STRUCTURAL AND KINETIC INSIGHTS INTO MEMBRANE PROTEIN FOLDING

INAUGURAL DISSERTATION

FOR THE ATTAINMENT OF THE TITLE OF DOCTOR IN THE FACULTY OF MATHEMATICS AND NATURAL SCIENCES AT THE HEINRICH HEINE UNIVERSITY DÜSSELDORF

PRESENTED BY

SHANTHA THERESA ELTER

FROM SNEHALYA

DÜSSELDORF, JULY 2020

from the Institute for Physical Biology

at the Heinrich Heine University Düsseldorf

Published by permission of the Faculty of Mathematics and Natural Sciences at

Heinrich Heine University Düsseldorf

Supervisor: Dr. Manuel Etzkorn

Mentor: Prof. Dr. Georg Groth

Date of the oral examination: 02.10.2020

1 EIDESSTATTLICHE VERSICHERUNG

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist.

Ich versichere, dass die eingereichte schriftliche Fassung der auf dem beigefügten Medium gespeicherten Fassung entspricht.

Signed: _____

Shantha Theresa Elter

Date: _____

2 ABSTRACT

How do membrane proteins get into shape? How do they fold in their natural environment to function without errors?

The aim of this work was to find answers to these questions by using cell-free protein expression, kinetic analysis and NMR studies of the 7-transmembrane (7-TM) α-helical protein bacteriorhodopsin (bR) from the archaebacterium *Halobacterium salinarium*. Due to its 7-TM structure, bR also serves as a model protein for similarly structured proteins, like the important family of G protein-coupled receptors (GPCRs).

In this work a stable, reliable *E. coli*-based cell-free expression system was established, enabling the production of bacterioOpsin (bO) and other proteins. In addition the established cell-free expression system was used for the production of previously inaccessible (membrane) proteins enabling a subsequent biophysical characterization. Furthermore, the system facilitates efficient and economic isotope labelling strategies, such as combinatorial- or stereospecific methyl Leu labelling. A characterization of the unfolded state of bO in sodium dodecyl sulfate (SDS) was achieved by partial, sequential resonance assignment. Assignments were enabled by combinatorial labelling through the cell-free expression system and the combination of various 2D NMR experiments. In addition, amino acid-specific interactions of bO with SDS were analyzed in greater detail with 3D TROSY-NOESY- and [^{15}N-^{1}H]-HSQC experiments.

The folded state of bR was analyzed using Atomic Force Microscopy (AFM) and NMR spectroscopy. NMR characterization of bR and GPCRs in amphipol (APOL) environment showed that resolution and dispersion of acquired spectra highly depend on particle size and sample homogeneity, which in turn influences refolding properties. Experiments to optimize bR samples for NMR-studies showed, that a constant compromise must be made between spectral quality and refolding efficiency. For a more detailed analysis, a new absorption based assay in a 384-well format was developed, facilitating extensive screening of various folding conditions of bR. The Combination of kinetic data of the most important parameters for folding (such as bO-, APOL-, ligand- and SDS concentration) with the acquired NMR data showed that the ratio of bO to SDS, as well as the presence of monomeric SDS molecules are key factors for the folding process. For the first time, insights into the folding process of bR via real time-resolved NMR measurements were obtained, indicating fast transitions into a NMR concealed state.

3 ZUSAMMENFASSUNG

Wie kommen Membranproteine in Form? Wie falten sie sich in ihrer natürlichen Umgebung, um fehlerfrei zu funktionieren?

Ziel dieser Arbeit war die Beantwortung dieser Fragen, mithilfe von zellfreier Proteinexpression, kinetischer Analysen und NMR-Studien des 7-transmembranen (7-TM) α-helikalen Proteins Bacteriorhodopsin (bR) aus dem Archaebakterium *Halobacterium salinarium*. Durch seine 7-TM-Struktur dient bR auch als Modellprotein für andere Proteine ähnlicher Struktur, wie z.B. der wichtigen Familie der G-Protein-gekoppelten Rezeptoren (GPCRs).

In dieser Arbeit wurde ein stabiles *E. coli*-basiertes zellfreies Expressionssystem etabliert, das die Produktion von bacterioOpsin (bO) und anderen Proteinen ermöglicht. Das neu etablierte zellfreie Expressionssystem wurde zudem dazu verwendet einige bislang nicht zugängliche (Membran)Proteine herzustellen und biophysikalisch zu untersuchen. Zusätzlich können Isotopenmarkierungen (z.B. kombinatorisch und stereospezifisch Methyl-Leu) mithilfe des Systems effizient und kostengünstig realisiert werden. Durch eine partielle, sequentielle Resonanzzuordnung wurde eine weitere Charakterisierung des ungefalteten Zustands von bO in Natriumdodecylsulfat (SDS) erreicht. Dies wurde unter anderem durch kombinatorische Isotopenmarkierungen mithilfe des zellfreien Expressionssystems in Kombination mit verschiedenen 2D NMR-Experimenten ermöglicht. Weiterhin wurden die aminosäurespezifischen Interaktionen von bO mit SDS mithilfe von 3D TROSY-NOESY- und [^{15}N-^{1}H]-HSQC-Experimenten genauer analysiert.

Der gefaltete Zustand von bR wurde mittels Rasterkraftmikroskopie (AFM) und NMR-Spektroskopie untersucht. Die NMR Charakterisierung von bR und GPCRs in Amphipol-Umgebung (APOL) zeigten, dass die Auflösung und Dispersion der Spektren in hohem Maße von der Partikelgröße und Homogenität der Proben abhängt, welche wiederum die Refolding-Eigenschaften beeinflusst. Experimente zur Optimierung der bR-Proben für NMR-Studien zeigten, dass stets ein Kompromiss zwischen spektraler Qualität und Refoldingeffizienz eingegangen werden muss. Für eine genauere Analyse wurde ein neuer absorptionsbasierter Assay im 384-well-Format entwickelt, der ein umfassendes Screening verschiedener Faltungsbedingungen von bR ermöglichte. Die Kombination kinetischer Daten zu den wichtigsten Parametern der Faltung (z.B. bO-, APOL-, Liganden- und SDS-Konzentration) mit den erhaltenen NMR-Daten zeigte, dass die Ratio von bO zu SDS sowie das Vorhandensein monomerer SDS Moleküle die Schlüsselfaktoren des Faltungsprozesses sind. Erstmalig konnten zudem auch Einblicke in den Faltungsprozess mittels Echtzeitaufgelösten-NMR-Messungen gewonnen werden, welche die schnelle Umwandlung in einen NMR-verborgenen Zustand andeuten.

4 ACKNOWLEDGEMENTS

I thank Dr. Manuel Etzkorn for always supporting me with everything, and for giving me the chance to develop myself further under his supervision. I also thank Manuel for showing me the great opportunities of NMR-spectroscopy. Thank you, Manuel!

I thank Prof. Georg Groth for giving me the chance to make my first steps in research and for giving me a lot of skills and knowledge to proceed. Thank you, Georg!

I thank Prof. Vlado Gelev for providing great, pure and variable amphipols, for great collaboration and inspiration. Thank you Vlado!

I thank the whole AG Etzkorn, namely Manuel Etzkorn, Aldino Viegas, Thibault Viennet, Sabine Schriek, Jan Borggräfe, Marcel Falke and Laetitia Heid. It was a great time with you all in the lab and in Jülich. I keep you all in my heart. Thank you!

I thank Barbara Schulten, Bernd Esters, Elke Reinartz, Astrid Wies and Reinhild Wurm for always supporting and helping me. Thank you!

I thank Nadine Rösener for everything! Thank you, Nadine!

I thank Prof. Dieter Willbold, Prof. Henrike Heise Mario Schneider, Stefanie Bungert-Plümke, Lothar Gremer, Olli Bannach, Andreas Kulawik, Gerhard Steger, Luitgard Nagel-Steger, Ingrid Span, Rudolf Hartmann, and Wolfgang Hoyer for inspiration, collaboration and fresh input whenever needed. Thank you!

Furthermore I thank Aldo Vacco, Agnes Papala, Angelina Schliebs, Christine Arndt, Adrian Glatzel and Sebo Kork for inspiration. Thank you!

Last but not least I thank Violetta for always being at my side! I love you!

Thank you, Violetta!

5 LIST OF TABLES

6 LIST OF FIGURES

7 LIST OF ABBREVIATIONS AND ACRONYMS

Entries appear by order of appearance in the dissertation. Names of chemicals, proteins and methods which are commonly known are not included here but defined in the text.

MP - integral membrane protein
GPCR - G protein-coupled receptors
ER - endoplasmic reticulum
SI - sequence identity
7-TM - seven-transmembrane
ECL - extracellular loop
ICL - intracellular loop
GAP - GTPase-activating protein
PM - purple membrane
SB - Schiff base
PTM - posttranslational modification
CHO - Chinese Hamster Ovary cells
HEK - Human Embryonic Kidney cells
CFPS - cell-free protein synthesis system
CFCF - continuous-flow cell-free translation
CECF - continuous-exchange cell-free expression
RM - reaction mix
FM - feeding mix
T7RNAP - T7RNA polymerase
P-CF - precipitate-based cell-free expression
D-CF - detergent-based cell-free expression
L-CF - lipid-based cell-free expression
PSC - protein-surfactant complex
CMC - critical micelle concentration
CMT - critical micelle temperature
MLV - multilamellar vesicle
GUV - giant unilamellar vesicle
LUV - large unilamellar vesicle
SUV - small unilamellar vesicle
NLP - nanolipoprotein
MSP - membrane scaffold protein
APOL - amphipol
PDC - protein-detergent complex
PAA - poly acrylic acid
bO - bacterioOpsin
bR - bacteriorhodopsin
tBLM - tethered bilayer lipid membrane
AFM - Atomic Force Microscopy
TROSY - Transverse Relaxation Optimized Spectroscopy
HSQC -Heteronuclear Single QuantumCoherence
SEC - size exclusion chromatography
IMAC - immobilized metal affinity chromatography
NOESY - Nuclear Overhauser Enhancement Spectroscopy
NOE - Nuclear Overhauser Effect
MD - molecular dynamics simulations
Vpr - HIV-1 Viral protein R
AIDS - Acquired Immunodeficiency Syndrome
bHLH - basic helix-loop-helix
FPLC - Fast Protein Liquid Chromatography

8 CONTENTS

9 INTRODUCTION

9.1 MEMBRANE PROTEINS

Membranes compartmentalize cells and act as an important permeability barrier. Biological membranes are generally composed of proteins, lipids and carbohydrates in variable proportions. The ability of integral membrane proteins (MPs) to interact with the aqueous and lipid environment is the requirement for their appearance in lipid bilayers. Like all other proteins, MPs are synthesized by ribosomes and subsequently make their way to different membrane locations.

Membrane proteins are as versatile as their functions and build a group of very fascinating structures in biology. They are of crucial importance for cell-to-cell communication, sensory stimuli transduction, transport of solutes and secretion of substances and additionally for energy conversion e.g. in respiratory and photosynthetic systems. To allow rapid responses to the varying demands on a cell, rapid synthesis, turnover and degradation of a typical MP are important to maintain their normal dynamic function.

Due to the amphipathic character of MPs, they are insoluble in aqueous media, which can make their analysis quite challenging. Although in the last decades a lot of effort and success has been made in understanding the function of MPs and gaining new insights in their structural organization, intensive research is still required to complete the puzzle.

The structure of MPs provides a rich source of information in biology. In order to maintain lipid bilayer integrity, protein structure must enable stable integration of MPs and support function of the protein in its ambiguous environment. The predominant secondary structures are α-helices or β-barrels. Their tertiary complex subunit structure and quaternary structure involved in building of larger protein complexes are fundamental aspects in the molecular description of their functionality.

Since 20-30 % of human genes encode membrane proteins[1] further research on their structure enables the finding of better drugs which can improve human and animal health.

9.1.1 G protein-coupled receptors

The largest protein superfamily in mammalian genomes are comprised by G protein-coupled receptors (GPCRs)[2,3]. GPCRs represent the largest and most ubiquitous family of membrane receptors. They play important roles in the perception and transduction

of diverse signals by activation of various and complex pathways. Research has shown that GPCRs can be found in numerous types of membranes, for example in mitochondria[4], melanosomes[5], endoplasmic reticulum (ER) membranes[6], lysosomes[7,8], and nuclear membranes[9]. Nuclear localization of GPCRs may favor *in situ* interactions with chromatin to regulate gene transcription, chromosome remodeling, and genomic integrity[9].

So far there are more than 800 known human GPCRs, which exhibit a wide-ranging structural diversity and can be classified into five major families and several subfamilies on the basis of their shared sequences and structural features[10]:

1. Rhodopsin-like family
2. Secretin receptor family
3. Glutamate receptor family
4. Adhesion receptor family
5. Frizzled/Taste2 receptor family

The rhodopsin-like family represents the largest and most diverse group of GPCRs and includes roughly 700 human GPCRs[10]. Its subgroups α–δ share a sequence identity (SI) of ≥ 25%, which indicates their great structural diversity[10]. Receptors for peptides, biogenic amines, metal ions, nucleotides, signaling lipids, visible light and odorants are subsumed by the rhodopsin-like family. The secretin receptor family includes receptors for the hormones glucagon and parathyroid hormone, among others, and respond to long α-helical peptide ligands. The glutamate receptor family includes large extracellular domains that contain the entire ligand-binding site. They are obligate dimers. This group encompasses the metabotropic receptors for gamma-amino butyric acid (GABA) and glutamate, in addition to receptors involved in the perception of sweet and bitter tastes. While the adhesion receptor family remains poorly understood until these days, the Frizzled/Taste2 receptor family comprises Frizzled and Smoothened receptors, which play important roles in developmental biology.

GPCR-mediated signal transduction is of fundamental importance for most physiological processes, including sensory perception, like vision, smell and taste, but also reproductive, cardiovascular and neurological functions. As GPCRs play important roles in physiological and pathological processes, they provide major targets for therapeutic intervention[11–13].

Current drug discovery efforts are aiming to improve therapies for more than 50 established GPCR targets and to broaden the list of targeted GPCRs[14,15]. GPCR-targeting drugs

represent 33% of presently marketed small-molecule drugs[16], covering common therapeutics like antihistamines, bronchodilators, antihypertensive agents, and many others.

9.1.2 Structure of G protein-coupled receptors

All GPCRs are characterized by two key features[10]:

1. Seven-transmembrane (7-TM) topology with an extracellular N-terminus and an intracellular C-terminus
2. Cytosolic interaction with a G protein.

The seven-transmembrane domains (TM1-TM7) are connected via three extracellular loops (ECLs) and three intracellular loops (ICLs)[17]. Even though all GPCRs exhibit a similar 7-TM topology, the five major families of human GPCRs[10,18] possess different N-terminal domains and share little sequence identity.

Notably, structural variations between different GPCR subfamilies and groups are much more significant than within subfamilies. Crystallography revealed a remarkable heterogeneity in the extracellular loop region, that shows a diverse collection of secondary structures and disulfide crosslinking patterns[19]. Variations in extracellular loops, TM helices and side chains generate an enormous variety of shapes, sizes and electrostatic properties of the ligand binding pockets in different GPCR subfamilies, which reflects the diversity of their associated ligands.

9.1.3 G proteins

Compared with the enormous collection of different GPCR genes encoded in the human genome, only four major G protein families ($G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$, $G\alpha_{12/13}$) have been classified based on sequence homology between at least 21 human Gα-isoforms encoded by 16 genes[20,21].

G proteins form a heterotrimeric complex, consisting of the subunits Gα, Gβ and Gγ. With the condition that guanosinediphosphate (GDP) is bound to Gα, this subunit can associate with the Gβγ dimer, forming an inactive heterotrimer. GPCR activation promotes binding of the GDP-bound heterotrimer, which leads to dissociation of GDP from Gα. This process also represents the rate-limiting step in G protein activation[22]. The high intracellular guanosinetriphosphate (GTP) concentration promotes fast GTP binding to the nucleotide-binding site of the G protein. Finally, the Gα subunit experiences conformational changes followed by the dissociation of Gα and Gβγ subunits. Both subunits act as

activity modulators of various downstream effector proteins. As a result of the intrinsic GTPase activity of the Gα subunit GTP is hydrolyzed to GDP, which terminates the cellular response. Afterwards Gα reassociates with the Gβγ dimer, completing the G protein-activation circle.

In addition to that, interaction of GTPase-activating proteins (GAPs) with activated Gα subunits (GTP-bound) increases their GTP hydrolysis rate up to more than 2000-fold. In this way GAPs represent important modulators of the duration and the intensity of GPCR-mediated signaling (reviewed in [23,24])

9.1.4 Bacteriorhodopsin

The retinal protein bacteriorhodopsin (bR) was discovered almost four decades ago in the plasma membrane of the halophilic archaeon *Halobacterium salinarium*[25,26]. Soon afterwards bR was characterized as a light-driven proton pump[26,27], which converts the energy of "green" light (500-650 nm, max 568 nm) into an electrochemical proton gradient via unidirectional proton translocation out of the cell. This gradient is used subsequently for adenosine triphosphate (ATP) production by ATP synthase.

Due to bacteriorhodopsins fascinating properties, including high stability and its light-driven function, it serves as a model system of various topics in biophysics. BRs structure and function were the focus of much interest and have been characterized extensively using multitudes of different experimental techniques along with the use of all forms of vibrational spectroscopy (reviewed in[28]).

BR gave the very first structural information for any membrane protein. Its structure was initially revealed in 1975 by electron diffraction to 7 Å resolution[29]. These results were obtained from isolated bR crystals, which are naturally occurring in the plasma membrane of *Halobacterium salinarium*. The bR molecule forms trimers, which cluster to hexagonal 2D crystals, which results in a crystalline lattice, called purple membrane (PM). The protein:lipid ratio of the PM is 75:25 in percent. A landmark in further understanding of bRs structure was made 15 years later, with the publication of the first atomic level model[30]. BR consists of seven transmembrane helices (A-G) surrounding a retinal molecule, which is bound covalently to the side chain of Lys216 (helix G) via Schiff base (SB) linkage. Ligand-free apoprotein is termed bacterioOpsin (bO). The proton path is formed by charged residues such as Arg82, Asp85, Asp96, Glu194, Glu204, Asp212 and SB[31]. Due to its

heptahelical structure, bR also serves a model system for other heptahelical membrane proteins, like GPCRs.

Upon light activation, bR and its ligand retinal traverse structural changes, which are described by its photocycle. Considerable effort has been put into further analysis of bR in its resting state, as well as the various structures of its photocycle intermediates, which are summarized in a review[32] by Wickstrand *et al.*. Furthermore, a recent publication of Hasegawa *et al.*[33] gives detailed X-ray based structure analysis of bacteriorhodopsin at a resolution of 1.3 Å.

9.2 *IN VIVO* EXPRESSION OF MEMBRANE PROTEINS

The expression of MPs is not an easy task. In practice it can be very challenging and the expression of functional MPs in preparative scales often needs a lot of patience as seen from the operator's point of view.

The complex topic of MP expression offers lots of possibilities that may or may not lead to the desired result. During the preliminary stages of performing the expression a system for MP production has to be chosen according to the needs of the target protein.

Some questions to keep in mind while choosing the right system: What makes my target protein unique? Where is it from (organism, organ, cell type, type of membrane)? Does the encoding gene have any special features e.g. special codons? Are there special attributes in case of protein structure (primary, secondary, tertiary, quaternary)? Are there any posttranslational modifications (PTMs)?

Under consideration of these questions the operator can choose from a great variety of bacterial or eukaryotic hosts for homo- or heterologous target protein expression, which will be discussed briefly in the following section.

9.2.1 Bacterial expression systems

Among all bacterial expression systems *Escherichia coli* (*E. coli*) is the most frequently used one. *E. coli* requires only simple culture conditions, which are easily scalable. Additionally, *E. coli* is easy to manipulate genetically. Expression with *E. coli* furthermore doesn't depend on a high budget or "fancy" laboratory equipment and can be very quick, compared to other expression systems. The strain *E. coli* BL21 (DE3) combines the advantages of rapid replication, biological safety, inexpensive-cost, time-saving operation

and easy genetic manipulation[34]. *E. coli* C43 (DE3) and C41 (DE3) were especially developed for over-expression of membrane proteins[35,36].

However, heterologous expression of functional MPs in prokaryotic expression systems can be challenging, because of the absence of special lipids, molecular chaperons and posttranslational modifications. This can lead to accumulation of expressed proteins as inclusion bodies, degradation, misfolding and also endotoxin accumulation. Some of these unwanted aspects might improve by codon optimization[37], fusion of tags for solubility or tags to promote protein expression[38,39] or co-expression of posttranslational machineries[40,41]. *E. coli* is a truly powerful expression system, but it definitely has its limits.

9.2.2 Eukaryotic expression systems

9.2.2.1 Yeast cells

This system can combine some of the advantages of both: prokaryotic and eukaryotic expression systems. Analogous to *E. coli* expression systems, it is a non-expensive system with simple media requirements. *Pichia pastoris* (*Pichia*) and *Saccharomyces cerevisiae* (*S. cerevisiae*) have been genetically well characterized and can be used for overexpression of eukaryotic MPs[42,43]. Additionally *Schizosaccharomyces pombe* can be used[44]. It is worth mentioning that eukaryotic protein processing, like complex protein folding, which may involve special chaperones and attachment of posttranslational modifications, is possible with the yeast expression system.

Potential drawbacks of this system include hyperglycosylation of proteins. The system offers both N- and O-linked oligosaccharides on proteins, which differ from the original oligosaccharides on mammalian proteins. The hard and thick cell walls of yeast might lead to difficulties in cell disruption, which can be another potential drawback of the yeast expression system. For high-yield protein expression the system has to be optimized for fermentation. Optimizing growth conditions for fermentation can be time- and chemical-consuming at times.

9.2.2.2 Baculovirus/ Insect cells

Baculovirus vectors are used to transfect insect cells with the genes of the target protein. It is the dominant heterologous expression system for obtaining eukaryotic MPs[45]. This system is most similar to mammalian expression systems, but compared to mammalian viruses baculoviruses require less safety considerations. For MP expression there are two

most popular insect-cell lines: *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Hi5) that can be used both in adherent and suspension cultures.

Most eukaryotic protein-processing methods are possible with the insect-cell system, like protein folding or the attachment of posttranslational modifications. Unfortunately, glycosylation patterns differ from that of mammalian systems analogous to the glycosylation patterns of the yeast system.

Even though this system offers high quality of the expressed proteins in most cases, it also has some non-deniable drawbacks. It should be noted that the cloning procedure for generating recombinant baculoviruses is time-consuming. Mostly it is based on the site-specific transposition of an expression cassette into a baculovirus shuttle vector (bacmid), which is amplified in *E. coli*[46].

Furthermore the insect-cell system needs much more demanding culture conditions, compared to *E. coli* and yeast expression systems. Cultures are also very sensitive to contaminations. The costs for cultural medium can represent a serious hurdle for some laboratories. Additionally expression conditions are difficult to scale up.

9.2.2.3 Mammalian cells

In this expression system viral based vectors as well as plasmid based vectors are used to transfect target genes into cultured mammalian cells. Two broadly used cell lines are Chinese Hamster Ovary cells (CHO) and Human Embryonic Kidney cells (HEK293[47]), which can be both cultivated adherently or in suspension. Target-protein expression with these cells can be achieved by transient or stable transfected cell lines[48].
The outstanding advantages of this expression system are the most authentic attachment of posttranslational modifications, human protein-like structure assembly[49] and folding of expressed proteins.
But where there is light, there is also shadow.

High costs of protein production, because of expensive media and culture conditions, as well as slow cell growth are the main drawbacks of this expression system. Additionally, it is challenging to scale up the conditions. Furthermore, protein expression with this system is very time-consuming and even more time-consuming by the use of stable cell lines (one month at least)[50,51]. Gram per liter yields are only possible in suspension cultures[52].

9.3 *IN VITRO* EXPRESSION OF MEMBRANE PROTEINS VIA CELL-FREE PROTEIN EXPRESSION

Even though a lot of progress has been made in the development of *in vivo* expression systems, there are still a lot of undeniable drawbacks occurring by using these. Therefore, there is still the need for a less complex, less time-consuming and less expensive expression system for challenging proteins like MPs or intrinsically disordered proteins. In addition to that, it would be favorable to have more control over the expression parameters in order to make changes in experimental conditions easily feasible. These requirements could be matched by cell-free expression systems.

9.3.1 Basics of cell-free protein expression

The basic principle of cell-free systems was introduced by Eduard Buchner in 1897[53]. He explored the alcoholic fermentation of sugars into ethanol and carbon dioxide in yeast extracts. This system was not primarily developed to synthesize proteins. It took again more than 60 years until Nirenberg and Matthaei developed a cell-free protein synthesis system (CFPS) based on *E. coli* for the investigation of translational processes[54].

One fundamental necessity for CFPS is a cell extract harboring the highly active components of the cellular transcription/translation machinery. Notably, a cell extract for CFPS can be produced from every living organism on the planet, which creates theoretically endless possibilities for protein synthesis.

Protein synthesis in CFPS systems is initiated by the addition of an appropriate template for gene expression to the reaction. Circular or linear DNA and mRNA can serve as templates[55,56]. Reaction mixtures containing DNA templates are referred to as “coupled reactions”, due to the simultaneous performance of transcription and translation, whereas reaction mixtures supplemented with mRNA are termed “linked reactions”.

9.3.2 Membrane protein production via CFPS - the ultimate solution?

Membrane protein expression can be challenging. We all know that. Is CFPS the ultimate solution?

Despite the fact that an ultimate solution for every occurring protein-related problem does not exist, CFPS is already legitimately close and will come even closer in near future. So let us provide some evidence for this bold statement.

One fundamental concern of MP expression is the toxicity of the produced proteins for the host cells. In order to maintain sufficient and successful MP expression the host's regular internal cell metabolism must be preserved[57–59]. Successful CFPS does not depend on healthy cellular metabolism (the only exception is the cell-extract preparation). By using this technique, all key problems, related to cell toxicity are instantly eliminated, such as aggregation, inclusion-body formation, degradation, overloading of transport mechanisms or toxic effects upon membrane insertion[60]. Notably, *in vivo* protein-synthesis speed rates are achieved by CFPS, while enabling high accuracy of protein synthesis[61,62].

Another clear advantage of CFPS over classical *in vivo* protein expression is the open nature of the system. Due to this characteristic, systems can be easily modified and adapted to the special needs of individual MPs or following applications. In addition to that these systems are - once established - very stable and most of them surprisingly tolerant for a wide range of substances, such as detergents[63], even in high concentrations[60]. Molecular conversion of supplemented substances are minimized as a result of the extensively reduced metabolism in the cell extract[60].

Moreover it should be highlighted that special protein labelling is easily feasible and amazingly efficient in CFPS systems[60]. This includes special labelling via incorporation of non-natural amino acids in general[64–67], as well as isotope labelled amino acids for NMR experiments[68], or incorporation of labels for protein crystallography[69].

NMR often requires special isotope labelling. This can be inefficient and equally costly *in vivo*, due to limited uptake and amino-acid scrambling. Hence attempts to perform dual amino selective labelling techniques for NMR[70] can become troublesome *in vivo*. *In vitro* protein expression simplifies extensive labelling procedures as a result of reduced amino-acid metabolism, absence of bilayer barriers and small reaction volumes. For these reasons, CFPS offers a convenient and economical way for efficient isotope labelling. Furthermore, often no additional purification of expressed proteins is necessary, because heteronuclear single quantum coherence spectra can be recorded directly from the reaction[71].

In addition to that efficient selenomethionine labelling can be accomplished by CFPS, as used by protein crystallographers to allow protein phasing by multi-wavelength dispersion[69]. This kind of labelling often leads to poorly substituted proteins via *in vivo* gene expression, as a consequence of cytotoxicity of selenomethionine[69].

Regarding CFPS with focus on the aspect protein folding, promising options are possible. Reactions can be supplemented with a variety of molecules supporting correct folding, such as chaperones[72,73], nanodiscs[74], lipids[75] or microsomes[76,77]. Moreover, it is also possible to co-express proteins, on the occasion that some MPs require a partner for correct folding and assembly[78].

Last but not least CFPS is fast and straightforward. Problems arising from cell handling, as also any sterility problems are instantly eliminated by using CFPS. Additionally, no considerations about background expression, induction of expression or membrane extraction have to be made. Reactions can be set up in the afternoon and can be harvested after some hours or in the next morning, depending on the type of used CFPS, which will be elucidated in the following paragraphs.

9.3.3 Types of cell-free systems

As mentioned above, a cell extract utilizable for gene expression via CFPS can be produced from every living organism – in principle. As a result many different cell-free protein expression systems evolved since the first findings of Eduard Buchner[53] and Nirenberg and Matthaei[54]. A short overview of existing systems will be briefly presented in the following section.

9.3.3.1 Prokaryotic cell-free expression systems

CFPS systems based on cell extracts from *E. coli* are one of the most common used cell-free expression systems nowadays. These systems will be elucidated later in more detail. But systems based on other prokaryotic organisms, like archaea bacteria have also great potential for cell-free protein expression. Archaeal cell extracts were primarily developed to study how different antibiotics affect the archaeal translational machinery[79].

Exploring the assembly of ribosomal subunits in cell extracts originating from sulfur-dependent thermophilic archaea led to better understanding of the translation mechanism[80]. In 1993, a linked CFPS system was reported[81]. It was based on cell extracts from the extreme thermophilic *Sulfolobus solfataricus* and enabled protein expression at high temperatures[81]. In 2006, Endo *et al.* introduced a coupled CFPS system based on cell extracts derived from the hyperthermophilic *Thermococcus kodakaraensis* [82].

Systems based on archaeal cell extracts can help to express thermostable proteins, which might need high temperatures (up to 80°C) for correct folding. One advantage of using

cell-free systems at high temperatures based on thermophilic organisms resides in the reduction of secondary mRNA structures, which can be a major cause for translational inhibition[60].

9.3.3.2 Eukaryotic cell-free expression systems

Protozoa:

Since the end of the 1950s protozoan cell extracts were used aiming to gain new insights into the translational process[83–86]. These studies investigated the incorporation efficiency of amino acids, depending on energy sources, buffer composition, pH, ion concentrations and supplementation with spermidine or other compounds[83]. Organisms such as *Tetrahymena pyriformis*[84], *Crithidia oncopelti*[83], *Paramecium aurelia*[85] and *Entamoeba histolytica*[86] were used for the required cell-extract preparation. In the more recent past a CFPS system based on *Leishmania tarentolae*[87] was established to study and characterize proteins and their influence on parasitic biogenesis[88].

Parasite protein expression can be ineffective when using conventional *in vivo* systems, because transcription and translation initiation mechanisms differ between organisms. Protozoa-based cell-free systems could be a possible way of expressing these parasite proteins, even though further optimization of these systems concerning efficiency is still needed. Furthermore, it remains unclear until today which types of posttranslational modifications can be realized in protozoan cell-free systems.

Fungi:

Yeast extracts for protein expression based on *S. cerevisiae* were developed in the 1970s[89,90]. This technique was used in the following years to deepen the expertise of eukaryotic translation initiation[91,92]. Subsequently, cap-dependent translation initiation was discovered[93].

CFPS systems based on *S. cerevisiae* can combine the ability of high-yield protein synthesis and the competence to form correctly folded proteins. In addition, yeast extracts enable the attachment of (non-mammalian patterned) posttranslational modifications including glycosylation[94].

With further optimization, yeast-based cell-free systems could become a potent candidate for industrial protein production.

Plants:

There are two more or less popular CFPS systems based on plant extracts, namely the wheat germ system and the tobacco BY-2-system.

After Roberts and Paterson identified an efficient translation mechanism using tobacco mosaic virus mRNA in 1973, the development of wheat germ extraction started[95]. Since that time several optimization procedures were realized, such as the integration of a suitable energy-regeneration system and the adjustment of cation and amino-acid concentrations[96,97]. Albeit these optimizations, the wheat germ extract preparation is more expensive and laborious in comparison to *E. coli*, because of inhibitory effects caused by nucleases and proteases in the endosperm. For the complete removal of these inhibitory enzymes several intense washing steps of wheat germ embryos are needed[98]. During extract preparation the endoplasmic reticulum has to be removed as well. This impedes the attachment of posttranslational modifications and results in limited options of possible modifications, excluding glycosylation. Yet it is noteworthy, that the addition of microsomes or liposomes can mimic the missing compartment. Supplementation of wheat germ extracts with microsomes from dog pancreas was a frequent practice for protein expression using this system[99].

CFPS systems based on wheat germ extract are now well-established and probably the most advanced eukaryotic system nowadays, with applications in protein screening, analysis and engineering.

During the last 20 years another promising plant-based eukaryotic system based on tobacco BY-2 cells was established[100–102]. For extract preparation tobacco cells are only treated four to five hours in contrast to the time-consuming four to five days wheat-germ extract preparation procedure. Unfortunately only a few evaluations of this system exist until now. This is surprising, because by taking a first glimpse at the initial results, such as the synthesis of a functional and active full-size antibody[103], these findings indicate that even complex posttranslational modifications are feasible, including disulfide-bond formation and glycosylation. With further optimization steps and functional analysis after the expression of complex proteins a robust expression system might evolve, that can compete with the wheat germ system.

Insects:

The baculovirus/insect-cell system comprises one of the most efficient *in vivo* options for protein production[104], due to its remarkable ability to carry out a variety of posttranslational modifications. This mostly leads to highly functional produced proteins, which are almost similar to their native counterparts.

With the attempt to get rid of the immanent drawbacks of this system in the early 2000s a cell-free system based on *Spodoptera frugiperda* cell extracts was developed[105,106]. Due to the fact that the ER is not completely removed by the cell-extract preparation procedure, remaining structures reorganize themselves into distinct microsomes. These endogenous microsomes are translocationally active and can provide a first natural lipid membrane for the produced proteins[107,108], which is especially beneficial for the expression and functional folding of membrane proteins.

Another outstanding skill of this CFPS system lies in its ability to realize posttranslational modifications without any further supplementation of the cell extract with enzymes or cofactors. The possible modifications include lipidation[109], glycosylation[106,107], peptide cleavage[110], phosphorylation[105] and even the very important disulfide-bond formation[111,112]. As these modifications are regularly present in eukaryotic proteins, their correct attachment is a crucial factor for proper folding and functionality[113]. Furthermore, some labelling strategies, that can be accomplished by the *E. coli* cell-free system, such as the introduction of non-natural amino acids, are also feasible[114] with the insect-cell CFPS system.

Albeit these promising options, only low protein yields are obtained until now as a result of the rapid exhaustion of energy sources and the accumulation of inhibitory metabolites, like free phosphates[115].

Mammalian:

The attempt to use mammalian cell extracts as the “ultimate” tool for the expression of eukaryotic and particularly human proteins seems natural. This attempt is aiming at being able to produce proteins with native structure and functionality by the application of mammalian-like posttranslational modifications[48]. With these goals a variety of mammalian cell-extract-based CFPS systems have been developed[116,117].

One of the most commonly known mammalian CFPS systems is based on cell extracts of rabbit reticulocytes and has been known since the 1960s[118]. Rabbits were made anemic by

injections with acetylphenylhydrazine prior to the isolation of reticulocytes required for cell-extract preparation[119]. Microsomal structures are missing in rabbit-reticulocyte lysates. As mentioned before these structures can have a crucial function for the synthesis of functionally folded proteins. Therefore, cell extracts have to be supplemented with microsomes required for posttranslational modifications, like glycosylation, lipidation or membrane embedding. This is practically achieved by the addition of heterogeneous canine microsomal membranes to the reticulocyte lysate[120]. Another alternative is the addition of microsomal membranes from other mammalian systems[121,122]. It is noteworthy to say that even with this complex procedures protein yields of the rabbit-reticulocyte system remain comparatively low.

But luckily there are other mammalian extract-based CFPS systems available, for instance a recently developed system based on CHO-cell extract[117,123,124]. CHO cells represent the most commonly used cell line for the production of therapeutic proteins *in vivo*[125]. CHO-cell extract is produced by mild disruption of cultivated CHO cells. It still harbors translationally active microsomal structures from the ER[124]. As a result the CHO-CFPS system supports manifold posttranslational modifications, including glycosylation and co-translational insertion of MPs into biological membranes[123]. Several improvements have been made to overcome key limitations of most eukaryotic CFPS systems leading to low protein yields[126]. With further optimization concerning achieved protein yields CHO-lysate-based CFPS systems have great potential for the productive and profitable production of a wide range of diverse proteins.

Last but not least: CFPS systems based on human cell lines have been developed likewise. Contemporarily, a great variety of cultured human cell lines is available for research or industrial protein production[127]. CFPS systems based on HEK293 cells[128] and HeLa cells[129] are currently used.

The features of human cell-free systems are comparable to those of the CHO-CFPS systems. Endogenous microsomal membranes are available, as well as the ability to realize various posttranslational modifications[116,124]. One significant advantage of human CFPS systems is the natural codon usage. This enables the synthesis of even high-molecular-weight human proteins[130].

But as all eukaryotic CFPS systems suffer from low protein yields, mostly due to difficulties of translation initiation, this drawback is also manifesting in human-based cell-free systems. Aiming to overcome these issues, various strategies have been applied, like supplementation

with enzymes[131–133], usage of factor-depleted cell lines for cell-extract preparation or the alteration of DNA templates by introduction of special sequences enabling better ribosomal binding[126].

9.3.4 *E. coli*-based CFPS

Since Nirenberg and Matthaei[54] established the first CFPS system based on *E. coli* in the 1960s intensive optimization happened to the expression system. Despite the fact that not everything is possible with *E. coli*-based CFPS, it has unconquerable advantages among all other expression systems, which will be elucidated in the following paragraphs[60].

E. coli-based CFPS systems stand out due to their robustness, high reliability and their immense versatility. These systems already have demonstrated effective high-level production of prokaryotic and eukaryotic MPs[60], which makes them interesting for structural and functional studies. Various types of cell-free reaction formats have been designed, from classical batch reactions to more complex dialysis systems. Cell-free reactions in batch formats have a rather simple setup, but are also characterized by short reaction times and limited protein yields[134]. This is primarily caused by the limited life-time of their translational activity, in consequence of the expeditious depletion of the energy supplying phosphate pool. This occurs even in the absence of protein expression[134]. Subsequently, this leads to an accumulation of inorganic phosphates, which increasingly form complexes with magnesium ions and a further inhibition of protein synthesis takes place[135].

This dilemma was overcome first by Spirin and coworkers with the development of a continuous-flow cell-free translation method (CFCF)[136]. This method enables the continuous supply of energy and substrates, while protein-expression inhibiting reaction products are removed continuously as well[136]. An extension of the reaction time to 20 hours leads to a massive increase of protein yield compared to classical batch reactions, which only last a few hours. But even with this colossal improvements in protein yields the elaborate operational setup, necessary for CFCF reactions, is hardly a straightforward technology.

But the technique was later simplified and improved by establishing a semi-continuous or continuous-exchange cell-free (CECF) method[137,138], which is characterized by a passive rather than an active exchange of by-products and substrates resulting in an extended reaction life-time[137,139]. By operating the popular *E. coli* cell-free system in a continuous-exchange mode highest productivity can be obtained, namely milligram amounts of expressed protein in 1 ml of reaction[60,72,136,140].

9.3.5 State-of-the-art in *E. coli*-based high-yield production of membrane proteins

The CECF system harbors all necessary cellular components for effective high-yield protein expression. Continuous supply of substrates and removal of inhibiting reaction products[141], such as inorganics phosphates[135] is achieved by compartmentalization of the cell-free reaction. The key principle of CECF is the use of a reaction mix (RM) and a feeding mix (FM), which are separated by a semipermeable membrane. The volume ratio of RM to FM is normally between 1:10 and 1:30 for best results[60]. This setup maintains an extended supply of fresh precursor molecules into the RM, while inhibitory breakdown products are removed from the reaction and being regenerated. Subsequently, the reaction time is prolonged to 12-24 hours, which is an enormous improvement compared to batch-format cell-free reactions. They only consist of one compartment, namely the RM and protein expression already stops after 0.5–2 h[60].

As valid for all types of cell-free systems, the performance of successful high-yield protein production is dependent on active cell extract (see[142] for a detailed review of cell-extract preparations). *E. coli*-cell extracts are produced from a soluble fraction with a sedimentation coefficient of 30S (S30 fraction: 30×10^{-13}s), that harbors all cellular components of the bacterial transcription/translation machinery. This S30 fraction contains ribosomes, tRNAs, amino acyl-tRNA synthetases, T7RNA polymerase (T7RNAP) and essential translation factors.

Various strains of *E. coli* are applicable for cell-extract preparation, such as A19, D10 or BL21[68,74,140]. All of them offer unique features, which can further improve expression efficiency. The use of RNAse I-deficient *E. coli* strains for cell-extract preparation, like A19 or D10, can promote expression productivity by eliminating RNAse I-mediated mRNA degradation. But even by using cell extracts from these *E. coli* strains for CECF reactions, it is absolute inevitable to use additional strategies, such as RNAse inhibitors, to prevent mRNA degradation.

Furthermore, the use of protease inhibitors is of crucial importance for efficient high-yield protein expression. It must be noted that high levels of amino acid-metabolizing enzymes are present in *E. coli* S30 fractions, because of the cell harvesting in the exponential growth phase. This can affect the life-time and availability of arginine, cysteine, tryptophan, methionine, glutamate and aspartate in particular[134,143], even if the quantity of degradation might vary from

S30 batch to batch. The knowledge about the presence of amino acid-metabolizing enzymes is of crucial importance when performing labelling strategies for proteins of interest, because it can result in 'scrambling' of the label among amino acids. To overcome this effect, various types of amino acid metabolism inhibitors can be added to CECF reactions or special mutant *E. coli* strains can be employed for cell-extract preparation[144]. In the case of NMR labels in proteins biased by scrambling, the analysis of NMR experiments can be adapted accordingly.

Prolonged reaction times, leading to high-yield protein expression by CECF, require the installation of an effective energy-regeneration system, which can rebuild ATP and GTP. This can be effectively achieved by using the pyruvate- and acetate kinase system[60].

Another crucial point regarding efficiency is the generation of reaction conditions, which are as close to natural cellular conditions as possible. One important factor to achieve this goal is the application of polyethylene glycol (PEG) to CECF reactions. The viscosity of *E. coli* cytoplasm can be mimicked effectively by PEG 8000. In addition, PEG is able to stabilize mRNA and promote macromolecular crowding effects in cell extracts[60]. Natural polyamines, like spermidine or putrescine can be employed as substitutes for PEG. They might have positive effects on DNA and RNA stability as well as on efficiency of the translational process[145].

Natural pH conditions are implemented by buffering CECF reactions at pH 7.0 with magnesium- and potassium acetate. The supplied Mg^{2+} and K^{+} ions are also necessary for enzyme activity. CECF reactions have a distinct optimum for the concentration of both ions, which subsequently correlates with the highest protein-expression yields[146].

The use of appropriate DNA templates also influences the performance of CECF reactions. Vector DNA and linear polymerase chain reaction (PCR) products can be used in reactions. It is attractive in some cases to use linear PCR products as templates because elaborate cloning steps can be circumvented[60]. The transcription of these templates is under control of T7RNAP[76], which gives the advantage of strict promoter specificity. Furthermore, T7RNAP possesses high activity on supercoiled and linear DNA templates[60]. As a result potential background expression in CECF reactions is limited to residual *E. coli* RNA polymerase activity, which originates from the S30 extract.

The usual setup for CECF reactions is the precipitate-based approach (P-CF), which works well for all kinds of membrane proteins. This method requires the implementation of

following refolding strategies as expressed proteins precipitate in unfolded states during reaction times. But other multiple setup options are possible for CECF reactions, which are not yet possible or remain ineffective for other expression systems. The approach of detergent-based cell-free expression (D-CF) is unique to the CF technology[60]. *E. coli* CF systems are outstandingly tolerant to various commonly used detergents. Another option is the lipid-based cell-free expression (L-CF). These reactions are performed in presence of defined artificial liposomes, lipid monomers or preformed micelles or bicelles. The use of nanodiscs is also possible and efficient. These molecules form a hydrophobic environment, which can surround the nascent polypeptide chains. Note that the S30 extract still contains a small background of residual *E. coli* lipids, which can potentially interact with freshly synthesized MPs. D-CF and L-CF are not yet possible to this extent with any other expression system[60].

