Cygler M (2010) Application of dynamic light scattering in protein crystallization. In: Coligan JE (ed) Current protocols in protein science. Wiley Interscience, Hoboken
52. Gast K, Fiedler C (2012) Dynamic and static light scattering of intrinsically disordered proteins. Methods Mol Biol 896:137–161
53. Christian DA, Cai S, Bowen DM, Kim Y, Pajerowski JD, Discher D (2009) Polymersome carriers: from self-assembly to siRNA and protein therapeutics. Eur J Pharm Biopharm 71:463–474
54. Mabrouk E, Cuvelier D, Brochard-Wyart F, Nassoy P, Li MH (2009) Bursting of sensitive polymersomes induced by curling. Proc Natl Acad Sci USA 106:7294–7298
55. Jain JP, Ayen WY, Kumar N (2011) Self assembling polymers as polymersomes for drug delivery. Curr Pharm Des 17:65–79
56. Hupfeld S, Holsæter AM, Skar M, Frantzen CB, Brandl M (2006) Liposome size analysis by dynamic/static light scattering upon size exclusion-/field flow-fractionation. J Nanosci Nanotechnol 6:1–7

57. Josephy P, Eling T, Mason R (1982) The horseradish peroxidase-catalyzed oxidation of 3,5,3′,5′-tetramethylbenzidine. Free radical and charge transfer complex intermediates. J Biol Chem 257:3669–3675
58. Dworeck T, Petri AK, Muhammad N, Fioroni M, Schwaneberg U (2011) FhuA deletion variant Δ1-159 overexpression in inclusion bodies and refolding with polyethylene-poly(ethylene glycol) diblock copolymer. Protein Expr Purif 77:75–79
59. Güven A, Dworeck T, Fioroni M, Schwaneberg U (2011) Residue K556-a light triggerable gatekeeper to sterically control translocation in FhuA. Adv Eng Mater 13:B324–B329
60. Güven A, Fioroni M, Hauer B, Schwaneberg U (2010) Molecular understanding of sterically controlled compound release through an engineered channel protein (FhuA). J Nanobiotechnol 8:14
61. Ihle S, Onaca O, Rigler P, Hauer B, Rodrıguez-Ropero F, Fioroni M, Schwaneberg U (2011) Nanocompartments with a pH release system based on an engineered OmpF channel protein. Soft Matter 7:532–539
62. Elson EL, Magde D (1974) Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers 13:1–27
63. Magde D, Elson EL, Webb WW (1972) Thermodynamic fluctuations in a reacting system: measurement by fluorescence correlation spectroscopy. Phys Rev Lett 29:705–708
64. Neher E, Sakmann B (1992) The patch clamp technique. Sci Am 266:44–51
65. Neher E, Sakmann B (1976) Single channel currents recorded from membrane of denervated frog muscle fibres. Nature 260:799–802
66. Neher E, Sakmann B, Steinbach JH (1978) The extracellular patch clamp: a method for resolving currents through individual open channels in biological membranes. Pflugers Arch 375:219–228
67. Hamill OP, Marty A, Neher E, Sakmann B, Sigworth FJ (1981) Improved patchclamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflugers Arch 391: 85–100
68. Okada Y (2012) Patch clamp techniques: from beginning to advanced protocols, Springer protocols handbooks. Springer, Tokyo/New York
69. Petrov E, Palanivelu D, Constantine M, Rohde PR, Cox CD, Nomura T, Minor DL, Martinac B (2013) Patch-clamp characterization of the MscS-like mechanosensitive channel from Silicibacter pomeroyi. Biophys J 104:1426–1434
70. Aimon S, Manzi J, Schmidt D, Poveda Larrosa JA, Bassereau P, Toombes GE (2011) Functional reconstitution of a voltage-gated potassium channel in giant unilamellar vesicles. PLoS One 6: e25529
71. Morera FJ, Vargas G, González C, Rosenmann E, Latorre R (2007) Ion-channel reconstitution. Methods Mol Biol 400:571–585
72. Stava E, Yu M, Shin HC, Shin H, Rodriguez J, Blick RH (2012) Mechanical actuation of ion channels using a piezoelectric planar patch clamp system. Lab Chip 12:80–87
73. Brüggemann A, Farre C, Haarmann C, Haythornthwaite A, Kreir M, Stoelzle S, George M, Fertig N (2008) Planar patch clamp: advances in electrophysiology. Methods Mol Biol 491:165–176
74. Freundlich MM (1963) Origin of the electron microscope. Science 142:185–188
75. Kumar P, Sanjeev Kumar S, Shubhra S (2012) Uses of electron microscope in modern practical techniques: novel techniques of electron microscope. LAP Lambert Academic, Colne
76. Ruozi B, Belletti D, Tombesi A, Tosi G, Bondioli L, Forni F, Vandelli MA (2011) AFM, ESEM, TEM, and CLSM in liposomal characterization: a comparative study. Int J Nanomed 6:557–563
77. Tihova M, Tattrie B, Nicholls P (1994) Cytochrome c oxidase in proteoliposomes visualised by platinum-carbon and by tungsten-tantalum shadowing: image analysis. Biochem Biophys Res Commun 203:331–337
78. Yakataa K, Hiroakia Y, Ishibashic K, Soharad E, Sasakid S, Mitsuokae K, Fujiyoshia Y (2007) Aquaporin-11 containing a divergent NPA motif has normal water channel activity. Biochim Biophys Acta 1768:688–693
79. Andreeva-Kovalevskaya Z, Solonin AS, Sineva EV, Ternovsky VI (2008) Pore-forming proteins and adaptation of living organisms to environmental conditions. Biochemistry (Mosc) 73:1473–1492
80. Mouro-Chanteloup I, Cochet S, Chami M, Genetet S, Zidi-Yahiaoui N, Engel A, Colin Y, Bertrand O, Ripoche P (2010) Functional reconstitution into liposomes of purified human RhCG ammonia channel. PLoS One 5:e8921
81. Wang L, Sigworth FJ (2009) Cryo-EM structure of the BK potassium channel in a lipid membrane. Nature 461:292–295
82. Trépout S, Taveau J-C, Benabdelhak H, Granier T, Ducruix A, Frangakis AS, Lambert O (2010) Structure of reconstituted bacterial membrane efflux pump by cryo-electron tomography. Biochim Biophys Acta 1798:1953–1960
83. Muhammad N, Dworeck T, Schenk A, Shinde P, Fioroni M, Schwaneberg U (2012) Polymersome surface decoration by an EGFP fusion protein employing Cecropin A as peptide “anchor”. J Biotechnol 157:31–37
84. Zhao H, Kinnunen PK (2002) Binding of the antimicrobial peptide temporin L to liposomes, assessed by Trp fluorescence. J Biol Chem 277:25170–25177
85. Sood R, Domanov Y, Kinnunen PK (2007) Fluorescent temporin B derivative and its binding to liposomes. J Fluoresc 17:223–234
86. Powl AM, East JM, Lee AG (2003) Lipid-protein interactions studied by introduction of a tryptophan residue: the mechanosensitive channel MscL. Biochemistry 42:14306–14317