9.3.6 Pushing the limits of *E. coli*-based CFPS

Disulfide bonds are of crucial importance for the natural function and structure of many eukaryotic proteins. Protein-intrinsic disulfide bonds are formed *in vivo* under oxidizing conditions in extra-cytoplasmic compartments, namely the prokaryotic periplasm or the lumen of the endoplasmic reticulum of eukaryotes. Cell-free systems hold two features that potentially prevent disulfide bond-formation. One of them is the reducing agent Dithiothreitol (DTT), which stabilizes the transcription/translation machinery, while the other feature is the absence of compartments with oxidizing redox potential. This hurdle can be overcome by eliminating DTT from the cell extract prior to the translation reaction, which can lead to proper dual disulfide-bond formation[147]. Another approach is the combination of alkylation of the cell extract by iodoacetamide and the addition of a disulfide-bond isomerase to the cell-free reaction. This can affect the formation of multiple disulfide bonds and lead to the production of active proteins[148,149].

Protein folding and refolding can be a tough issue in MP production. Several strategies have proven to be beneficial for folding concomitant to the translational process. The addition of chaperones, such as DnaK, DnaJ, GroEL and GroES to cell-free reactions is one strategy that can lead to success in some cases[148,150,151]. Providing a suitable membrane-mimicking environment is another strategy. The various possibilities of providing suitable environments will be presented in the following paragraphs.

9.4 MEMBRANE-MIMICKING ENVIRONMENTS

As already described, biomembranes are comprised of phospholipid bilayers, that can also include cholesterol and MPs. Prokaryotic membranes do not contain cholesterol and in fungi, algae and plants it is replaced by ergosterol, sitosterol and lanosterol[152]. Hence, there exists a great diversity of biomembranes in different cell types, organisms and organelles. Even between two monolayers and even in different regions of the monolayer the lipid composition may vary[153]. The great variability of biomembranes is also demonstrated by Gram (-) and Gram (+) bacteria. They share an inner phospholipid membrane, but Gram (-) bacteria possess an additional outer membrane containing lipopolysaccharides[154].

The investigation of MPs requires the use of membrane-mimicking environments. They have to be chosen carefully, because membrane mimetics can change the structure and activity of MPs[155–157]. Firstly, they should efficiently mimic the natural environment of the investigated MP to allow MP-environment interactions while maintaining the native structure of the protein. Secondly, mimetics should be suitable for the subsequent experiments.

When using NMR experiments for MP investigation membrane mimetics should be chosen by the type of wanted information and used type of experiment. Additionally, parameters like temperature, pH, experiment duration, hydration and surfactant-to-MP ratio must be taken into account. Small and fast-tumbling protein-surfactant complexes (PSCs) up to 100 kDa are amenable by solution NMR techniques[158–164], while the use of solid-state NMR can enable studies on bigger or slow reorienting PSCs or aggregates[165,166].

Membrane mimetics exist in almost any size and lamellarity and can be self-orienting or mechanically aligned, lipid- or detergent-based or even customizable to experimental demands, such as bicelles. Traditionally, detergent micelles were used for studies on integral MPs[167,168], but nowadays lipid-detergent-bicelles[169,170], amphipathic polymers[171,172] and nanodiscs[173–175] are also proficient mimetics of hydrophobic membrane environments.

The following paragraphs will give a short introduction into the properties and possibilities of each type of membrane mimetic.

9.4.1 Detergents

Detergents are the most frequently used membrane mimetics for biophysical studies, due to their outstanding capabilities in terms of protein solubilization, refolding, reconstitution and crystallization[176].

The formation of micelles occurs at a specific detergent-dependent concentration, namely the critical micelle concentration (CMC). Micelles can be characterized by an aggregation number (N), which can range from four up to several hundreds of molecules[163,168,177]. Therefore, the molecular weight of a micelle is N times the molecular weight of the detergent monomer. Micelles are formed by self-assembly of amphiphilic monomers in aqueous solutions in favor of minimizing the contact of their hydrophobic tails with the surrounding environment.

The micelle morphology strongly differs from that of biomembranes. Note that a detergent micelle is never fully able to mimic a biomembrane, because biomembranes are lipid bilayers[178]. In contrast, micelles consist of small spherical monolayers with a rough surface. When using high detergent concentrations or weakly polar surfactants, micelles can also appear in elliptical or rod-like shapes[177].

Micelle formation takes place at a critical micelle temperature (CMT). Sodium dodecyl sulfate (SDS), for instance, precipitates at temperatures below 4 °C. The incorporation of MPs into micelles and following experiments must be performed above the Krafft point, which is also optically traceable, because turbid detergent solutions become clear at that temperature due to micelle formation.

But even with the phenomenal features of detergents in several aspects of MP investigation there is a hitch somewhere. Unluckily, detergents can reduce or abolish the functionality and dynamics of membrane proteins[163,169,179,180]. Furthermore, they also interfere with functional assays at higher concentrations[181]. The loss of MP activity in detergents might be due to the fact that internal protein dynamics can be reduced and that protein activity correlates with lipid flexibility under native conditions. Even though structural integrity seems to be protected in detergents for many MPs, their functional activity is frequently impaired. This can often be restored in a lipid environment.

9.4.2 Lipids

The mechanical properties of biomembranes are determined by the composition of their lipids, which can also influence protein structure and function substantially[155,182]. As a result, the model membrane's properties, such as curvature, thickness, fluidity and stiffness can modulate the function, folding and equilibrium of membrane proteins[182–184]. The characteristics of lipids, like acyl chain length, charge and degree of unsaturation have a high impact on the propensity to form a bilayer[185,186]. In addition, lipid composition can alter

the orientation of transmembrane helices in a membrane. Studies have proven that the tilt of helices is changed, when the model membrane lipid chain lengths differ from the hydrophobic length of the protein[187,188].

By using lipids as membrane mimetics various options are possible. Lipids in aqueous solution organize as closed bilayers in form of vesicles or liposomes. They can appear in different sizes and lamellarity, dependent on their preparation. Lipids in water form large multilamellar vesicles (MLVs) spontaneously. Even though MLVs are inhomogeneous in composition and can possess up to a dozen bilayers[189], they have the advantage of being very easy to prepare and any type of lipids can be used to mimic a native membrane environment.

Various methods are applicable for producing unilamellar vesicles: Mostly, reverse-phase evaporation of organic solvents, sonication or extrusion through polycarbonate filters are being used for the production[189]. Almost any type of lipid can be used for vesicle preparation. Properties depend on the vesicle diameter, which can range from 10 μm for Giant Unilamellar Vesicles (GUVs) down to 100 nm for Large Unilamellar Vesicles (LUVs), and even 20 nm for Small Unilamellar Vesicles (SUVs).

The smallest possible type of vesicles are SUVs. They are produced by sonication of an aqueous lipid-dispersion. SUVs are characterized by a high curvature, imposing an inhomogeneous distribution of lipids in the outer and inner monolayers, creating a strong pressure on incorporated MPs. Despite the fact, that SUVs maintain the advantages of a real lipid bilayer and are amenable to solution NMR, they have the drawback of being diluted, large and unstable.

Notably, MPs normally do not insert automatically into lipid membranes, even though vesicles have many advantages for the investigation of membrane proteins. Therefore, incorporation of MPs requires an additional reconstitution step[190,191], which can be delicate.

9.4.3 Bicelles and mixed micelles

Bilayered micelles -also called bicelles - have been announced in the 1990s as new membrane mimetics[192,193]. They quickly grew in popularity for their outstanding possibilities to study membrane proteins and peptides. Bicelles are transparent and amenable to optical spectroscopic techniques, such as circular dichroism[194–199]. They are mostly composed of

short- and long-chain phospholipids, which form a planar bilayer. Bicelles can mimic various biomembranes, such as prokaryotic, mitochondrial, neuronal or skin membranes[200].

Bicelles can also be prepared from a mixture of lipids and detergents or other non-lipid surfactants such as bile salt analogues[192,201] or Triton X-100[202]. These binary or even more complex compositions are often referred to as mixed micelles. By using bicelle mimetics for MP investigation, reconstitution protocols for incorporation are mandatory. The mixture of lipids and non-lipid surfactants creates a wide spectrum of generating different membrane environments by combining mimetics.

9.4.4 Organic solvents

Small integral membrane proteins or helical membrane peptides are often soluble in organic solvents with a low viscosity. Mixtures or pure organic solvents, such as methanol, isopropanol, chloroform, tetrafluoroethylene (TFE) or dimethyl sulfoxide (DMSO) have been applied[203–205]. Even though organic solvents are easy and straightforward to use, they are not able to mimic key aspects of biomembranes. The interactions in biomembranes are rooted in complex contacts between hydrophobic and hydrophilic elements, which are lost in organic solvents. As a result MPs will not preserve their native structure. This approach is only amenable to small MPs or peptides and needs careful case by case validation.

9.4.5 Nanodiscs

Nanodiscs - also called nanolipoproteins (NLPs) - are complete, variable and tunable membrane systems for the reconstitution of MPs[175,206–208]. Nanodiscs are composed of a phospholipid bilayer, which is surrounded and stabilized by two copies of amphipathic helical membrane scaffold proteins (MSPs). Lipids and MSPs form disc-shaped objects, which lead to their name “nanodiscs”. Nanodiscs are water-soluble and their thickness of ~4 nm is equivalent to that of biomembranes. In addition, they are very stable and can easily be diluted, concentrated and dialyzed, if a buffer exchange is required. While micelles possess a strong curvature, nanodiscs offer a planar lipid bilayer, which is beneficial compared to micelles[175,208]. It should be pointed out that bicelles are only stable in a small temperature interval, while nanodiscs remain stable on a much wider temperature range (depending on their composition) as a result of the stable MSP/lipid-interaction.

As mentioned before, nanodiscs are extraordinarily tunable and adaptable to requirements of MPs and experiments. Almost every lipid can be used for preparation, as well as many

different types and mutants of MSP. The use of different lipids and MSPs facilitate the adjustment of disc diameter. By choosing the components for the disc preparation wisely a wide spectrum of variability in nanodisc architecture and qualities open up. These qualities make nanodiscs an outstanding membrane environment for the investigation of even the touchiest MPs.

Compared to other membrane mimetics the preparation is more time-consuming, because it requires an *a priori* MSP expression and purification step. Nanodiscs can be prepared in two ways. One way is the sonication of purified MSP with phospholipid vesicles at the gel–liquid crystal phase transition of the lipid. The other way of preparation is by mixing phospholipids (with or without cholesterol) with MSP in a detergent solution, such as cholate or deoxycholate. This step is followed by a slow detergent removal via dialysis or biobeads[206,208–210]. Note that the detergent solution can also function as a first solubilization step before the incorporation of integral MPs into the nanodisc. Nanodiscs are also compatible with cell-free expression systems, which allows the co-translational incorporation of MPs into nanodiscs. Applicable systems are already commercially available[211].

9.4.6 Amphipols

Amphipols (APOLs) are short amphiphilic polymers, which excel through their high chemical stability. They were developed by the group of Popot[171] to provide an alternative for the use of detergents to keep integral MPs water soluble[171]. Now APOLs have been proven to be a useful alternative to the conventional use of detergents for almost 20 years[212–214].

Detergents can be used for various purposes, but their main function is the extraction of MPs from their membrane environment and to maintain solubility by covering hydrophobic parts of membrane proteins. APOLs cannot extract MPs from their membrane environment. This requires a reconstitution step, similar to that from lipids. MPs are transferred into APOLs from a protein-detergent complex (PDC) and in following steps the detergent is removed through usual procedures, such as biobeads or dialysis.

Amphipols have been designed to keep MPs in solution, while disassociation of subunits or cofactors are prevented, as well as the occupation of active sites or the alteration of MP function[171,215,216]. APOLs have been demonstrated to maintain solubility of a very broad range of MPs such as human GPCRs BLT1 and BLT2, bacterial outer MPs OmpA and OmpX, bacteriorhodopsin and even more. Furthermore, amphipols stabilize the MPs biochemically compared to detergent solutions[160,212,217].

APOLs carry a large number of hydrophobic chains, which can replace detergent molecules around MPs. They stand out through a very low rate of spontaneous desorption and through a high affinity for protein transmembrane surfaces. As a result MP/APOL-complexes present a low k_{off} and a small K_D. This means that, even at extreme dilutions, they should not dissociate and if they do, it would happen extremely slowly. In protein-detergent complexes protein-bound detergent molecules are in rapid equilibrium with free micelles and detergent monomers. In addition, the PDC dissociates upon dilution below the detergent's CMC. This is only one of the advantages of APOLs over detergents.

The almost full absence of free APOLs reduces the viscosity and inhomogeneity of the MP-harboring solutions. This is beneficial for all experiments. Concerning the NMR-point of view, protein correlation time is reduced as compared to the detergent option.

APOLs are produced in two major steps. The reaction starts with a flexible, short precursor polymer, namely poly acrylic acid (PAA). Some of its' carboxylates are then first randomly modified with octylamines followed by another modification through isopropylamines[171]. Carboxylates can also be altered aiming at producing zwitterionic, sulfonated or non-ionic versions[212]. Different modifications of the PAA-backbone enable phenomenal diversity and tunability of APOLs, which makes them outstanding membrane mimetics for almost every application. In addition to their stabilizing features APOLs are suitable with many biophysical techniques, including NMR spectroscopy[171].

9.5 OBJECTIVES

The key objectives of this thesis are:

- Establishing a homebuilt *E. coli*-based cell-free expression system enabling fast and reproducible protein expression and labelling strategies.
- Cell-free expression of ligand- and lipid-free bacterioOpsin to enable bias-free insights of the protein's states.
- Cell-free expression of difficult to express (membrane) proteins.
- Characterization of the folding transition of bO to bR using tailored membrane mimetics.
- Enhancing NMR-based insights via comprehensive combinatorial labelling as well as novel isotope labelling strategies.
- Generating new insights into the unfolded state of bO including the interactions with the environment and its consequences for self-assembly into the active state.

10 RESULTS AND DISCUSSION

10.1 SETUP OF CELL-FREE PROTEIN SYNTHESIS

The setup of the continuous-exchange cell-free (CECF) protein synthesis system was mainly conducted following the protocol given by Schwarz *et al.*[60]. Reagents and solutions were prepared as stated in the mentioned publication with some minor changes. The changes include T7RNAP-, RNase-inhibitors and DNA concentrations (see chapter 14.1.3). A volume-ratio of 1:10 between RM and FM was used for all reactions.

As the S30 extract plays a crucial role in the efficiency of the CECF synthesis, meticulous, thoughtful preparation can lead to successful protein synthesis. For preparation of the S30 extract protocols from Schwarz *et al.*[60], Kigawa *et al.*[218] and James Zawada[219] were used. Different types of *E. coli* strains can be used for cell-extract preparation, such as BL21 (DE3), A19 or D10. Bacterial growth is characterized by five different stages (reviewed in Pletnev *et al.*[220]), starting with a lag phase, followed by logarithmic phase, stationary phase, death phase and finally long-term stationary phase. Each phase involves major changes in the composition of the cellular machinery in order to adapt to changing environmental conditions. In logarithmic phases genes are expressed that lead to rapid growth of bacterial cells, including ribosomal genes[221]. Ribosomal synthesis stops by transition to the stationary phase, when stringent control starts and general cell metabolism decreases[220]. As functional ribosomes are the most crucial factor for efficient cell-free protein synthesis, it is of high importance to harvest cells during the logarithmic phase.

The *E. coli* strains A19 and D10 carry relA and spoT mutations, which are involved in the mediation of stringent response[222–224]. As crucial mediators of the stringent response are missing in these strains, the logarithmic phase is elongated, which can theoretically lead to increased cellular ribosome concentration. Additionally, *E. coli* strain A19 is deficient of RNase I, which can lead to increased stability of transcribed mRNA in the following CECF reaction[60]. In order to facilitate the decision, which *E. coli* strain should be used for most efficient cell-extract preparation, growth curves of *E. coli* strains BL21 (DE3) and A19 were compared (Fig. 1). Cells were cultivated as described in chapter 14.1.2.

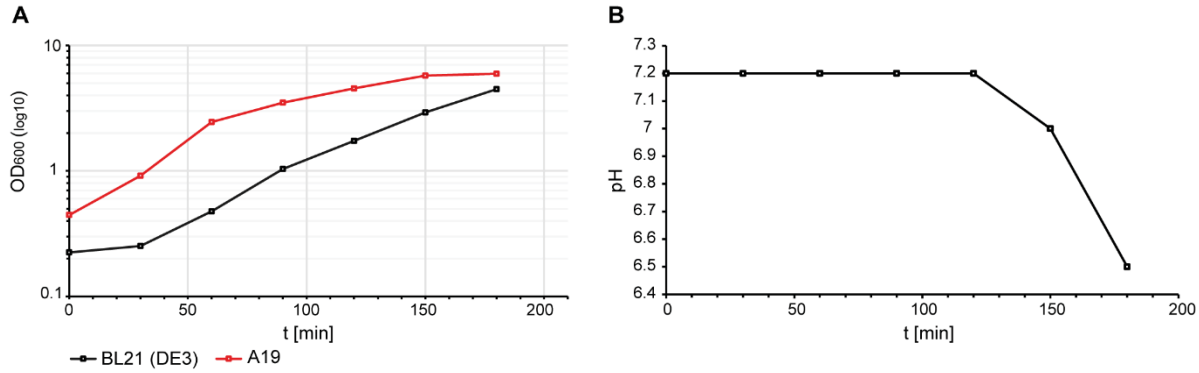


Figure 1: Cultivation of *E. coli* cells. (A) Growth curves of *E. coli* strains BL21 (DE3) (black) and A19 (red). Cells were cultivated in 2YTPG media, suitable for cell-extract preparation, at 37 °C and 180 rpm. (B) PH changes during cultivation of BL21 (DE3) in 2YTPG.

Both cultures of *E. coli* BL21 (DE3) and A19 were harvested after 180 min and used for cell-extract preparation (see section 14.1.2). Afterwards both cell extracts were tested in CECF reactions for a test expression of bacterioOpsin (bO) (see section 14.1.5). In the following steps bO was refolded according to the protocol of Etzkorn *et al.*[225], and the absorbance of the unpurified solution was measured at 555 nm (Fig. 2), suitable to quantify the amount of produced protein.

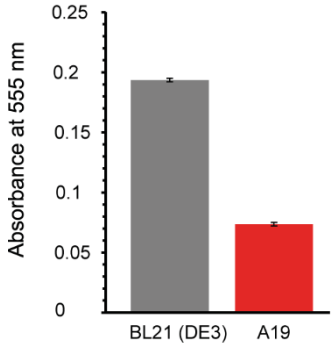


Figure 2: Comparison of the translation efficiency of different cell extracts. Cell extracts of *E. coli* strains BL21 (DE3) (grey) and A19 (red) were used in CECF reactions for the expression of bacterioOpsin (bO). BO was refolded as described by Etzkorn *et al.*, 2013. After refolding, absorbance of the unpurified solution was measured at 555 nm, indicative for functional, refolded bR.

The absorbance at 555 nm of refolded bR, which was produced using the cell extract of *E. coli* BL21 (DE3), is more than twofold higher than that of A19-cell extract expressed bR. Considering the data shown in Fig. 1 (A), A19 cells were already in the stationary phase, when cells were harvested for cell-extract preparation. As a result cellular ribosome-concentration was decreased, which complies with the data shown in Fig. 2, leading to decreased efficiency in CECF reactions.

Taken together, the results shown in Fig. 1 and Fig. 2 suggest that cultivation of *E. coli* A19 cells is more suitable for the use of fermenters. Their high growth rate, caused by their relA and spoT mutations may lead to a rapid decrease of surrounding media pH. As a consequence, the pH stress abolishes the mutational advantages by leading to the stationary phase. This is avoidable with constant adjustment of the pH under fermenting conditions. In addition, BL21 (DE3) cells are more easy to handle under the given conditions, without taking the risk of contaminations by adjusting the pH manually and furthermore delaying bacterial growth by doing so. As a consequence, *E. coli* BL21 (DE3) was used for cell-extract preparations.

In addition to the cell extract, other factors can be optimized in a homebuilt cell-free system to achieve high efficiency in protein synthesis, while maintaining the economic advantages over commercially available systems. The protocol of Schwarz *et al.*[60] for cell-free systems recommends the preparation and expression of T7RNAP. In order to specify the T7RNAP concentration for best protein-synthesis results and for a reliable comparison between commercially available T7RNAP and own preparations, different amounts (1-4 U) in CECF reactions expressing bO were tested (see section 14.1.6). In the following steps bO was refolded and the absorbance of the unpurified solution was measured at 555 nm (Fig. 3).

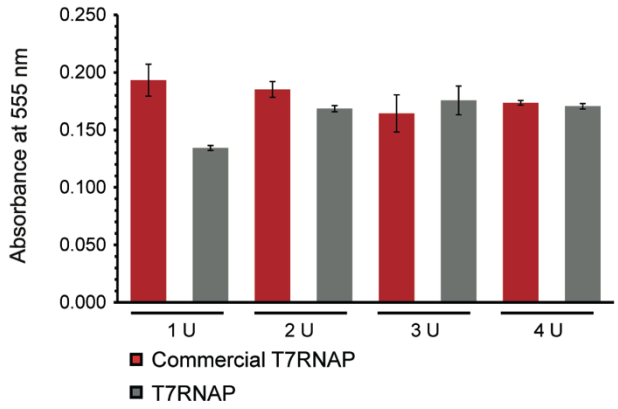


Figure 3: Comparison between different T7RNAPs. Commercial T7RNAP (red) and own preparations of T7RNAP (grey), (kindly provided by Bernd Esters and Reinhild Wurm). Different quantities (1 U-4 U) of T7RNAP were used in CECF reactions expressing bacteriorhodopsin (bR). BR was refolded as described by Etzkorn *et al.*,2013. After refolding, absorbance of the unpurified reactions was measured at 555 nm, indicative for functional, refolded bR.

Fig. 3 shows a comparison between commercially available T7RNAP and own preparations. Refolded bR serves as an indicator of expressed bO in each CECF reaction. By correlating different quantities of T7RNAP (1-4 U), with the absorbance of unpurified, refolded bR,

data can be interpreted in relation to efficiency of protein expression at varying T7RNAP amounts. A difference of 0.06 in absorbance was observed using 1 U T7RNAP. At this concentration the commercial T7RNAP leads to more efficient protein expression than our own preparation. At quantities of 2 U the difference in absorbance between the different T7RNAPs decreases to 0.02. Amounts of 3 or 4 U T7RNAP do not show a clear difference. The protocol by Schwarz *et al.*[60] recommends final concentrations of 6 U T7RNAP, but other experiments[225] showed that such immense quantities are not necessary for efficient cell-free protein expression. Fig. 3 shows no obvious changes in absorbance by increasing the T7RNAP amount from 3 U. As a result constant final concentrations of 3 U T7RNAP were used in all CECF reactions.

Based on the results shown in Fig. 1-3 and meticulous, careful preparation of reagents and CECF reactions, a stable, reliable cell-free expression system was established, facilitating the production and labelling of various kinds of proteins.

10.2 LABELLING OF METHYL GROUPS IN PROTEINS FROM CELL-FREE EXPRESSION

The established CECF system was directly used to implement a new isotope labelling strategy for NMR application.

The following chapter reflects the content of the publication:

Title: Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems

Authors: Abhinav Dubey, Nikolay Stoyanov, Sandeep Chhabra, Thibault Viennet, Shantha Elter, Jan Borggräfe, Aldino Viegas, Radosław P. Nowak , Nikola Burdziev, Ognyan Petrov, Eric S. Fischer, Manuel Etzkorn, Vladimir Gelev and Haribabu Arthanari

Journal: submitted

Abstract

Therapeutically relevant proteins such as GPCRs, antibodies and protein kinases face clear limitations in NMR spectroscopic studies due to the challenges in site-specific isotope labelling and deuteration in eukaryotic expression systems. Here we describe an efficient and affordable method to observe by NMR the methyl groups of leucine residues in proteins expressed in bacterial, eukaryotic or cell-free expression systems without modification of the expression protocol. The method relies on affordable stereo-selective ^{13}C-labelling and deuteration of leucine that alleviates the need for additional deuteration of the protein. The spectroscopic benefits of 'local' deuteration are examined in detail through Forbidden Coherence Transfer (FCT) NMR experiments and simulations. The utility of this labelling method is demonstrated in the cell-free synthesis of bacteriorhodopsin and in the insect-cell expression of the RRM2 domain of human RBM39.

Results and discussion

NMR spectroscopy provides a rich source of information about the structure, dynamics and interactions of proteins, and is especially useful in characterizing weak interactions and subtle conformational equilibria that are increasingly recognized as essential for protein function. The side-chain methyl groups of Ile, Leu and Val form critical hydrophobic contacts and stabilize the structural cores and interaction grooves of proteins. Methyl groups in proteins are often used as indicators of more general behavior, especially in large molecular

weight systems[226,227]. The rapid rotation of the methyl hydrogen around the connected C-C bond introduces favorable relaxation properties leading to improved spectral properties that are less affected by the molecular weight of the system. Additionally, the three-fold symmetry of the methyl group ensures that all three hydrogens contribute to a single intense signal. In combination with the methyl- Transverse Relaxation Optimized Spectroscopy (TROSY) experiment, which exploits the interference between multiple ^{1}H–^{1}H and ^{1}H-^{13}C dipolar interactions to select ^{13}C-$^{1}H_3$ coherences that relax much slower, methyl resonances can be observed even in megadalton-size proteins, and the power of methyl groups has been previously demonstrated in the study of larger therapeutically relevant systems such as biologics[228]. Due to their unique NMR properties[229], the methyl groups of amino acids have arguably been the most successful tool for mapping protein-protein/protein-ligand interactions and for quantifying protein motion on a microsecond to millisecond timescale[230–232]. However, for large protein complexes, observation of the [^{13}C,^{1}H]-methyl signal requires replacement of surrounding protons with deuterons to reduce signal loss due to dipole-dipole relaxation[233]. The most common method for achieving specific labelling at methyl positions is to grow bacteria in perdeuterated medium supplemented with the appropriate $^{13}C^{1}H_3$-methyl-labelled biosynthetic precursors of the amino acids targeted for observation[234]. However, using this method, significant dipole-induced relaxation can also arise from intra-residue proton-proton interactions, e.g. between the geminal methyl groups in Leu or Val[233,235]. Until now the contribution of these interactions to signal relaxation has not been quantified. The introduction of pro-chiral amino acid precursors by Boisbouvier and co-workers[227] eliminated intra-residue dipole-dipole relaxation in bacterially expressed proteins. However, $^{13}CH_3$-methyl labelling of deuterated proteins remains a challenge in eukaryotic cells. The latter cannot utilize isotope-labelled precursors and high concentrations of D_2O, but require pre-synthesized isotope labelled amino acids to grow. *In vitro* protein expression also requires a pre-synthesized set of labelled amino acids, albeit in much smaller quantities.

Therapeutically relevant targets, including most protein kinases, integral membrane proteins (GPCRs, ABC transporters and TCRs), and therapeutic antibodies often cannot be functionally expressed in bacteria. *In vitro* protein expression with ribosomal extracts represents an increasingly viable alternative for obtaining functional proteins; however the [^{13}C,^{1}H,^{2}H]-labelled amino acids required for production of methyl-labelled proteins by this method remain expensive[236–238] and/or require elaborate experimental protocols[237,238] to

achieve methyl labelling. Recently, Lazarova *et al.* used labelled Leu and Val deaminated precursors together with exogenous trans-aminase to achieve stereoselective methyl labelling of cell-free expressed proteins. While the valine precursor, ketobutyrate is quite inexpensive, the authors report the cost of the corresponding leucine precursor as a major obstacle[237]. Furthermore, expressing functional G protein-coupled receptors (GPCRs) *in vitro* remains a challenge[239]. Recently, yeast strains able to produce functional GPCRs, survive deuteration, and process the metabolic precursor from Ile have been developed. Met[240–242] or Ala[243] methyl groups have been used to probe GPCR dynamics, and the NMR signal was boosted by judicious "local deuteration", i.e. addition of deuterated amino acids of the types that occur in the spatial vicinity of the methyl group of interest. Alternatively, chemical modifications such as ^{13}C-methylation of Lys[244,245] and reacting Cys with trifluoroethanethiol[246,247] have been employed to introduce NMR observable methyls in GPCRs, but these result in non-natural probes. Despite progress, there remains an unmet need for simple and broadly applicable labelling of protein methyl groups, e.g. through the use of affordable stereo-specifically labelled amino acids[241,243].

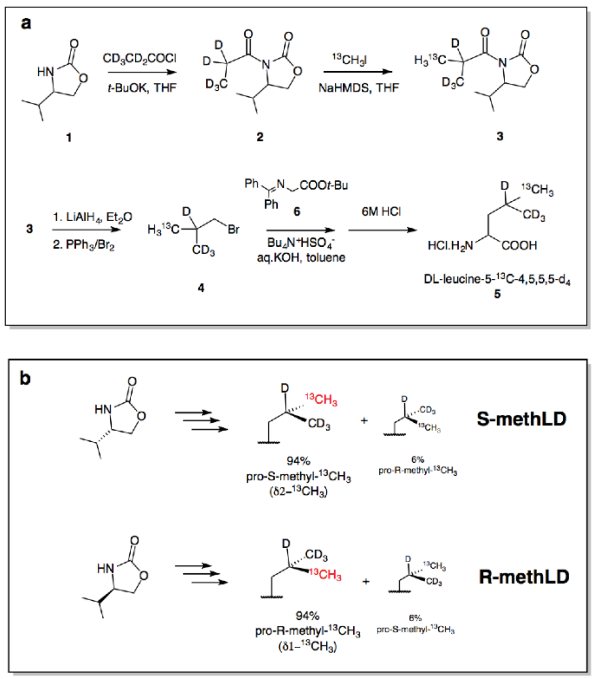


Figure 4: Synthesis of $^{13}CH_3$-methyl leucine. a) Simple and inexpensive synthesis of $^{13}CH_3$-methyl leucine that is "locally" deuterated ($meth_{LD}$, 5) to enhance the NMR spectra of non-deuterated proteins; b) The pro-R-$^{13}CH_3$ or pro-S-$^{13}CH_3$ methyl configuration of 5 is obtained with 94% stereoselectivity starting from the corresponding Evans' chiral auxilliary 1.

The stereo-selective isotope enrichment of branched amino acids is amply documented in the chemistry literature[248–255], and that work has been largely driven by the need for new NMR probes for protein structural studies. An impressive set of [^{13}C,^{15}N,^{2}H]-labelled amino acids optimally designed for automated NMR structure determination was developed by Kainosho and co-workers more than 10 years ago[236,256]. A simplified version of Leu specifically designed for methyl labelling of perdeuterated proteins was later reported by the same group[257]; however due to the complex chemistry and expensive starting materials, these amino acids remain prohibitively expensive for applications other than cell-free protein expression. In the light of this, we set out to produce a more affordable version of Leu, containing the essential features needed for observing methyl resonances in large proteins.

Here we report a straight-forward synthesis of Leu in which the desired methyl group is [^{13}C,^{1}H]-labelled while some of the proximal hydrogens are deuterium-enriched (Fig. 4). The 'locally' deuterated, methyl ^{13}C-labelled leucine (5, Fig. 4) is hereafter referenced as Leu-meth$_{LD}$. We present a comprehensive and comparative analysis of the benefits of this labelling configuration by studying the effects of intra- and inter-residue proton-proton interactions on transverse relaxation. Our data demonstrate that the isotope pattern of Leu-meth$_{LD}$ can significantly increase the NMR spectral resolution and sensitivity even in an otherwise protonated background. We successfully test incorporation of Leu-meth$_{LD}$ into the 42-kDa maltose binding protein, the 12-kDa human RBM39245-332 encapsulating RRM2 domain, and the 25-kDa membrane protein bacteriorhodopsin using *E. coli*, insect cell and cell-free protein expression, respectively, without modification of the original expression protocols. This method of labelling is flexible and affordable and promises to expand the range of isotopically labelled proteins available to NMR-based structural studies.

The synthesis shown in Figure 1a is a simplified version of the method reported by Siebum *et al*.[258] that was originally designed to produce all possible amino acid isotopomers of leucine. Briefly, deuterated propionic acid was converted to the acyl chloride and used to acylate Evans' isopropyl oxazolidinone (1). Stereo-selective methylation with less than two-fold excess $^{13}CH_3$-methyl iodide proceeded with an enantiomeric excess of 94%. Reductive cleavage with $LiAlH_4$ (or $LiAlD_4$ for deuteration of the β positions), followed by treatment with PPh3/Br2 yielded the stereoselective methyl-labelled isobutyl bromide (4). Chiral alkylation of the activated glycine (6) in the presence of chiral quaternary ammonium salts yielded the labelled L-Leu. However, the latter reaction proved to be quite sluggish unless stoichiometric quantities of expensive Maruoka catalysts[259] were used. To save cost,

we opted for the racemic DL-Leu which retains the stereochemistry of the labelled methyl groups. This was obtained by classic biphasic alkylation in the presence of a standard quaternary ammonium salt (Fig. 4a). Final deprotection with hydrochloric acid gave the HCl salt of 5-^{13}C-5,5,5,4-d_4-DL-Leu (5) containing 94% of the desired $^{13}C^{1}H_3$-methyl leucine and 6% of the methyl-inverted labelling configuration (Fig. 4b). The stereoisomer mixture of Leu-meth$_{LD}$ with predominantly pro-R-isomer is hereafter called as R-meth$_{LD}$. The overall yield was 24% from d5-propionic acid and 18% from $^{13}CH_3$-methyl iodide. The cost of reagents per milligram of Leu (5) was less than €1. By comparison, a labelled ketoacid used to generate an analogous Leu labelling pattern in cell-free expression was reportedly obtained for €20/mg[237]. A summary of the cost of Leu-meth$_{LD}$ (5) used to label the proteins in this study is given in the supplementary information (see chapter 15.2.1).

Using theoretical calculations, we analyzed the relative contributions of intra-residue versus inter-residue dipole-dipole interactions to transverse relaxation. The major source of transverse relaxation in methyl groups is due to ^{1}H-^{1}H dipole interactions. We compared the contribution of intra- and inter-residue ^{1}H to dipole-dipole relaxation of Leu-methyl proton atoms in the 42-kDa protein MBP. We used the atom coordinates from a previously deposited NMR structure (PDB ID: 1EZO) and calculated all inter-proton distances between leucine methyl protons and the rest of the protons in MBP. The dipole-dipole contribution was computed using these distances averaged over all the conformations in 1EZO. The details of the computation are described in the supplementary information. Interestingly in MBP, deuterating four of the closest intra-residue protons, namely Hδ21, Hδ22, Hδ23 and Hγ, translates to a 37% reduction in dipole-dipole interactions (see supplement Fig. 40) leading to narrower linewidths. This is comparable to the dipole-dipole relaxation caused by all the remaining inter-residue protons (~38%), indicating that proximal protons have a more prominent effect on relaxation than numerous more distant protons. This suggests that local deuteration, including stereo-specific methyl deuteration could sufficiently reduce the rate of transverse relaxation and improve resolution and sensitivity in the absence of complete protein deuteration.

To experimentally validate the simulated findings and quantify the effect of local deuteration, we used forbidden coherence transfer (FCT) experiments[260]. FCT experiments use the ratio of signal intensities between the multiple quantum and single quantum coherences of the methyl proton, measured as a function of evolution time. This ratio is plotted as a build-up curve in the form of a hyperbolic tangent, where the slope reports on local dynamics

(order parameter S2) and the plateau reports on the local ^{1}H density that contributes to relaxation (see supplement Fig. 42). Since we directly compared the same residues within the same protein (MBP, with different labelling schemes), the contribution from dynamics should be identical leaving the level of deuteration as the only variable. A lower plateau value indicates a higher ^{1}H density in the vicinity of the methyl proton[260]. As examples, we highlight Leu7 and Leu121 in MBP that was (i) δ2-[^{13}C,^{1}H]-methyl-labelled and fully protonated, and (ii) δ1-[^{13}C,^{1}H]-methyl, δ2-[^{12}C,^{2}H]-methyl, γ-[^{2}H]-labelled in an otherwise fully protonated background. The samples were produced in *E. coli* with the addition of (i) ^{13}C-acetolactate[227] or (ii) R-meth$_{LD}$ (Fig. 4b) to the aqueous M9 minimal medium 1 h before induction (see supplement Fig. 43). The geminal methyl group makes a significant contribution to the transverse relaxation rate (Fig. 5), and thus a gain in resolution, upon its deuteration (Fig. 6). The ratio of values to which the two FCT curves of each residue with and without local deuteration in Fig. 5b plateaus, gives us an estimate of the linewidth reduction (~50%) due to the reduced dipole-dipole interaction. We present additional FCT curves for other Leu residues in MBP in supplementary Figure 41.

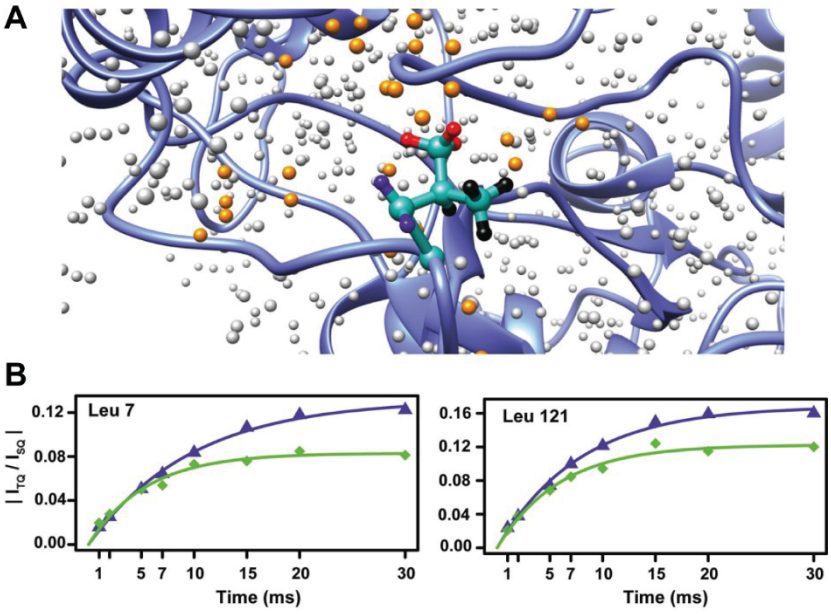


Figure 5: Local deuteration reduces the transverse relaxation rate of Leu methyl groups. (a) Enlarged view of the NMR structure of maltose binding protein (MBP, PDB:1EZO) around Leu121. Leu121 is shown as a ball and stick model with deuterated protons in black, Hβ2 and Hβ3 in blue, and the methyl protons being studied in red. The orange spheres are the inter-residue protons within 6 Å of the Cδ1 atom of Leu121, and the grey spheres are the rest of the protons in MBP. (b) Experimentally determined FCT curves of Leu7 δ1[$^{13}CH_3$] and Leu121 δ1[$^{13}CH_3$] in natural hydrogen abundance MBP (green curve) and MBP with δ1[$^{13}CH_3$] that is "locally" δ2 and γ-deuterated with Leu-meth$_{LD}$ (blue curve). The lower the value to which the FCT curve converges, the greater the proton density around it, which causes faster transverse relaxation.

These results suggested that "local deuteration" of Leu will reduce transverse relaxation sufficiently to produce usable NMR spectra from large, otherwise non-deuterated protein complexes. To further validate the use of Leu-meth$_{LD}$ we expressed the seven-transmembrane-helix protein bacteriorhodopsin (bR) in an *E. coli*-based cell-free system. Uniformly ^{13}C-labelled non-deuterated L-Leu (8.5 mg) or R-meth$_{LD}$ (Fig. 4b) (17 mg) was used in dialysis mode, 3-ml reactions. No difference in expression yield was observed between the samples. BR was refolded and purified in n-dodecyl-β-D-maltoside (DDM) micelles, leading to a final protein-detergent complex of approximately 100 kDa[225]. Heteronuclear multiple quantum correlation (HMQC) spectra were acquired on both samples (see Fig. 6a black vs red). The $^{1}J_{CC}$ couplings in the [U-^{13}C,^{1}H]-Leu sample were absent in the Leu-meth$_{LD}$ sample, obviating the ^{13}C constant-time evolution component of the pulse sequence[261]. A projection of the 2D HMQC (along ^{1}H dimensions, Fig. 6c, e) showed an ~30% increase in resolution and 5.5-fold increase in intensity. An approximately five-fold increase in intensity was also measured, when comparing the methyl-labelled Leu signal to the individual components of the doublet in the uniformly ^{13}C-labelled sample. The absence of ^{13}C-^{13}C ^{1}J-coupling (Fig. 6d) and the reduced ^{1}H-^{1}H dipolar couplings in the Leu-meth$_{LD}$ (5) sample significantly contributed to the enhanced spectral features. Overall, use of Leu-meth$_{LD}$ in the otherwise natural abundance expression system improved the quality of methyl HMQC spectra sufficiently to allow NMR studies of binding interactions[237,256], dynamic and allosteric mechanisms[238] of large membrane protein complexes.

Insect-cell culture is frequently the method of choice for the production of proteins that have stringent requirements for proper folding and posttranslational modifications. However, higher eukaryotic cells are not viable in $^{2}H_2O$ concentrations higher than 30%. Still up to 75% protein deuteration has been achieved by the addition of deuterated amino acids to the growth medium[242,262,263]. Local deuteration surrounding the desired [^{13}C,^{1}H]-methyl signals was found to improve spectral quality significantly in studies of the μ-opioid and P2X4 purinergic receptors[242,262]. We expressed the 12-kDa human RBM39245-332 encapsulating RRM2 domain in Sf-9 cells in medium prepared simply by replacing uniformly labelled Leu with 75 mg/L of R-meth$_{LD}$.

The Heteronuclear Single Quantum Correlation (HSQC) spectrum (Fig. 6f) displayed the expected eight Leu pro-R (δ1) methyl peaks. The quality of the spectrum was sufficient to detect the much weaker signals arising from the residual 6% pro-S (δ2) $^{13}C^{1}H_3$-methyl groups.

In line with the improvements observed in the spectrum of bR, Leu-meth$_{LD}$ (5) human RBM39245-332 (Fig. 6h-j, red) also exhibited a roughly six-fold improvement in the signal intensity ratio and a two-fold improvement in resolution in the proton dimension, as compared to that of the uniformly ^{13}C-labelled, Leu-labelled human RBM39245-332 (Fig. 6h). Due to the absence of $^{1}J_{CC}$ couplings, the spectrum obtained from the R-meth$_{LD}$ sample (Fig. 6i, red) shows no splitting in contrast to the doublet visible in the [U-^{13}C,^{1}H]-Leu-labelled human RBM39245-332 sample (Fig. 6i, black). Note that, as with bR, this absence renders the use of constant time evolution[261] in the ^{13}C dimension unnecessary and results in increased sensitivity, especially for large molecular weight proteins. It further facilitates acquisition of high-resolution spectra by extended sampling in the ^{13}C dimension using non-uniform sampling.

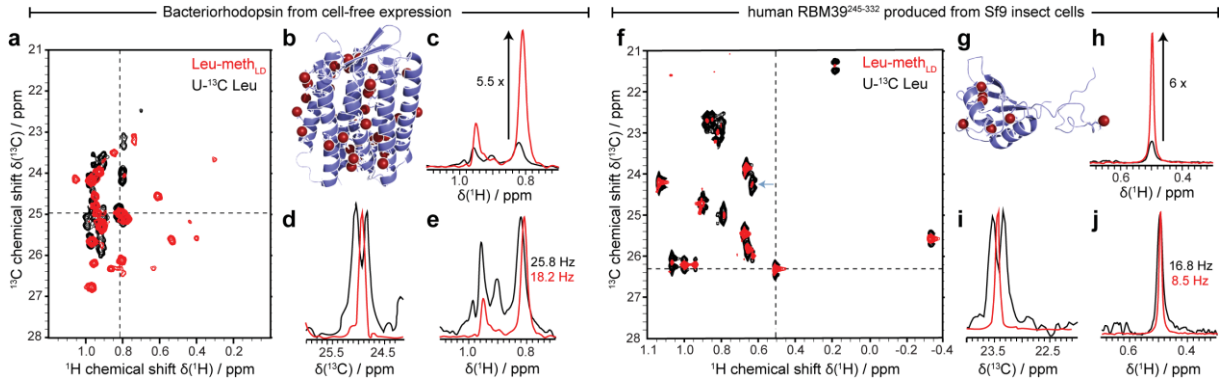


Figure 6: Stereo-specific CD$_3$/^{13}CH$_3$-labelled Leu with Leu-meth$_{LD}$ facilitates NMR characterization of proteins expressed *in-vitro* and in insect cells. (a) [^{13}C,^{1}H]-HMQC spectra of cell-free expressed bR in DDM micelles with [U-^{13}C,^{1}H]-Leu (black) or Leu-meth$_{LD}$ (red). (b) Structure of bacteriorhodopsin (bR, PDB ID:1R84) with [^{13}C, ^{1}H]-labelled Leu methyl groups highlighted as red spheres. One-dimensional slices at indicated frequency in the ^{1}H/^{13}C dimension are shown in (c)-(e). (c) Signal intensity is increased by 5.5 folds. The comparison of normalized data in the ^{1}H/^{13}C shows that a singlet is obtained instead of doublet facilitating high-resolution acquisition in the ^{13}C dimension (d) and that linewidth at half-height is reduced by 29% in the ^{1}H dimension (e). (f) [^{13}C,^{1}H]-HSQC spectra of the human RBM39$^{245-332}$ encapsulating RRM2 domain expressed in insect cells with [U-^{13}C, ^{1}H]-Leu (black) or Leu-meth$_{LD}$ (red). (g) Structure of the CAPER- RRM2 domain (PDB ID: 2JRS) with [^{13}C, ^{1}H]-labelled Leu methyl groups highlighted as red spheres. Some of the residual 6% of inverted methyl-labelled leucines are also observable (an example marked by a blue arrow). Analog as for bR (c-e), one-dimensional slices at indicated frequency in the ^{1}H/^{13}C dimension are shown in (h)-(j). (h) Signal intensity is increased by six folds. (j) Linewidth at half height is reduced by 49% in the ^{1}H dimension.

Further improvements in the protein spectrum would be expected if Leu-meth$_{LD}$ is additionally deuterated at the beta protons, e.g. by using $LiAlD_4$ (an additional 20-40% increase in the cost of reagents) in step 3 (Fig. 4). We estimated that 70% of the Leu residues

in MBP have a 50% or more contribution from intra-residue Leu protons towards the dipole-dipole relaxation of their intra-residue methyl protons (see supplement Fig. 40). Additional intra-residue deuteration at the beta and alpha protons would further contribute to narrowing linewidths and increasing sensitivity. One current bottleneck in leveraging the power of methyl resonances is in obtaining the resonance assignments, particularly in high molecular weight systems. One way of assigning methyl resonances is by making targeted mutants. However, recent work has shown that, with a known structure, the methyl resonances can be assigned in automated manner from NOESY data, which suffers less from relaxation losses than traditional side chain experiments[264,265]. These new assignment methods dramatically increase the utility of methyl-NMR to characterize structure and dynamics of large protein systems. Interestingly, the fixed local deuteration pattern of Leu-meth$_{LD}$ (5) in combination with its FCT analysis of surrounding proton density, may also be exploited as sensor of structure and dynamics of the surrounding proton network.

Conclusions

The method reported here provides a relatively inexpensive route to produce stereospecific methyl labelling of Leu residues, which is directly applicable for most commonly used protein production systems, including insect cells and *in vitro* expression without a need for modification of the existing expression protocols. Our simulations and experiments further clarify the effects of local deuteration on transverse relaxation rates supporting the notion that the approach expands the range of pharmacologically interesting proteins that are amenable to NMR characterization including proteins expressed in eukaryotic systems. The frequency of Leu in the eukaryotic proteome is around 9%, making it one of the most abundant amino acids and providing multiple NMR probe sites while maintaining a less crowded spectrum. Analogous site-selective [^{13}C,^{1}H/^{2}H]-enrichment of Val and Ile should produce similar gains in the quality of protein NMR spectra, including characterizations via solid-state NMR[266]. Simplified protein spectra and improved resolution enable better use of non-uniform data sampling[267], which will lead to further gains in sensitivity for megadalton protein complexes.

10.3 STRUCTURAL CHARACTERIZATION OF BACTERIORHODOPSIN FOLDING (BO/BR)

Where do we start if we want to investigate the process of protein folding for our favored protein? The investigation of protein folding needs precise characterization of different conformational states during the folding process. Detailed analysis of basic conformational states, i.e. folded and unfolded states, are the fundament to also understand the transitions between those states and to complete the picture of the folding process.

10.3.1 Folded state of bacteriorhodopsin (bR)

Even though the folded state of bR has been studied extensively[32,268,269], its folding process has not been fully understood at the present day. NMR offers a powerful tool to investigate different conformational states. The presented results in the following sections provide useful information about the structural characterization of bR by NMR techniques and AFM Force Spectroscopy. These findings are not only relevant in terms of creating a deeper knowledge of bR folding and structure, but also in terms of investigating folding and structure of 7-TM proteins in general[214,225]. Insights gained based on bR results are offering new perspectives for the investigation of membrane proteins with regard to the selection of membrane mimicking environment and accessibility of complexes by solution NMR techniques.

10.3.1.1 The use of amphipols for NMR structural characterization of 7-TM proteins

The established CECF system facilitated the efficient production of bO and the GPCRs MC4R and MC4R. Amphipols were used as non-conventional membrane mimetics for refolding and subsequent NMR characterization of the proteins.

The following chapter reflects the content of the publication:

Title: The use of amphipols for NMR structural characterization of 7-TM proteins

Authors: Shantha Elter, Thomas Raschle, Sabine Arens, Aldino Viegas, Vladimir Gelev, Manuel Etzkorn, Gerhard Wagner

Journal: The Journal of Membrane Biology (2014)

Keywords: Amphipathic polymers, solution-state NMR, 7-TM proteins, bacteriorhodopsin, melanocortin receptor

Abstract

While amphipols have been proven useful for refolding of seven-transmembrane helical (7-TM) proteins including G protein-coupled receptors (GPCRs) and it could be shown that an amphipol environment is in principle suitable for NMR structural studies of the embedded protein, high-resolution NMR insights into amphipol refolded and isotopically labelled GPCRs are still very limited. Here we report on recent progress towards NMR structural studies of the melanocortin-2 and -4 receptor, two class A GPCRs which so far have not been reported to be incorporated into an amphipol environment. Making use of the established 7-TM protein bacteriorhodopsin (bR) we initially tested and optimized amphipol refolding conditions. Most promising conditions were transferred to the refolding of the two melanocortin receptors. Analytical scale refolding experiments on the melanocortin-2 receptor show very similar behavior to results obtained on bR. Using cell-free protein expression we could generate sufficient amounts of isotopically labelled bacteriorhodopsin as well as melanocortin-2 and -4 receptors for an initial NMR analysis. Upscaling of the amphipol refolding protocol to protein amounts needed for NMR structural studies was, however, not straight forward and impeded detailed NMR insights for the two GPCRs. While well resolved and dispersed NMR spectra could only be obtained for bacteriorhodopsin, a comparison of NMR data recorded on the melanocortin-4 receptor in SDS and in an amphipol environment indicates that amphipol refolding induces larger structural modifications in the receptor.

Introduction

Amphipathic polymers have demonstrated great potential as suitable membrane substitutes[270]. Among the reasons that favor amphipols over conventional detergent based surfactants are their ability to increase stability of the embedded membrane protein in certain cases as well as their good refolding properties in particular for seven-transmembrane helical (7-TM) proteins[271,272]. These properties make amphipols very promising for the investigation of 7-TM proteins including G protein-coupled receptors (GPCRs)[273,274]. In addition, it could be demonstrated that amphipol stabilized membrane proteins including β-barrel[275] and 7-TM proteins[225] are in general accessible by solution-state NMR techniques[276].

In regard to GPCRs, amphipols were used to determine the structure of leukotriene B4 (LTB4), a small molecule ligand, bound to an amphipol refolded and stabilized BLT2 receptor[277]. Ligand binding capabilities of amphipol refolded GPCRs could be

demonstrated and high-resolution insights of the LTB4 ligand while bound to perdeuterated BLT2 receptor could be obtained using homonuclear ^{1}H spectroscopy of the ligand. However, NMR spectra of the respective GPCR in amphipols still suffer from limited sensitivity, resolution and dispersion[277].

Here we report on initial results obtained using amphipols for the structural investigation of the melanocortin-2 and the melanocortin-4 receptors (MC2R and MC4R respectively). The signaling pathways of the MC4R are of great pharmaceutical relevance due to their role in the control of body weight and appetite, regulation of blood pressure and the inhibition of inflammation[278,279]. The native melanocortin agonists comprise the adrenocorticotropic hormone (ACTH) as well as the α, β and γ-melanocyte-stimulating hormones (MSHs). Current efforts have mainly focused on ligand binding studies, see e.g.[280,281] for recent reviews. However, no experimental high-resolution structural information is available for these two receptors so far.

In the following we provide a biophysical characterization of the MC2R and MC4R receptors in an amphipol environment. These initial results will guide our efforts towards the characterization of structural details underlying hormone-receptor communication. NMR sample preparation including refolding strategies was optimized using the well-established 7-TM protein bacteriorhodopsin (bR).

Materials and Methods

Cell-free protein expression

The bacterioOpsin and melanocortin receptors were expressed using an *E. coli*-based cell-free expression system following established procedures[60,74]. The wt-MC2R and wt-MC4R sequence was cloned into a pIVEX2.4d expression vector containing an N-terminal His10-tag followed by a Factor Xa cleavage site. Dialysis mode reactions were carried out at 28°C in the absence of ligands and surfactants. After 12-16 h the reaction mix was centrifuged for 10 min at 12000 x g. The resulting pellet was stored at -20°C or directly refolded. Pellets for NMR sample preparation were additionally washed with buffer (10 mM Tris-acetate (pH 8.2), 14 mM Mg^{2+} acetate, 0.6 mM K^{+} acetate). Note that residual Mg^{2+} could lead to aggregation of amphipols and the addition of EDTA prior to refolding could be beneficial[282].

Amphipol refolding

Amphipols were synthesized following published procedures for A8-35[283,284]. Note that NMR spectra of the amphipol batch used in this study show an increase in the ratio of free carboxyl to isopropylamine as compared to A8-35. Based on our NMR data the amphipol side-chain composition of the batch was determined to be 57% free carboxyls, 12% isopropylamine and 31% octylamine. Refolding into amphipols was done by initially resuspending the protein pellets in SDS-buffer (50 mM sodium phosphate (pH 7.5), 20 mM SDS). For bR refolding no additional purification step was carried out prior refolding with amphipols, for the GPCRs cell-free protein pellet was purified using IMAC in SDS buffer before refolding. For bR we could not observe differences in the refolding yield when refolding was carried out before or after IMAC purification. To induce folding four different strategies for SDS removal were tested:

KCl precipitation: Amphipols (to 2.2 % w/v) and (unless otherwise stated) the ligand (5 fold molar excess) were added, the mixture was kept at room temperature for 15–30 min, SDS was precipitated by the addition of KCl to a final concentration of 150 mM and kept at room temperature with occasional shaking for additional 1-2 h. Residual SDS was not removed using a dialysis step, instead the refolded protein was directly purified using a Ni-NTA agarose column. Buffer A (50 mM sodium phosphate buffer (pH 8), 150 mM NaCl) supplemented with 0.08 % w/v amphipols and 20 mM Imidazole was used for washing (3 steps of two column volumes), buffer A supplemented with 0.15 % amphipols and 250 mM Imidazole was used for elution (5 steps of one column volume). Protein containing fractions were pooled and centrifuged for 5 min at 12000 x g. 100 µl of the supernatant was directly analyzed by analytical gel filtration using a Superdex 200 10/300 gl (GE) column equilibrated in amphipol-free buffer A using a flow rate of 0.5 ml/min at 4°C.

Dilution: Amphipols (to 2.2 % w/v) and the ligand (5 fold molar excess) were added; the mixture was diluted 1:10 by fast addition of SDS-free buffer A and kept at room temperature for 1-2 h. Ni-NTA purification and gel filtration were carried out as described above.

Bio-Bead preparations: Amphipols (to 2.2 % w/v) and the ligand (5-10 fold access) were added. Washed and buffer equilibrated Bio-Beads (Bio-Beads SM-2; Biorad) were added (up to 80% w/v) and kept at room temperature for 8-14 h under constant shaking. Bio-Beads

were removed using a centrifugation step. Ni-NTA purification and gel filtration were done as described above.

Refolding on Ni-NTA agarose matrix: SDS solubilized protein was loaded onto Ni-NTA beads prior to addition of amphipols or ligands. After protein was bound to the Ni-NTA matrix (15-30 min at room temperature) refolding by SDS removal using KCl precipitation, dilution or Bio-Beads was carried out as described above.

No residual SDS was detected in the bR-amphipol NMR samples by 1D and 2D ^{1}H correlation spectroscopy after refolding into amphipols followed by immobilized metal affinity chromatography (IMAC) and SEC (data not shown). The ligands used in this study comprise *all-trans* retinal (Sigma-Aldrich) in 10 mM stock solution in ethanol for bR refolding as well as versions of the adrenocorticotropic hormone (ACTH 1-39) for MC2R and MC4R refolding. For MC4R also the high affinity antagonist SHU 9119 (Tocris Bioscience) was used.

Small-scale refolding was typically carried out in 50–150 µl sample volumes while large-scale refolding was done in sample volumes between 1–6 ml. Protein concentrations during refolding and SEC were in the order of 10–20 µM. Purification of NMR samples as well as all GPCR preparations were carried out with protease inhibitor supplemented buffer (Complete EDTA free, Roche).

NMR sample preparation and experimental setup

Isotope-labelled proteins were produced by using double (^{2}H and ^{15}N) or triple (^{2}H, ^{15}N, ^{13}C) labelled ALGAL isotope mixtures (Cambridge Isotope Laboratories or Sigma Isotec). The ALGAL isotope mixtures contain all amino acids (in different concentrations) except for the four amino acids Cys, Trp, Gln, and Asn. The missing amino acids were added in natural abundance. All NMR samples were expressed under >90 % D_2O conditions. Typical bR concentrations were in the range of 100-300 µM, the melanocortin receptors NMR samples had protein concentrations between 40-80 µM. Measurements were carried out at 25°C (melanocortin receptors) and 38°C (bR), at proton resonance frequencies of 750 or 800 MHz Duration of 2D experiments was in the order of 4 – 12 h for bR and 24-48 h for the melanocortin receptors.

Results and Discussion

Amphipol refolding of bacteriorhodopsin (bR) and the two melanocortin receptors MC2R and MC4R was studied using cell-free protein expression, size exclusion chromatography (SEC) and nuclear magnetic resonance (NMR) spectroscopy.

All proteins used in this study were produced using an *E. coli* extract based cell-free expression system in the absence of any membrane mimicking surfactants. Figure 1a shows SDS-PAGE results of cell-free (CF) expressed bacterioOpsin (bO). As reported before, the resulting CF-pellet predominantly contains the expressed protein[285,286]. Still Ni-NTA purification in SDS buffer clearly improves sample purity (Fig. 7a).

Protein refolding into amphipathic polymers

We investigated refolding into amphipols in particular with regard to a subsequent solution-state NMR spectroscopic study, which, similar to crystallization attempts[287] should preferentially be carried out with a sample consisting of small and homogeneous particles. Therefore the refolding product was assessed using size exclusion chromatography to optimize refolding with respect to homogeneity and particle size. To simplify efficient removal of the SDS, which is used to solubilize the cell-free expressed protein pellet, we reduced the SDS concentration from 0.8 % (w/v) as reported in[271] to 0.6 % (w/v).

Noteworthy, cell-free expression enables a well-defined starting condition which is not biased by the presence of coordinated lipids[225] and also allows the easy production of ligand-free bO. Figure 7b shows the SEC profile of purified cell-free expressed bO in SDS. Figure 1c-m summarizes our SEC results for refolding of bO from SDS into amphipols. As reported before[288,289] the amphipol-solubilized receptor can be diluted to a large extend with amphipol-free buffer, however, removing all non-protein attached amphipols, e.g. by washing amphipol-solubilized bR with amphipol-free buffer while being immobilized on a Ni-NTA column, leads to formation of very large protein-amphipol particles (Fig. 7c) (also see [290]). In agreement with previous findings these large particles can be converted into regular sized particles to some extent by the addition of free amphipols (Fig. 7d).

Refolding of bO in the absence of retinal significantly reduces the amount of protein that is incorporated into amphipols while increasing the protein aggregation as visible by very strong absorbance in the SEC void volume (Fig. 7e). Our results on cell-free expressed bO, i.e. bO that never came into contact with its native lipids, are consistent with previous studies on bO

which was expressed in *Halobacterium salinarum* and successively delipidated[291]. Refolding in the absence of the ligand was improved when bO was immobilized on a Ni-NTA matrix during refolding (Fig. 7f). However, both SEC profiles reveal a rather heterogeneous size distribution when bO is refolded in the absence of retinal. Interestingly, in the presence of retinal the efficiency of refolding is not enhanced by immobilizing bR (Fig. 7g,h), suggesting that the ligand has a strong positive effect on the refolding of bR.

Refolding of proteins in general is accomplished by reducing the concentration of denaturant that keeps the protein of interest in an unfolded state. It is known that the method for reducing the denaturant concentration can influence refolding yields. In our study, refolding was originally initiated by precipitating the denaturant SDS as its potassium salt by the addition of KCl. In the following we also tested different methods of SDS removal including the addition of Bio-Beads as well as dilution below the critical micelle concentration (CMC).

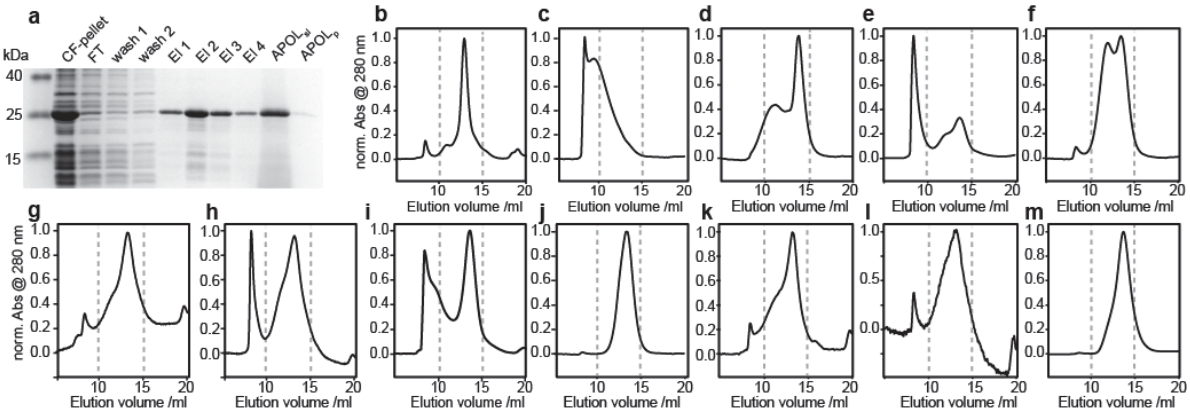


Figure 7: Biochemical and biophysical characterization of amphipol refolding of bacteriorhodopsin. (a) Coomassie blue stained SDS-PAGE showing cell-free expression, SDS purification and amphipol solubilization. (b-m) Normalized size exclusion chromatograms of different amphipol-bacteriorhodopsin preparations. (b) unfolded bacterioOpsin in SDS micelles, (c) amphipol refolded bR purified without excess of free amphipols, (d) same as in (c) but after incubation in 1 % w/v amphipol buffer for 12 h, (e-l) amphipol refolding from 15 mM SDS using different strategies as summarized in Table 1. In short: (e,f) without retinal, (g,h) using KCl precipitation, (i,j) using dilution, (k,l) using Bio-Beads. (m) Reinjection of separated bR amphipol peak after 48 h (see text and Table 1 for more details). Note that SEC results in SDS buffer (b) were obtained at room temperature, whereas all other results were recorded at 4°C.

For both methods we again also tested refolding of bR while being immobilized (see table 1 for an overview of the different methods used). In general all strategies produced correctly refolded bR (as indicative by the characteristic color of the samples, data not shown). Refolding of immobilized bR by dilution produced a very homogeneous SEC profile (Fig. 7j) offering an attractive alternative method to the precipitation of SDS by KCl. SDS removal

using Bio-Beads results in a SEC profile (Fig. 7k) that is similar to the precipitation method (Fig. 7g). The simultaneous use of Bio-Beads and Ni-NTA beads is in general feasible (Fig. 7l), however it makes sample handling more difficult and in our case resulted in significant sample loss. This may be partially improved by the use of magnetic beads. Note that the use of Bio-Beads may interfere with refolding in the presence of ligand due to the adsorption of ligand. For the bR-retinal system this was clearly the case as evidenced by the yellow color of the Bio-Beads after refolding of bR in the presence of *all-trans* retinal.

Table 1: Different conditions tested for amphipol refolding of bR. Letters in top row represent labels of resulting SEC profiles as shown in Figure 7.

	b	c[1]	d[2]	e	f	g	h	i	j	k	l	m[3]
bacterioOpsin	+	+	+	+	+	+	+	+	+	+	+	+
Amphipol	-	+	+	+	+	+	+	+	+	+	+	+
Retinal present	-	+	+	-	-	+	+	+	+	+	+	+
Bound to Ni-NTA	-	-	-	-	+	-	+	-	+	-	+	-
KCl-precipitate		+		+	+	+	+					
Dilute								+	+			
Bio-Beads										+	+	

[1] after Ni-NTA purification with amphipol-free buffer
[2] same as c) but after 12 h incubation with free amphipols
[3] reinjection after first SEC separation (using method in g) and collection of BR containing fraction

We also tested reinjection of the fractions containing the regular sized bR-amphipol particles. The respective elution fractions (i.e. elution volumes 12.5-15.5 ml of the conventional KCl precipitation shown in Fig. 7g) were pooled, concentrated and stored at 5°C for 48 h. The clean SEC profile (Fig. 7m) of this sample suggests that largely homogeneous and stable particles can be generated and isolated.

The insight gained from the optimization of bR refolding was used to study the amphipol refolding of two human GPCRs. Cell-free expression enabled the expression of the melanocortin-2 and the melanocortin-4 receptor in sufficient yields for further biophysical characterization as well as potential NMR structural studies. Purification of SDS solubilized receptor using a Ni-NTA column results in relatively pure unfolded receptor. In the following we investigated the transfer of SDS-solubilized melanocortin receptor into amphipols.

No protein bands are detected by SDS-PAGE in the insoluble pellet fraction after refolding into amphipols followed by 10 min of centrifugation at 12000 x g (Fig. 8a,b) suggesting that the amphipols efficiently solubilize both melanocortin receptors. Figure 8c-f shows SEC profiles of different amphipol refolding strategies for the MC2-receptor. The behavior observed for the MC2-receptor under the different refolding conditions closely resembles the

results obtained for bR. Note that the additional peak with an elution volume of approximately 19 ml matches SEC results obtained on free ACTH (data not shown).

In accordance with our observations for bO, refolding of the MC2-receptor can be improved when the receptor is immobilized on Ni-NTA beads (Fig. 8f). Due to the potential interference of Bio-Beads with the receptor ligands, the KCl precipitation method (in the presence of Ni-NTA beads) was selected as a straight forward method for refolding of larger quantities as required for NMR structural studies.

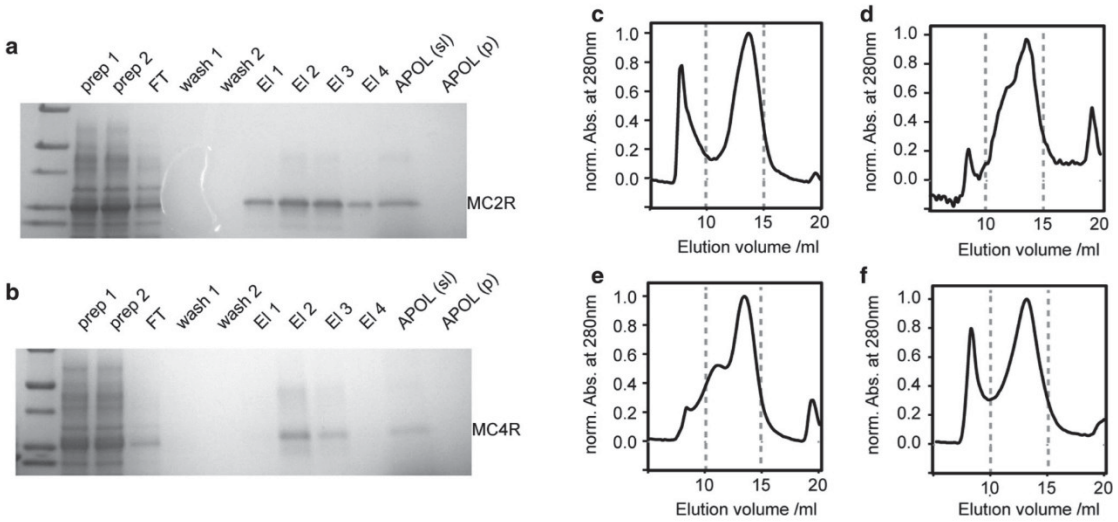


Figure 8: Biochemical and biophysical characterization of amphipol refolding of the two cell-free expressed melanocortin receptors MC2R and MC4R. (a,b) Coomassie blue stained SDS-PAGE results of expression, purification and amphipol solubilization of MC2R (a) and MC4R (b). (c-f) SEC profiles of amphipol incorporated MC2R using different transfer strategies form SDS: (c) Bio-Beads, (d) dilution, (e,f) KCl precipitation in the absence (e) and presence (f) of Ni-NTA beads.

NMR characterization of amphipol refolded 7-TM proteins

Figure 9 shows resulting SEC profiles after upscaling of amphipol refolding for bR as well as the MC2 receptor (see material and methods for more details). In the case of bR, linear upscaling resulted in the predominant occurrence of the desired (smaller-sized) population (Fig. 9a, bR-APOL).

However, non-ideal preparations can also contain a sizable population of a larger fraction (Fig. 9a, bR APOL2). (In this preparation IMAC was carried out with amphipol-free buffer (see Figure 7c) and free amphipols were added after elution (see Figure 7d)). Notably, this preparation allowed the comparison of NMR spectral quality of the smaller-sized and the larger-sized fractions (Figure 9b and 9c, respectively). While the smaller-sized fraction gives

rise to a very well resolved and dispersed TROSY-HSQC spectrum (indicative for a well-structured and homogeneous sample), the NMR data of the larger-sized fraction shows very limited dispersion (indicative for random coil segments). Importantly, it is evident by the characteristic color of the samples that both fractions are well folded. Quantitative absorbance measurements of bR purified with and without excess of free amphipols show that the amount of well-folded bR is similar in both preparations although SEC profiles differ significantly (see supplement Fig. 44).

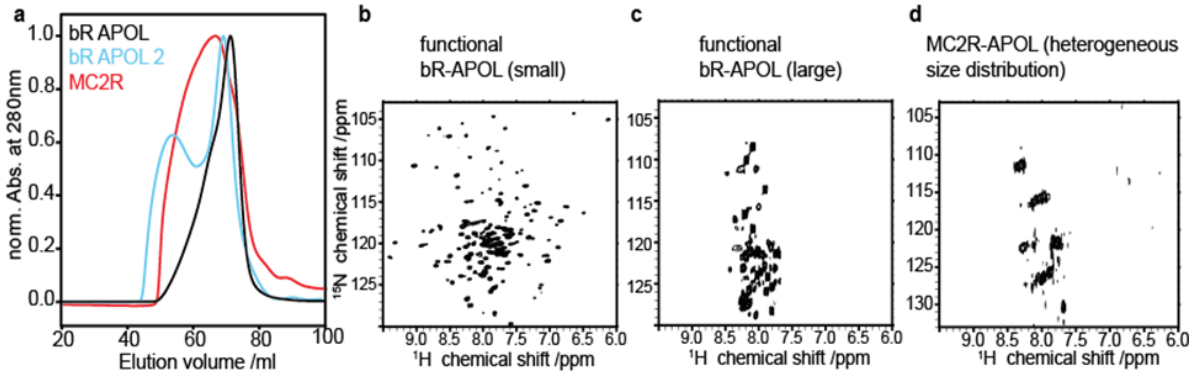


Figure 9: Preparative scale SEC profiles and NMR results of amphipol refolded bR and MC2R. Black chromatogram in (a) shows optimal refolding of bR, blue line shows additional species after not optimal folding of bR. Red curve shows chromatogram of amphipol refolded MC2R after upscaling. (b-d) 2D [^{15}N,^{1}H]-TROSY spectra of optimal refolded bR species (b), only the larger bR species (c) and of MC2R (d). Note that the spectra in (b) and in (c) both contain well-folded bR as indicative by the characteristic color of the sample (see also supplement Figure 44).

This strongly suggests that only the highly flexible parts of bR, including the protein termini, are visible in the spectrum of the larger-sized fraction (Fig. 9c). Upscaling of amphipol refolding for the MC2 receptor results in a broad distribution of different sizes as shown by SEC (most of which too large for high-resolution solution-state NMR detection). At this point it is not clear why upscaling changes the MC2R SEC profile after refolding and why the MC2R behavior differs in this point from bR upscaling. In line with the broad SEC profile, the NMR spectrum of amphipol refolded MC2R shows very limited dispersion and resolution (Fig. 9d).

The observed spectral properties of amphipol-refolded MC2R would be commonly attributed to a not appropriately folded receptor. However, as evident by our NMR data recorded on the larger-sized fraction of correctly folded BR (Fig. 9c), it is difficult to judge the quality of refolding based on a badly dispersed NMR spectrum.

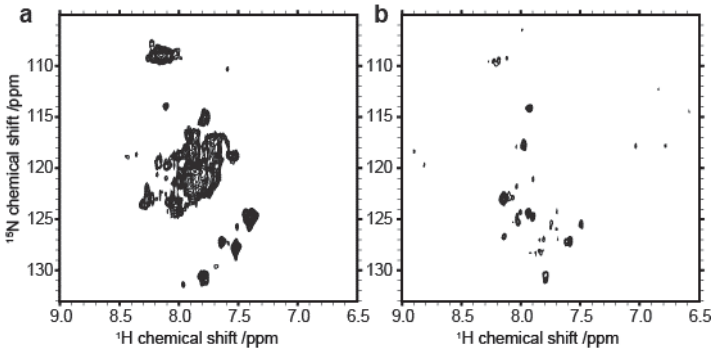


Figure 10: Initial NMR results of the melanocortin-4 receptors. 2D [^{15}N,^{1}H]-TROSY spectra of MC4R in SDS (a) and after amphipol refolding (b).

In the next setup we therefore first recorded a spectrum of the MC4-receptor solubilized in SDS to have a reference for its unfolded state (Fig. 10a). Transferring the receptor into an amphipol environment clearly changes the NMR spectrum (Fig. 10b). Although several well resolved and partly dispersed peaks are visible it is clear that not all resonances expected for the MC4R are present. Based on the rather unconventional dispersion of the peaks as well as the presence of small residual impurities visible on the SDS gel, it cannot be fully excluded at this point that parts of the spectrum in Figure 10b represent degradation products and/or signals not originating from the receptor. However, arguments speaking in favor of a NMR spectrum representing the MC4R in amphipols include the fact that cell-free expression does not produce larger amounts of isotope labelled impurities as well as that protein expression and purification was carried out in the presence of protease inhibitors. In addition the spectrum of MC4R in amphipols was recorded after refolding the receptor directly from the MC4R-SDS sample which showed expected peak dispersions (spectrum in Fig. 10a). Hence the major difference between Figure 10a and 4b is the absence of SDS as well as the presence of amphipols and the ligand in Figure 10b. Since both compounds, amphipols and the ligand (SHU9119), were added in relative large quantities, we additionally recorded 2D TROSY-HSQC spectra of isolated amphipols and SHU9119 to exclude contributions of natural abundance signal in Figure 10b (see also supplement Fig. 45).

Therefore, our NMR data would be most consistent with a partly folded receptor as well as with an amphipol-receptor complex that is too large for a detailed NMR spectroscopic study. The spectrum shown in figure 10b strongly suggests that replacing SDS with amphipols induces larger structural rearrangements in the receptor.

Conclusion

We investigated the effects of amphipol assisted refolding of the 7-TM protein bacteriorhodopsin as well as the melanocortin-2 and melanocortin-4 receptors using size exclusion chromatography and solution-state NMR. We measured biophysical properties and recorded NMR spectra of the proteins after they were solubilized by SDS and refolded using amphipathic polymers. We found that the choice of the method for removal of the SDS plays a critical role in obtaining a decent NMR spectrum. In addition upscaling the refolding protocol for the required amounts to produce NMR samples was not always straight forward and in the case of the two tested GPCRs largely interfered with the production of suitable NMR samples.

In general, all resulting NMR spectra are consistent with their underlying SEC profile, suggesting that sample optimization towards small, homogeneous particles is of fundamental importance for high-resolution NMR studies of amphipol stabilized membrane proteins in solution. Amphipol refolding of MC4R induces clear changes in the NMR spectrum, however only parts of the receptor are visible in the spectrum and further sample optimization towards smaller homogeneous particle sizes has to be carried out to obtain high-resolution insights of MC2R and MC4R in an amphipol environment. For bR it became clear that amphipol refolding can be very efficient and in generally permits high-resolution NMR insights into a fully folded 7-TM protein. On the other hand, our data of well-folded bR in larger amphipol particles (Fig. 9c) also show that badly dispersed NMR spectra of 7-TM proteins alone are not sufficient to exclude well-structured proteins.

Here we could optimize SEC profiles for amphipol refolded bacteriorhodopsin as well as for small scale refolding of the MC2-receptor. Refolding in the presence of the ligand and/or while attached to a Ni-NTA matrix can have significant positive effects on the refolding product.

We could also generate mg amounts of isotope labelled MC2 and MC4 receptors for NMR investigation using a home build cell-free expression system. The straightforward upscaling of the refolding protocol did not yield the homogeneous and small particle distribution required for high-resolution NMR studies. Nevertheless, based on our small scale refolding results, we anticipate that amphipol refolding of NMR quantities of cell-free expressed MC2 and MC4 receptor in respect to a smaller and more homogeneous size distribution should be feasible. A thorough optimization of the upscaling conditions has to be

carried out which naturally would strongly benefit from a reliable functional assay. Our current NMR comparison of the MC4 receptor in SDS and in an amphipol environment already provides first evidence that amphipol refolding induces major structural rearrangements.

10.3.1.2 BR reconstitution into tethered lipid membranes for AFM Force Spectroscopy

As an alternative to NMR, Atomic Force Microscopy (AFM) offers interesting means to study multi span TM proteins. However the used membrane system suffers from limitations. We therefore set out to test new membrane mimetics for AFM.

The following chapter reflects content of the publication:

Title: Oriented Membrane Protein Reconstitution into Tethered Lipid Membranes for AFM Force Spectroscopy

Authors: Anna M. Bronder, Adeline Bieker, Shantha Elter, Manuel Etzkorn, Dieter Häussinger, Filipp Oesterhelt

Journal: Biophysical Journal (2016)

Abstract

Membrane proteins act as a central interface between the extracellular environment and the intracellular response and as such represent one of the most important classes of drug targets. The characterization of the molecular properties of integral membrane proteins, such as topology and interdomain interaction, is the key to a fundamental understanding of their function. Atomic Force Microscopy (AFM) and force spectroscopy have the intrinsic capabilities of investigating these properties in a near-native setting. However, atomic force spectroscopy of membrane proteins is traditionally carried out in a crystalline setup. Alternatively, model membrane systems, such as tethered bilayer membranes, have been developed for surface-dependent techniques. While these setups can provide a more native environment, data analysis may be complicated by the normally found statistical orientation of the reconstituted protein in the model membrane. We have developed a model membrane system that enables the study of membrane proteins in a defined orientation by single molecule force spectroscopy. Our approach is demonstrated using cell-free expressed bacteriorhodopsin coupled to a quartz glass surface in a defined orientation through a protein anchor and reconstituted inside an artificial membrane system. This approach offers an effective way to study membrane proteins in a planar lipid bilayer. It can be easily transferred

to all membrane proteins that possess a suitable tag and can be reconstituted into a lipid bilayer. In this respect, we anticipate that this technique may contribute important information on structure, topology, and intra- and intermolecular interactions of other seven-transmembrane helical receptors.

Introduction

The characterization of transmembrane proteins is the key to a better understanding of essential processes in life. Transmembrane proteins account for ~30% of all proteins[292] and act as sensors, catalysts, receptors, transporters, and channels. Thus, they play an important role in almost all cellular processes and are associated with a broad range of different diseases[293–296]. Among the membrane proteins, the class of seven-transmembrane helical (7-TM) proteins, which includes G protein-coupled receptors (GPCRs), have a very central part in a variety of sensing and signaling pathways, as well as physiological responses, making them a prominent target for drug development.

Transmembrane proteins easily lose their functionality and denature when removed from their natural membrane environment. One approach to study functional membrane proteins is the development of simple artificial membrane systems that reduce the complexity but still mimic the most important properties of biological membranes. A promising system is a tethered bilayer lipid membrane (tBLM). Lipids of the first layer of the membrane are anchored covalently to a solid substrate through a spacer. This spacer, e.g., a polymer, acts as a cushion that compensates surface roughness, mimicking a cytoskeleton, and can additionally create an ion reservoir beneath the membrane[297–299]. The spacer can be bound to the substrate first, binding the lipid to the spacer in a subsequent step[300]. tBLMs have proven to be stable for days and can even be used for weeks when covered by a hydrogel[300]. Binding of the membrane to the spacer occurs through an anchoring molecule, a lipid or hydrophobic chain. Important while binding the anchoring molecule is the grafting density, i.e., the ratio of anchored to not-anchored lipids. Although a high grafting density leads to high electrical resistance, it can also hinder the incorporation of proteins[299] or the diffusion of incorporated proteins through the membrane[301]. In addition to tBLMs, so-called protein tethered membranes can be formed as well. In this case, the anchoring lipid or hydrophobic chain is directly substituted with the membrane protein, which is coupled to the surface, e.g., through a complex between a surface bound nitrilotriacetic acid (NTA) and the protein histidine-tag (His-tag)[302].

Alternatively, membrane proteins can be studied in other artificial membrane systems, e.g., liposomes[303], nanodiscs[304], black or bilayer lipid membranes (BLMs)[299], and solid supported BLMs (sBLM)[299].

The formation of a tBLM can be done by self-assembly[305]. One way of self-assembly is to adsorb and then spread out whole vesicles over the surface[306,307]. This way, proteins can be reconstituted already into the prepared vesicles before the final bilayer formation[308] or can be added to the finished bilayer on the surface[309].

Proteins embedded in a tBLM can be studied by Atomic Force Microscopy (AFM). AFM-based single-molecule force spectroscopy can be used to obtain information on dissociation rates[310], energy barriers[311,312], Gibbs free energy[310,312], the form of a binding potential[312], and inter- and intramolecular interactions[311,313–315], as well as the folding of proteins and their constitution inside a membrane[311].

In this study, we demonstrate a method of forming a tBLM on quartz glass with the incorporated 7-TM model membrane protein bacteriorhodopsin (bR) in a defined orientation. BR has been studied extensively[32,268,269] and offers high stability. In addition, the covalently bound retinal allows direct insight into its folding state through absorption measurements (e.g., at 555 nm). We investigated the conformation of bR inside the presented tBLM by AFM-based force spectroscopy. Our data are in accordance with expected force-distance curves of a 7-TM membrane protein[304,316]. The orientation of bR in purple membranes can be predetermined through AFM imaging[317,318]. Thus, the characteristics of force-distance curves of bR pulled from the extracellular side and the cytoplasmic side are already known, which in turn can be used to process the data of statistically oriented bR. In general, knowledge of the orientation will substantially facilitate data analysis and interpretation. In our model system, predefined orientation is achieved through first functionalizing the quartz glass by silanization. Afterwards, a polyethylene glycol (PEG) is coupled as a spacer. To reconstitute the protein on the surface in a defined orientation, a protein anchor is bound to the spacer in addition to the lipid anchors. We make use of cell-free protein expression, a method that is suitable to produce 1) a broad range of different membrane proteins[319], and 2) large amounts of protein in the presence or absence of cofactors such as ligands, detergents, or lipids. We could show before that cell-free expression of bR enables analysis of the pure protein that is not biased by copurification of, e.g., coordinated lipids, which are normally present when bR is extracted from the native purple membrane[225]. To test our model system, we use cell-free expressed bR that contains a His-tag and a surface-coupled protein anchor

consisting of Tris-NTA (trisNTA). Our results show the specific coupling of the protein to the surface and the successful formation of a tBLM containing the membrane protein in a defined orientation.

Materials and Methods

All water used was purified by the Milli-Q Integral Water Purification System (Merck Millipore, Darmstadt, Germany). Specific buffer compositions are described in the Supporting Material. All given cantilever values are nominal values.

Lipids, detergents, and stock solutions

1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC; >99% purity), 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-(succinyl) Sodium salt (SuccinylPE; >99% purity), and n-dodecylphosphocholine (DPC; >99% purity) were purchased as powders from Avanti Polar Lipids (Alabaster, AL). n-Dodecyl-β-D-maltopyranoside (DDM; Anagrade) was purchased from Affymetrix (Santa Clara, CA). Carboxy-functionalized tert-butyl ester (OtBu)-protected trisNTA[320] was provided by the group of Prof. Dr. Jacob Piehler of the University of Osnabrück. Stock solutions in chloroform (for spectroscopy, 99+% purity, stabilized with amylene; Acros Organics, Geel, Belgium) were created as follows: 0.026 g SuccinylPE, 0.022 g DPPC, and 0.005 g trisNTA were dissolved in 1 mL chloroform and stored at −80°C.

AFM

AFM images were taken with an AFM Nanowizard III (JPK Instruments, Berlin, Germany). Imaging was done at room temperature (RT) using contact mode with the probe OMCL-TR400PSA (Cantilever 2, Asylum Research, Mannheim, Germany) with a spring constant (kc) of 0.02 N/m, a resonance frequency (f) in water of 11 kHz, and a tip radius (r) of 15 nm, or using tapping mode with the probe SNL-10 (Cantilever A, kc = 0.35 N/m, f = 65 kHz in air, r = 2 nm; Bruker, Billerica, MA) in buffer P1.

AFM-based force spectroscopy

Force-distance curves were measured using a Nanowizard III (JPK Instruments). The same probes as for imaging in contact mode (OMCL-TR400PSA, Cantilever 2) were used. All experiments were performed at RT in buffer P1. Force-distance curves were acquired as follows. Starting away from the surface, the cantilever is moved downward at a constant speed (200 nm/s). When the surface is reached, the cantilever is moved farther until

it reaches a defined maximal deflection (0.4 mV). This position is held for 1 s before retraction of the cantilever (200 nm for force-distance curves unfolding bR, 100–400 nm for force-distance curves measuring adhesion). This cycle is repeated at different positions on the surface (acquiring ~500 force-distance curves per position for experiments unfolding bR and ~100 force-distance curves per position during adhesion experiments). The position on the surface was changed two to three times per experiment. Experimental conditions were repeated in two independent measurements. A fresh cantilever was used for each measurement. As force measurements require the precise knowledge of the cantilever's spring constant, k_c, the thermal noise method[321] is used for calibration. Additionally, the cantilever's sensitivity (deflection → force) is calibrated by measuring the deflection of the cantilever approached against the surface. Calibration was performed at the beginning of each experiment and between positions on the surface.

X-ray photoelectron spectroscopy

Cleaned quartz glass surfaces and silanized quartz glass surfaces were characterized with x-ray photoelectron spectroscopy (XPS) using an ESCALAB MK II spectrometer (Thermo Fisher Scientific, Oberhausen, Germany).

Functionalization steps

Cleaning of quartz glass according to a previously described protocol[322] was followed by silanization with (3-aminopropyl)trimethoxysilane (APTES; 97%, Sigma Aldrich, St. Louis, MO) and subsequent binding of the spacer Mmt-NH-PEG12-COO-Tfp (PEG; Iris Biotech, Marktredwitz, Germany). To this spacer was coupled the protein anchor trisNTA alone, the lipid anchor SuccinylPE alone, and a mixture of both anchors 1:2 (mol/mol). The detailed coupling steps are described in the Supporting Material (see chapter 15.2.3).