87. Yoshimura K, Sokabe M (2010) Mechanosensitivity of ion channels based on protein-lipid interactions. R Soc Interface 7:S307–S320
88. Moon CP, Fleming KG (2011) Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes, vol 492, Methods in enzymology. Elsevier, Baltimore
89. Kumar MNVR (2008) Handbook of particulate drug delivery, vol 2. American Scientific, Stevenson Ranch
90. Sawant RR, Torchilin VP (2012) Challenges in development of targeted liposomal therapeutics. AAPS J 14:303–315
91. Torchilin VP (2010) Passive and active drug targeting: drug delivery to tumors as an example. Handb Exp Pharmacol 197:3–53
92. Kim MS, Lee DS (2010) Biodegradable and pH-sensitive polymersome with tuning permeable membrane for drug delivery carrier. Chem Commun 46:4481–4483
93. Pawar PV, Gohil SG, Jainc JP, Kumar N (2013) Functionalized polymersomes for biomedical applications. Polym Chem 4:3160–3176
94. El Fray M, Prowans P, Puskas J, Altstädt V (2006) Biocompatibility and fatigue properties of polystyrene-polyisobutylene-polystyrene, an emerging thermoplastic elastomeric biomaterial. Biomacromolecules 7:844–850
95. Webster R, Didier E, Harris P, Siegel N, Stadler J, Tilbury L, Smith D (2006) PEGylated proteins: evaluation of their safety in the absence of definitive metabolism studies. Drug Metab Dispos 35:9–16
96. Puskas J, Chen Y, Dahman Y, Padavan D (2004) Polyisobutylene-based biomaterials. J Polym Sci Polym Chem 42:3091–3109
97. Reitz S, Cebi M, Reiss P, Studnik G, Linne U, Koert U, Essen LO (2009) On the function and structure of synthetically modified porins. Angew Chem Int Ed Engl 48:4853–4857
98. Howorka S, Cheley S, Bayley H (2001) Sequence-specific detection of individual DNA strands using engineered nanopores. Nat Biotechnol 19: 636–639
99. Storm AJ, Storm C, Chen J, Zandbergen H, Joanny JF, Dekker C (2005) Fast DNA translocation through a solid-state nanopore. Nano Lett 5: 1193–1197
100. Sigalov G, Comer J, Timp G, Aksimentiev A (2008) Detection of DNA sequences using an alternating electric field in a nanopore capacitor. Nano Lett 8:56–63
101. Luchian T, Shin SH, Bayley H (2003) Single-molecule covalent chemistry with spatially separated reactants. Angew Chem Int Ed 42:3766–3771
102. Zhao Q, Jayawardhana DA, Guan X (2008) Stochastic study of the effect of ionic strength on non-covalent interactions in protein pores. Biophys J 94:1267–1275
103. Shim JW, Tan Q, Gu L (2009) Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity. Nucleic Acids Res 37:972–982
104. Talaga DS, Li J (2009) Single-molecule protein unfolding in solid state nanopores. J Am Chem Soc 131:9287–9297
105. Wang G, Wang L, Han Y, Zhou S, Guan X (in press) Nanopore stochastic detection: diversity, sensitivity, and beyond. Acc Chem Res, Epub ahead of print
106. Coulter WH (1953) Means for counting particles suspended in a fluid. US Patent 2,656,508
107. DeBlois RW, Bean CP (1970) Counting and sizing of submicron particles by the resistive pulse technique. Rev Sci Instrum 41:909–916
108. Howorka S, Siwy Z (2009) Nanopore analytics: sensing of single molecules. Chem Soc Rev 38:2360–2384
109. Schmidt J (2005) Stochastic sensors. J Mater Chem 15:831–840
110. Howorka S, Siwy ZS (2012) Nanopores as protein sensors. Nat Biotechnol 30:506–507
111. Nestorovich EM, Danelon C, Winterhalter M, Bezrukov SM (2002) Designed to penetrate: time-resolved interaction of single antibiotic molecules with bacterial pores. Proc Natl Acad Sci USA 99:9789–9794
112. Chimerel C, Movileanu L, Pezeshki S, Winterhalter M, Kleinekathoäfer U (2008) Transport at the nanoscale: temperature dependence of ion conductance. Eur Biophys J 38:121–125
113. Chen M, Khalid S, Sansom MSP, Bayley H (2008) Outer membrane protein G: engineering a quiet pore for biosensing. Proc Natl Acad Sci USA 105: 6272–6277
114. Mohammad M, Howard KR, Movileanu L (2011) Redesign of a plugged β-barrel membrane protein. J Biol Chem 286:8000–8013
115. Bayley H, Cremer PS (2001) Stochastic sensors inspired by biology. Nature 413:226–230
116. Howorka S, Movileanu L, Lu X, Magnon M, Cheley S, Braha O, Bayley H (2000) A protein pore with a single polymer chain tethered within the lumen. J Am Chem Soc 122:2411–2416
117. Beacham DW, Blackmer T, O'Grady M, Hanson GT (2010) Cell-based potassium ion channel screening using the FluxOR assay. J Biomol Screen 15: 441–446
118. Bezrukov SM, Vodyanoy I, Parsegian VA (1994) Counting polymers moving through a single ion channel. Nature 370:279–281
119. Kang X, Gu L, Cheley S, Bayley H (2005) Single protein pores containing molecular adapter at high temperature. Angew Chem Int Ed 44:1495–1499
120. de Zoysa RSS, Krishantha DMM, Zhao Q, Gupta J, Guan X (2011) Translocation of single stranded DNA through the alpha-hemolysin nanopore in acidic solutions. Electrophoresis 32:3034–3041