Lipid suspension

30 mM DPPC stock (230 µL) was transferred into a sample vial and the chloroform was evaporated under nitrogen. Buffer P1 (1 mL) with 0.1% DPC (w/v) was added to create the final lipid suspension.

Binding of bR and membrane formation on functionalized quartz glass

After rinsing the functionalized slides (1 cm2) 10 times with water, the surface was covered with 50 mM EDTA (disodium salt dehydrate, molecular biology grade; AppliChem, Darmstadt, Germany) solution for 5 min. After removal of the EDTA solution,

a 10 mM nickel(II) chloride (98%, for analysis; Grüssing, Filsum, Germany) solution was added to the surface for 5 min. The solution was removed from the surface and replaced with buffer DDM-W1. The surface was rinsed once with protein buffer DDM-P1, and 20 µL of a 10 µM bR solution in buffer DDM-P1 was added to 180 µL lipid suspension. This solution was placed onto the surface and allowed to incubate for 2 h. The sample was rinsed 10 times with 250 µL buffer P1 and covered with 50 µL of the same buffer for AFM measurements.

Membrane formation on mica

Lipid suspension (50 µL) was incubated on a round mica surface (Ø 10 mm, Ted Pella, Redding, CA) for 2 h. The surface was rinsed five times with 100 µL P1 and covered with 50 µL of the same buffer for AFM measurements.

Membrane formation with reconstituted BR on mica

A 10 µM bR solution in buffer DDM-P1 (5 µL) was added to 45 µL lipid suspension. The mixture was incubated on a round mica surface (Ø 10 mm) for 2 h. The surface was rinsed five times with 100 µL P1 and covered with 50 µL of the same buffer for AFM measurements.

Dot blot

Specifically bound bR should only be rinsed from a functionalized surface with a buffer containing imidazole. After rinsing the surface once with buffer missing imidazole and once with buffer containing imidazole, a simplified western blot (dot blot) was utilized to test the specific binding of bR to the functionalized surface. The procedure is described in detail in the Supporting Material (see chapter 15.2.3).

Cloning and cell-free expression

Cell-free expression of bR was carried out as described before[225]. An *Escherichia coli*-based system following previously published protocols[60] was used. Dialysis-mode reactions were carried out in the absence of retinal and detergents. The resulting protein pellet was washed with S30 buffer and directly refolded or stored at −20°C. The refolding procedure is described in more detail in the Supporting Material (see chapter 15.2.3).

Results

In our study, we aimed to reconstitute bR as a model membrane protein in a defined orientation in artificial tBLMs to perform single-molecule force spectroscopic studies. The process is depicted schematically in Figure 11. The clean quartz glass surface (Fig. 11a)

is coated with APTES for amino functionalization (Fig. 11b). After this, a PEG with an activated carboxyl group and a protected amino group is coupled to the surface (Fig. 11c). Following the deprotection of the PEG's amino group, the binding of an anchoring lipid (SuccinylPE) (Fig. 11d) and a protein anchor (trisNTA) (Fig. 11e) were investigated both separately and as a mixture of both anchors (Fig. 11f). The protein is reconstituted on the surface together with the free lipid (DPPC) supported by detergent (DPC).

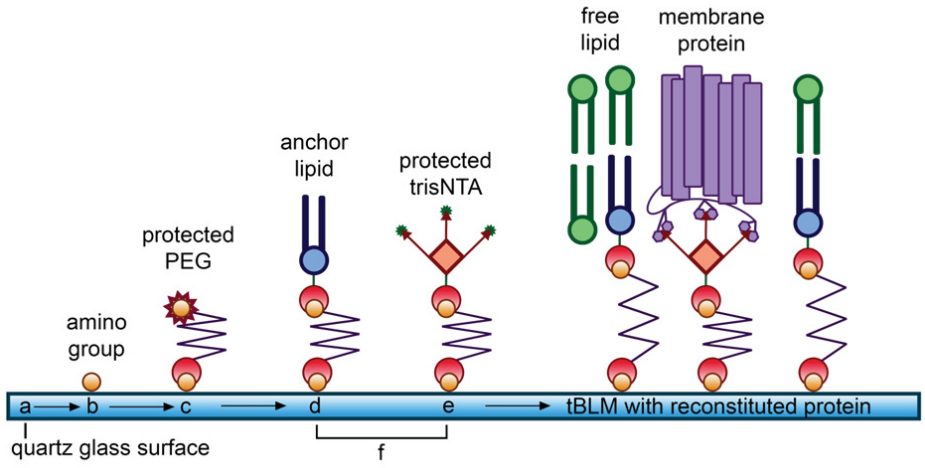


Figure 11: Schematic representation of the functionalization and coupling steps. The cleaned quartz glass surface (a) is first functionalized using APTES (b) to create amino groups. To these amino groups a PEG linker (c) is coupled with its activated carboxyl group and protected amino group. After deprotection of the PEG linker amino group, the surface can be functionalized in three ways, with only the lipid anchor bound to PEG (d), with only the protein anchor (protected trisNTA) bound to PEG (e), or with both anchors bound simultaneously to PEG (f). Afterward, a mixture of the protein with free lipid is applied to the surface, forming the final tBLM with reconstituted protein.

Surface composition before and after silanization

The atomic composition of a pure quartz glass surface was compared to a silanized surface using XPS. In the survey spectrum of cleaned quartz glass (see supplement Fig. 46a), binding energies corresponding to fluorine (F), sodium (Na), and carbon (C) levels are detected apart from expected silicon (Si) and oxygen (O) levels. A survey spectrum after silanization with APTES shows binding energies corresponding to levels in O and Si (see supplement Fig. 46b). Additionally, there is an increase of C from 12.8 atomic percent (at.%) to 41.6 at.% and an appearance of 7.2 at.% N.

Changes in surface topography

The functionalization of quartz glass surfaces was controlled by AFM to be able to observe the changes in surface topography after each functionalization step. The surface roughness is calculated as the arithmetic mean (i.e., the average roughness, Ra) of absolute height values for each surface. The AFM image of quartz glass (Fig. 12a) after the cleaning procedure shows the standard surface topography of quartz glass. The average surface roughness is 511 picometers (pm). Holes and scratches, which are typical for quartz glass, are also visible. The surface topography of cleaned quartz glass is similar after silanization (Fig. 12b), which is also represented by the average surface roughness. A surface functionalized with PEG is shown in Figure 12c. The average surface roughness has decreased to 154.1 pm and there are no longer holes or scratches visible.

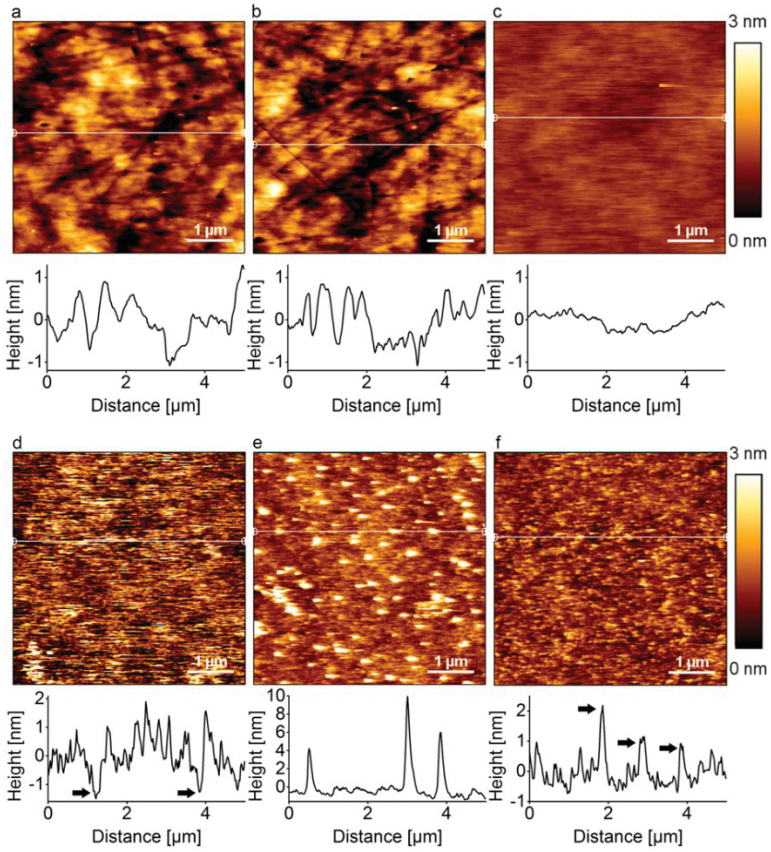


Figure 12: AFM scans in contact mode in buffer P1 (5 µm × 5 µm), with corresponding height profiles, represented as white horizontal lines in each image, below each scan. (a) The cleaned quartz glass surface shows a roughness of 511 PM. (b) Silanized quartz glass with APTES shows no change in surface roughness. (c) Binding PEG to silanized quartz reduces the roughness to 154.1 PM. (d) A surface functionalized only with the anchoring lipid SuccinylPE. (e) A surface functionalized only with the protein anchor trisNTA. (f) A surface functionalized with a mixture of 2:1 (mol/mol) SuccinylPE:trisNTA.

The surface was functionalized further with the lipid anchor SuccinylPE only (Fig. 12d), with trisNTA only (Fig. 12e), and with a mixture of 2:1 (mol/mol) SuccinylPE/trisNTA (Fig. 12f). The surface with the lipid anchor alone shows an increased average roughness of 585 pm compared to the previous step (Fig. 12c). Some small regions indicate an additional depth of 1 nm (arrows). The surface functionalized solely with trisNTA also shows a generally higher average roughness and round structures of 4–10 nm height and 200 nm diameter distributed over the whole surface. A mixture of SuccinylPE and trisNTA yields an average roughness of 332.6 pm, but distributed protruding structures of 1–3 nm height can be seen (arrows).

Changes in adhesion

Force-distance curves were performed between surface functionalization steps to observe the changes in cantilever-surface interaction (adhesion). This was done to observe whether adhesion on the functionalized surfaces remained constant over the majority of the surface, which would be an indication of complete and homogeneous functionalization. In Figure 13, the mean adhesion of the cantilever on the surface is depicted for all functionalization steps, and for surfaces functionalized with SuccinylPE only, trisNTA only, or a mixture of the two. For quartz glass (Fig. 13a), no adhesion can be observed. APTES-functionalized surfaces (Fig. 13a) show a strong increase in adhesion force, from −25 pN ± 8 pN to 12,420 pN ± 1039 pN. After the second functionalization step, the strong adhesion forces diminish to 212 pN ± 125 pN for PEG (Fig. 13c). From this step on, three functionalization ways were chosen. Surfaces functionalized with the lipid anchor SuccinylPE (Fig. 13d) show an increase of adhesion forces to 556 pN ± 262 pN. Surfaces functionalized only with the protein anchor trisNTA (Fig. 13e) show a higher increase in adhesion to 4160 pN ± 1152 pN, whereas surfaces with both anchors in a 2:1 (mol:mol) SuccinylPE:trisNTA mixture (Fig. 13f) show an adhesion force of 669 pN ± 300 pN. The detailed distribution of adhesion forces as well as exemplary force-distance curves for each functionalization step can be seen in supplementary Figure 47.

Protein complex formation

To test the successful immobilization of bR through a complex between the protein's N-terminal deca-histidine-tag (His10-tag) and the trisNTA on the surface, specific elution of the protein by imidazole-containing buffer was performed and visualized by dot blot. To accomplish this, bR was first immobilized on the surface functionalized with the protein

anchor and the lipid anchor. Then the surface was rinsed first with a buffer without imidazole. The volume of this rinsing step was collected as the rinsing fraction. Afterward, the surface was rinsed again with the buffer containing imidazole. The volume of this elution step was collected as the elution fraction. Together with a buffer control of the buffer containing imidazole, the fractions were distributed on a dot blot. As can be seen in supplementary Figure 48, only the elution fraction (C) shows luminescence.

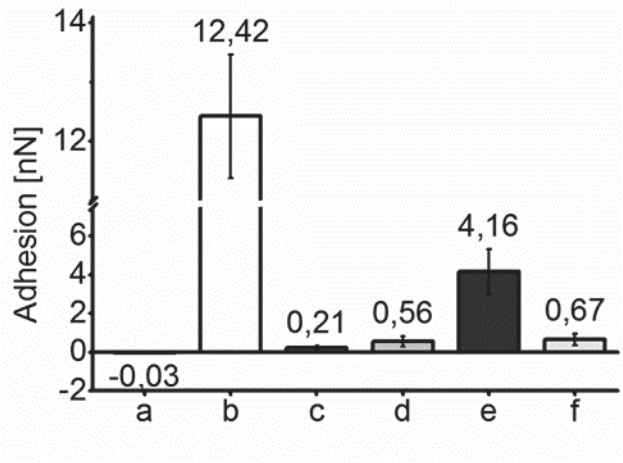


Figure 13: Mean adhesion measured through force-distance curves with AFM. An average of 410 ± 38 force-distance curves were analyzed per experimental condition. (a) Mean adhesion for cleaned quartz glass (−0.025 nN ± 0.008 nN). (b) Mean adhesion for silanized quartz glass (12.42 nN ± 1.04 nN). (c) Mean adhesion for silanized quartz glass functionalized with PEG (0.21 nN ± 0.13 nN). (d) Mean adhesion for SuccinylPE bound to PEG (0.56 nN ± 0.26 nN). (e) Mean adhesion of trisNTA bound to PEG (4.16 nN ± 1.15 nN). (f) Mean adhesion of a 2:1 (mol/mol) mixture of SuccinylPE:trisNTA bound to PEG (0.67 nN ± 0.30 nN).

BR in membranes

Vesicle spreading on mica is an often-used method of creating solid sBLMs[306,307,309]. A trial of the lipid mixture with bR was performed on mica to investigate possible effects of bR on bilayer formation. Figure 14a shows a mica surface after incubation of the lipid alone. The surface shows a homogeneous coverage with a lipid bilayer membrane (light gray area) with a few defects (dark gray areas). The corresponding height profile shows a membrane height of 5 nm and a low roughness of the membrane surface. Compared to an sBLM without bR, the bilayer containing bR (Fig. 14b) has a much higher average roughness of 709.4 pm. All in all, the surface shows a complete coverage with no membrane defects. When spreading the lipid mixture with bR onto a functionalized quartz glass surface (Fig. 14c), this high average roughness not only persists but increases to 1094 pm. However, the protein-lipid layer is spread over the whole surface. It has to be noted that during the scan process (bottom to top),

the structures on the surface of the protein-containing membrane appear to increase in size, which is likely due to lipid adsorption to the cantilever tip.

After scratching with the cantilever tip on a surface covered with the protein-lipid mixture (Fig. 14d), a depth of 10 nm can be measured.

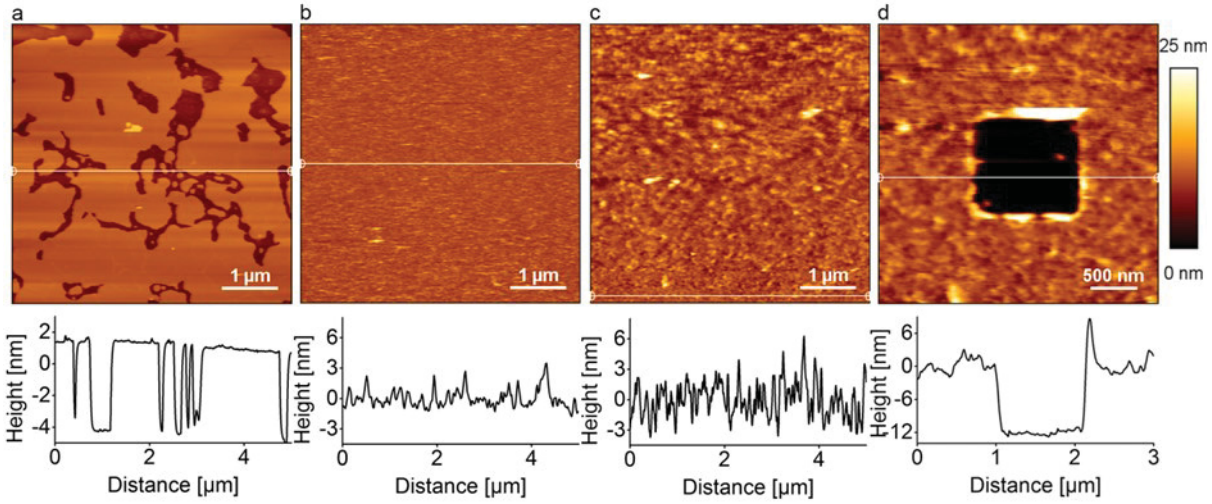


Figure 14: AFM scans in tapping mode in buffer P1 (5 µm × 5 µm for a–c; 3 µm × 3 µm for d), with corresponding height profiles, represented as white horizontal lines in each image, below each scan. (a) Membrane formation on a mica surface. (b) A mica surface covered with membrane containing bR. (c) A functionalized glass surface covered with the membrane containing bR. (d) Scratching was performed with the cantilever on a surface functionalized with lipid and protein anchor and incubated with the protein lipid mixture.

Force spectroscopy

Force spectroscopy measurements were performed as described in Materials and Methods on surfaces with bR bound to the protein anchor and reconstituted into the lipid/detergent membrane. Only force-distance curves showing more than two force peaks and the last force peak at a peak position of >60 nm with a measured force of >100 pN were selected, to assure that only force-distance curves on specifically oriented and complex-coupled bR were used for analysis. Supplementary Figure 49 shows three representative force-distance curves that were selected according to these criteria. Force spectroscopic data was smoothed using moving-average filtering. In Figure 15, selected force curves were overlaid to show the four characteristic peaks for the unfolding of bR as a seven-α-helical protein[268,269,304,316]. Although the first peak (Fig. 15a) is inhomogeneous, the following peaks (Fig. 15b–d) can be distinguished. The second and third unfolding peaks remain at forces <150 pN, whereas the last peak shows a force of ~200 pN.

Discussion

Substrate characterization by XPS

To analyze the quartz glass surface and verify the first functionalization step, the surface was measured before and after silanization using XPS. A pure quartz glass surface shows the presence of a few undesired elements, namely fluorine, carbon, and sodium. As even quartz glass has some impurities, this is not surprising. However, any sign of these elements disappears after silanization with APTES. Given the structure and composition of APTES, and considering the theoretical distribution on the surface (see supplement Fig. 50a), a theoretical composition of the silanized surface is shown in supplementary Figure 50b. It is also taken into account that XPS measures not only the topmost layer of a surface but can penetrate further. Thus our XPS results of the silanized surface (see supplement Fig. 50b) are in accordance with a successful coverage of the surface with APTES.

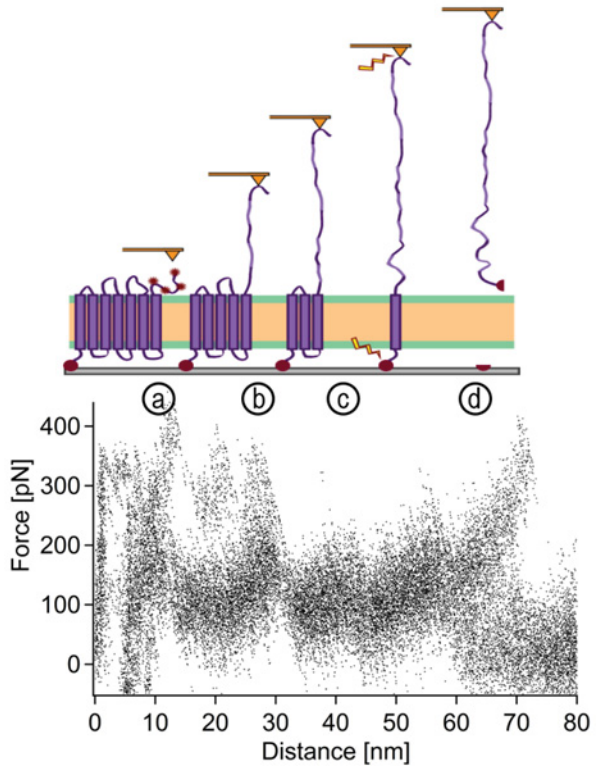


Figure 15: BR unfolding curves from the model tBLM. As a 7-TM membrane protein, the unfolding of bR will lead to four characteristic peaks in a force-distance curve. (a) The first peak shows the stretching of the terminus and unfolding of the first two transmembrane helices (TMs). As the cantilever can adsorb to the terminus at different positions, the first peak is not resolved clearly. This is followed by the pairwise unfolding of the next four TMs during the next two peaks (b and c). The last peak (d) shows the pulling of the seventh TM until bR dissociates from the tip or from the membrane. In our case, this means the dissociation between the His10-tag and the trisNTA.

Surface characterization by topography and adhesion

Surface topography The surface functionalization steps were further characterized by AFM. AFM imaging offers the possibility of observing the changes in surface topography. A comparison of AFM imaging data of a clean quartz glass surface (Fig. 12a) with a silanized quartz glass surface (Fig. 12b) shows no change in average surface roughness and characteristics of the surface, in line with the small molecule size and flexibility of APTES. The short PEG linker that is coupled subsequently to the surface can have a length of 4 nm in its stretched conformation. This linker is flexible and will either coil on the surface or stretch to accommodate for surface roughness and differences between the bilayer and protein. AFM data of this stage (Fig. 12c) shows that the holes and scratches found on quartz glass disappear, suggesting that the linker can compensate for surface roughness. The homogeneity and absence of defects also indicates a complete coverage with PEG. Coupling of only the anchoring lipid SuccinylPE to PEG yields a surface with an increased average roughness (Fig. 12d). Additionally, buffer repelling properties, observed during the experiments, suggest strong hydrophobic characteristics of the surface. This can be explained by the hydrophobic chains of the anchoring lipid orientating up and away from the substrate, thereby creating a hydrophobic surface. The roughness of the surface is due to the length of the anchoring lipid (2 nm). The surface is again homogeneous and shows no defects. Coupling only the protein anchor, trisNTA, to the surface leads to the appearance of round structures ~4–10 nm in height and 200 nm in diameter. This can indicate the aggregation of an excess amount of protein anchor. When the protein anchor is diluted with the lipid anchor 1:2 (mol/mol), the aggregates are no longer present. This and the uniform surface structure indicate a homogeneous mixture of lipid and protein anchor.

Adhesion The differences in adhesion after each step are indicative of a successful functionalization process. Furthermore, a homogeneous adhesion force throughout the investigated surface implies a homogeneous surface functionalization. The strong adhesion forces for a surface functionalized only with protein anchor can relate to the presence of aggregates on the surface. Thus, even though the surface functionalized with the mixture shows only slightly higher mean adhesion forces than the surface only functionalized with the lipid anchor, a homogeneous and well distributed protein anchor is assumed due to the previously described changes in surface topography.

Specificity of protein binding

To test the specificity of protein binding to the surface and exclude unspecific adsorption, bR was coupled to a functionalized surface. The surface was then rinsed with a buffer without and with imidazole. As only imidazole-containing buffer can elute proteins with a poly-histidine-tag from NTA groups, bR will only be detected (using a dot blot system with anti-His antibodies) for the elution fraction containing imidazole. To exclude a possible binding of the antibody to the imidazole, imidazole-containing buffer was applied to the membrane as a buffer control. The results of the dot blot clearly show that the protein could only be eluted from the surface with imidazole. These data verify not only the specific binding of the protein to the surface anchor, but also the integrity of the previous coupling step leading to accessible trisNTA.

Membrane and protein reconstitution

Artificial lipid membranes have been observed to undergo phase transitions from more liquid unordered phases to ordered gel/crystalline phases at specific temperatures. The phase transition temperature depends on the membrane composition. As the integration of proteins into lipid membranes and their functionality is highly dependent on the membrane fluidity, we aimed to provide a lipid membrane system that is stable but still offers enough fluidity for the integration of membrane proteins. Future studies may benefit from the possibility of a careful selection of lipids that are compatible with the setup. We chose DPPC as the tBLM main component due to its good properties to form stable membranes on mica. DPPC has been used previously to create stable sBLMs[323,324]. However, as the phase transition occurs at a temperature of 42°C, those sBLMs are in the ordered gel-to-crystal-like phase at RT[325,326]. Adding a detergent like DPC weakens lipid-lipid interactions and increases the permeability of the bilayer to proteins[327]. A membrane mimetic system consisting of a lipid and detergent mixture has been described previously[328]. The formation of a sBLM on mica by a mixture of DPPC and DPC was tested, and a homogeneous bilayer formation with a few defects could be detected by AFM. The bilayer height of 5 nm is in accordance with bilayer heights in previous studies[329,330]. Reconstitution of bR into the lipid/detergent mixture and applying this mixture to a mica surface leads to coverage of the surface with a homogeneous layer (Fig. 14b). Defects could not be observed. However, compared to an sBLM without bR (Fig. 14a), the roughness of the membrane surface has increased. It is known that membrane-integrated proteins can protrude out of the membrane[330]. The roughness of the membrane could thus indicate protrusions of bR out of the membrane.

The protrusions persist on membranes formed on functionalized quartz glass (Fig. 14c). The roughness is even more pronounced, as functionalized quartz glass in itself provides a rough substrate (Fig. 12f). The increase in size of the structures observed on the surface with scan direction (bottom to top) is a broadening effect of the cantilever tip. Especially soft materials, like membranes and proteins, in our case, most likely the free lipid DPPC, can adsorb to the tip during scanning and increase the effective tip diameter. Every structure will then be broadened by the tip diameter. This is why for the first 500 nm the surface structure appears comparable to a membrane with protein on mica and becomes increasingly dissimilar during the scan process. For this reason, the height profile was measured during the first 500 nm of the scan. As a control, the membrane was also formed on a functionalized surface without bR (see supplement Fig. 51a). Here, the surface was again homogeneously covered with the tBLM with only a few small defects. The surface roughness of the membrane was only influenced by the roughness of the quartz glass and the underlying functionalization, indicating that the high roughness of Figure 14c is caused by protrusions of bR. In contrast, a membrane on a surface functionalized only with trisNTA and no lipid anchor (see supplement Fig. 51b) could not be formed on quartz glass under the same conditions.

Scratching of the membrane with bR on quartz glass (Fig. 14d) revealed a height of 10 nm and protein clusters protruding 1–2 nm out of the membrane. When taking into account the length of the PEG linker and the height of the sBLM on mica (4 nm + 5 nm), the measured height of 10 nm is in accordance with expectations.

Force spectroscopy

Force-distance curves of force spectroscopy measurements performed on bR reconstituted into tBLMs showed the typical pattern of four peaks, consistent with a functional reconstitution of bR. In supplementary Figure 52, a representative bR force-distance curve from bR purple membranes is compared to the measured force-distance curves of Figure 15. The first peak of the measured force curves of bR in the tBLM was inhomogeneous, which can be attributed to the unspecific adhesion positions of the terminus to the tip[316]. The second peak seems to favor lower forces and follows the shape of the intermediate peaks found in force-distance curves of purple membranes[318]. Considering the average noise of the recorded force-distance curves (±30 pN) the third peak is in accordance with unfolding forces of bR from purple membranes and nano-discs[304]. During the last unfolding event, bR can either dissociate from the tip or be pulled out of the membrane completely. In the case of being pulled out completely, bR would have to also dissociate from the protein anchor.

It has been shown previously on soluble proteins that the dissociation forces between triNTA and His6 can be between 100 and 400 pN[331]. Our results indicate that the last peak of our force-distance curves represents these dissociation forces. Although this influence of the protein anchor on the last unfolding event should be considered during analysis, it also shows that bR reconstituted into the here-presented tBLM is coupled to the protein anchor. In addition, in the case of bR not being inserted in a defined orientation, two sets of force-distance curves are expected (one pulled from the N- and one from the C-terminus), which can be the case for reconstitution of bR into nanodiscs[304] and force spectroscopic measurements performed on bR crystals[318]. As bR force-distance curves from both termini are indeed very similar, the high forces detected for the last peak indicate a complex formation between the His-tag and trisNTA, thus supporting a defined orientation of bR. Should force-distance curves be measured that do not show the dissociation forces of the complex, they can be sorted out before analysis of the data.

Apart from the influence of the trisNTA, there could also be an influence of the PEG linker on the last unfolding event. As we expect a typical stretching behavior of the PEG linker[332,333], we do not analyze the PEG-linker elasticity directly. However, it can be assumed that the influence of the PEG-linker elasticity is minimal due to its short length.

Even though higher forces for the last peak can be seen in our experimental setup, we also have to note the stronger noise. As this high noise can indicate misfolding or denaturation of bR, we cannot clearly determine whether bR is still in its functional state after reconstitution. The functional form of the protein has been checked before reconstitution into the membrane through ultraviolet-visible spectroscopy (described in section 15.2.3). An example absorbance spectrum of a 1:10 diluted solution of cell-free expressed and purified bR is shown in supplementary Figure 53. Due to the very low protein concentration and the experimental setup, we can, however, only assume the functional form of the protein via the measured force curves. However, we anticipate that further optimization of tBLM composition, which has to be adjusted to every new protein of interest, will allow for further stabilization of the protein conformation inside the tBLM.

Conclusions

Gaining knowledge about the structure of membrane proteins is an important endeavor to reliably study the cause of diseases and facilitate drug development. In this respect, single-molecule AFM force spectroscopy offers unique possibilities in particular for the

investigation of multipass membrane proteins. We here showed that AFM force spectroscopy can be carried out in a tBLM system, and our method allows for filtering out force-distance curves specific to a defined orientation. Although the tBLM system offers a near-native environment that can be adjusted to accommodate different membrane proteins, the predefined orientation can simplify data analysis by removing a second contribution arising from different orientations and by better defining the starting conditions for each protein. In our work, we show a five-step functionalization and reconstitution process that we successfully apply to the cell-free expressed 7-TM protein bR that has been frequently used as a model for GPCRs. We anticipate that the presented approach can be transferred to a broad range of target membrane proteins and surface-based techniques and may help to provide new insights into structure, function, and interactions of these important biological systems.

10.3.2 Unfolded state of bacterioOpsin (bO)

Characterizing the endpoint of the folding process, i.e. the conformational state of a functionally folded protein, does not on its own provide comprehensive information of the folding process itself. In this respect a great need of gathering information also about the start point of folding, namely the unfolded conformational state, exists in order to create and test folding hypothesis.

NMR provides a powerful tool to investigate conformational states. But obtaining high-resolution data can be challenging, especially when it comes to more complex systems, such as multi-span transmembrane proteins. Therefore, it is of key importance to first optimize sample conditions according to the requirements of the applied NMR techniques.

10.3.2.1 Optimization for NMR studies

Here the aim of optimizing the system for NMR studies is to gather high-resolution insights into the folding process of bR. Rigorous characterization of basic conformational states can open up the possibility to examine and understand the whole folding process in detail. Particular attention should be paid on the choice of membrane mimicking environments (see section 9.4), as large complexes exhibit slower tumbling-rates, which are not benefitting high-resolution NMR measurements. Furthermore, surfactants can have severe impact on protein structure, which can enable or disable the folding process.

Based on detailed characterization of the unfolded and folded state of bR, we are aiming at acquiring highly resolved spectra for real-time measurements of bR folding. To realize this,

compromises need to be found in terms of spectral quality and "protein-happiness", meaning that bO is well solubilized, while maintaining its ability to fold into a functional state.

A first characterization of purified SDS-solubilized bO by analytical size exclusion chromatography (SEC) is already presented in section 10.3.1.1, Fig.7b. The analytical SEC profile shows a rather homogeneous particle-size distribution of the sample, as e.g. visible by occurrence of only a small peak in the void volume, which represents larger protein aggregates. The general peak shape of the main elution peak is symmetrical, even though the main bacterioOpsin peak overlaps with a smaller one, which indicates a small population of particles in heterogeneous size.

As already mentioned in section 10.3.1.1, spectral quality and dispersion is highly dependent on particle size and sample homogeneity[214,225]. By consequence, homogeneous samples are of crucial importance in order to obtain highly resolved NMR spectra. Figure 16 shows the result of a preparative SEC conducted with the same conditions as in 10.3.1.1, Fig.7b (where an analytical column was used).

This SEC result shows a much better peak separation, than the profile of the analytical SEC. By collecting small fractions, and joining only those for NMR measurements, which are not overlapping with the smaller peak, sample homogeneity is maintained as effectively as possible.

But is purification really necessary for obtaining high-quality NMR spectra for cell-free expressed bO? Purification by immobilized metal affinity chromatography (IMAC) and SEC are doubtlessly effective methods to increase sample purity, but it is time-consuming and sample losses cannot be avoided. Protein pellets derived from cell-free expression particularly contain the protein of interest[60] without any bias of co-purified lipids. Depending on the following analytical methods, elaborate steps of further purification might therefore not be required, when using CECF. Especially when it comes to investigation of isotope labelled proteins by NMR techniques, low concentrations of non-labelled proteins, with incorporated isotopes in natural abundance, should not interfere with most experiments.

The result of a 2D (^{1}H-^{15}N)-TROSY-HSQC experiment with the pooled fractions of the result presented in Figure 16a is shown in Figure 16b.

Figure 16b shows a comparison of TROSY-HSQC experiments with purified and non-purified uniformly labelled bO in SDS micelles. The black spectrum corresponds to the

sample of purified bO (Fig. 16a), the red one to the sample of non-purified bO. Note that amino acids QNW were not labelled in the purified sample of bO, in line with the additional peaks visible in the red spectrum.

While the overall peak pattern is very similar in both samples (indicative of similar structural features), the spectral resolution is slightly better for the sample of purified bO, even though side chains of tryptophan, asparagine and glutamine are not visible in the spectrum, due to the applied isotope labelling. For the unpurified sample, side chains of tryptophan are visible in the spectrum (~ 10 ppm ^{1}H chemical shift), as well as asparagine and glutamine side chains (~ 7 ppm ^{1}H chemical shift). Note that the red spectrum of non-purified bO was recorded in a higher field (700 MHz) than the red one (600 MHz), further emphasizing a reduced spectral quality.

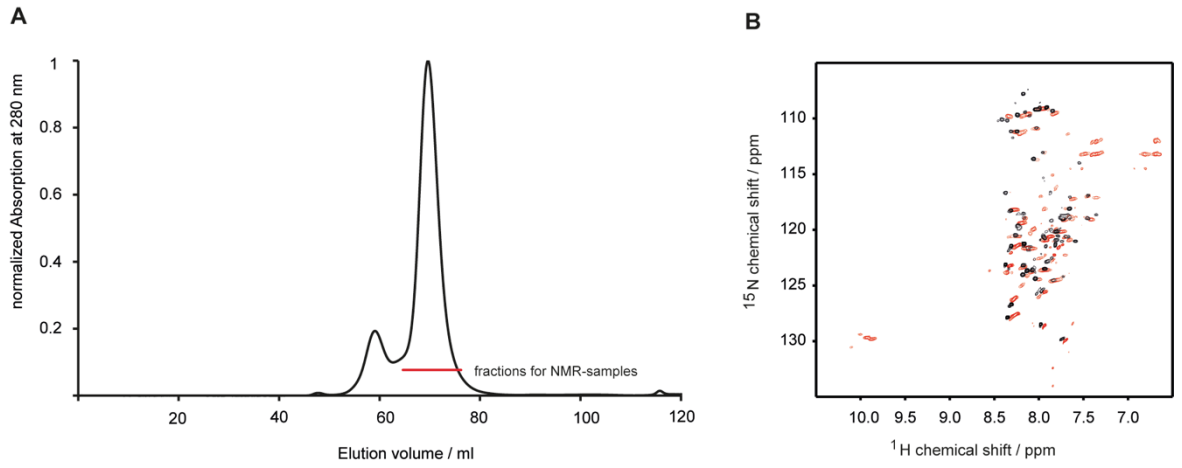


Figure 16: Optimization for NMR studies. (A) Normalized preparative size exclusion chromatogram of bacterioOpsin in SDS micelles; (B) Comparison of TROSY-HSQC spectra of purified (black) and unpurified bO (red). Both were produced via cell-free expression. Concentrations of 100 µM bO were used. NMR data were acquired at 600 MHz (black) and 700 MHz (red) at 35°C.

Even though we could detect some negative effects in terms of spectral quality, when using unpurified, cell-free expressed bO, we decided to proceed with our NMR analysis by using non-purified samples of bO. The reason is that refolding of bO was considerably less effective for purified bO. As discussed in more detail in section 10.3.3.1, several attempts to optimize refolding conditions of purified bO, including various crowding agents, did not yield sufficient amounts for further NMR studies under the applied conditions. This is in strong contrast to non-purified bO which refolds at very high yields. Therefore, using unpurified cell-free expressed bO is a compromise between moderately reduced spectral quality and strongly increased refolding efficiency.

In addition to sample purity, the detergent to protein ratio can have strong effects on NMR spectral quality as well as on refolding yield. To counter the anticipated spectral crowding due to reduced structural dispersion of the unfolded state, we further explored different isotope labelling patterns (see the following chapter 10.3.2.2 for detailed information about the applied labelling schemes).

In general, the results presented in Figure 17 show that our applied conditions permit the acquisition of high-resolution insights for isotope enriched bO in SDS micelles. Interestingly, the data show a large dependency of spectral quality on the SDS concentration and it can be stated that high-quality NMR data can only be recorded at significantly higher SDS concentrations as required for basic bO solubilization. Furthermore, the inserts in Figure 17 show the presence of larger fractions of soluble monomeric SDS molecules in the samples that provide well-dispersed bO spectra. It can therefore be concluded that NMR spectral quality correlates with the availability of monomeric SDS molecules. In addition, our data suggests that the occurrence of soluble monomeric SDS molecules does not follow the expected critical micelle concentration (CMC ~ 5 mM) in presence of bO.

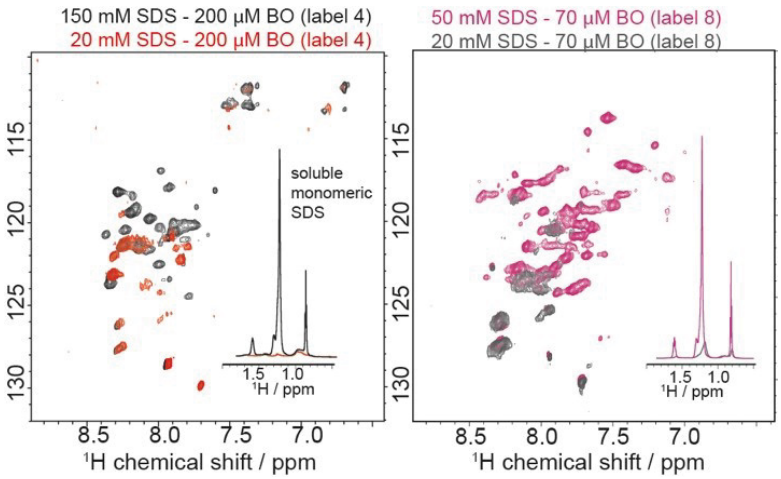


Figure 17: Effects of SDS concentration and SDS-to-bO ratio on NMR spectral quality and refolding properties. Overlay of identical isotope labelled samples and NMR experiments, but using different SDS concentrations and therefore SDS-to-bO ratios. Inserts show 1D ^{1}H spectral region of soluble monomeric SDS for the respective samples scaled for same protein intensities. NMR data were acquired at 600 MHz (Label 4) and 700 MHz (Label 8) at 35°C.

Our data would be most consistent with a scenario, in which SDS monomers preferentially interact with bO, and saturate its interaction sites. Our data show that in the range of 1:280 a small fraction of the SDS molecules do not interact with bO (Fig. 17 right insert, grey). Under these conditions the NMR spectrum, however, is still not well resolved.

When further increasing the ratio to above 1:700 the spectrum becomes well resolved, correlating with the presence of considerable fraction of monomeric non-interacting SDS (Fig. 17 right insert, purple).

This provides interesting, new insights into the molecular mechanisms of surfactants for the stabilization of hydrophobic proteins in solution. In consequence of these results we used a minimum bO-to-SDS ratio of 1:700 for further NMR experiments, enabling resonance assignments by combinatorial labelling. Sequential assignment via combinatorial isotope labelling.

10.3.2.2 Sequential assignment via combinatorial isotope labelling

Since the relaxation properties of bO in SDS micelles prevent long magnetization pathways, normally required for resonance assignments based on 3D correlation NMR spectra, we pursued the concept of combinatorial isotope labelling. For the procedure of sequential resonance assignments via combinatorial isotope labelling we used a strategy based on the approach by Parker *et al.*[334]. This strategy uses only simple 2D experiments, with distinct samples, which harbor different labelling patterns. Amino acids of each pattern can be only ^{15}N or ^{13}CO labelled, as well as uniformly labelled with ^{15}N and ^{13}C. Comparisons of 2D experiments are leading to identification of the present amino-acid type and ideally to subsequent residue-specific resonance assignments. Figure 18 illustrates the applied 2D NMR experiments necessary for the used strategy.

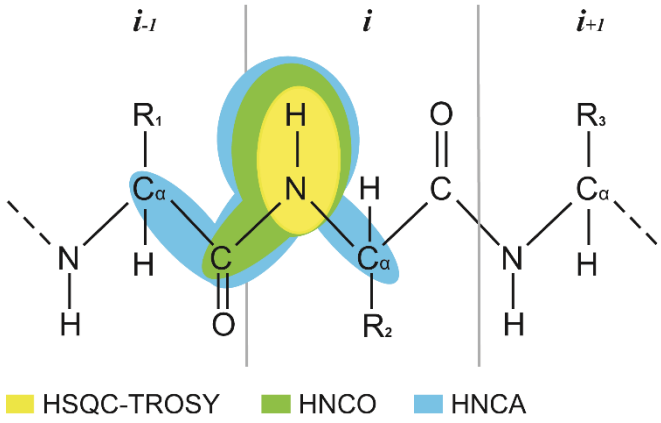


Figure 18: 2D NMR experiments for sequential resonance assignments. The figure shows the applied NMR experiments and the detected atoms of labelled amino acids along the protein backbone.

Specific isotope labelling was achieved by the use of cell-free expression (see chapter 14.1). Two main features make cell-free expression the method of choice for our strategy: Cell-free expression supports custom labelling strategies[335–337], while displaying a very low

level of metabolic scrambling[338] compared to cell-based methods for introducing isotope labels into proteins. Furthermore, cell-free expression offers a well-defined, clean starting point, because expressed proteins are not biased by coordinated and co-purified lipids[337], which is the case when isolating bO from the purple membrane of *Halobacterium salinarium*[339] .

As surfactants decrease the tumbling rate by forming protein-surfactant complexes (PSCs), high-resolution NMR insights can be challenging. In addition, the NMR signals of flexible parts of proteins, such as extra-membrane loop regions or termini can appear very strong in spectra, because of their NMR-supportive properties. This can result in an unfavorable broad range of peak intensities.

Table 2: Assignment table. Assigned resonance frequencies are listed with their corresponding amino-acid type and residue number in the primary sequence of bO.

Residue-No.	Assigned amino acid	^{1}H chemical shift /ppm	^{15}N chemical shift /ppm
34	W	7.658	116.240
105	Y	7.442	118.882
124	D	7.994	122.571
130	I	8.050	121.70
153	Y	8.086	123.760
198	N	7.858	118.035

However, even complex PSC structures can be analyzed with a smart strategy. The strategy of combinatorial labelling strongly reduces the complexity of spectra, and through combination and comparison of different spectra resonance assignments can be made, even with the portrayed challenges.

In the first steps cell-free expressed [^{13}C-^{15}N]-labelled bO was solubilized in SDS buffer (see section 13.3.4). Afterwards, a ^{1}H-^{15}N-correlation TROSY-HSQC spectrum was acquired as a reference. In the following steps 2D spectra of the different labelling patterns of bO were acquired by using [^{1}H-^{15}N]-HSQC-TROSY, ^{13}CO-filtered HSQC and ^{13}Cα-filtered HSQC experiments. For the creation of efficient, bO-specific labelling patterns (see chapter 15.1.1) the UPLABEL algorithm[340] was used. Figure 19 shows the results of sequential resonance assignments.