121. Meller A, Nivon L, Branton D (2001) Voltage-driven DNA translocations through a nanopore. Phys Rev Lett 86:3435–3438
122. Kuyucak S, Bastug T (2003) Physics of ion channels. J Biol Phys 29:429–446
123. Cheley S, Gu LQ, Bayley H (2002) Stochastic sensing of nanomolar inositol 1,4,5-trisphosphate with an engineered pore. Chem Biol 9:829–838
124. Fromherz P, Stett A (1995) Silicon-neuron junction: capacitive stimulation of an individual neuron on a silicon chip. Phys Rev Lett 75:1670–1673
125. Jenkner M, Fromherz P (1997) Bistability of membrane conductance in cell adhesion observed in a neuron transistor. Phys Rev Lett 79: 4705–4708
126. Patolsky F, Timko BP, Yu G, Fang Y, Greytak AB, Zheng G, Lieber CM (2006) Detection, stimulation, and inhibition of neuronal signals with high-density nanowire transistor arrays. Science 313:1100–1104
127. Misraa N, Martineza JA, Huanga S-CJ, Wang Y, Stroeve P, Grigoropoulos CP, Noya A (2009) Bioelectronic silicon nanowire devices using functional membrane proteins. Proc Natl Acad Sci USA 106:13780–13784
128. Noy A, Artyukhin AB, Huang SC, Martinez JA, Misra N (2011) Functional integration of membrane proteins with nanotube and nanowire transistor devices. Methods Mol Biol 751:533–552
129. Noy A (2011) Bionanoelectronics. Adv Mater 23:807–820