The [^{1}H-^{15}N]-HSQC spectrum of cell-free expressed [^{13}C-^{15}N]-labelled bO in SDS micelles (Fig. 19A) shows limited dispersion, indicative for unfolded proteins. For the bO construct 271 peaks were expected (Fig. 19B), but only a subset could be resolved in the present spectrum. The peak overlap leads to heterogeneous peak-shapes, which may be a result of line-broadening, heterogeneous conformations of bO and different particle sizes.

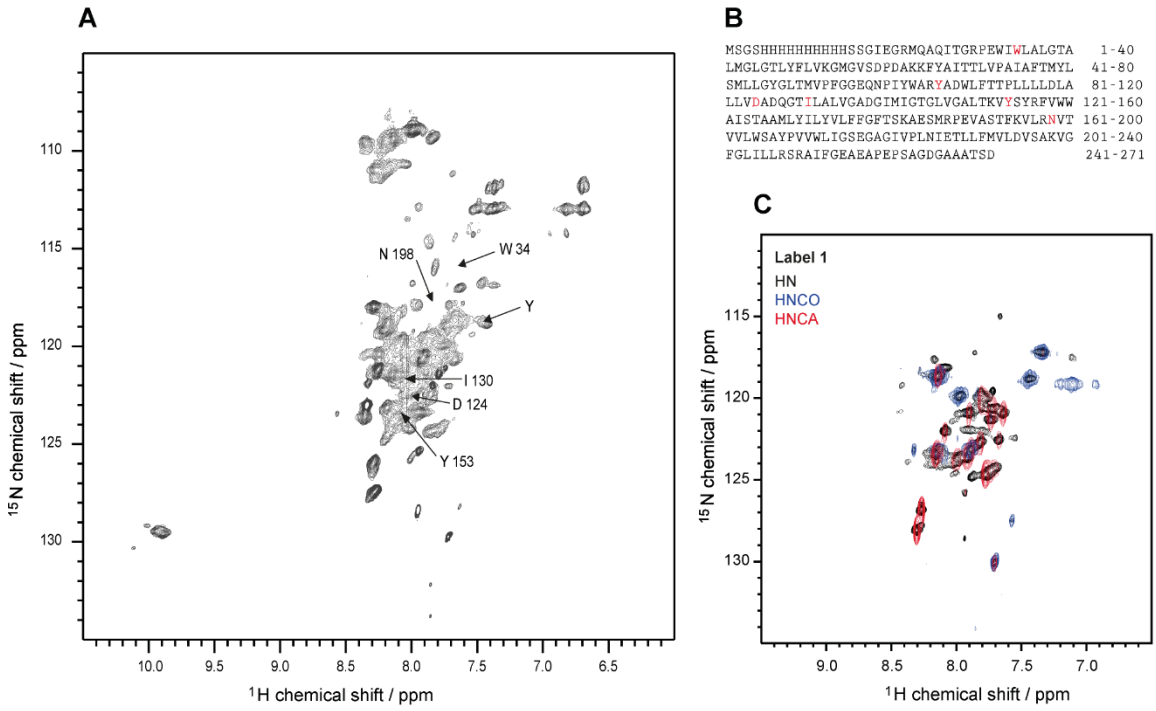


Figure 19: Partial resonance assignment of bO in SDS micelles. (A) [^{1}H-^{15}N-]-HSQC spectrum of cell-free expressed [^{13}C-^{15}N]-labelled bO. Assigned residues are marked with arrows. (B) Primary sequence of the used wildtype bO construct. Highlighted residues show assigned resonance frequencies using combinatorial labelling (red). (C) Overlay of a [^{1}H-^{15}N]-HSQC-TROSY (black) and the [^{1}H-^{15}N]-dimension of a HNCO (blue) and HNCA (red) acquired on SDS-solubilized bO, harboring labelling pattern 1 (see chapter 15.1.1). Labelled amino acids according to labelling pattern 1: [^{13}C] = FPV; [^{15}N] = Y; [^{13}C-^{15}N] = RAM. Spectra were acquired at 35 °C and 600 MHz (C) or 700 MHz (A) (see chapter 15.1.3 for details).

While the uniformly labelled [^{1}H-^{15}N-]-HSQC-TROSY spectrum provides information about the occurrence of all ^{15}N-labelled amino acid in the sample (Fig. 19C, black), this information is specified through the acquisition of an HNCO spectrum (Fig. 19C, blue). Visible peaks in the shown HNCO spectrum derive from the carbonyl-group of ^{13}C-labelled residues (i-1), preceding a ^{15}N-labelled residue (i). To extend the information of these two spectra, an additional HNCA was acquired. Peaks in the HNCA spectrum originate from [^{13}C-^{15}N]-labelled (uniformly labelled) amino acids i and i-1. By applying this strategy using 9 different labelling schemes (see section 15.1), it was possible to assign resonances from 6 residues of the 271-amino acid SDS-solubilized bO with high certainty (Tab. 2).

Additionally, several other amino acids could be identified, but were not assignable with certainty.

In conclusion, the approach of combinatorial labelling could provide access to (a limited number) of specific resonance assignments. Since resonance assignments are the prerequisite for site-resolved NMR insights and since the SDS-solubilized state of bO would elute conventional methods for resonance assignment, even the limited number of obtained resonance assignments are the key to a further understanding of the fundamental properties of unfolded bO. This also holds true in respect to understanding the protein folding process, including the occurrence of transition states, which requires profound knowledge of the starting point and means to follow the folding process with time- and site-resolved experiments.

10.3.2.3 Solvent interactions

Solvent interactions may be the most striking feature of the unfolded state of bO. In order to understand the transition of the unfolded state to the folded state of bO, it is necessary to characterize the starting point of the folding process including the environmental effects. This might also give important information about the failure or success of some refolding strategies, especially when it involves transferring bO from one membrane mimetic into another one.

Folded states of bR have already been subject to analyzing its solvent interactions in different environments, such as detergents, amphipols and nanodiscs[225,341]. In addition, several attempts have been made to characterize the unfolded or partially unfolded state of bO[342–345]. Unfortunately, all these studies are based on extraction of bR or bO from the membranes of expression hosts, for instance *Halobacterium species* or *E. coli*. As a result, findings of these studies are biased by co-purified lipids, even though elaborate procedures for delipidation of extracted bO exist[346]. Furthermore, purified bR or retinal-free bO need to be transferred into an unfolded, retinal-free state by denaturants or other procedures. In case of bR, this is not easy, because of its known high stability. Denaturants, which completely eradicate the structure of most soluble proteins, have only slight effects on bR. BR resists treatments with 8 M guanidinium chloride[347] and random coil conformation is only reported for treatments with trifluoroacetic acid or formic acid[347]. In addition, so far no high-resolution structural study of unfolded bO was attempted.

For the studies shown in this work cell-free expressed bO was used to maintain ligand- and lipid-free conditions. Furthermore, the formation of tertiary structures, prior to any conducted experiments, should be prevented by the cell-free precipitation expression method. This should facilitate the investigation of bO's unfolded state.

For the investigation of the solvent interactions of unfolded, SDS-solubilized bO a Nuclear Overhauser Enhancement Spectroscopy (NOESY)-spectrum (Fig. 20) was acquired, which shows residue resolved interactions of the protein with water and SDS molecules.

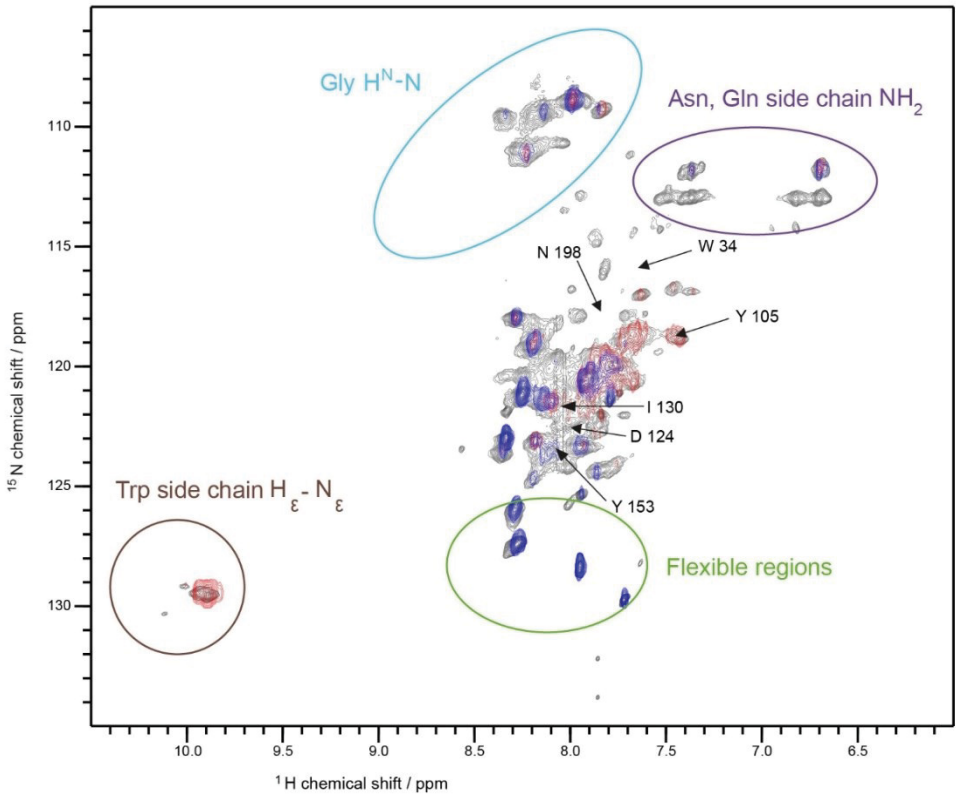


Figure 20: Solvent interactions of unfolded bO with SDS and water molecules. Overlay of the [^{15}N-^{1}H]-HSQC spectrum of cell-free expressed [^{13}C-^{15}N]-labelled bO with 2D H-N-planes extracted from the 3D TROSY-NOESY at the water frequency at 4.7 ppm (blue) and SDS's hydro-carbon frequency at 1.7 ppm (red), respectively. Assigned residues are marked with arrows. Spectra were acquired at 35 °C and 700 MHz (HSQC) or 750 MHz (TROSY-NOESY) (see chapter 15.1.3 for details).

Due to the different properties of amino acids, it is expected, that hydrophobic amino acids would tend to interact with the hydro-carbon chain of the SDS molecules, while charged residues, as well as polar or neutral amino acids are expected to interact with water molecules or the sulfate groups of SDS.

Tryptophan side chains appear in the region of ~10 ppm ^{1}H chemical shift and ~130 ppm ^{15}N chemical shift in H-N correlation-based spectra. The Nuclear Overhauser Effect (NOE) peak in this region shows that some of bO's tryptophan side chains are covered by SDS under the applied conditions. As SDS is known to bind to hydrophobic side chains, this interaction follows the expectations. The NOE peaks in the glycine region (~8 ppm ^{1}H chemical shift and ~110 ppm ^{15}N chemical shift) show an interaction of glycines with water and SDS. This interaction differs for different glycines pointing to sequence-specific interactions. Asparagine- and glutamine side chains are visible in the spectrum by two peaks appearing at the same ^{15}N chemical shift, but with differing ^{1}H chemical shifts. In the presented spectra, these side chains are visible at ~6.7- 7.5 ppm ^{1}H chemical shift and 112 ppm ^{15}N chemical shift. In this region NOE-peaks for water- and SDS-interaction could be detected, as also observed for the glycine residues.

Flexible regions, such as the C-terminus and loops are represented by strong peaks at 7.5-8.5 ppm ^{1}H chemical shift and 125-130 ppm ^{15}N chemical shift. Because of the NMR-favorable properties, flexible regions often lead to dominant peaks in the spectrum. In line with their expected solvent interactions, these residues show strong NOE peaks with the H_2O molecules in the presented spectrum.

In order to create a deeper understanding of the observed interactions the 2D H-N-planes, extracted from the 3D TROSY-NOESY, at the water frequency (4.7 ppm) and the SDS frequency (1.7 ppm) were analyzed, using four of the previously assigned residues (Fig. 21). The data allows to obtain residue-specific quantitative information of the environmental interactions of the unfolded state. The corresponding NOE signal intensities of the assigned residues are shown in Figure 21B.

The extracted 2D H-H-planes for Y105 show a clear NOE peak at the SDS frequency (Fig. 21A) and only very weak NOEs for the interaction with water molecules that are below the shown contour level. Due to the polar nature of tyrosine residues at physiological pH, its side chains are expected to interact either with water molecules or sulfate-groups of SDS. However, the peak under consideration reports on the tyrosine's backbone (amid proton). This result demonstrates the good spatial resolution of the applied NMR methods that is well capable to distinguish the interaction of the Y105's backbone with the hydrocarbon chain from the expected solvent exposed side chain. When neglecting potentially different relaxation properties, the data further suggest that the SDS interaction are about five-fold more frequent during the course of the experiment as the interaction with water molecules.

On the contrary, the NOE signal intensities of Y153 account for a predominant interaction with water molecules, as the intensity for water is almost two-fold higher, than for SDS. This result shows that the environmental interactions of bO are not solely based on the amino-acid type but also depend on its position within the protein, pointing to a preferential arrangement of 'unfolded' bO in the given environment.

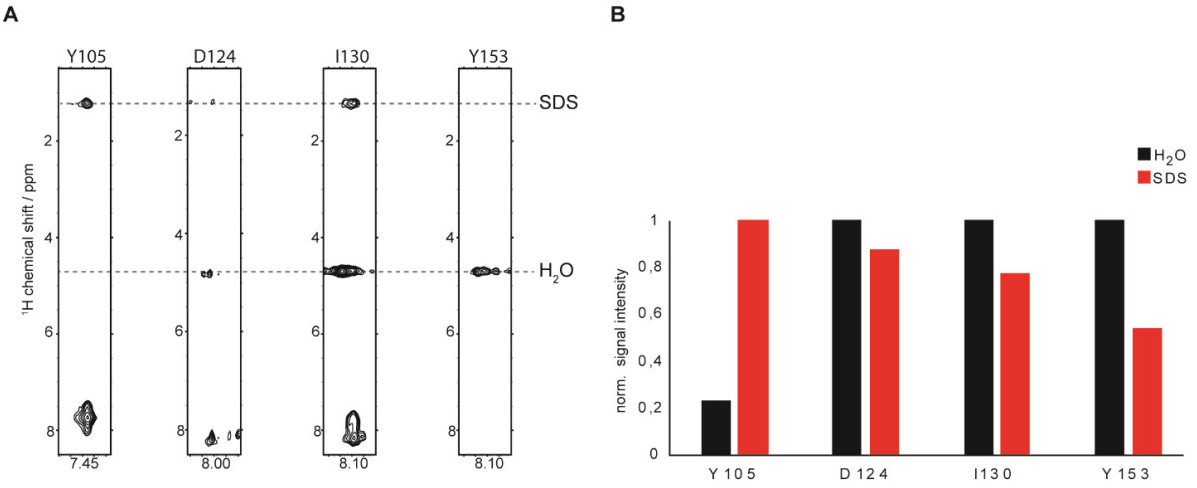


Figure 21: Residue-specific interactions with water and SDS. (A) 2D H-H-planes, extracted from the 3D TROSY-NOESY, at the respective ^{15}N frequencies of 4 assigned and resolved residues of bO. (B) Normalized signal intensities of the detected NOE crosspeaks for water and SDS interactions of 4 assigned residues of bO.

For D124 an interaction with water molecules was observed, as well as with SDS from the detected NOEs (Fig. 21A). The comparison of both signal intensities reveals a slightly higher interaction with water molecules than with SDS. We observed the same interactions as for D124 for I130, while the signal intensity for the SDS interaction is slightly lower for the shown isoleucine residue. Since simultaneous interactions with H_2O and SDS are unlikely due to steric restrains, the obtained data either point to a mixed heterogeneous conformational state in which about 50% of the bO molecules interact at the given position with water and the rest with SDS. Alternatively, a dynamic exchange between water- and SDS-exposed states could also explain the data. The latter is further supported by the absence of peak splittings that would be expected in case of a static two-state ensemble.

There are existing manifold theories and postulations about the mechanism of protein solubilization by SDS, but the exact mechanism remains unclear and probably varies depending on protein structure and properties.

Our experiments regarding the optimization of bO conditions for NMR studies, revealed that the behavior of monomeric SDS molecules does not follow the expected CMC. This suggests that SDS monomers preferentially interact with bO and saturate its binding sites, resulting (only) in soluble, monomeric SDS molecules after saturation of bO with approximately 280 SDS molecules per bO molecule. Under the assumption that the approximated 280 SDS molecules are constantly in contact with bO and considering the expected size of (protein-free) SDS micelles, it is likely that bO is not inserted in a 'standard' spherical SDS micelle, which would require considerably less SDS molecules. However, a dynamic exchange of SDS molecules cannot be excluded and could also explain the rather high number of SDS molecules required for interaction saturation of bO.

For now, a rough overview about the residue specific interactions of bO with the solubilizing surfactant SDS and water has been given. But for a closer interpretation of the presented findings, it helps to take a look at the location of these residues in the secondary structure of folded bO. Although the occurrence of secondary structure elements is unclear in the investigated cell-free expressed bO, some important conclusions may be drawn concerning the success of refolding strategies for SDS solubilized bO and the structure of the PSC in general. Figure 22 gives a schematic overview for our assigned residues with regard to their location in helices or loop regions of bR.

Krishnamani *et al.* investigated the secondary and tertiary structure of bO in the SDS denatured state via pulsed electron paramagnetic resonance spectroscopy (double electron-electron spin resonance, DEER)[344]. Through analysis of intrahelical distance distributions they found that much of bOs helical content is retained in SDS, but suggesting heterogeneous populations of disrupted and intact segments[344]. Earlier experiments indicated an amount of overall helicity of roughly ~50% for bO in SDS, while it was unclear which segments were preserved in detail[348]. These findings are in contrast to photo-oxidation experiments, which revealed an unfolding of helix A and D in SDS, but not for the other helices[349].

In terms of tertiary structure Krishnamani and co-workers could show, through measurements of interhelical distances, that very little of bOs tertiary structure remains in SDS[344]. They observed that SDS displaces hydrophobic protein-protein contacts with detergent-protein contacts, which leads to the loss of most tertiary contacts of bOs helical pairs in presumed SDS micelles[344]. This is generally in accordance with previous findings in NMR experiments. The group of Pervushin analyzed the isolated helical pairs A and B,

which showed almost native-like α-helical content, while tertiary contact is abolished, suggesting that the two helices become separated through SDS solubilization[342,343].

Furthermore, Krishnamani and co-workers conducted molecular dynamics simulations (MD) of the unfolding of individual helices of bacteriorhodopsin in SDS[345]. Their results indicate that partitioning of helices in micelles and their stability are dependent on the hydrophobicity of the TM segments. Additionally, their *in silico* experiments strongly indicate that overall structural stability of each segment of bO correlates with its partitioning to the micellar surface and interaction with polar groups[345]. Helices A, B, and E were found to be stable and retain their secondary structure during the simulation time. During simulations these helices were observed to be embedded in the micellar core, while helices C, D, F and G participate to the SDS-water interface and display structural perturbations immediately. Notably, these structural instabilities were detected near charged residues in TM segments.

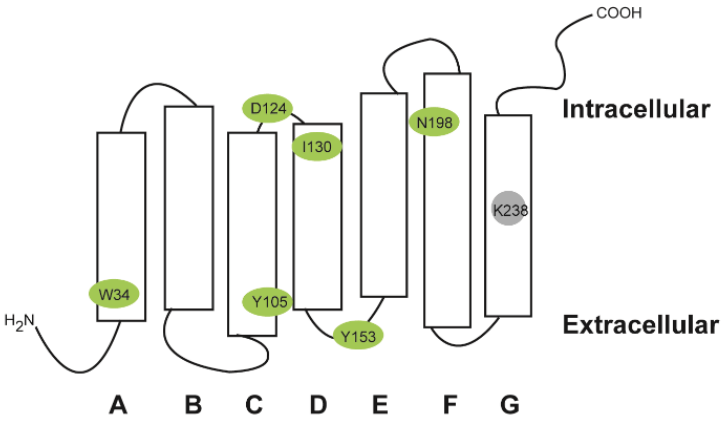


Figure 22: Schematic overview of the secondary structure of bacteriorhodopsin. Alignment and length of the seven-helical transmembrane segments are based on the structure of bR obtained by X-ray and electron diffraction[350,351]. Residues assigned by combinatorial sequential assignment strategies are circled in green. Lys238 (circled in grey) of the used construct is equivalent with Lys216 in the native sequence of bO, which is located in helix G and forms the site of attachment of retinal via Schiff base linkage.

Unfortunately, all of the mentioned experimental studies are based on membrane extracted bO and in this context we cannot exclude bias of co-purified lipids on the structural properties of bO.

Additionally, studies show that transformation of bacterial-expressed, folded bR into a completely unfolded random-coil state of bO is not possible using SDS[342–344,348,349]. Even though these studies provide useful information about the behavior of partially unfolded bO in SDS, which might be viewed as a transition state between the folded and completely unfolded state, they lack information about the unbiased unfolded state of bO in SDS.

Our measurements are aiming at filling this gap. In summary, presented results reveal that side chains of tryptophan are predominantly interacting with the hydro-carbon chain of SDS under the used experimental conditions, while flexible regions are predominantly interacting with water. Most residues, including all assigned residues, as well as glycines and side chains of asparagine and glutamine interact with water and SDS molecules. Analysis of NOE signal intensities for some of the assigned residues enables a closer look at the interactions regarding their partitioning between water and SDS molecules. D124 and Y153 are located in conjunction loops between segments of bO. Both residues show a high intensity for interaction with water molecules, while the NOE intensity of SDS interaction differs, suggesting that D124 is located in the interface of water and surfactant, while Y153 is more exposed to water molecules than to SDS. The NOE signal intensities of I130, which is sequentially close to D124 and located in the N-terminal part of TM segment D, likewise account for a location at the interface of water and SDS. Y105 is located in the N-terminal part of helix C and exhibits the strongest NOE signal intensity for SDS interaction, indicative for a more "water-shielded" position in the solubilizing environment.

The results shown in this work may create additional evidence towards the previously suggested scenario in which SDS fully saturates bO's binding sites. This would also be in accordance with the findings that protein-protein contacts are displaced with SDS-protein contacts by Krishnamani *et al.*[344]. In a random-coil structure binding sites would also be available for SDS interactions, which are normally shielded by helical interactions or involved in helix formation through H-bonding. This might also give a possible explanation for the difference in spectral quality in dependence of the SDS concentration in the measured samples. Full saturation of possible binding sites of bO with SDS might lead to more homogeneous PSCs which could result in higher quality NMR spectra. On the other hand, not fully saturated binding sites of bO would lead to lower resolution and dispersion in NMR spectra by formation of heterogeneous PSCs in this scenario. It is conceivable that some of these generated, heterogeneous PSCs could facilitate the transfer into other membrane mimetics enabling refolding strategies.

For future studies, it would be interesting to perform further NMR experiments, which could reveal the structure of the SDS environment in this context and their residue-specific interactions with bO in its unfolded state. But in order to conduct these studies a more comprehensive assignment of bO under these conditions is necessary. Furthermore, ^{13}C-based experiments could lead to a deeper understanding of the exact structure bO adopts under

these conditions. Additionally, techniques to generate unfolded bO from folded bR, such as thorough bleaching in appropriate membrane mimetics, could be applied in order to compare the unfolded forms and, as a future perspective, facilitate the examination of the unfolding process via NMR techniques.

10.3.3 From bO to bR - Refolding of bacteriorhodopsin

After investigation of the unfolded as well as functionally folded state of bacteriorhodopsin, the possibilities to obtain high-resolution NMR insights of the full folding process were further explored. Several *in vitro* strategies have been published so far for the refolding of bO, which mostly involve the transfer of SDS-solubilized bO (partially denatured) into renaturing membrane mimetics, such as bicelles[347,352], micelles[225,352], vesicles[353,354], nanodiscs[225] and amphipols[214,225]. Even though refolding of bO can be a straightforward procedure nowadays, folding kinetics using most common membrane mimetics are usually much too fast to be captured in detail by high-resolution techniques. As a result folding kinetics must be optimized towards the NMR timescale in order to acquire high-quality spectra. The next section will summarize obtained results concerning refolding kinetics of bO and their utilization for further measurements.

10.3.3.1 Kinetic insights into refolding of bR

Several folding pathways of bO are conceivable. One of them is postulated and refined based on the two-stage-model for α-helical membrane proteins[355], involving diverse intermediate states[356,357]. But all published folding pathways agree on the key role of retinal in the folding process. But whether binding of retinal enables or stabilizes helical contacts[358] or allows the helices to pack around the bound retinal[359] remains unclear.

Detection of Schiff base formation via the characteristic shift in retinal absorbance is an established procedure to follow folding of bO. The observable color shift to an absorbance maximum at 555 nm is indicative of retinal binding and formation of the correct binding pocket and therefore a reliable reporter on formation of fully folded, retinal-bound bR.

To be able to screen a larger set of conditions, a refolding assay based on absorbance measurements in a 384-well format was established first. The assay enables fast screening of various folding conditions using small volumes of only 20 µl, allowing the efficient use of produced protein and screening reagents. During the folding process

the absorption is monitored using three wavelengths in parallel, i.e. 380 nm, 555 nm and 800 nm. 380 nm corresponds to free retinal, while the absorption at 800 nm corresponds to the baseline for each experiment, facilitating the correction of measured spectra. Additionally, a full absorption spectrum ranging from 260–-900 nm is detected before and after the refolding assay, further aiding subsequent data processing and analysis. A detailed experimental-setup description of our established assay is given in chapter 14.2.

We used this assay to investigate several hundred different refolding conditions, including the effects of pH, salts, protein purity, crowding reagents as well as protein, retinal and surfactant concentrations. In the following section the most important factors will be discussed. The first one is the ligand concentration of *all-trans* retinal, which binds covalently to Lys216 of helix G (Fig. 22) during the folding process via Schiff-base linkage[25,352,353,360]. Another important parameter is the concentration of the used renaturant, namely the amphipol (APOL) concentration. Additionally, the concentration of denaturant and bO also play important roles in the refolding process. The denaturant should enable refolding by keeping the MP soluble, while enabling the transfer into other membrane mimetics. Properties of denaturants can also be modified by altering salt concentrations of the tested samples. In chapter 10.3.2 we already pointed to the importance of SDS to bO ratio for the folding process and spectral quality. Figure 23 shows the impact of the retinal concentration on the refolding kinetics of bO.

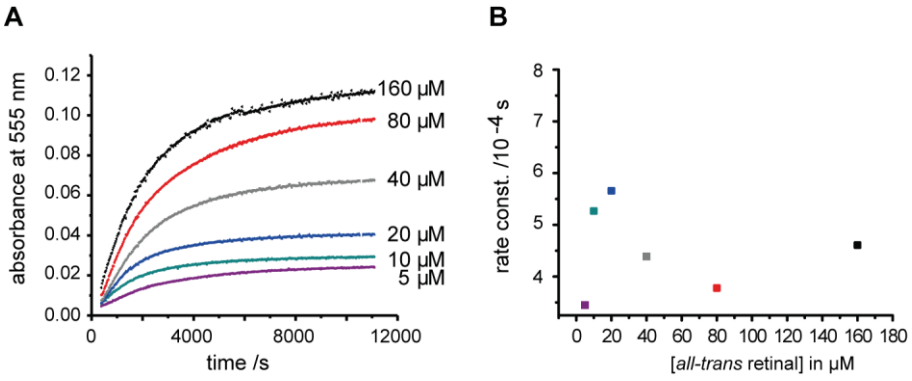


Figure 23: Impact of the retinal concentration on bO's refolding kinetics. (A) *All-trans* retinal concentrations (5-160 µM) were tested using 10 µM bO, 10 mM SDS and 2% APOL (v/v). (B) Corresponding kinetic rate constants for every tested retinal concentration.

The data shown in Figure 23 illustrate the dependence of bO's refolding yield and speed on availability of *all-trans* retinal. Retinal concentrations were ranging from 5 to 160 µM, while the concentration of bO was kept constant at 10 µM. The experimental data show an increase in refolding yield with increasing retinal concentration (Fig. 23A). Equimolar ratios of bO to retinal resulted in very low refolding yields, which increases with higher retinal concentrations. Highest refolding yields were obtained for retinal concentrations at 80 to 160 µM. These findings suggest that an excess of retinal is beneficial for bO refolding. Even though only one retinal molecule binds covalently via Schiff-base linkage to Lys216[361].

It can be assumed that an excess of retinal could increase the propensity to bind successfully into bO's binding pocket, resulting in functionally folded bR. It is known that retinal binds in different stages to bO[361–363]. The folding pathway of bO involves a loosely attached retinal, which is then transferred into a covalently bound ligand at Lys216[360]. The loosely attached retinal may pass multiple cycles of attaching and detaching in the folding process before its covalent bond formation. Additionally, there might be a general competition between SDS and retinal regarding potential binding sites. This would also include residues forming the retinal binding pocket.

The influence of retinal concentration on the speed of refolding kinetics is reflected by the determined rate constants (Fig. 23B). A retinal concentration below the bO concentration leads to the lowest rate constant under the applied conditions, while ligand concentrations in the range of equimolar to 16-fold ratios to bO display higher rate constants of 4.4- 5.7 /10^{-4}s. The only exception is the rate constant at a retinal concentration of 80 µM, which is below 4 /10^{-4}s, similar to that at 5 µM retinal. These findings neither support nor oppose the assumption that an excess of retinal could increase the propensity to bind successfully into bO's binding pocket, leading to functionally folded bR. Due to clear beneficial effects of an excess of retinal and the lack of a clear correlation between retinal concentration and refolding rate, the data indicate that a considerable amount of retinal is trapped irreversibly (on the investigated time scale) in a misfolded protein or surfactant state.

In addition to the importance of retinal, another key role in the refolding process is played by the concentration of membrane mimetics, i.e. amphipols and SDS. To investigate their influence on bO refolding kinetics, additional experiments were conducted, which are summarized in Figure 24-28.

Tested APOL concentrations ranged from 1 to 4% (v/v) and show a big influence on the speed of the bO refolding process, which is illustrated by Figure 24.

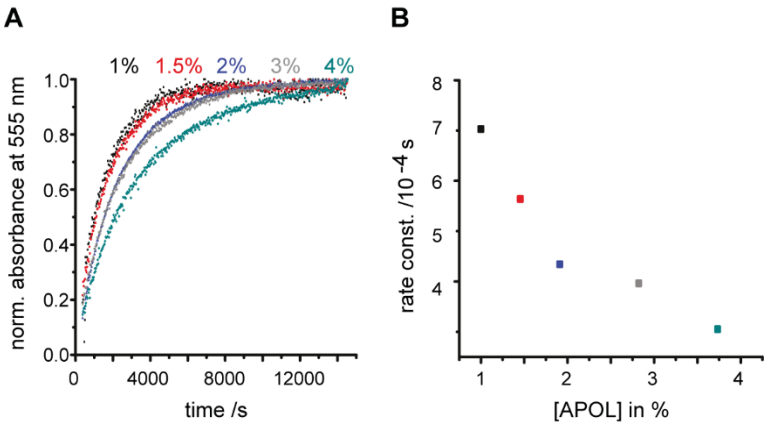


Figure 24: Impact of the APOL concentration on bO's refolding kinetics. (A) APOL concentrations of 1-4% (v/v) were tested using 10 µM bO, 10 mM SDS and 80 µM *all-trans* retinal. (B) Corresponding kinetic rate constants for every tested APOL concentration.

Determined rate constants decrease with increasing presence of APOLs, meaning that the refolding speed is lowered by higher APOL concentrations. In terms of optimizing bO's folding kinetics for the NMR timescale conditions are favored, which exhibit slow rate constants while maintaining high refolding yields. High APOL concentrations may, however, lead to an increase in particle size, which is compromising NMR spectral quality though affection of the tumbling rate.

In addition to the refolding rate, the refolding efficiency can be considered the most important parameter. Refolding yields for each tested APOL concentration are shown in Figure 25.

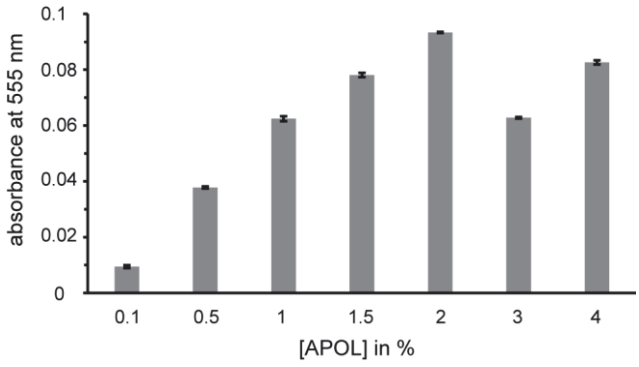


Figure 25: BR absorbance after refolding with different APOL concentrations. APOL concentrations of 1-4% (v/v) were tested using 10 µM bO, 10 mM SDS and 80 µM *all-trans* retinal.

Total refolding yields of bO increase with higher APOL concentrations until a concentration of 2% APOL is reached, according to the applied absorption measurements (Fig. 25).

Afterwards, refolding yields decrease again. Similar absorbance values were detected for concentrations of 1 and 3% APOL and respectively for 1.5 and 4% APOL. The findings show a correlation between APOL concentration and bO-refolding kinetics in the tested concentration range.

As APOL concentrations of 2.2% have been proven useful in terms of spectral quality in the past[214,225], new data (Fig. 24 and 25) provide more evidence to maintain a concentration range between 1-2% APOLs for further NMR-based studies of bR.

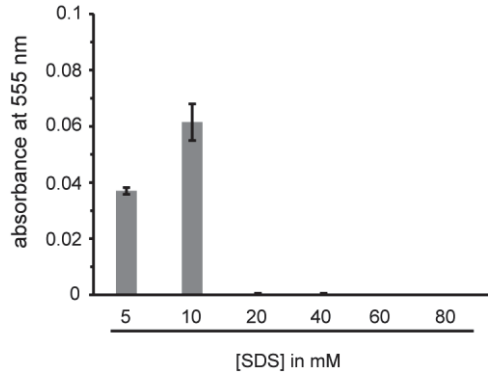


Figure 26: BR absorbance after APOL refolding with different SDS concentrations. SDS concentrations from 5-80 mM were tested using 10 µM bO, 80 µM *all-trans* retinal and 2% APOLs (v/v).

To investigate the effect of the SDS concentration on the bO-refolding process, refolding absorbances of constant 10 µM bO with increasing SDS concentrations were measured (Fig. 26).

The data clearly show that usage of 10 mM SDS results in most effective refolding. While at higher concentrations the absorbance drops very strongly, a lower SDS concentration (5 mM) also supports effective refolding but seems to have negative effects on the total refolding yield. To test whether the SDS concentration or the SDS:bO ratio is the relevant parameter for refolding, subsequently the bO concentration was varied at a constant SDS concentration of 10 mM SDS (Fig. 27).

The optimal value of 10 mM SDS was chosen. A comparison between the total refolding yields (Fig. 27A) and their normalized values (Fig. 27B) reveals that refolding efficiency is lowered by increasing bO concentrations. The measured absorbances show a strong dependence on the bO concentration suggesting that indeed the SDS:bO ratio is an important parameter.

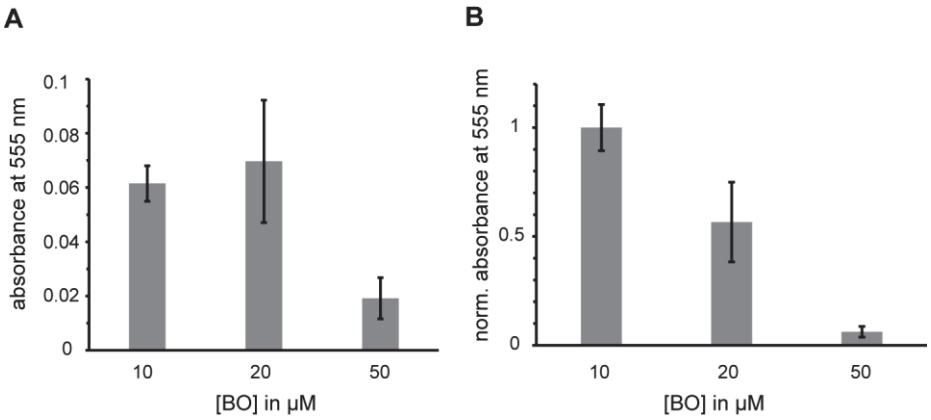


Figure 27: BR absorbance after APOL refolding using different bO concentrations. (A) BR absorbance after refolding in APOLs. BO concentrations of 10-50 µM were tested using 10 mM SDS, 2% APOLs (v/v) and 160 µM *all-trans* retinal. (B) Normalized absorbance of (A).

Based on previous NMR observations, it was already found that bO in particular interacts with monomeric SDS (see chapter 10.3.2). Interestingly, the strong dependence of the CMC of SDS on the ionic strength of the buffer can be exploited to modify the concentration of monomeric SDS without changing the total SDS concentration. To test whether the availability of monomeric SDS is also correlated with the refolding of bO, additional refolding assays were performed at two different constant SDS concentrations (5 mM and 2.5 mM) and varying NaCl concentrations (Fig. 28).

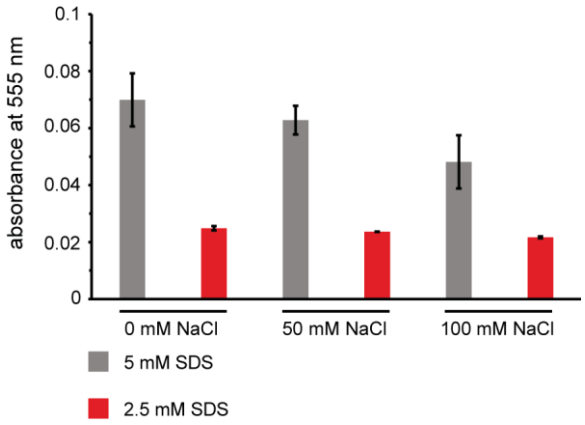


Figure 28: BR absorbance after APOL refolding using different SDS and NaCl concentrations. NaCl concentrations of 0, 50 and 100 mM were tested using 10 µM bO, 80 µM *all-trans* retinal and 2% APOLs (v/v). SDS concentrations of 2.5 mM (red) and 5 mM (grey) were used.

The value of 5 mM roughly reflects the expected CMC of SDS in the used buffer without addition of NaCl. In the absence of bO it can hence be assumed that nearly all SDS molecules are monomeric. When increasing the NaCl levels the CMC is reduced and consequently less monomeric SDS is present in the sample. When assessing the total refolding yield of bO under these conditions it can be seen that indeed the yield drops when less monomeric SDS

is present. Since the CMC will not drop below 2.5 mM for the used NaCl conditions, the monomer concentration should be rather constant in the case for the 2.5 mM condition, rendering this measurement series a suitable control. In line with this picture, the refolding yields under these conditions are constant for the tested NaCl concentrations (Fig. 28, red).

Taken together the data demonstrate that the ratio of monomeric SDS to bO is the central parameter of the refolding process. It was consistently observed that this ratio should be above 700:1 (Fig. 27). This value is well in line with the number of SDS monomers required to saturate the bO's SDS-binding sites as estimated by NMR (Fig. 17). When increasing the SDS concentration considerably above the CMC, the refolding yield is strongly reduced. This could either be a direct consequence of the used refolding assay that relies on dilution of the SDS to induce refolding. Alternatively, increasing amounts of SDS micelles may also directly interfere with the refolding process. Overall, the presented results suggest that bO is best capable of refolding, when it is in a state in which the SDS interaction sites are saturated by monomeric SDS but it is not in a micellar environment. Interestingly, in both directions (lower and higher SDS concentrations) protein-protein interactions should be increasing, either due to presence of unsaturated SDS binding sites or condensation within a micelle. It is therefore tempting to speculate that under the tested conditions the refolding process is most effective when it is not biased by preformed protein-protein contacts.

10.3.3.2 Real-time refolding of bR via NMR studies

To investigate the folding transition in more detail a series of NMR experiments were carried out, to capture the folding event in real-time (see section 14.3.5). In this series a reference spectrum of bO in SDS was recorded at first, then the refolding was initiated via addition of 2% APOL and 80 µM *all-trans* retinal to the sample and a second spectrum was recorded during the refolding event. After the transition an additional spectrum was recorded serving as reference of the folded state of covalently retinal-bound bR in APOLs. The results are shown in Figure 29.

While the limited spectral quality of the data does not allow a residue-specific analysis of the refolding process, the time-resolved NMR data still reveals a number of important insights into the folding process. In general, a change of the nuclei frequency (due to the refolding event) during the course of the measurements should lead to a characteristic change of the resulting peak shapes[364–366]. Since the refolding time scale is only captured in the indirect dimension these effects should only be visible in the ^{15}N dimension of the spectrum.

Assuming a simple two-state transition the peak shape of the start configuration (unfolded state) should significantly increase in linewidth, whereas the final state (folded bR) should show narrow lines which are flanked by negative shoulders. Interestingly, these negative shoulders are indeed captured in the real-time spectrum for a limited number of peaks, validating the general setup. Due to the availability of a partial resonance assignment of the APOL folded bR[225], a few peaks of the final state can be sequentially assigned. Noteworthy, all of these peaks belong to the protein's C-terminal region. This region appears to be the only region that shows peak shape modifications that are consistent with a simple two-state refolding transition. While peak overlap prevents a more detailed analysis of the underlying kinetics, the data strongly indicate that this region undergoes a rather homogeneous transition from the SDS unfolded state to the folded bR state.

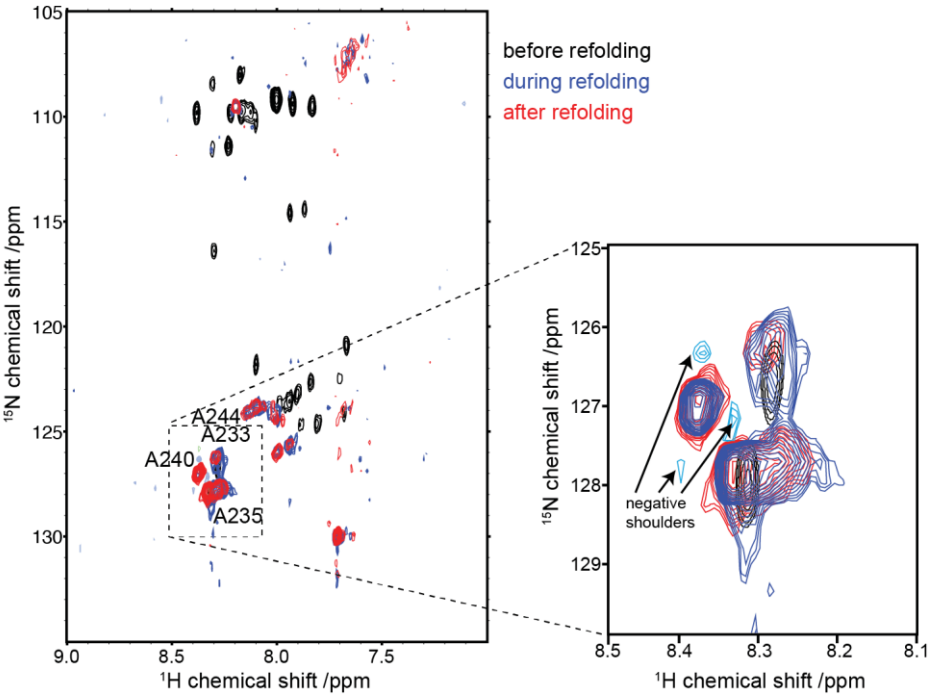


Figure 29: Investigation of bO refolding in real-time via NMR spectroscopy. Comparison of three BEST-TROSY spectra of ^{15}N-AGS-labelled bO. All spectra were acquired at 35 °C and 750 MHz. The black spectrum represents bO in 10 mM SDS and was recorded before refolding was initiated. The blue spectrum was recorded directly after refolding was initiated via addition of 2 % APOL (v/v) and 80 µM *all-trans* retinal (final concentrations). The duration of the spectrum was adjusted to coincide with the refolding process (~3 h), positive and negative contours are shown in dark and light blue, respectively. Due to low signal-to-noise level, in total 16 experiments were recorded and added together. A third spectrum (red) was recorded directly after the refolding process and serves as reference for the APOL refolded retinal-bound bR configuration. Refolding was verified via the characteristic color change visible after the measurement.

On the contrary, nearly all other peaks are not captured in the time-resolved spectrum. Since the signal-to-noise ratio of the SDS unfolded state is rather good and the time-resolved spectrum does not show detectable signal intensities remaining in this state, the data reveals that basically all regions of bO (except the C-terminus) undergo a fast transition into an NMR invisible transition state. In this respect, NMR invisibility could be caused either by a heterogeneous transition including e.g. multiple intermediate states[367] appearing sequentially or in parallel, or by a transition state with considerably increased particle size.

Overall, the time-resolved data reveal that the unfolded state disappears considerably faster as compared to the overall refolding rate constant. Furthermore, the large resemblance of the real-time spectrum to the final spectrum indicates that a considerable three-dimensional fold is also realized fast in the refolding process. In this respect the presented data may provide initial insights into the molecular details underlying the membrane protein folding process.

One possible rate limiting step consistent with this data could therefore be the correct Schiff-base assembly in the retinal binding pocket[368] as well as the precedent non-covalent attachment of retinal.

10.4 APPLICATIONS OF CECF

Cell-free protein synthesis is generally a great and efficient tool for the expression of all kinds of proteins. Even "touchy" proteins like GPCRs, other MPs or intrinsically disordered proteins can be expressed with stable efficiency, circumventing the drawbacks of *in vivo* expression (see chapter 9.3.2). Due to the defined and open nature of cell-free systems, they can be specifically modified and adapted to the special needs of individual proteins and following applications. Especially *E. coli*-based continuous-exchange cell-free (CECF) protein expression is a fantastic, versatile tool for protein expression (see chapter 9.3.5).

In the following, side projects of this thesis will be presented that focus on expression of different kinds of proteins with high pharmacological relevance, which have been formerly proven to be very stubborn and complicated in cellular expression systems.

10.4.1 Soluble proteins

10.4.1.1 HIV-1 Viral protein R (Vpr)

Although Acquired Immunodeficiency Syndrome (AIDS) is a result of infection with the lentiviruses HIV type1 and 2, HIV-1 is the predominant virus worldwide[369]. HIV-1 encodes for the structural proteins Gag, Pol and Env and furthermore six auxiliary proteins including Vpr, Vif, Nef (virion associated), Tat, Rev (essential regulatory functions) and Vpu.

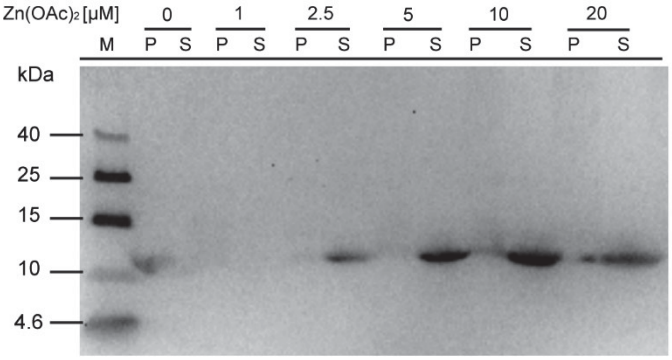


Figure 30: Vpr solubility increases with zinc concentration. Western blot of cell-free expressed, centrifuged HIV-1 Vpr WT with increasing zinc acetate concentrations (0 to 20 µM, S: supernatant, P: pellet). Vpr was specifically detected by using the HIV-1 Vpr antibody vN-20 (Santa Cruz) as a primary antibody.

The accessory protein Viral protein R (Vpr) is conserved in all immunodeficiency viruses (HIV type 1 and 2 as well as SIV)[370]. Vpr plays several important roles in terms of infection,

such as host-cell cycle progression and apoptosis, nuclear import of viral DNA and reverse transcription of HIV-1[371,372]. As a result Vpr is a promising target for the development of pharmacological drugs treating HIV infections[373]. Due to its conserved zinc binding motif as reported for Vpx[374], it is hypothesized that Vpr is a zinc-binding protein.

We tested cell-free expression of HIV type 1 Viral protein R under various conditions. Furthermore, the impact of zinc on the solubility of Vpr during expression was tested. Expression of WT Vpr with zinc acetate ranging in concentrations from 1-20 µM is shown in Figure 30. A negative control without zinc acetate was used as reference. Our results indeed show a clear effect of zinc in solubility of Vpr (Fig. 30). Detailed descriptions of conducted experiments and results can be found in the PhD thesis of Sabine Schriek, "*MC4R ligands: Molecular features, membrane interactions and their potential role in modulation of hormone signaling*", 2017.

10.4.1.2 Enhanced filamentous growth protein 1 (Efg1)

Candida albicans (*C. albicans*) colonizes the human gastrointestinal and oral tract and can be either a commensal or an opportunistic fungal pathogen[375]. It can become an invasive pathogen, causing life-threatening systemic mycoses[376–378].

One major factor, that plays an essential role in the formation and progression of *C. albicans* infections, is the change of cellular morphology, including yeast-hypha morphogenesis[379,380]. One of the various signaling pathways, leading to activated hyphal development is regulated by Enhanced filamentous growth protein 1 (Efg1)[381,382]. Efg1 is thought to be a major morphogenetic regulator in *C. albicans*[381].

Ernst and coworkers found that Efg1 contains a basic helix-loop-helix (bHLH) domain with significant sequence similarity to morphogenetic regulators from other fungi[381,382]. Some transcription factors exhibit characteristic bHLH domains, which can promote dimerization and DNA binding[383,384]. As a result, most models of morphogenetic regulation of *C. albicans* are based on the presumption that Efg1 is a transcription factor[379,381,382]. More detailed information is available in reviews by Lassak *et al*.[385] and Pierce *et al*.[375].

Our collaboration partner, Prashant Desai (AG Ernst, HHU Düsseldorf), failed to produce Efg1 using conventional expression strategies. We produced native Efg1 (68 kDa) with a polyhistidine-tag via cell-free expression. The results are shown in Figure 31.

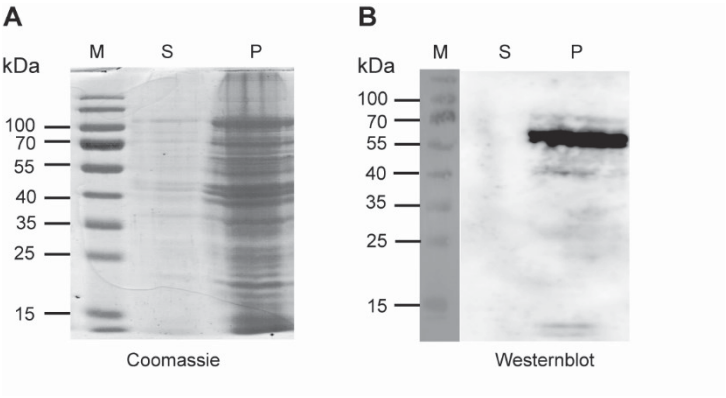


Figure 31: Cell-free expression of Efg1 from *C. albicans*. (A) Coomassie blue-stained SDS-PAGE gel of cell-free expressed Efg1 after centrifugation, (S: supernatant, P: pellet); (B) Western blot of cell-free expressed Efg1. Efg1 detection was enabled by Anti-His-Antibody (Fisher Scientific). SDS-PAGE and western blot executed by Prashant Desai (AG Ernst, HHU Düsseldorf).

10.4.2 GPCRs

10.4.2.1 G protein-coupled bile acid receptor (TGR5)

The bile acid receptor TGR5, also known as GPBAR1, was discovered in 2002[386]. Classified as the founding member of the bile acid receptor subclass of GPCRs[387], high levels of TGR5 mRNA were found in various organs, such as small intestine, liver, lung, stomach and especially placenta and spleen[388,389]. Activation of TGR5 by bile acids induces cyclic adenosine-monophosphate (cAMP) production[386]. As TGR5 plays an important role in several essential cell-signaling pathways, its agonists serve as potential drugs for the treatment of digestive, metabolic or inflammatory disorders[390,391]. Further detailed information about TGR5 can be found in a recent review by van Nirop *et al*.[392].

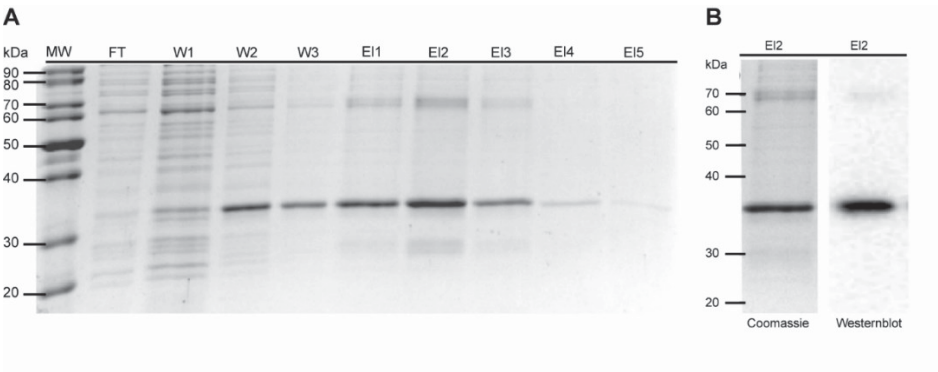


Figure 32: Cell-free expression and purification of human TGR5. (A) Coomassie blue-stained SDS-PAGE gel of cell-free expressed TGR5, purified via denaturing IMAC in SDS, (FT: flow-through, W: washing fractions 1-3, E: elution fractions 1-5); (B) Coomassie blue-stained SDS-PAGE gel of TGR5-purification with corresponding western blot. TGR5 detection was enabled by Anti-His-Antibody (Fisher Scientific).

We synthesized native TGR5 (38.5 kDa) with a hexa-histidine-tag via CFPS. Figure 32 shows the results of cell-free TGR5 expression and SDS purification via IMAC, verified by western blotting.

10.4.2.2 Reconstitution and NMR characterization of the ion-channel accessory subunit barttin in detergents and lipid-bilayer nanodiscs

The following chapter reflects the content of the publication:

Title: Reconstitution and NMR Characterization of the Ion-Channel Accessory Subunit Barttin in Detergents and Lipid-Bilayer Nanodiscs

Authors: Thibault Viennet, Stefanie Bungert-Plümke, Shantha Elter, Aldino Viegas, Christoph Fahlke, Manuel Etzkorn

Journal: Frontiers in molecular biosciences (2019)

Keywords: Barttin, ion channel, lipid bilayer nanodisc, detergent micelle, nuclear magnetic resonance

Abstract

Barttin is an accessory subunit of ClC-K chloride channels expressed in the kidney and the inner ear. Main functions of ClC-K/barttin channels are the generation of the cortico-medullary osmotic gradients in the kidney and the endocochlear potential in the inner ear. Mutations in the gene encoding barttin, BSND, result in impaired urinary concentration and sensory deafness. Barttin is predicted to be a two helical integral membrane protein that directly interacts with its ion channel in the membrane bilayer where it stabilizes the channel complex, promotes its incorporation into the surface membrane and leads to channel activation. It therefore is an attractive target to address fundamental questions of intermolecular communication within the membrane. However, so far inherent challenges in protein expression and stabilization prevented comprehensive in vitro studies and structural characterization. Here we demonstrate that cell-free expression enables production of sufficient quantities of an isotope-labelled barttin variant (I72X Barttin, capable to promote surface membrane insertion and channel activation) for NMR-based structural studies. Additionally, we established purification protocols as well as refolding strategies in detergent micelles and phospholipid bilayer nanodiscs. Stability, folding and NMR data quality are reported as well as a suitable assignment strategy, paving the way to its structural characterization.

Introduction

ClC-K channels form a subgroup of chloride channels within the ClC family of chloride channels and chloride/proton antiporters. They are expressed in the kidney and in the inner ear, and are essential for NaCl re-absorption in the loop of Henle and for potassium secretion by the stria vascularis[393,394]. Barttin is the accessory β-subunit of ClC-K channels. It is the gene product of BSND, the disease gene of a rare human disease, called Bartter syndrome IV. Patients suffer from a severe impairment of urinary concentration ability as well as sensorineural deafness[395–398]. Barttin has been shown to be necessary for the function of both chloride channels ClC-Ka and ClC-Kb[399] present respectively in inner ears and in kidneys. Barttin modifies protein stability, promotes its insertion into the plasma membrane and turns the channel into a conductive state[400,401].

The predicted transmembrane topology of barttin consists of a short cytoplasmic N-terminus of eight amino acids, two transmembrane helices encompassing the amino acids between 9 and 54 and a large cytoplasmic C terminus. Whereas the transmembrane core of barttin including the short cytoplasmic N terminus is sufficient for fulfilling the effects on channel stability and intracellular trafficking, amino acids in the C-terminus are required for normal function of ClC-K/barttin channels[401–404]. Co-immunoprecipitation studies and confocal microscopy indicate binding of barrtin to the B- and/or J-helix at the outer surface of the pore-forming subunit of ClC-K channel[405]. Despite this model and features, the structural characteristics of barttin remain elusive, pointing to the need for proper investigation of the protein, aiming to better understand the pathophysiology, the effect of BSND mutations[403,404] and hoping to discover targets for drug development[405]. While detergent micelles are most commonly used for structural studies of membrane proteins, they exhibit severe destabilization[327] and denaturation effects[178]. Phospholipid bilayer nanodiscs (NDs) have proven to be a useful alternative membrane mimetic[406–408]. They have the advantage of allowing the use of numerous types of lipids, and have the power to mimic key properties of membranes thus providing potentially more native environments to membrane proteins. The nanodisc system has been successfully applied in numerous NMR-based studies[409], including dynamic and structural investigations of β-barrel proteins[410,411] as well as the tumor necrosis factor receptor p75NTR[412] which contains a single transmembrane domain or to the seven-transmembrane helices protein bacteriorhodopsin[225]. Although NMR-optimized constructs have been developed[411], NDs are still a challenging environment for high-resolution NMR studies due to their overall large size. As such, pursuing a combined

approach using both detergent micelles and nanodiscs in order to exploit the good NMR properties of micelles and the presumably more native environment of the nanodiscs can be a valid approach. This was done in the case of Opa60, where restraints derived from NMR in both dodecylphosphocholine (DPC) and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) bilayer NDs were used[410].

Here we report on expression, purification and refolding properties of barttin, which so far eluded a comprehensive in vitro characterization. The usage of an *E. coli*-based cell-free protein expression system enabled production of larger quantities of barttin that we used for refolding procedures both in detergent micelles and in phospholipid bilayer nanodiscs. For the latter the influence of lipid and detergent types as well as detergent removal procedures on the quality of resulting nanodiscs was systematically evaluated. Using the optimized protocols, barttin could be successfully solubilized in detergent micelles as well as refolded into nanodiscs in sufficient quantities for subsequent structural and biophysical characterization. As expected NMR spectral quality is largely improved in detergent micelles, however our data show that barttin is considerably less stable in the used micelle system as comparted to nanodiscs. While barttin adopts a secondary structure in detergent micelles consistent with its predicted behavior, the NMR spectra in micelles and nanodiscs differ more than expected for the changing environments, indicating that the two environments do not necessarily stabilize the same barttin structure.

Materials and Methods

Cell-free expression

A barttin construct containing the N-terminal 72 residues of barttin preceded by a His-tag (His-I72X) was cloned into a pET21a vector and plasmid DNA was produced. Barttin was expressed in an *E. coli*-based cell-free expression system following established protocols[60,285]. For NMR samples expression was carried out using deuterated buffers (> 90 % D_2O) and either triple (^2H, ^{13}C, ^{15}N)-labelled Algal extract supplemented with the missing four amino acids or custom-made mix with selected amino acids with different isotope labelling suitable for combinatorial assignments was used. The combinatorial test sample contained the following labelling pattern: ^{15}N-Val, ^{15}N-Cys, ^{15}N-Leu, ^{15}N-Ser, $^{13}C'$-Met, u-(^{13}C-^{15}N)-Val and all other amino-acid types at natural abundance. Dialysis mode reactions were carried out at 28 °C with or without presence of nanodiscs. After 12–16 h, the reaction mix was centrifuged for 10 min at 12000 x g. The pellet was washed once with 5 to 10 volumes of

buffer containing 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl (buffer A, all reagents from Sigma Aldrich if not stated otherwise) and Complete protease inhibitors (Roche). The resulting pellet was stored at −20 °C until further use.

Purification and detergent solubilization

For detergent solubilization tests, pellets were directly solubilized in 50 mM NaH_2PO_4 pH 8.0, 150 mM NaCl supplemented with either 20 mM sodium dodecyl sulfate (SDS), 100 mM decylphosphocholine (FOS-10, Cube Biotech) 100 mM dodecylphosphocholine (FOS-12 or DPC, Cube Biotech), 100 mM N,N-dimethyldodecylamine N-oxide (LDAO, Cube Biotech), 42 mM lyso-myristoylphosphatidylglycerol (LMPG, Avanti polar lipids), 250 mM n-decyl-β-D-maltoside (DM, Anatrace) or 196 mM n-dodecyl-β-D-maltoside (DDM, Cube Biotech) in a thermomixer (Eppendorf) at 37°C, 800 rpm for 2 h, without further purification.

For small-scale NMR samples and combinatorial-labelled samples, pellets were directly solubilized with NMR buffer (20 mM NaPi pH 7.0, 100 mM NaCl, 2 mM TCEP, 0.2% (v/v) NaN_3 and 10% (v/v) D_2O) supplemented with either 100 mM DPC, 42 mM LMPG or 100 mM LDAO in a thermomixer (Eppendorf) at 37°C, 800 rpm for 2 h without further purification.

For large-scale triple-labelled (^2H,^{13}C,^{15}N) samples, pellet was solubilized at room temperature for 30 min with 10 volumes of buffer A supplemented with 2 mM DTT, 50 mM LDAO and Complete protease inhibitors, then centrifuged at room temperature at 16000 x g for 30 min. The supernatant was incubated with previously washed Ni-NTA agarose chemical beads (Macherey-Nagel) at room temperature for 1 h. The slurry was transferred to a column, washed with 10 column volumes of buffer A supplemented with 2 mM DTT and 10 mM LDAO. Barttin was eluted using buffer A supplemented with 2 mM DTT, 10 mM LDAO and 300 mM imidazole, fractions containing barttin were pooled and applied to a desalting column equilibrated with NMR buffer supplemented with 10 mM LDAO. Finally, the eluate was concentrated in a 10 kDa cutoff Vivaspin concentrator.

Reconstitution in lipid bilayer nanodiscs

Membrane scaffold protein preparation

E. coli BL21 (DE3) were transformed with the MSP1D1 or MSP1D1Δ5 plasmid DNA in a pET28a vector as reported in[411,413]. In short, cells were grown in LB medium. Protein was resuspended with 6M Gu-HCl and purified by IMAC (without denaturating agent).

The elution fractions were pooled and dialyzed in order to remove imidazole. N-terminal His-tag was cleaved using TEV protease incubated overnight at 4°C. ΔHis-MSP1D1 or ΔHis-MSP1D1Δ5 was separated from MSP1D1 or MSP1D1Δ5 by IMAC.

Barttin purification in SDS

Barttin was purified from washed CFE pellets to SDS following similar procedure as in LDAO. In short, pellets were solubilized in buffer containing 20 mM SDS, supernatant was diluted to 10 mM SDS previous to binding to Ni-NTA agarose beads. Either the slurry or the pooled elution fractions was used for nanodiscs assembly.

Nanodiscs assembly

In all other cases, barttin in SDS, 6-fold ΔHis MSP1D1 or MSP1D1Δ5 and respectively 450-fold or 270-fold lipids solubilized in 60 mM Na-cholate were mixed together in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM DTT, 10 mM SDS. Different lipids (all from Avanti polar Lipids) were used including 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), a mixture of 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol (POPG) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) in the ratio 1:4 and a mixture of 1,2-distearoyl-sn-glycero-3-phospho-L-serine (DSPS), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and 1,2-diarachidoyl-sn-lycero-3-phosphocholine (DAPC) in ratio 1:3:3:3[414,415].

Detergent removal procedure was empirically optimized, different approaches including usage of Biobeads SM-2 (Biorad) on-column or in the purified product, dialysis through a 10 kDa cutoff membrane (Thermo Scientific), on-column fast and stepwise wash of the detergent were tested. For more information see Results section.

For all off-column procedures an additional IMAC purification was done in order to separate nanodiscs containing barttin from those which did not. Finally, barttin-containing NDs were purified by SEC on a HiLoad 16/600 Superdex 200 pg (for big scale) or Superdex 10/300 gl (for small scale) column (GE Healthcare) equilibrated with 20 mM sodium phosphate pH 7.0, 100 mM NaCl, 2 mM DTT using a ÄKTA pure device running at 1 ml/min. The quality of nanodiscs preparation was inspected by the SEC chromatogram. The main peak from SEC was pooled and concentrated using a Vivaspin centrifugal device of 10 kDa MWCO, 2 mM TCEP, 10% w/w D_2O and 0.01% w/w NaN_3 were finally added to the samples.

For co-translational incorporation of barttin into NDs, empty nanodiscs assembled with DMPC according to established protocols[413,416] were used in the cell-free system's reaction mix at a concentration of 58 µM.

SDS-PAGE

Proteins were analyzed using denaturing, non-continuous tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis according to published methods[417]. The gel consists of three parts, a 17% acrylamide separation gel, a 10% intermediate gel and a 5% stacking gel. The gel is suited to separate proteins of low molecular masses or peptides. To separate the proteins in an electric field, the following anode buffer (200 mM Tris-HCl pH 8.9) and cathode buffer (200 mM Tris-HCl, 100 mM Tricine and 0.1% SDS) were used. The PageRuler Plus prestained protein ladder (ThermoFisher) was used as protein standard. Proteins were visualized after Coomassie blue staining.

Circular Dichroism

A sample of 10 µM barttin in LDAO micelles was investigated using circular dichroism (Jasco J-715) in 20 mM NaPi pH 6.8, 100 mM NaCl, 2 mM TCEP and at room temperature. Measurement from 260 to 200 nm was done at a constant bandwith of 1 nm, with a data pitch of 0.2 nm and repeated 10 times for averaging. Data was converted to mean-residue molar ellipticity and convoluted using the K2d protocol.

NMR data acquisition

NMR experiments were performed on Bruker Avance III HD+ spectrometers operating either at 600 or 700 MHz, both equipped with a triple resonance TCI (^{1}H, ^{13}C, ^{15}N) cryoprobe. Data was collected at 35 ºC in 20 mM NaPi pH 6.8, 100 mM NaCl, 2 mM TCEP, 0.2% (v/v) NaN_3 and 10% (v/v) D2O. For experiments using LDAO micelles, the detergent concentration after concentration was estimated to roughly 100 mM. All NMR experiments contained Transverse Relaxation Optimized Spectroscopy (TROSY) components[418,419]. The 2D ^{13}CO-and ^{13}Cα-filtered TROSY-HSQC pulse sequences were designed from their respective 3D experiments. All NMR spectra were processed with TOPSPIN 3.2 (Bruker) and analyzed with CARA[420] and CCPN[421].

Results and Discussion

Barttin can be expressed in a cell-free setup and incorporated into nanodiscs

In general, structural biology aspires to obtain data in a native environment of the target system in order to ensure relevant folding and necessary cofactors for function, without introducing bias due to in vitro handling procedures. While some techniques theoretically have the power to do so, such as in-cell NMR[422] or dynamic nuclear polarization NMR[423], they often face limitations in sensitivity and/or resolution, which is particularly true for membrane systems. For membrane proteins, the presence of a phospholipid bilayer with similar chemical composition and physical properties as the native membrane is evidently one of the most important environmental factors. When pursuing *in vitro* investigations, nanodiscs, as compared to other membrane mimetics such as detergent micelles, bicelles or amphipols, exhibit several key advantages[206]: (i) they provide homogeneous lipid-bilayer particles, (ii) both sides of the bilayer are accessible in solution, (iii) they have high stability and low exchange rates, (iv) the presence of the scaffold protein may mimic a crowded environment as often found in natural membranes and (v) a large range of lipids (charge, length, unsaturations, non-phospholipids, etc.) can be incorporated allowing mimicking of properties of various types and states of membranes[409].

Structural studies of membrane proteins are often impeded by limitations in protein expression due to cytotoxic effects of the target protein in its heterologous overexpression systems. Unfortunately, barttin belongs to those membrane proteins that are difficult to produce in large quantities in conventional expression systems (unpublished observations). Therefore, so far, all published barttin data resulted from in cell experiments where protein amount is not critical[394,401,402]. Cell-free expression (CFE) offers a valuable alternative, which also has the advantage of supporting custom labelling strategies thanks to the direct use of amino acids and the very low level of metabolic scrambling[424]. We tested expression of barttin in an *E. coli*-based cell-free expression system. Following established protocols[60] we could obtain sufficient quantities of protein for subsequent characterization.

We initially carried out CFE in the absence of membrane mimetics. This strategy requires refolding of the resulting protein pellet into a suitable membrane mimetic. While canonical protocols for detergent refolding or incorporation into nanodiscs exist, they need to be empirically optimized for each target protein. For the latter factors such as the target protein

to scaffold protein to lipid ratio, the detergent type and removal procedure may strongly differ for different target proteins[409].

Table 3: Summary of tested conditions for barttin refolding in NDs. Type of scaffold protein, detergents lipids as well as detergent removal procedure are indicated together with the respective yields (x = not incorporated into NDs).

Scaffold	Detergent for Barttin	Detergent for lipids	Lipids	Detergent removal	Yield (%)
D1	SDS	Na-cholate	DMPC	Biobeads	x
D1	SDS	SDS	DMPC	Biobeads	x
D1	SDS	Na-cholate	DMPG:DMPC	Biobeads	x
D1	SDS	SDS	DMPG:DMPC	Biobeads	x
D1	DPC	Na-cholate	DMPC	Dialysis	x
D1	DPC	Na-cholate	DMPG:DMPC	Dialysis	x
D1	SDS	Na-cholate	DMPC	Dialysis	< 1%
D1	SDS	Na-cholate	DMPG:DMPC	Dialysis	< 1%
D1	SDS	Na-cholate	DMPC	On-column step dilution	x
D1	SDS	Na-cholate	DMPC	On-column fast dilution	4.7
D1	SDS	Na-cholate	DMPC	On-column biobeads	8.2
D1	SDS	Na-cholate	DMPC	Dialysis	5.8
Δ5	SDS	Na-cholate	DMPC	On-column biobeads	14
Δ5	SDS	Na-cholate	"Native" mix	On-column biobeads	9
Δ5	SDS	Na-cholate	DMPC	Dialysis	19.3
Δ5	SDS	Na-cholate	"Native" mix	Dialysis	3.7
Δ5	SDS	Na-cholate	DMPC	Dialysis / Upscale	8.2
Δ5	SDS	Na-cholate	"Native" mix	On-column biobeads / Upscale	6.2
Δ5	Cell free	-	DMPC	-	12.8

To test barttin incorporation into nanodiscs we started with the following setup: SDS was used to solubilize barttin CFE pellets, MSP1D1 (D1) was used as scaffold protein, DMPC lipids were solubilized in sodium cholate, and absorbent polystyrene beads were used to remove detergents. Although empty NDs assembly was detected, no detectable levels of barttin were incorporated with this setup (data not shown, see Table 1). Subsequently we changed the detergent in which lipids are solubilized to SDS, and the lipid mixture to DMPG:DMPC 1:4 (20% negatively charged head groups). While previously shown to largely improve quality of the data[411], it did not lead to detectable improvements for barttin. The most probable explanation is that barttin, a rather hydrophobic and low molecular weight (10.1 kDa) protein, could adsorb to the polystyrene beads.

In the next step we therefore used dialysis to remove detergent for ND assembly. The same conditions concerning lipids and detergents were tested and indeed successful incorporation

of barttin in nanodiscs was detected (Fig. 33a, supplementary Fig. 54a and Table 3). Unfortunately, the yield, defined as the recovery of barttin from the input to the purified NDs, remained very low (less than 1%; see Table 3). Changing detergent for barttin solubilization to DPC micelles, which provided good results in other cases[410,411], did not improve barttin incorporation. In addition (and possibly in line) with the low yield, the size exclusion profiles also reveal rather poor homogeneity (Fig. 33a).

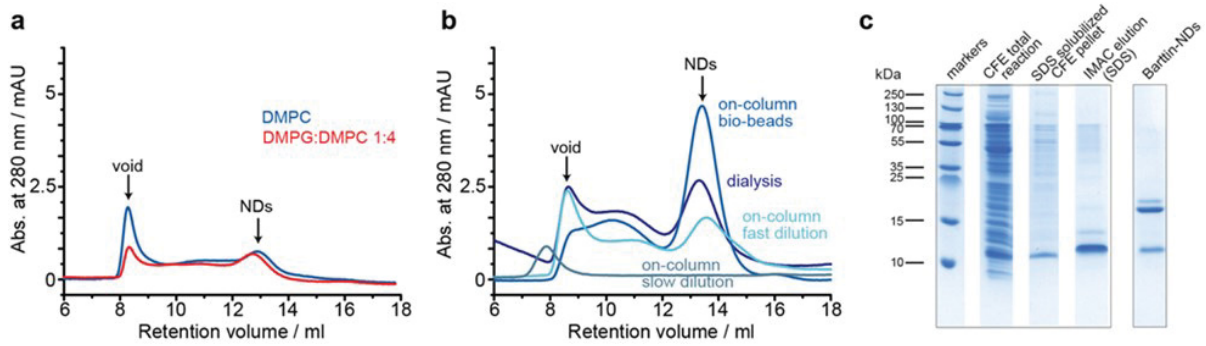


Figure 33: Optimization of barttin refolding into D1 NDs. a) SEC profiles of barttin NDs (MSP1D1) assembled with DMPC (blue) and DMPG:DMPC 1:4 (red) upon sodium cholate dialysis. b) SEC chromatograms of barttin NDs (MSP1D1) assembled with DMPC upon detergent dialysis (dark blue), on-column biobeads adsorption (blue), on-column fast dilution (cyan) and on-column step dilution (grey-blue). c) Coomassie blue-stained SDS-PAGE of CFE expressed barttin (total reaction), solubilized CFE pellet, SDS-purified barttin and purified (IMAC, before SEC) barttin NDs (MSP1D1) assembled with DMPC upon on-column biobeads adsorption (minimal variance in apparent Mw of barttin is due to running variations, see supplementary Fig. 54 for full gels).

Having found initial conditions allowing barttin incorporation into nanodiscs, additional parameters were screened in order to improve yields and homogeneity of the preparations. Since the detergent removal step seemed critical, three different approaches were tested using on-column procedures successfully reported before[424]. In short, barttin in SDS was immobilized through its histidine tag to a Ni-NTA agarose column, incubated with MSP and lipids and the detergents were removed either by introduction of adsorbent beads, fast dilution of the slurry, or stepwise washing procedure. The latter did not allow any ND assembly, which we attribute to a too low detergent removal rate, kinetically favoring lipid aggregation over nanodisc formation as seen from the SEC profile (Fig. 33b). The other protocols were successful in producing substantial amounts of barttin-containing NDs, with an overall higher homogeneity and yield for the on-column polystyrene beads adsorption of detergents (Fig. 33b and Table 3).

In addition to size homogeneity, another important factor in ND refolding is the resulting oligomeric state of the membrane protein, which is difficult to control as well as to measure accurately[409,425]. While, due to the lack of larger extra-membranous domains, incorporation of multiple copies of barttin into the NDs not necessarily changes the resulting SEC profile, the amount of barttin per ND can be estimated from the relative band intensities of MSP and barttin on a SDS-PAGE gel. The respective band intensities, after IMAC purification to remove empty NDs, are in line with single copy of barttin in D1 NDs (Fig. 33c).

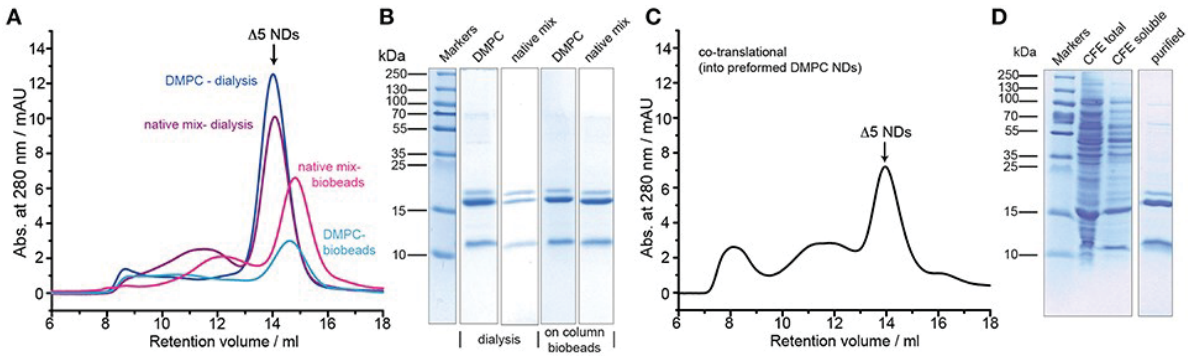


Figure 34: Optimization of barttin incorporation into Δ5 NDs. a) SEC profiles of barttin NDs (MSP1Δ5) assembled with DMPC (blue) and 'native-lipid mix' (purple) upon detergent dialysis, respective profiles for on-column biobeads adsorption for DMPC (cyan) and lipid mix (magenta). b) Coomassie blue-stained SDS PAGE gel of the corresponding barttin-NDs samples, purified via IMAC, before SEC. c) SEC profile of barttin NDs formed via co-translational refolding of barttin into Δ5 NDs during CFE. d) Coomassie blue-stained gel of total CFE reaction in the presence of preformed NDs (MSP1Δ5, DMPC), soluble fraction and IMAC purified barttin NDs.

Smaller nanodiscs, obtained by using shorter MSP variants[411], are known to provide better NMR spectral quality and may provide better control of the oligomeric state for small to medium-sized membrane proteins. We therefore tested the membrane incorporation of barttin also using MSP1D1ΔH5 (Δ5)-based NDs. In this setup we additionally tested the possibility of refolding barttin in a more 'native-like' phospholipid mixture based on a conserved aliphatic chain lengths distribution and negative charge content[414,415] using a lipid composition of DPPC:DSPC:DAPC:DSPS in a 3:3:3:1 molar ratio. For both standard DMPC and the 'native-lipid mix', we tested the refolding using either dialysis or on-column adsorbent beads removal. In all conditions barttin incorporation was successful and provided decent homogeneity of the resulting NDs (Fig. 34a). Usage of the 'native-lipid mix', however, provided lower yields (Fig. 34a and Table 3). The oligomeric state of barttin in Δ5 NDs was also estimated to one barttin per ND in all tested conditions (Fig. 34b).

Noteworthy the usage of a cell-free expression system also allows to add preformed nanodiscs to the reaction mix to promote co-translational incorporation of the expressed protein directly into the NDs[409,426]. We tested this approach using preformed DMPC Δ5 nanodiscs, which yielded fair quality and amount of barttin-containing NDs (Fig. 34c, d). However, overall homogeneity and yield was lower as for other refolding protocols (see Table 3).

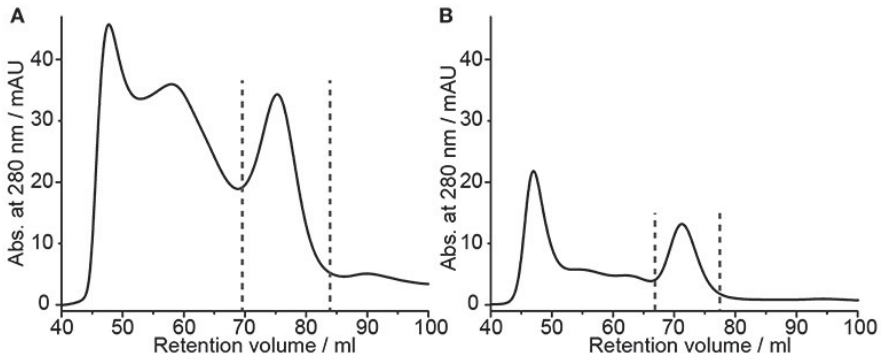


Figure 35: NMR sample preparation of barttin in Δ5 NDs. a) SEC profile of up-scaled barttin NDs (MSP1Δ5) assembled with 'native-lipid mix' upon on-column biobeads adsorption. b) SEC profile of up-scaled barttin NDs (MSNMR sample preparation of barttin in Δ5 NDs. a) SEC profile of up-scaled barttin NDs (MSP1Δ5) assembled with 'native-lipid mix' upon on-column biobeads adsorption. b) SEC profile of up-scaled barttin NDs (MSP1Δ5) assembled with DMPC upon dialysis. Dotted line mark fractions collected in the corresponding NMR samples.P1Δ5) assembled with DMPC upon dialysis. Dotted line mark fractions collected in the corresponding NMR samples.

All previous experiments were carried out using a rather low amount of barttin (from two small scale 50 µl cell-free expressions). Since NMR studies require higher amounts of material (in the mg range), we tested up-scaling of the most promising protocols, i.e. 'native-lipid mix' together with on-column detergent adsorption and DMPC together with detergent dialysis. For both setups cell-free pellets from 3 ml CFE reactions were used. In line with general observations in upscaling, in both cases the yields dropped by 30% and 60% respectively (Fig. 35a,b and Table 3). Nevertheless sufficient amounts of ^{15}N-barttin in Δ5 NDs for NMR characterization could be obtained.

Summarizing the screening of the tested conditions, several key points in barttin expression and refolding can be concluded, including: (i) CFE offers a convenient approach to prepare high quantities of barttin, (ii) cell-free expressed barttin can be easily purified, (iii) barttin co-translationally inserts into preformed NDs during cell-free expression, (iv) barttin is likely to adsorb into polystyrene beads (v) detergent removal setups and resulting removal rates during barttin-ND self-assembly are very critical parameters, and (vi) various types of phospholipid mixtures lead to successful incorporation of barttin into NDs.

Barttin in detergent micelles

Although providing clear advantages, nanodiscs can still be tedious in NMR-based studies due to their larger size as compared to detergent micelles (approx. 30 kDa for barttin in LDAO micelles as compared to 110 kDa for barttin in Δ5 NDs). Therefore, in cases where protein structure and function are not corrupted by the detergent, detergent micelles can facilitate NMR-based characterization.

To obtain barttin in detergent micelles, we initially performed detergent screening by directly solubilizing barttin from CFE protein pellets following previous approaches[285]. A range of detergents suitable for NMR studies were tested, including FOS-10, DPC, LDAO, LMPG, DM and DDM (Fig. 36a). Interestingly, all tested detergents with the exception of DM and DDM lead to similar solubilization levels as seen from the intensities of the corresponding bands in SDS-PAGE. Three of them were selected for initial NMR screening for their capacity to form small micelles (LDAO, LMPG) or their reported good NMR properties (LDAO, DPC)[411,427]. 1D ^{1}H-spectra of (not isotope labelled) barttin solubilized in the respective micelles without further purification were acquired and compared (Fig. 36b). In general, the NMR spectra show an overall similar shape of the amide signals for barttin in the three different detergent micelles. However, the spectrum in LDAO shows several more intense and in particular also better resolved peaks on the flanking region of the amid signals. The latter is an important indicator for NMR-spectral quality[409]. In addition, the appearance of tryptophan side-chain resonances was only observed in LDAO micelles (Fig. 36b, Trp-label). We therefore decided to use LDAO micelles for the following steps.

Barttin in LDAO micelles

In order to further evaluate the characteristics of barttin in LDAO micelles we recorded circular dichroism (CD) data (Fig. 37b). As expected for folded barttin the CD spectrum shows a double minimum around 225 and 215 nm, indicative for the presence of α-helical secondary structure. Deconvolution of the CD spectrum (K2d method) indicates an α-helical propensity of 37%. This value is in line with the proposed topology models of barttin[394], which predict an α-helical content of 44% for the used construct (Fig. 37a).

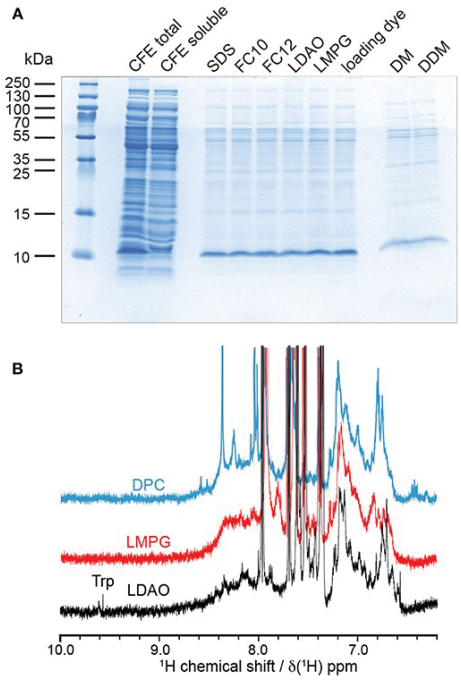


Figure 36: Solubility and NMR properties of barttin in different detergent micelles. a) Coomassie stained SDS-PAGE solubility screen using CFE pellets directly solubilized with indicated detergents (for different detergents only the soluble fraction after centrifugation at 20000 x g is loaded into the gel). b) 1D ^{1}H NMR screening of resulting most promising barttin containing micelles.

We additionally recorded 2D [^{15}N-^{1}H]-HSQC NMR data of isotopically labelled barttin in LDAO micelles. The spectrum shows dispersed and mostly resolved peaks (Fig. 37c). From the 77 expected peaks about 65 are resolved. Note that cysteine, glutamine, asparagine and tryptophan are not ^{15}N-labelled in the used sample. The overall peak positions and dispersion is also in line with the expected features of barttin. To obtain more comprehensive NMR insights into barttin in LDAO micelles we prepared a triple (^{2}H, ^{13}C, ^{15}N) labelled sample of barttin at protein concentration of 250 µM. Note that amide proton back exchange was carried out during the refolding process to enable standard ^{1}H-detected TROSY-based NMR experiments.

Since ^{1}H and ^{15}N chemical shift are less reliable secondary structure indicators as compared to ^{13}C chemical shifts, we additionally recorded a ^{13}C-^{13}C correlation spectrum (Fig. 37f). Note that the spectrum was obtained for free during the acquisition of a 3D NOESY spectrum using the UTOPIA setup[428]. Evaluation of the amino acids specific ^{13}Cα-^{13}Cβ peak positions indicates that most residues predominantly occur in random coil and α-helical secondary structure, in line with the CD data and the expected topology model. Taken together these results suggest that barttin adopts a proper secondary structure in LDAO micelles.

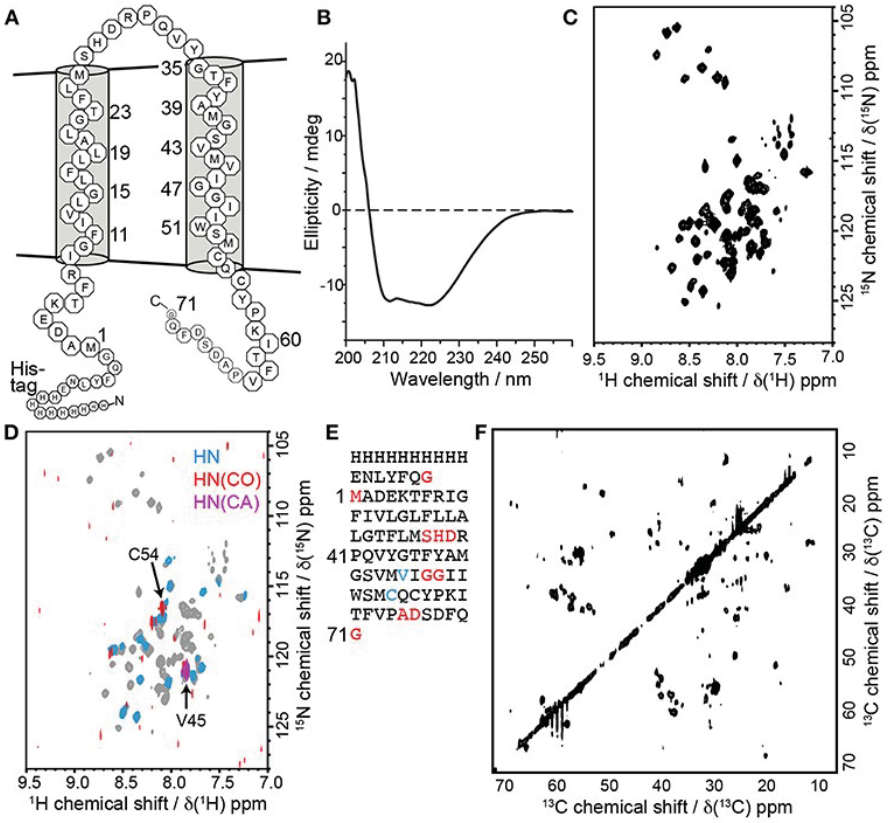


Figure 37: Barttin in LDAO micelles. a) Expected topology model of barttin according to Wojciechowski *et al.*, 2015. b) Circular dichroism spectrum of barttin in LDAO micelles. c) [^{15}N-^{1}H]-HSQC spectrum of [U-^{2}H-^{13}C-^{15}N]-labelled barttin in LDAO micelles (Cys, Asn and Trp are unlabelled). d) Overlay of [^{15}N-^{1}H]-HSQC acquired on [^{2}H-^{13}C-^{15}N]-labelled barttin in LDAO micelles (grey) and on a selectively labelled sample (see methods for labelling pattern, blue). In addition, the [^{15}N-^{1}H] dimension of a HNCO (red) and of a HNCa (purple) are shown. The labelling pattern directly allows assignments of indicated residues. e) Primary sequence of the used barttin construct highlighting residues with assigned resonance frequencies from conventional 3D experiments (red) and using selective isotope labelling (blue). f) 2D ^{13}C-^{13}C FLOPSY spectrum acquired using an UTOPIA setup (Viegas *et al.*, 2016).