Final Considerations

7

The present book is intended as an introductory journey to the biology, biotechnology, characterization and nano-technological applications of the β-barrel outer membrane proteins (OMPs).

The hope of the Authors is that the interested Reader is able to acquire some valuable new information on this ongoing as well as challenging research field, while at the same time enjoying a pleasant read.

The number of books and articles dedicated to OMPs is exponentially increasing and, most probably, will further increase as soon as certain issues that currently still prevent a high level production of this class of proteins will be overcome, allowing β-barrel outer membrane protein channels to be applied as nano-channel components to technologies such as drug delivery, stochastic sensors and bio-nanoelectronics to mention few.

In general, OMPs are well known among biologists, known among biotechnologist and less known among scientists with backgrounds more remote from the biological life sciences, though OMPs are slowly introduced as potent and flexible nano-components also to chemists and physicists with an interest in nanotechnology; a process that this book hopefully will facilitate further.

The interest OMPs spark is mainly due to their unique structural features and due to their folding behavior, as these proteins are exceptionally robust (*e.g.* temperature stable and resistant against low to moderate concentrations of organic solvents) and have been shown to refold *in vitro* from a completely or partially unfolded state. Furthermore they are known among molecular biologists for their tolerance toward vast sequence mutations rendering them valuable targets for genetic engineering purposes.

However, the quality jump β-barrel outer membrane proteins must overpass before being considered of value for any real world technological application must deal with building on and expanding their ability to be easily engineered and obtained through a mass scale production, which might turn out to be a difficult job, as:

The nature of membrane proteins in general and the resulting special demands they make in terms of hydrophobicity of their environment and presence of a boundary surface (hydrophobic/hydrophilic) render their industrial large-scale production in the range of bulk chemicals very unlikely and scale-up intentions will most likely reach at most to the scale of fine chemical or pharmaceutics production rates, as required for nano-channel applications such as drug-delivery from nano-containers or OMP use as stochastic nano-sensors.

To be industrially considered significant increases of product yield per liter of culture are necessary, such yield increases might in the future be possible employing an alliance between conventional protein expression methods with metabolic/cellular engineering (currently a maximum protein yield in the mg/ml range is reached

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2_7,

by expression into inclusion bodies or by the use of cell-free expression systems).

The range of nano-technological applications in which engineered β-barrel outer membrane protein channels can be used is limited further due to their "biological" origin, implying that in spite of their previously mentioned robustness they can be used in a very limited range of temperatures, pH, ionic strength as well as presence of organic solvents.

In synthesis, to reach the high level of technological sophistication that allows the use of OMPs as valid nano-components still necessitates the need of a deeper understanding on the origin of their structural stability, as well as their biogenesis, folding and reconstitution ability.

All of these obstacles can be overcome only through a consistent use of several theoretical (*i.e.* simulations and structural prediction) and experimental techniques (genetic engineering and production) requiring the use of advanced analytical methods.

Though there is still a long way to go before such targets can be reached, the way is already partly leveled and first model systems have been reported.

The present book can be seen as a first introductory work pointing out the potentials of OMPs for the bio-nanotechnology field. It shows the state of the art of the involved research underlining where it still lacks knowledge and understanding on both the protein based nano-channels as well as on the hybrid systems of biological and non-biological (*e.g.* protein functionalized polymersomes) components.

In conclusion the authors hope that this first introduction will help to increase the attention, this fascinating class of proteins (*i.e.* β-barrel outer membrane proteins) obtains from the scientific community and that β-barrel outer membrane proteins will be considered not only as interesting targets for model studies but also as tools and components that can be actually put to use in bio-nanotechnological applications.

Index

M. Fioroni et al., *ß-barrel Channel Proteins as Tools in Nanotechnology*, Advances in Experimental Medicine and Biology 794, DOI 10.1007/978-94-007-7429-2,