Additionally, a set of 3D experiments for resonance assignment was carried out. Unfortunately and reproducibly, NMR spectra of barttin in LDAO micelles change their appearance rather fast (within a few days) indicative of protein degradation and/or aggregation (*vide infra*). The limited stability of barttin in micelles largely reduces the NMR spectral quality due to the alterations of the resonance frequencies of the protein over the time course of the 3D experiments. Consequently, the resulting 3D spectra are difficult to analyze (see supplementary Fig. 55) and have lower as expected signal-to-noise ratios, especially when ^{13}C-^{13}C INEPT transfers are involved. This is often observed in large systems due the unfavorable relaxation properties and may indicate

aggregation of barttin micelles. Albeit the spectral challenges, we could obtain resonance assignments for 10 residues (Fig. 37e, red).

While the limited sample stability prevents usage of time-intensive 3D NMR experiments, in the case of barttin in LDAO it is still sufficient to obtain 2D spectra, which can be recorded sufficiently fast. Therefore one possibility to obtain more information in this conditions is to use a combinatorial isotope labelling approach that only requires usage of simple 2D experiments[334]. The principle is to produce several samples with different types of amino acids that are isotopically labelled with either only ^{15}N, only ^{13}CO or ^{13}C and ^{15}N. Comparison of 2D spectra from [^{15}N-^{1}H]-HSQC, ^{13}CO-filtered HSQC and ^{13}Cα-filtered HSQC experiments lead to amino acid type assignment and ultimately to residue-specific assignments. This is made easy by the use of cell-free expression, which allows to introduce single amino-acid types with the desired labelling[335,336]. While a full assignment is not attempted here, we investigated whether this strategy would in general be feasible for barttin in LDAO. We therefore prepared a selectively isotope labelled version of barttin in LDAO micelles and recorded in addition to the 2D [^{1}H,^{15}N]-HSQC spectrum (Fig. 37d, blue) also 2D versions of HNCO (Fig. 37d, red) and HNCA (Fig. 37d, purple) experiments. By using this single sample two additional unambiguous assignment can be made (Fig. 37e, blue). We therefore anticipate that usage of a substantial number of differently isotope labelled samples should enable a near complete resonance assignment of the amid resonances. However, since it is not fully clear whether the LDAO embedded state resembles the functional relevant barttin structure (*vide infra*), it is at this point not clear whether this effort would be justified.

Comparison of barttin in LDAO micelles and nanodiscs

The lack of a reliable functional in vitro assay for barttin renders it difficult to judge, whether the protein adopts a relevant state after refolding in any membrane mimetic. In this respect it is only possible to compare structural integrity such as secondary structure with the expected behavior as well as to compare NMR spectra obtained in different membrane mimetics. Under the (in general not necessarily valid) assumption that the NDs are more likely to induce a native-like protein fold, NMR offers the possibility to use spectra obtained in NDs as reference for the protein fold and compare it to the data obtained in detergent micelles[429,430].

One disadvantage of the nanodisc system is that the presence of the scaffold protein impedes light absorption-based experiments such as concentration determination at 280 nm or

CD spectroscopy[409]. Thus, the same characterization of barttin secondary structure in NDs is not feasible. However, for NMR measurements signal separation of target and scaffold protein is easily realizable via isotope enrichment of barttin in the CFE setup. We thus carried out an initial NMR characterization of isotope labelled barttin in nanodiscs. Two different ND preparations, i.e. using either 100 % DMPC lipids or using the 'native-lipid mixture' as described above, were used. 2D [^{15}N-^{1}H]-HSQC spectra of barttin in each of the two ND preparations show a rather small number of resolved peaks indicative of homogeneous line broadening of bigger particles and/or heterogeneous barttin conformations. Only about 15 out of the expected 77 peaks are visible. The most probable explanation for this is that only signals from residues situated in the protein termini or in the extra-membrane loop appear because of their NMR-favorable dynamic properties. However, without residue-specific assignments, only assumptions can be made.

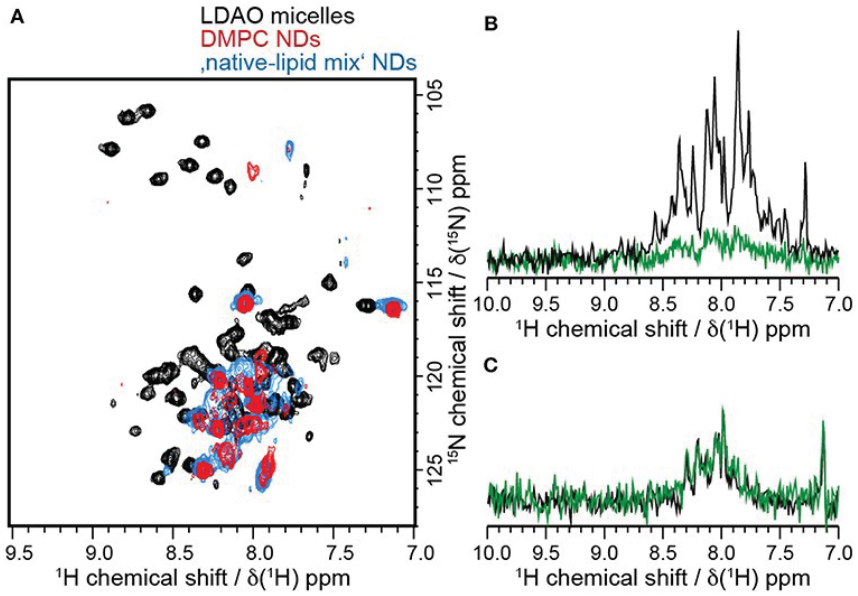


Figure 38: Barttin has different properties in LDAO and NDs. a) [^{15}N-^{1}H]-HSQC spectra of ^{15}N-labelled barttin in LDAO micelles (black), DMPC NDs (red) and 'native-lipid mix' NDs (blue). b) ^{15}N-filtered ^{1}H 1D spectra of ^{2}H-^{13}C-^{15}N-labelled barttin in LDAO micelles at the beginning of the NMR measurements (black) and after 14 days (green). c) ^{15}N-filtered ^{1}H 1D spectra of ^{15}N-labelled barttin in DMPC NDs at beginning of NMR measurements (black) and after 18 days (green).

When comparing the NMR results obtained in the different NDs to the results in LDAO micelles, it is apparent that the positions of the visible peaks of barttin differ considerably between micelles and nanodiscs (Fig. 38a). While the absence of peaks could be explained by the difference in particle sizes, the observed peak shifts indicate that the conformation of the protein is altered in the corresponding region of the protein. Under the assumption that only

the extra-membrane residues are observed in the ND samples, this observation would be in line with the denaturation effect of detergents at the micelle-water interface as it was postulated in the case of OmpX[431]. Since the conformation and dynamics of this region in general plays an important physiological role, it can at this point not be assumed that barttin in LDAO represents the same conformation as barttin in nanodiscs.

Interestingly, a large overlap between the NMR spectra obtained on barttin in pure DMPC and the 'native-lipid mix' NDs are visible, pointing to rather similar structures. Small changes do arise, which are in line with chemical shift perturbation induced by different membrane mimetic that do not alter protein function as e.g. observed for bacteriorhodopsin in detergent and nanodiscs[225].

While NMR data acquired for barttin in LDAO micelles are clearly of much better quality, the stability of the sample is much lower than in NDs. Indeed, as discussed above we could observe gradual degradation of the HSQC signal over time when performing long-time measurements. Over the time course of 14 days a nearly complete disappearance of the signals was observed (Fig. 38b). On the contrary barttin in nanodiscs showed no signal decay after 18 days (Fig. 38c). This data demonstrates that barttin in the ND-system is considerably more stable as compared to LDAO micelles.

Conclusion

Our study could identify and overcome several key challenges in the *in vitro* characterization of barttin. We confirm the usefulness of cell-free expression systems for membrane proteins that are challenging to express in bacterial systems[338]. Conditions were found for which cell-free expressed barttin is successfully incorporated into nanodiscs. While the ND incorporation must be empirically optimized for each target, our screening of conditions reveals several features that may be of general relevance for other systems.

A detergent screening for barttin solubilization and NMR spectral quality identified LDAO as best micelle environment for our study. CD and NMR data indicate that barttin in LDAO micelles adopts a similar secondary structure as predicted, it is however not possible to conclude that it adopts its native structure. Comparison of NMR data of barttin in LDAO micelles or in lipid-bilayer nanodiscs indeed reveals considerable differences reflecting most likely on variations in structure and dynamic of the extra-membrane residues. Interestingly, a comparison of NMR data obtained in nanodiscs formed by the model DMPC lipids are very similar to data obtained on a heterogeneous lipid mixture reflecting

the most prominent lipids in barttin's cellular environment. This observation also indicates that predominantly the extra-membranous regions are visible in the NMR spectra of barttin in the used NDs.

Our data represent the first *in vitro* results of recombinantly expressed and purified barttin and reports on very initial structural results including insights into the effects of different membrane mimetics. While we have invested considerable efforts to overcome several of the major obstacles for the system, our results also suggest that structural characterization of barttin in nanodiscs is still substantially more challenging as expected from its size or as observed for other comparable systems. While barttin in detergent micelles is significantly better accessible via conventional solution NMR approaches, our data shows that the protein is not stable in the used LDAO micelles. We demonstrate that combinatorial labelling could overcome this challenge, however our data also shows that the barttin structure in micelles likely differs from the one in nanodiscs. Without a reliable *in vitro* functional assay, it can at this point not be decided, which membrane environment is suitable to support barttin's native structure. It is however clear that both tested systems, i.e. detergents and nanodisc, will present challenges in terms of stability or NMR spectral quality, respectively.

Overall several aspects of the presented approach should be transferable to other comparable systems where they may help to optimize sample preparation for structural studies. In respect to barttin we anticipate that our results enable future in vitro characterizations of the system, including interaction studies with the ClC-K channels.

11 OUTCOME AND KEY FINDINGS

In summary the main results of this thesis are:

I A homebuilt *E. coli*-based cell-free expression system was successfully established.

II Ligand- and lipid-free bacterioOpsin was efficiently produced in the cell-free expression system.

III GPCRs and soluble proteins, such as Vpr1, barttin, Efg1, MC4R, and TGR5 were produced using the cell-free expression system.

IV Contributed to the introduction of the first affordable stereospecific methyl labelling strategy compatible with all common expression system including cell-free expression.

V Amphipol-assisted refolding of 7-TM α-helical proteins (bR, MC2R, MC4R) was optimized and characterized via comparative NMR studies and SEC results illustrating the advantages of the non-conventional membrane mimetics.

VI Contributed to the introduction of the t-BLM system in combination with AFM spectroscopy as a powerful tool for investigation of structure, function and interactions of membrane proteins**.**

VII A fast and efficient absorption-based refolding assay for bO was established using small volumes in a 384-well format. The new assay enables screening of various folding conditions, such as denaturants, renaturants, buffers, pH, crowding reagents and ligand concentrations, allowing a detailed analysis of the influence of each factor on refolding efficiency of bO.

VIII A partial sequential resonance assignment of bO was obtained based on the approach of combinatorial isotope labelling. The combination of different 2D NMR experiments and isotope-labelling patterns facilitated the assignment of six residues with certainty.

IX High-resolution NMR-insights into bO's unfolded state were obtained using NOESY and HSQC spectra. Residue-specific interactions of bO with water or SDS molecules were identified. Results revealed that some of bO's tryptophan side chains are covered by SDS, while glycine, asparagine and glutamine side chains tend to interact with water and SDS. Flexible regions, such as the C-terminus and loops show a predominant interaction with H_2O molecules. Extracted 2D H-H-planes for four of the previously assigned residues (Y105, D124, I130 and Y153) allowed a quantitative analysis of signal intensities corresponding to interactions with the

surrounding environment. Results suggest that residues D124 and Y153 (located in conjunction loops between segments of bO) and the sequentially close I130 are located in the interface of water and surfactant, while Y153 is more exposed to water than to SDS molecules. In addition, Y105 (located in the N-terminal part of helix C) adopts a more “water-shielded” position in the solubilizing environment.

X Factors that determine refolding kinetics and efficiency of bO were characterized, such as ligand-, denaturant-, renaturant-, protein- and salt concentrations. Even though clear beneficial effects of an excess of retinal were detected, no clear correlation between retinal concentration and refolding rate could be deduced indicating that a considerable amount of retinal is irreversibly trapped in a misfolded protein or surfactant state. Concentrations between 1-2% APOLs represent a compromise regarding refolding efficiency and refolding speed for strategies of NMR RT-folding of bR. In addition, results underlined the role of the surfactant-to-bO ratio as well as the presence of monomeric SDS as key factors in the refolding process. The ratio of monomeric SDS to bO should be above 280:1. In summary, results suggest that for refolding of bO a state should be maintained in which the SDS interaction sites are saturated by monomeric SDS but it is not in a micellar environment.

XI First insights into NMR RT-folding of bR could be obtained and revealed that the C-terminal region of bO undergoes a rather homogeneous transition (assuming a simple two-state transition) from the SDS unfolded state to the folded bR state. Unfortunately, most of the peaks corresponding to other regions of bO are not captured in the time-resolved spectrum, suggesting a fast transition into an NMR invisible transition state.

12 LIST OF PUBLICATIONS

The content of this thesis and key findings will be/have been published in the following publications:

1. **Elter S**, Raschle T, Arens S, Viegas A, Gelev V, Etzkorn M, Wagner G, The use of amphipols for NMR structural characterization of 7-TM proteins, J. Membr. Biol. 247, 957-964 (2014)
    - This publication reflects key findings I-III, V (in part) and the content of chapter 10.3.1.1 and 10.3.3.

2. Bronder AM, Bieker A, **Elter S**, Etzkorn M, Haeussinger D, Oesterhelt F, Oriented membrane protein reconstitution into tethered lipid membranes for AFM Force Spectroscopy, Biophys. J. 111, 1925-1934 (2016)
    - This publication reflects key findings I, II,VI, IX (in part) and the content of chapter 10.3.1.2.

3. Viennet T, Bungert-Pümke S, **Elter S**, Viegas A, Fahlke C, Etzkorn M, Reconstitution and NMR characterization of the ion-channel accessory subunit barttin in detergents and lipid-bilayer nanodiscs, Frontiers in Molecular Biosciences 6, 13 (2019)
    - This publication reflects key findings I, III (in part) and the content of chapter 10.4.2.2.

4. Dubey A, Stoyanov N, Chhabra S, Viennet T, **Elter S**, Borggräfe J, Viegas A, Nowak RP, Burdziev N, Petrov O, Fischer ES, Etzkorn M, Gelev V, Arthanari H, Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems, submitted.
    - This publication reflects key findings I, II, IV (in part) and the content of chapter 10.2.

5. **Elter S**, Viennet T, Borggraefe J, Viegas A, Etzkorn M, Insights into Membrane protein folding from the unfolded state of BacterioOpsin, in preparation

Table 4: Contributions to published results listed above.

Publication	Contributions
1	• cell-free synthesis of bacterioOpsin and MC4R • refolding of bR in APOLs with various conditions • purification of bR via IMAC • size exclusion chromatography for bR • NMR sample preparation for bR
2	• cell-free synthesis of bO • refolding of bR in DDM micelles • purification of bR via IMAC • Dot blotting
3	• cell-free synthesis of barttin with different isotope labels for combinatorial resonance assignment approach
4	• cell-free synthesis of bacterioOpsin with different stereoselective methyl labelled Leucines • refolding of bR in DDM micelles • purification of bR via IMAC • NMR sample preparation for bR

13 OUTLOOK

For future studies, it would be interesting to conduct further NMR experiments, which could reveal the structure of the SDS environment in this context and their residue-specific interactions with bO in its unfolded state. But in order to conduct these studies a more comprehensive assignment of bO under the applied conditions is necessary. Furthermore, ^{13}C-based experiments could lead to a deeper understanding of the exact structure bO adopts. Moreover, it should be tested if the purified bO retains its capability to refold, because our results provide evidence that the structural conformation and the interactions with SDS of purified and non-purified bO may differ. Additionally, techniques to generate partially folded bO from folded bR, such as through bleaching in appropriate membrane mimetics, could be applied in order to compare the different states and, as a future perspective, facilitate the examination of the unfolding process via NMR techniques, such as time- and site-resolved experiments.

14 EXPERIMENTAL TECHNIQUES

14.1 CELL-FREE PROTEIN SYNTHESIS

14.1.1 BO vector

The bO gene was isolated from *Halobacterium salinarum* strain JW-3 and transferred into the pIVEX2.4d expression vector using restriction-free cloning. The resulting construct comprises a 10x N-terminal His tag followed by a Factor Xa cleavage site. The number of artificially introduced residues was minimized by restriction-free cloning to the following sequence (N-terminal): MSGSHHHHHHHHHHSSGIEGRM (followed by the WT $bO_{1\text{-}248}$ sequence) for the 10x His tag construct[225].

14.1.2 Cell-extract preparation

In this work variations and combinations of the protocols, described by Schwarz *et al.*[60] , Kigawa *et al.*[218] and James Zawada[219] were used for cell-extract preparation. Even though the usage of fermenters is suggested in most protocols, a conventional shaker with baffled-flasks setup for *E. coli* cultivation was used.

Media for *E. coli* cultivation

2YT agar plates, pH 7.2 (1 l)
5 g NaCl
10 g Yeast extract
16 g Tryptone
1.5 % (w/v) Agar

LB, pH 7.0 (1 l)
10 g NaCl
5 g Yeast extract
10 g Tryptone

2YTPG, pH 7.2 (2 l)
32 g Tryptone
20 g Yeast extract
10 g NaCl
6.07 g NaH_2PO_4
14.23 g Na_2HPO_4
100 mM Glucose (sterile filtered)

All media were autoclaved before use. Note that the 2.5 M Glucose solution needed for 2YTPG preparation was not autoclaved, but sterile filtered (0.2 µm filter). It was added under sterile conditions directly before usage of the medium.

Buffers for cell-extract preparation

Buffer A
10 mM Tris-Acetate, pH 8.2
14 mM $Mg(OAc)_2$
0.6 mM KCl
6 mM 2-Mercaptoethanol

Buffer B
10 mM Tris-Acetate, pH 8.2
14 mM $Mg(OAc)_2$
0.6 mM KCl
1 mM DTT
Protease inhibitor (EDTA-free, Roche)

S30 buffer
10 mM Tris-Acetate, pH 8.2
14 mM $Mg(OAc)_2$
0.6 mM KOAc

4 M NaCl

Preliminary preparations

Prior to inoculation of the cell-extract preculture, 5 µl of chemo-competent *E. coli* cells (strains: BL21 (DE3) or A19) were plated on 2YT agar plates and incubated over night at 37 °C and stored subsequently at 4 °C. The LB preculture was inoculated the next evening with one colony of the previously plated *E. coli* cells in baffled 500 ml-flasks. The preculture was incubated over night at 37 °C and 180 rpm.

***E. coli* cultivation**

For the cultivation of *E. coli* cells, baffled 5 l-flasks filled with 2 l 2YTPG medium were used. They were inoculated with 100 ml of the LB preculture. Afterwards the OD_{600} of every flask was measured under sterile conditions. The cultures were incubated at 37 °C and 180 rpm. As it is very important to harvest the cells in their logarithmic growth-phase, OD_{600} was measured every 30 min under sterile conditions. Note that sterile handling is even more important than in normal cell cultures, as no antibiotics can be used. After a time of averagely 150-180 min cultures were immediately chilled on ice for 15 min.

Cell-extract preparation

After chilling the *E. coli* cultures cells were harvested at 7500 x g for 10 min at 4 °C. Afterwards cells were resuspended and washed in 330 ml of buffer A on ice, and centrifuged at 6500 x g for 30 min at 4 °C. The supernatant was discarded. This cycle was repeated three times before measuring the cell weight. In the following step, cells were resuspended on ice in 110 % (w/v) of buffer B, referring to the total cell weight. Subsequently, cells were lysed by a French-Press (Glen Mills, kindly provided by Prof. Georg Groth, HHU Düsseldorf) maintaining a constant pressure of more than 1200 psi. The resulting lysate was then centrifuged at 30 000 x g for 30 min at 4 °C. The supernatant (2/3 from the top) was then

centrifuged again with the same settings as mentioned before. After the second centrifugation, again 2/3 from the top of the supernatant were filled into 15-ml tubes, referring as extract A. The same procedure was repeated for the remaining bottom supernatant, referring as extract B. To both extracts a volume of 10 % (v/v) 4 M NaCl was added prior to incubation for 45 min at 42 °C. In the following steps the extracts A and B were dialyzed separately thoroughly with a 10K MWCO-*SnakeSkin™ Dialysis Tubing* (Thermo scientific). The first dialysis-step was performed by dialyzing both extracts against 5 l S30-buffer for 2-3 h at 4 °C. The second dialysis-step was carried out by repeating the step mentioned before over night. On the next day both extracts were centrifuged at 30 000 x g for 30 min at 4 °C. The resulting supernatants of extract A and B were aliquoted separately on ice, immediately frozen in liquid nitrogen and stored at -80 °C.

14.1.3 Experimental setup

The setup of the continuous-exchange cell-free protein synthesis is based on the protocols published by Schwarz *et al.* in 2007[60]. Reagents and solutions were prepared as stated in the mentioned publication with some minor changes. The changes include T7RNAP-, RNAse-inhibitors and DNA concentrations. The final concentration of used T7RNAP (kindly provided by Bernd Esters and Reinhild Wurm) was 3 U/µl. RNAse inhibition was achieved by using *RiboLock RNAse Inhibitor* (Thermo Scientific). Final DNA concentrations of 8 ng/µl were used. For reaction mix volumes up to 50 µl 20K MWCO Slide-A-Lyzer™ MINI Dialysis Devices (Thermo Scientific) were used in sterile 48-well BD falcon cassettes (Beckton Dickinson, Falcon). For reaction mix volumes of ≥1 ml 10K MWCO *SnakeSkin™ Dialysis Tubing* (Thermo Scientific) was used in sterile 50 ml-tubes. Reactions were incubated at 28 °C and 100 rpm over night. In the following steps the reaction mix was aliquoted (50 – 500 µl) and synthesized proteins were harvested by centrifugation for 20 min at 16 100 x g and 4 °C. The supernatants were discarded afterwards. Pellets were washed with 10-fold volumes of 10 mM sodium phosphate buffer, pH 7.5 and centrifuged for 20 min at 16 100 x g and 4 °C. This cycle was repeated two times before pellets were stored at -20 °C or directly used.

14.1.4 Isotope labelling

Uniform isotope labelling was achieved by using [^{13}C-^{15}N-^{2}H]-Algal-mix (Sigma-Aldrich) with a final concentration of 2.75 mg/ml, instead of the normally used amino acid and RCWMDE-mix. As the Algal-mixture does not contain asparagine (N), glutamine (Q) and

tryptophan (W), an isotope labelled QNW-mix was added in the same concentrations as the normal RCWMDE-mix. Isotope labelling suitable for combinatorial assignments was achieved by using custom-made mixtures with selected amino acids (Sigma-Aldrich). Expression was carried out using deuterated buffers (> 90% D_2O).

14.1.5 Analysis of different cell extracts

For the analysis of cell extracts derived from different *E. coli* strains (BL21 (DE3), A19) cell-free reactions were performed as described in chapter 14.1.3 using different cell extracts. A bR-encoding plasmid, optimized for cell-free protein synthesis was used as DNA template (see section 14.1.1). After harvesting of expressed bO, the bO-containing pellets were refolded (see section 14.1.7) separately with 50 µl refolding buffer (RFB). An absorption spectrum (300 – 750 nm) of the protein-solution was measured after refolding in a UV-Vis spectrophotometer (V-650, Jasco). The total refolding yield was calculated by subtraction of the baseline at 700 nm from the value at 555 nm using bO's extinction coefficient (60 390 $M^{-1}cm^{-1}$).

14.1.6 Testing different T7RNAPs

Cell-free reactions for bO synthesis were performed as previously described (see section 14.1.3) and the used quantities of T7RNAP were varied from 1-6 U/µl in every reaction. Additionally, two different aliquots of T7RNAP were compared. One was a commercially available T7RNAP (Thermo Scientific) and the other one was kindly provided by Bernd Esters and Reinhild Wurm. After harvesting of synthesized proteins, pellets were refolded (see section 14.1.7) separately with 50 µl RFB. The absorbance of each refolded sample was measured in a plate reader (Spark 10M, TECAN) from 300-750 nm.

14.1.7 Refolding of bO

Cell-free expressed pellets of bO (50 -500 µl) were resuspended in equal volumes of RFB if not stated otherwise. Samples were protected from light for 1h at RT. An absorption spectrum (300-750 nm) of the protein-solution was measured in a UV-Vis spectrophotometer (V-650, Jasco). The total refolding yield was calculated by subtraction of the baseline at 700 nm from the value at 555 nm using bO's extinction coefficient (60 390 cm^{-1} M^{-1}).

RFB, pH 7.0

1	M	NaCl
0.04	M	sodium-phosphate buffer
5	% (w/v)	DDM
100	µM	*all-trans* retinal (Sigma-Aldrich)

14.1.8 Testing batches of cell-free expressed bO pellets

Samples of new batches of cell-free expressed bO were refolded as described in section 14.1.7. Yields were determined in duplicates by the described methods in section 14.1.7.

14.2 KINETIC ANALYSIS

For screening of a larger set of conditions we first established a refolding assay based on absorbance measurements in a 384-well format. The assay enables fast screening of various folding conditions, including the effects of pH, salts, protein purity, crowding reagents as well as protein, retinal and surfactant concentrations. The assay is efficient, straight-forward and economic by using small volumes of only 20 µl, allowing the efficient use of expressed protein and screening agents.

14.2.1 Refolding assay for bO

50 µl-pellets of cell-free expressed bO (see section 14.1.3) were washed twice with 500 µl sodium-phosphate buffer (50 mM, pH 7.5) by resuspension of the pellet followed by a centrifugation step (21 000 x g, 35 min, 4 °C). The washed pellets were then resuspended in 50 µl resuspension buffer (RSB) per pellet. If not stated otherwise RSB with concentrations as listed below were used. Target concentrations of bO were adjusted based on determined batch concentrations (see section 14.1.8). The wells of a clear 384-well plate (20 µl sample volume, Corning Life Sciences) were filled with a total volume of 10 µl with sodium-phosphate buffer (10 or 50 mM, pH 7.5) and 4 % (v/v) APOL333 (kindly provided by Prof. Vladimir Gelev, Sofia University, Bulgaria) . *All-trans* retinal (Sigma-Aldrich) was added to resuspended bO pellets to final concentrations of up to 160 µM from a 10 mM stock solution.

Refolding was subsequently initiated by adding 10 µl of the premixed bO-retinal sample to the wells containing sodium-phosphate buffer and APOL333. The plate was immediately sealed with UV-transparent film and transferred into a plate reader (Spark 10M, TECAN) preheated to 30 or 35 °C and the measurement of duplicates was started.

The exact time from adding bO into the first well until the start of the measurement was noted for each experiment and used for the following data processing (see section 14.2.3).

RSB, pH 7.5		
20	mM	SDS
50	mM	sodium-phosphate buffer

APOL 333 stock, pH 7.5		
10	mM	sodium-phosphate buffer
10	% (w/v)	APOL333
0.03	% (v/v)	NaN_3

14.2.2 Experimental plate-reader script

After insertion of the plate a full spectrum was acquired from 260-900 nm (step size: 1 nm, number of flashes: 1, settle time: 0 ms) directly followed by measurement of the refolding kinetic data. These were monitored at three wavelengths, one corresponding to functionally refolded bR (555 nm), one representative for free retinal (380 nm) and the last one as a baseline (800 nm). The kinetic measurements were recorded for 24 h with an interval time of 30 s (number of flashes: 10, settle time: 0 ms). After acquisition of kinetic data another full spectrum was recorded with the same parameters as stated above.

14.2.3 Processing of kinetic data

For processing of the kinetic data the time of the measured data was corrected by addition of the time from adding bO into the first well until the start of the measurement. Afterwards the values of the baseline at 800 nm were subtracted from the values at 555 nm, indicative for functionally refolded bR. In the following step duplicates were averaged by the arithmetic mean value. Depending on the conducted experiment, mean values were additionally normalized to saturation of the curves at 15 000 s.

14.2.4 Applied fits

Kinetic data were processed as described above (section 14.2.3). Data of single refolding curves were further analyzed with the software Origin. K-values were obtained by fitting refolding curves with a non-linear fit (MnMolecular1) based on the following equation.

$$y = A1 - A2 \cdot e(-k \cdot x)$$

14.3 NMR SAMPLE PREPARATION

14.3.1 Purification of bO in SDS via IMAC

Pellets of cell-free expressed [^{15}N-^{13}C-D]-bO (see section 14.1.4) were resuspended in equal volumes of RSB, pH 8.0 (see section 14.2.1). A Ni-NTA-agarose column (Macherey-Nagel) was equilibrated with buffer P. The sample was loaded on the column and incubated for 1 h. Buffer P supplemented with 20 mM Imidazole was used for washing (4 steps of two column volumes), buffer P supplemented with 250 mM Imidazole was used for elution (5 steps of one column volume).

Protein containing fractions were pooled and centrifuged for 15 min at 15 000 x g. The supernatant was directly analyzed by preparative SEC using a Superdex 200 pg 16/600 (GE healthcare) column equilibrated with NMR buffer 1. A flow rate of 1 ml/min at room temperature was used for FPLC (ÄKTA *pure*, GE healthcare). Protein-containing fractions were pooled and concentrated in 10K MWCO-*SnakeSkin™ Dialysis Tubing* (Thermo scientific) to a volume of ~300 µl. The bO concentration was determined using a Nanodrop 2000 (Thermo scientific).

Buffer P, pH 8.0		
150	mM	NaCl
50	mM	sodium-phosphate buffer
20	mM	SDS

NMR buffer 1, pH 7.5		
50	mM	NaCl
20	mM	sodium-phosphate buffer
20	mM	SDS

14.3.2 Unpurified bO samples

Pellets of tested batches (see section 14.1.8) of cell-free expressed [^{15}N-^{13}C-D]-bO (see section 14.1.4) were resuspended in standard NMR buffer (see section 14.3.1), if not stated otherwise, and centrifuged for 15 min at 15 000 x g. For some experiments the SDS concentration was varied. The supernatant was used for further analysis.

14.3.3 Acquisition of TROSY-HSQC spectra

Purified or non-purified (see sections 14.3.1 and 14.3.2) samples were supplemented with D_2O and NaN_3 to final concentrations of 10 % (v/v) and 0.03 % (v/v), respectively. NMR data were acquired at 600 MHz (purified bO) and 700 MHz (non-purified bO) at 35°C. Detailed parameters for NMR data-acquisition can be found in section 15.1.3.

14.3.4 Combinatorial resonance assignment

Labelling patterns enabling combinatorial resonance assignments were created using the UPLABEL algorithm[340]. Detailed information about the applied labelling patterns can be found in section 15.1.1. Isotope labelling was carried out using cell-free expression in deuterated buffers (see section 14.1.4). Protein pellets (800-1000 µl) were resuspended in 200 µl NMR buffer 2 supplemented with EDTA-free protease inhibitor (Roche). Concentrations were ranging between 100 and 200 µM bO. NMR data were acquired at 600 MHz or 700 MHz at 35 °C. All NMR experiments contained Transverse Relaxation Optimized Spectroscopy (TROSY) components[418,419]. The 2D ^{13}CO- and ^{13}Cα-filtered TROSY-HSQC pulse sequences were designed from their respective 3D experiments. NMR spectra were processed using TOPSPIN 3.2 (Bruker) and analyzed with CCPN[421]. Detailed parameters for NMR data acquisition can be found in section 15.1.3.

NMR buffer 2, pH 7.5

50	mM	sodium-phosphate buffer
50	mM	SDS
10	% (v/v)	D_2O
0.003	% (v/v)	NaN_3

14.3.5 Real-time refolding

Isotope labelling of bO (^{15}N-AGS) was carried out using cell-free expression in deuterated buffers (see section 14.1.4). Protein pellets (1600 µl) were resuspended in 400 µl NMR buffer 3 supplemented with EDTA free protease inhibitor (Roche). BO concentrations were ranging between 80 and 100 µM. Initially, a TROSY-HSQC spectrum was acquired using a 3 mm NMR tube. Afterwards the sample was transferred into a 1.5 ml-tube. Folding was initiated by addition of APOL333 to a final concentration of 2% (v/v) and *all-trans* retinal to a final concentration of 80 µM. The sample was immediately transferred into a fresh 3 mm NMR tube. The duration from initiation of folding to starting the NMR experiment was measured. All spectra were acquired at 35 °C and 750 MHz. Detailed parameters for NMR-data acquisition can be found in section 15.1.3.

NMR buffer 3, pH 7.5

50	mM	sodium-phosphate buffer
10	mM	SDS
10	% (v/v)	D_2O
0.003	% (v/v)	NaN_3

14.3.6 Acquisition of TROSY-NOESY spectra

Non-purified (see section 14.3.2) bO pellets supplemented with D_2O and NaN_3 to final concentrations of 10 % (v/v) and 0.03 % (v/v), respectively. NMR data were acquired at 750 MHz and 35°C. Detailed parameters for NMR data acquisition can be found in section 15.1.3.

14.3.7 NMR spectrometers

All NMR experiments were performed on Bruker Avance III HD^+ spectrometers operating at ^1H frequencies between 600-900 MHz Cryogenic probes with 5 mm inverse detection triple-resonance and z-gradient coils were used to maximize ^1H signal detection.

TOPSPIN 3.4 or 3.5 (Bruker BioSpin) was used for experimental design, data acquisition and processing.

15 SUPPLEMENT

15.1 NMR

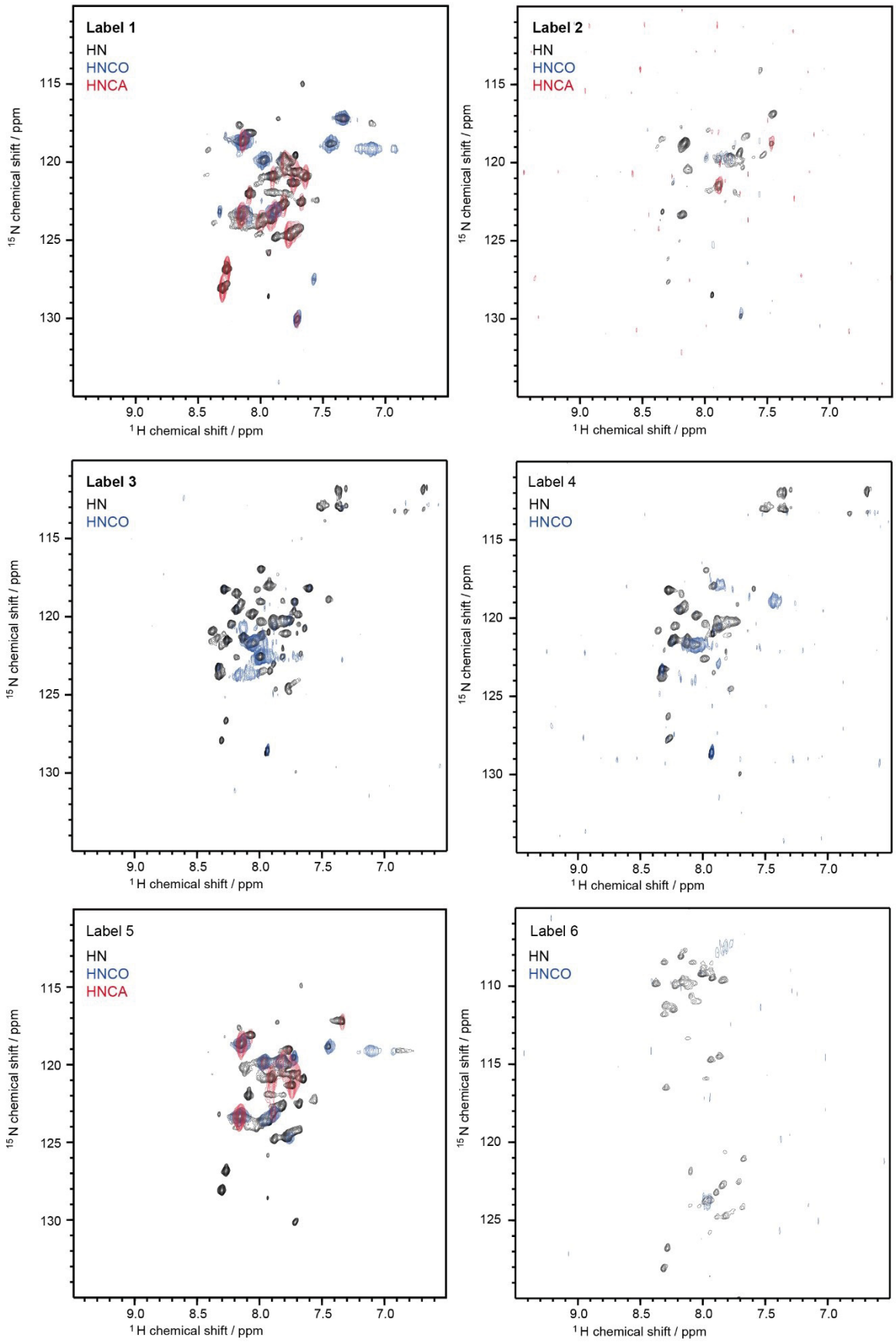
15.1.1 Combinatorial isotope labelling patterns

Table 5: Combinatorial isotope labelling patterns.

Amino acid	Label 1			Label 2			Label 3			Label 4			Label 5		
	13C	15N	U	13 C	15N	U	13C	15N	U	13 C	15N	U	13C	15N	U
Arginine (R)			R							R					R
Asparagine (N)								N			N				
Alanine (A)			A	A										A	
Aspartate (D)								D			D				
Glutamate (E)								E			E				
Glutamine (Q)								Q			Q				
Glycine (G)															
Histidine (H)															
Isoleucine (I)				I				I			I		I		
Leucine (L)															
Phenylalanine (F)	F			F									F		
Proline (P)	P														
Lysine (K)					K										K
Serine (S)															
Tryptophan (W)															
Threonine (T)							T			T					
Valine (V)	V				V		V								
Cysteine (C)															
Methionine (M)			M		M										M
Tyrosine (Y)		Y				Y		Y			Y			Y	

Amino acid	Label 6			Label 7			Label 8			Label 9		
	13C	15N	U	13C	15N	U	13C	15N	U	13C	15N	U
Arginine (R)				R				R		R		
Asparagine (N)												
Alanine (A)		A				A			A	A		
Aspartate (D)												
Glutamate (E)												
Glutamine (Q)												
Glycine (G)		G										
Histidine (H)												
Isoleucine (I)							I			I		
Leucine (L)												
Phenylalanine (F)									F		F	
Proline (P)						F						
Lysine (K)	K				K							
Serine (S)		S										
Tryptophan (W)											W	
Threonine (T)												
Valine (V)				V				V			V	
Cysteine (C)												
Methionine (M)									M			
Tyrosine (Y)												

15.1.2 NMR spectra of the combinatorial labelling approach



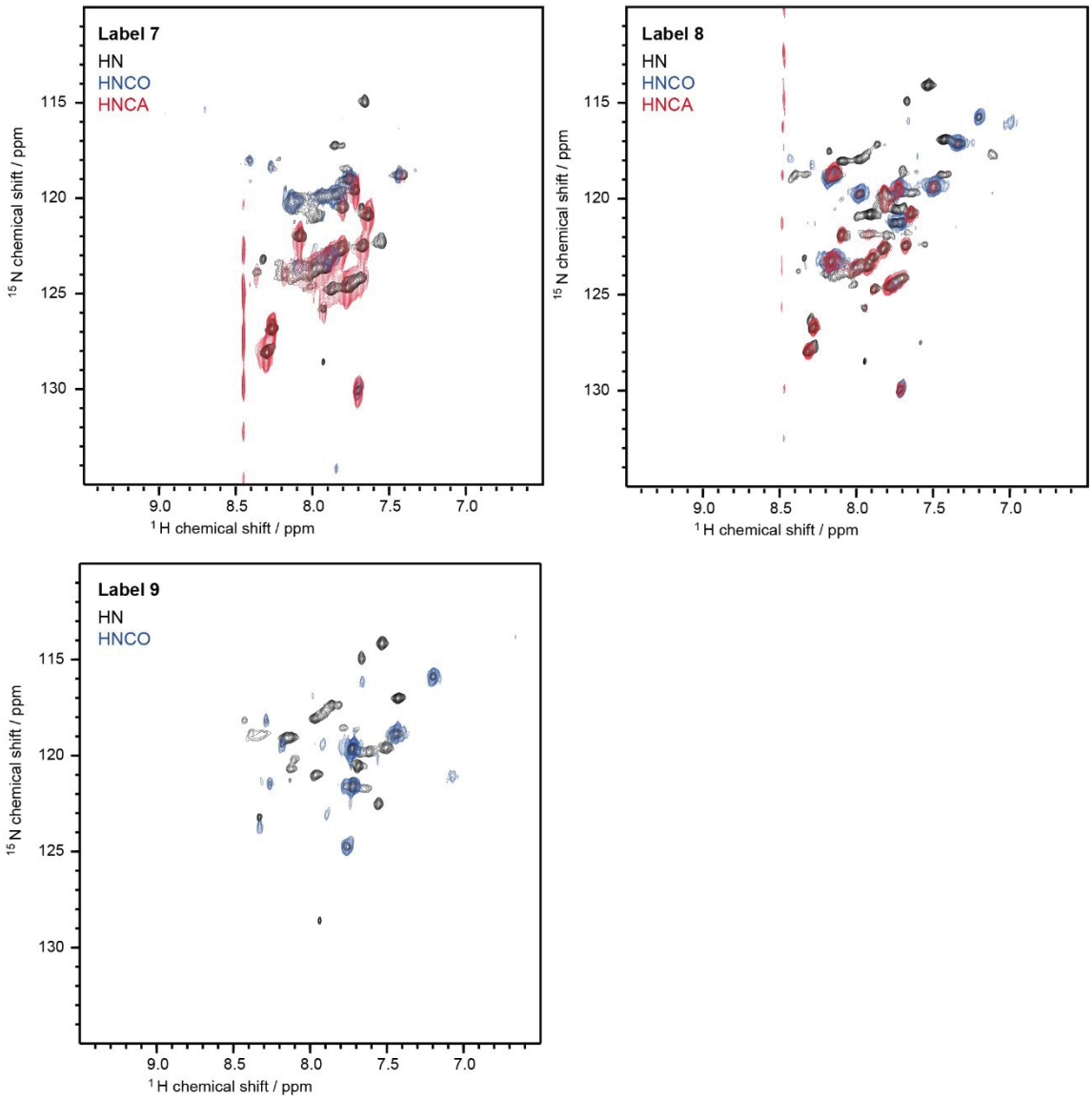


Figure 39: NMR spectra of the combinatorial isotope labelling approach. Overlay of a [^{1}H-^{15}N]-HSQC-TROSY (black) and the [^{1}H-^{15}N]-dimension of a HNCO (blue) and HNCA (red) acquired on SDS-solubilized bO, harboring labelling patterns 1-9 (see chapter 15.1.1). All spectra were acquired at 35 °C and 600 MHz or 700 MHz (see chapter 15.1.3).

15.1.3 NMR acquisition parameters

Table 6: Acquisition parameters of all conducted NMR experiments and labelling strategies in this work.

Labelling scheme	Experiment	Carrier freq. (^{1}H / ^{15}N / ^{13}CO or CA) in ppm			Spectral width (^{1}H / ^{15}N / ^{13}CO or CA) in ppm			No. of complex points (^{1}H / ^{15}N)		No. of scans	Magnetic field (^{1}H Larmor freq. in Hz)	Figure
Label 1	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	512	600	39,19C
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1120	600	39,19C
	HN(ca)	4.7	119.5	54	12	34	32	1024	32	1120	600	39,19C
Label 2	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	256	700	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	512	700	39
Label 3	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	1024	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	2048	600	39
Label 4	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	512	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1856	600	39
Label 5	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	512	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1376	600	39
	HN(ca)	4.7	119.5	54	12	34	32	1024	32	1376	600	39
Label 6	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	128	128	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	20	512	600	39
Label 7	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	128	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1120	600	39
	HN(ca)	4.7	119.5	54	12	34	32	1024	32	1120	600	39
Label 8	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	128	700	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1120	700	39
	HN(ca)	4.7	119.5	54	12	34	32	1024	32	1120	700	39
Label 9	[^{15}N-^{1}H] TROSY	4.7	119.5	-	12	34	-	1024	64	512	600	39
	HN(co)	4.7	119.5	176	12	34	22	1024	32	1024	600	39
Uniform - QNW	[^{15}N-^{1}H] TROSY (Fig. 16B, black)	4.7	119.5	-	18	40	-	2048	256	64	600	16B
Uniform	[^{15}N-^{1}H] TROSY (Fig. X, red)	4.7	119.5	-	12	34	-	2048	512	128	700	16B
Label 4	[^{15}N-^{1}H] TROSY (Fig. 17, red)	4.7	119.5	-	12	34	-	2048	128	512	600	17
Label 4	[^{15}N-^{1}H] TROSY (Fig. 17, black)	4.7	119.5	-	12	34	-	2048	128	512	600	17
Label 8	[^{15}N-^{1}H] TROSY (Fig. 17, purple)	4.7	119.5	-	12	34	-	2048	128	512	700	17
Label 8	[^{15}N-^{1}H] TROSY (Fig. 17, black)	4.7	119.5	-	12	34	-	2048	128	512	700	17
Uniform	[^{15}N-^{1}H] TROSY (Fig. 19A, black)	4.7	119.5	-	12	34	-	2048	512	128	700	19A,20

Uniform	[^{15}N-^{1}H-^{1}H] NOESY-TROSY (Fig. 20, blue/red)	4.7	119.5	-	12	34	14	2048	128/64	16	750	20
Before	[^{15}N-^{1}H] – BEST-TROSY	4.7	119.5	-	12	35	-	2048	64	64	750	29, black
After	[^{15}N-^{1}H] – BEST-TROSY	4.7	118.5	-	12	27	-	2048	128	128	750	29, red
Refolding	[^{15}N-^{1}H] – BEST-TROSY	4.7	119.5	-	12	35	-	2048	128	216	750	29, blue

15.2 PUBLICATIONS

15.2.1 Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems

Methods and experiment details

- Synthesis of isotope labelled amino acids was performed by Prof. Vladimir Gelev, Sofia University, Bulgaria.
- Expression of maltose binding protein (MBP) was carried out in *E. coli*
- Bacteriorhodopsin was produced via cell-free protein expression.
- Human recombinant RBM39$^{245\text{-}332}$ encapsulating RRM2 was expressed as N-terminal His_6 fusion in *Spodoptera frugiperda* Sf-9 insect cells using the baculovirus expression system (Invitrogen).
- Acquisition and processing of NMR spectra was carried out by Aldino Viegas and Thibault Viennet.

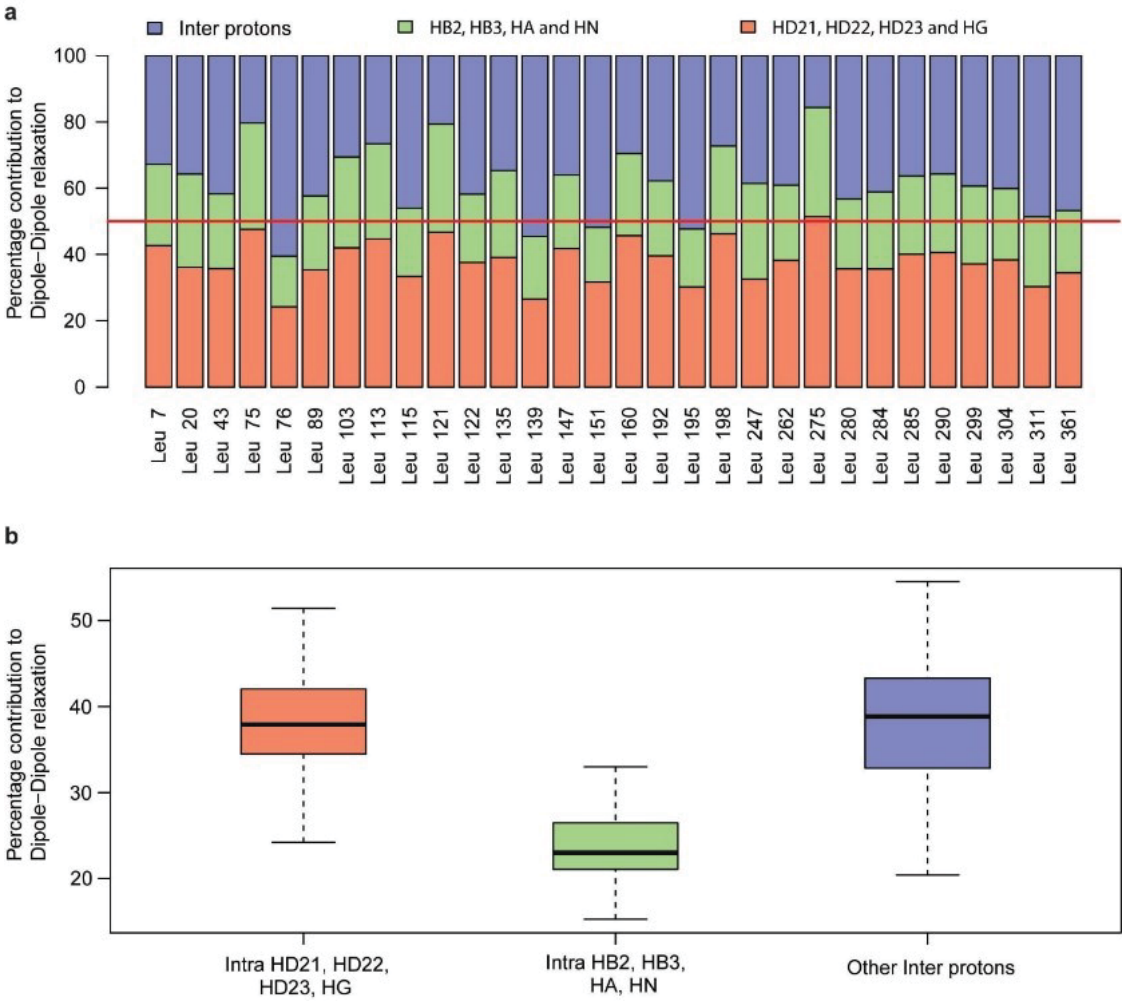


Figure 40: Estimation of the contribution of protons to the dipole-dipole relaxation of leucine-methyl protons in MBP. a) The stacked bars show the dipole-dipole contribution of three categories of protons to each leucine proton attached to Cδ1 in MBP. The orange bar is for intra-residue HD21, HD22, HD23 and HG atoms of leucine. The green bar is for intra-residue HB2, HB3, HA and HN atoms of leucine. The blue bar is for the rest of inter-residue protons. The red line indicates 50%. b) The statistical summary of the bar plot in (a) presented for 30 Leu residues in MBP as box plot.

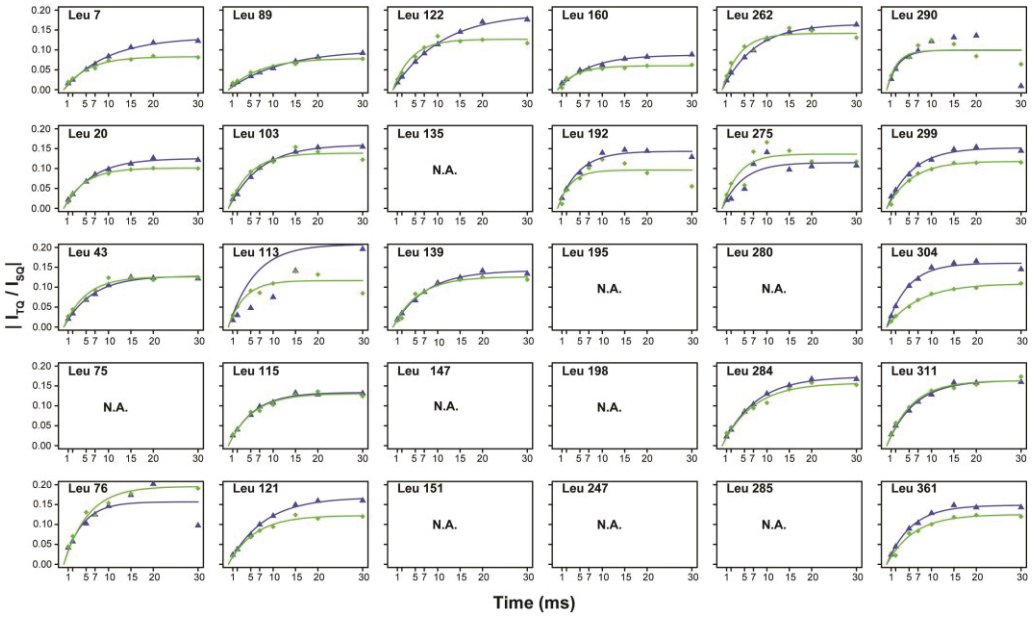


Figure 41: The Forbidden Coherence Transfer curve for all thirty Leu in MBP. The blue curve is FCT measured on the protein prepared using *Leu-meth*$_{LD}$ and green curve is FCT measured on fully protonated protein. 'N.A.' depicts data not available for the particular leucine either due to inability to unambiguously assign the peaks or inability to fit the FCT curve due to poor signal to noise ratio. The curves for Leu 76, and Leu 275 is unexpected and we attribute to peak integration errors due to poor sensitivity of triple quantum FCT experiment.

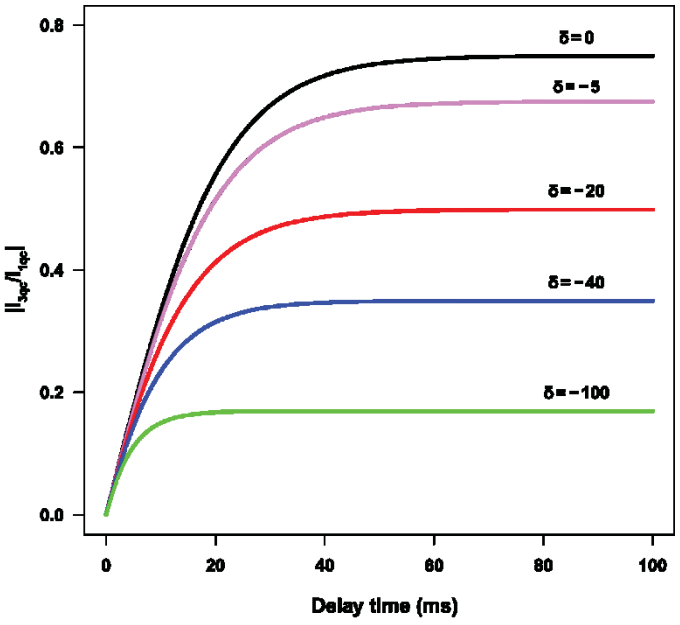


Figure 42: Simulation of FCT build-up curves. All the curves were simulated with identical parameters in equation S1, except for the values of δ. The δ value is indicative of the protons contributing towards the dipole-dipole relaxation of the methyl protons. The fitted value of δ on experimental data can be used for relative comparison of surrounding proton density.

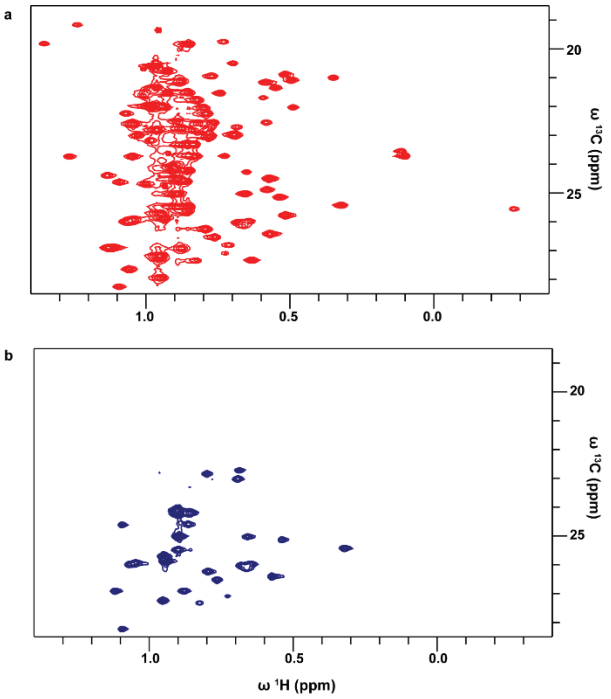


Figure 43: 2D [^{13}C, ^{1}H]-HSQC of MBP expressed in *E. coli* using (a) ^{13}C-acetolactate as precursor, which uniformly labels Leu and Val methyl and (b) *Leu-meth*$_{LD}$.

Table 7: Acquisition parameters for different NMR experiments done in the present study.

Experiments	Forbidden Coherence Transfer		2D [^{13}C, ^{1}H] T2 measurement	2D [^{13}C, ^{1}H] HSQC	2D [^{13}C, ^{1}H] HMQC
	Triple Quantum Coherence	Single Quantum Coherence		Insect Cell Sample	Cell free
Complex Points (F2 X F1)	2048 X 104	2048 X 104	1024 X 128	4096 X 2048	512 X 206
Acquisition Time (ms)	98.3 and 21.5	98.3 and 21.5	49.1 and 26.5	196.7 and 63.6	113.7 and 27.3
# Scans	200	80	80	16	336
Recycle Delay (s)	1.2	1.2	1.2	1	1
Experiment Time (hh:mm)	74 hours and 56 mins	30 hours	47 hours and 13 mins	11 hours and 20 mins	23 hours and 55 mins

15.2.2 The use of amphipols for NMR structural characterization of 7-TM proteins

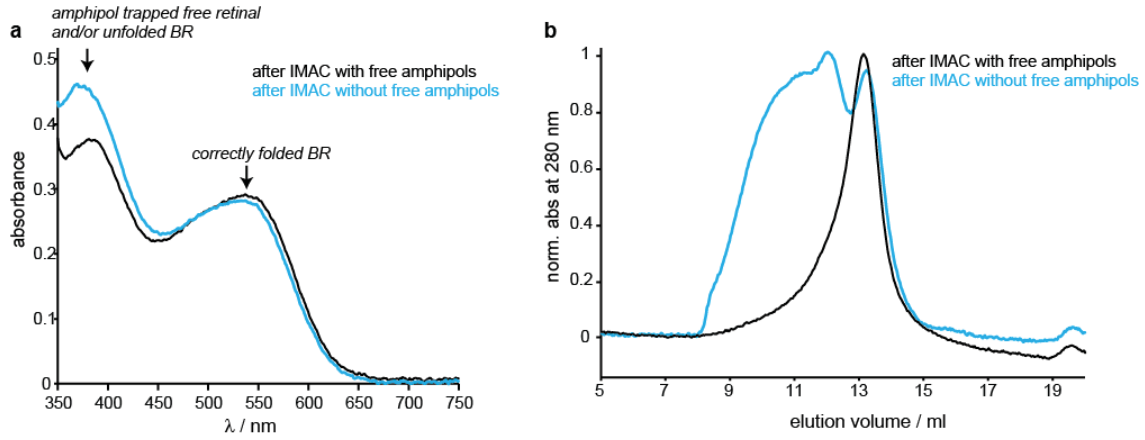


Figure 44: Analysis of different bR-APOL preparations. (a) Quantitative analysis of different bR-APOL preparations. Cell-free expressed protein pellet was refolded with amphipols as described in material and methods. Refolded bR-APOL was split into two and loaded onto a pre-equilibrated Ni-NTA agarose matrix. Identical volumes of bR-APOL as well as Ni-NTA agarose were used. IMAC purification of bR-APOL was carried out using either amphipol-free buffer (blue) or buffer supplemented with amphipols (0.08% in washing and 0.15% in elution buffer) (black). Figure 44a shows the resulting absorption spectra of the pooled elution fraction 2 and 3 after IMAC. The total amount of correctly folded bR is nearly identical when IMAC was carried out in the presence or absence of amphipols in the buffer. However, the different perpetrations have a significant effect on the resulting SEC profile as visible in (b). Note that normalized profiles are given in (b) and that larger fractions in blue profile also show absorption at 555 nm. The data suggest that IMAC with amphipol-free buffer does not unfold bR, however it leads to formation of larger aggregates (containing folded bR). Hence, NMR spectra of bR in these larger sized amphipols (Fig. 9c) likely reflect well-folded bR.

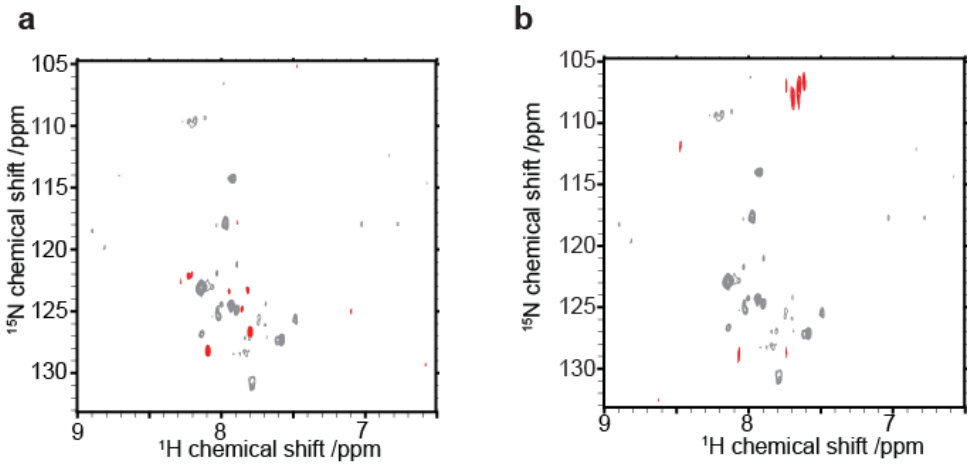


Figure 45: 2D [^{15}N, ^{1}H]-TROSY-HSQC natural abundance spectra of components present at larger quantities in NMR sample of amphipol refolded MC4R. Natural abundance spectrum of SHU9119 (a) and amphipols (b) are shown in red. Spectrum of isotope labelled MC4R-APOL as shown in Figure 10b is overlaid in grey.

15.2.3 bR reconstitution into tethered lipid membranes for AFM Force Spectroscopy

Materials

Buffers

1) Tris-buffered saline and Tween 20 (TBT): 20 mM Tris, 150 mM NaCl, 0.1 % Tween 20, pH 7.5

2) Tris-buffered saline (TBS): 20 mM Tris, 150 mM NaCl, pH 7.5

3) phosphate buffer (P1): 46.6 mM Na_2HPO_4, 3.4 mM NaH_2PO_4, 150 mM NaCl, pH 8

4) wash buffer (DDM-W1): P1 + 0,4 mM DDM + 20 mM Imidazole

5) protein buffer (DDM-P1): P1 + 0,4 mM DDM

6) elution buffer (DDM-E): DDM-P1 + 250 mM Imidazole

7) refolding buffer (RFB): 23.1 mM Na_2HPO_4, 16.9 mM NaH_2PO_4, 1 M NaCl, 5%

DDM, 100 mM Retinal, pH 7

8) S30 buffer: 10 mM Tris-acetate, pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 1 mM

DTT and 0.1 mM PMSF

Experimental Design and Methods

Dot blot

A functionalized quartz surface with bound bR was first rinsed with 100 µl DDM-P1 by pipetting up and down several times. The rinsing buffer was collected as the rinsing fraction. The same surface was then rinsed the same way with 100 µl DDM-E, which was again collected as the elution fraction. A Dot blot was prepared of 100 µl DDM-E (buffer control), 100 µl rinsing fraction and 100 µl elution fraction on a membrane. The membrane was swayed in 1 % (w/v) powdered milk solution at 4 °C, overnight. After washing with water the membrane was swayed 3 x 10 min in TBT and 1 x 10 min in TBS at RT. The membrane was shortly washed with water and swayed in histidin-tag (His-tag) antibody solution (1:5000 anti-His-HRP (Miltenyi Biotec, Bergisch Gladbach, Germany) in 1 % (w/v) powdered milk solution for 1 h, at RT. This was followed by swaying 2 x 10 min in TBT and 1 x 10 min TBS at RT and shortly rinsing with water. The membrane was covered with a 1:1 (v/v) solution of luminol enhancer solution:stable peroxide solution, (both SuperSignal West Pico, Thermo Fisher Scientific, Oberhausen, Germany) and chemoluminescence was measured with a ChemiDoc MP documentation system (Bio-Rad, Hercules, U.S.).

Protein refolding

The protein pellet from CFE was washed with S30 buffer and directly refolded or stored at -20 °C. 2 x 50 µl protein pellets were resuspended in 2 x50 µl RFB. The samples were wrapped in aluminum foil to be protected from light and let to incubate for 2 h, at RT. A Micro Bio-Spin(TM) gravity-flow column (Bio-Rad, Hercules, U.S.) was filled with 160 µl Ni-NTAAgarose (Macherey-Nagel, Germany) for a total column volume (CV) of 80 µl. After letting the agarose settle the column was washed with 5 ml water and equilibrated with 5 ml DDMP1. The resuspended pellets were combined and loaded onto the column. The agarose and bR sample were pipetted up and down to mix and increase binding. Covering the column from light it was let to incubate for 1 h while mixing every 20 min. The column was then washed with 6 CV DDM-W1 and 2 CV DDM-P1. BR was eluted using 4 CV DDM-E giving elution fractions 1-4. Fractions 2-4 were combined for concentration. A concentrator tube (10.000 MWCO, Vivaspin 4, PES membrane, Sartorius, Göttingen, Germany) was washed with 6 x 10 ml water, 4 x 10 ml P1 and 1 x 10 ml DDM-P1 by centrifugation at 4 °C, 4000 x g, 15 min each. 230 µl BR (fractions 2-4) were loaded into the concentrator and centrifuged at 4 °C, 4000 x g for 20 min until the sample was concentrated to 50 µl. The sample was washed with DDM-P1 by centrifugation at 4 °C, 4000 g for 20 min until again the total volume was concentrated to 50 µl. 10 µl of the sample were used to measure the concentration in a 1:10 (v/v) dilution in DDM-P1 at the spectrophotometer (V-650, Jasco, Groß-Umstadt, Germany). An absorption spectrum was measured between 260 nm and 750 nm. The peak at 555 nm corresponds to the adsorption of successfully refolded bR and is used to calculate the concentration with an extinction coefficient of $\varepsilon = 55000\ L \cdot cm^{-1} \cdot mol^{-1}$.

Functionalization steps

Plasma activation

Quartz Slides (3” x 1” x 1 mm, TED PELLA Inc.) were placed into a plasma cleaner (Plasma, low-pressure system, Femto, Diener electronics) and 500 µl of water were distributed as drops inside the chamber and allowed to evaporate under vacuum to saturate the atmosphere inside the plasma cleaner with H_2O. The surfaces were activated for 5 min at 100 %. Further functionalization was done immediately after.

Silanization

Adapted from method B of the protocol 2 (1). A desiccator was flooded with nitrogen for 5 min. The plasma activated surfaces were placed into the desiccator together with 100 µl of APTES and 33 µl of triethylamine (Et_3N; puriss., p.a. ≥99.5 % GC, Fluka, Seelze, Germany). After 5 min of constant nitrogen flow the desiccator was closed completely and the silanization was allowed to take place for 2 h. APTES and Et_3N were removed from the desiccator and it was again flooded with nitrogen for 5 min. After closing the desiccator the surfaces were allowed to cure for at least 48 h.

Functionalization with polyethylene glycol

Mmt-NH-PEG_{12}-COO-Tfp (PEG; Iris Biotech, Marktredwitz, Germany), Et_3N and dimethylsulfoxide (DMSO; A.C.S, spectrophotometric Grade ≥99.9 %, Sigma-Aldrich,

Munich, Germany) were mixed 20:1:200 (v/v/v). The silanized surfaces were taken out of the desiccator and placed into petri dishes. A small space between two silanized slides facing each other, created by clean cover slips, was filled with 10 µl/cm² of the solution followed by incubation for 2 h. After incubation the surfaces were placed into 50 ml centrifuge tubes (TPP, Gamma-sterilized, Sigma-Aldrich, Munich, Germany) and rinsed 10 x with water by filling and emptying the tube completely. After drying with nitrogen the functionalized surfaces were stored at -20 °C in fresh 50 ml falcon tubes. Figure 11E schematically shows the surface functionalized with PEG.

Binding anchoring lipids and trisNTA to PEG

To remove the monomethoxytrityl (Mmt) protection group from PEG the surfaces were covered with 80 % acetic acid (Ultra, ≥99.5 % GC/T, Fluka, Seelze, Germany) for 1 h inside 50 ml falcon tubes. Meanwhile a solution was prepared depending on the targeted surface functionalization. trisNTA surface: 10:10:1 (v/v/v) of trisNTA stock: *N,N'*-Diisopropylcarbodiimide (DIC; 99 %, Sigma-Aldrich, Munich, Germany): N,NDiisopropylethylamine(DIPEA; Reagent Plus, 99 %, Sigma-Aldrich, Munich, Germany). SuccinylPE surface: 2:10:1 (v/v/v) of SuccinylPE stock:DIC:DIPEA. Mixed surface: 1:5:10:1 (v/v/v/v) of SuccinylPE stock:trisNTA stock:DIC:DIPEA. After deprotection the slides were rinsed again ten times with water by filling and emptying the falcon tubes completely and then dried under nitrogen. A small space between two glass slides facing each other, created by clean cover slips (Microscope Cover Glasses, #1, 22 x 22 mm, Menzel Gläser, Braunschweig, Germany), was filled with 10 µl/cm² of the respective solution followed by incubation for 1 h at 75 °C. Two clean glass beakers were filled with chloroform and the functionalized surfaces were dipped and rinsed inside the beakers subsequently. The surfaces were dried with nitrogen and placed into glass petri dishes with the functionalized side facing upwards. They were covered with trifluoracetic acid (TFA; ≥99.9 % for synthesis, Carl Roth, Karlsruhe, Germany) and a few drops of water for 12 h to remove the OtBu protection groups from trisNTA.

Supplementary figures

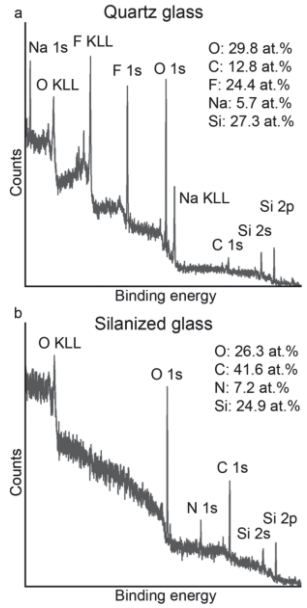


Figure 46: Quartz glass properties. (a) XPS survey spectrum of cleaned quartz glass. (b) A survey spectrum after silanization with APTES only shows binding energies corresponding to levels in O, C, N and Si.

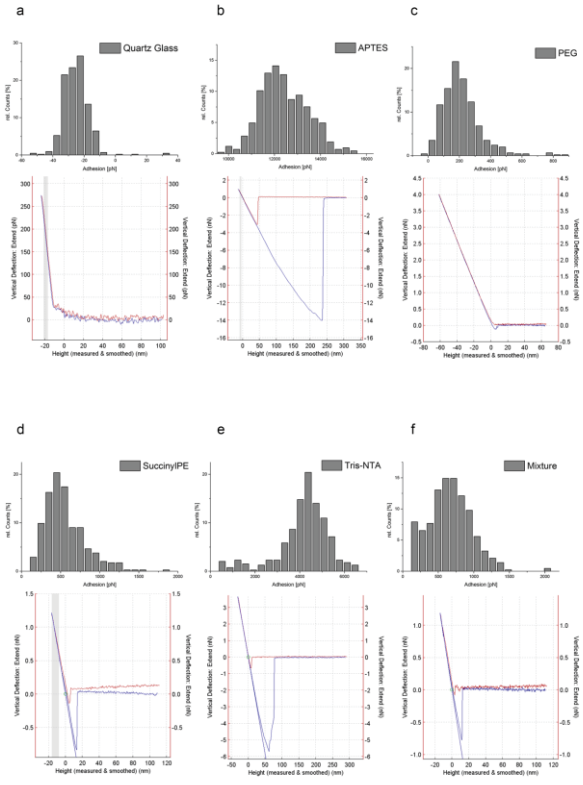


Figure 47: Functionalization steps. Histograms showing the characteristic adhesion forces for the surface functionalization steps a-f. Number of total counts: a) 419, b) 425, c) 454, d) 344, e) 392, f) 429. Below each histogram a typical force-distance curve is shown. The approach of the cantilever is represented by the red curve, while the retrace is shown in blue.

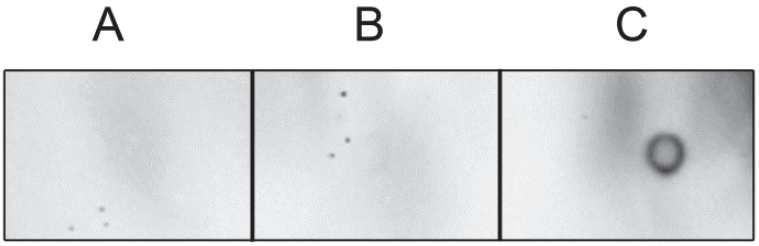


Figure 48: Dot blot of buffer control consisting of buffer DDM-E (buffer with imidazole) (A), rinsing fraction consisting of buffer P1-DDM (buffer without imidazole), which was used to rinse the surface (B) and elution fraction consisting of buffer DDM-E, which was used to rinse the surface and elute bR (C).

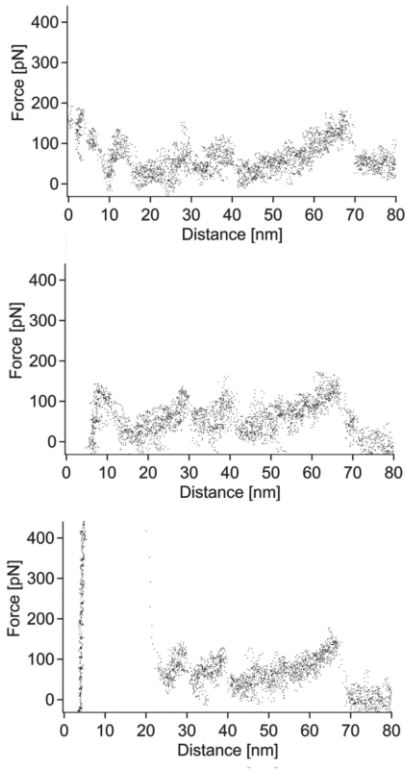


Figure 49: Single representative measured force distance curves of bR from the tBLM.

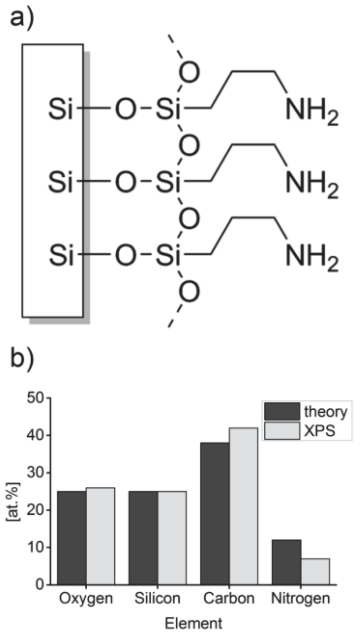


Figure 50: Silanization of quartz glass. (a) Schematic representation of the theoretical silanization of quartz glass. (b) Theoretical composition of silanized surface (theory) compared with XPS results (XPS).

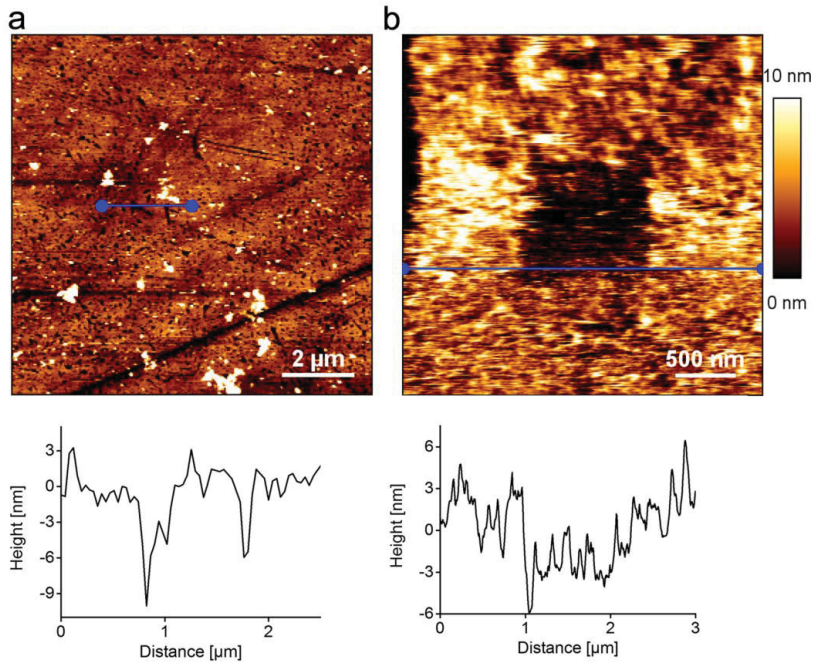


Figure 51: AFM scans (10 µm x 10 µm, contact mode for a) and 3 µm x 3 µm, tapping mode for b)) in buffer P1 with corresponding height profiles, represented as horizontal lines in each image, below each scan. (a) A membrane formed on a functionalized quartz glass surface without bR. (b) Attempted membrane formation on a quartz glass surface functionalized only with the protein anchor trisNTA and bR included in the lipid suspension. We could observe that no membrane was formed without the lipid anchor, as is indicated by the scratched area in the center.

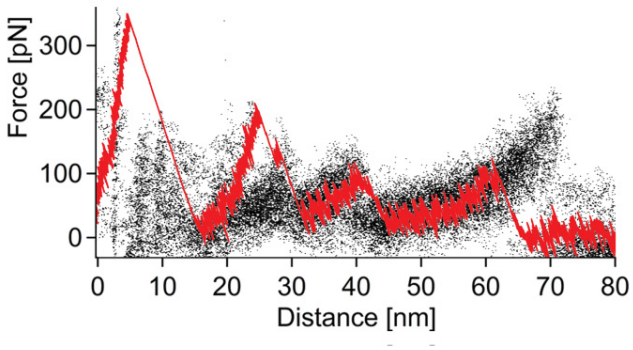


Figure 52: Distance curve of bR from PM. Representative force distance curve of bR from purple membranes pulled from the C-terminus (red force distance curve) placed over the measured force distance curves of the tBLM, also pulled from the C-terminus.

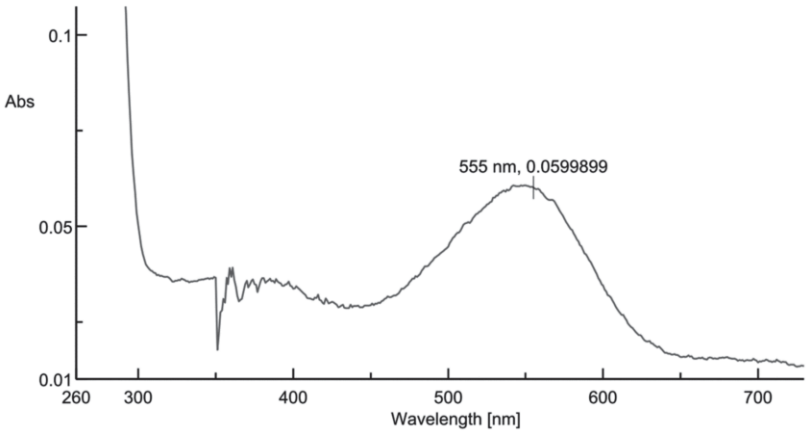


Figure 53: Example of absorbance spectrum of purified bR from CFE. 10 µl of the sample were used to measure the concentration in a 1:10 (v/v) dilution in DDM-P1 at the spectrophotometer. An absorption spectrum was measured between 260 nm and 750 nm. The peak at 555 nm corresponds to the adsorption of successfully refolded bR and is used to calculate the concentration with an extinction coefficient of $\varepsilon = 55000\ L \cdot cm^{-1} \cdot mol^{-1}$.

15.2.4 Reconstitution and NMR characterization of the ion-channel accessory subunit barttin in detergents and lipid-bilayer nanodiscs

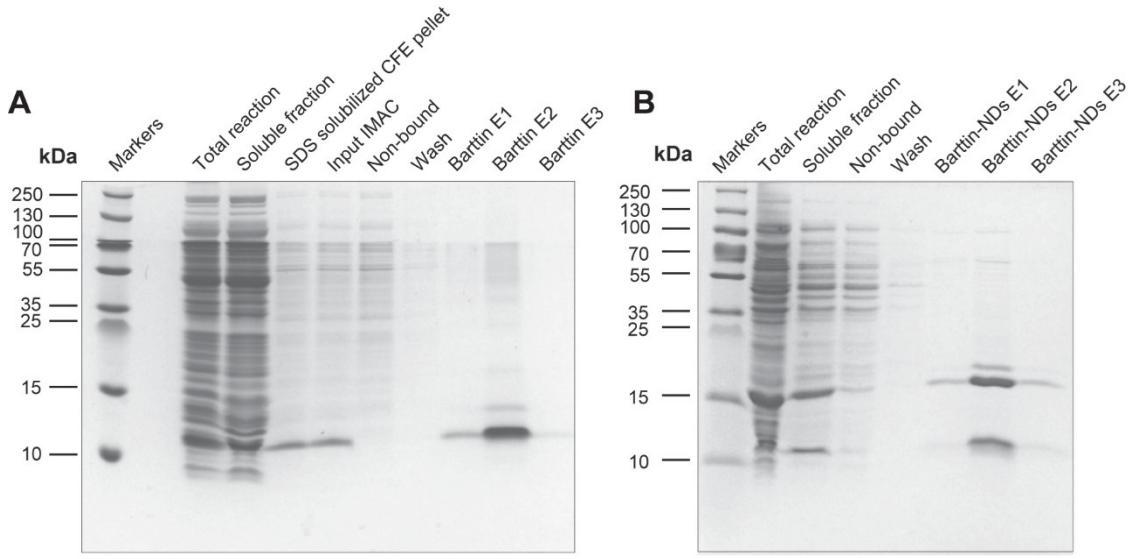


Figure 54: Refolding *vs.* co-translational incorporation of barttin into NDs. (A) Coomassie blue-stained SDS PAGE of CFE expressed barttin (total reaction), soluble fraction of the reaction, solubilized CFE pellet, input for IMAC purification in SDS, non-bound fraction, washing fraction and three elution fractions of SDS-purified barttin (full gel of Fig. 33). (B) Coomassie blue-stained SDS-PAGE of CFE-expressed barttin in the presence of DMPC Δ5 NDs (total reaction), soluble fraction of the reaction, input for IMAC purification, non-bound fraction, washing fraction and three elution fractions of barttin NDs (full gel of Fig. 36).

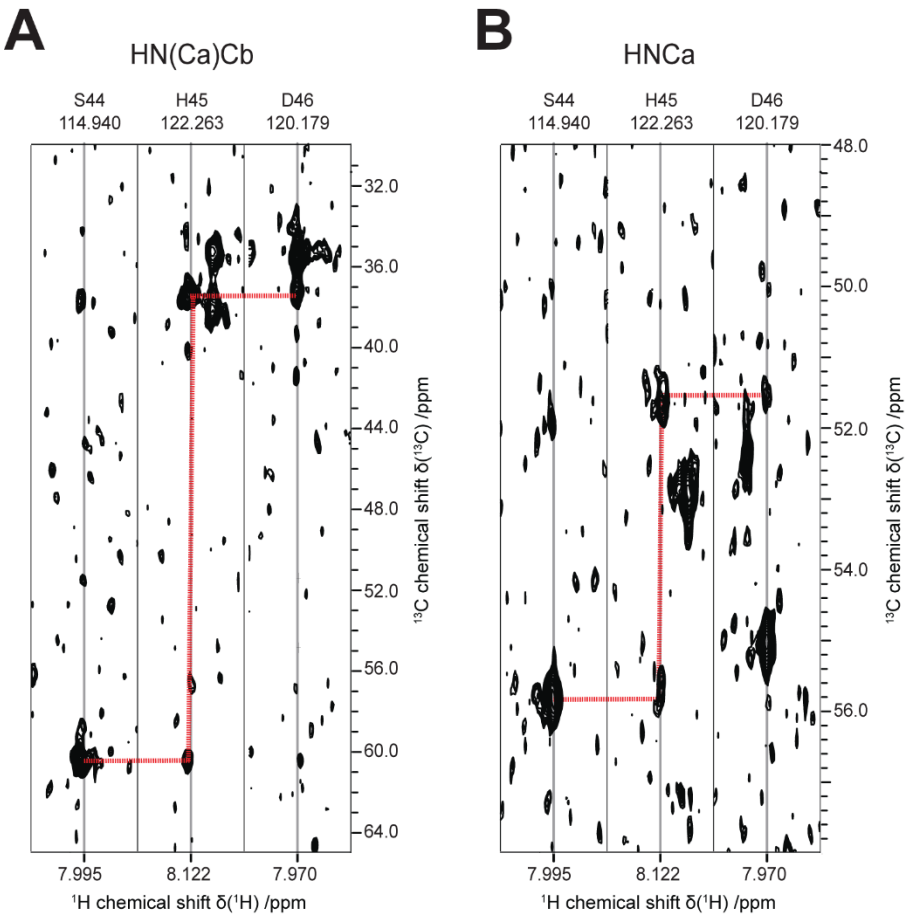


Figure 55: Barttin standard assignment procedure is impeded by the quality of NMR data in LDAO. a) Strips of the HNCa spectrum of [^{2}H-^{13}C-^{15}N]-labelled barttin in LDAO micelles corresponding to the S44-H45-D46 fragment, sequential links through Cα signals are exemplified by arrows. b) Strips of the HN(Ca)Cb spectrum of [^{2}H-^{13}C-^{15}N]-labelled barttin in LDAO micelles corresponding to the S44-H45-D46 fragment, sequential links through Cβ signals are exemplified by arrows.

16 BIBLIOGRAPHY

1. Vit, O. & Petrak, J. Integral membrane proteins in proteomics. How to break open the black box? *Journal of Proteomics* (2017). doi:10.1016/j.jprot.2016.08.006

2. Consortium, I. H. G. S. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

3. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–51 (2001).

4. Bénard, G. *et al.* Mitochondrial CB1 receptors regulate neuronal energy metabolism. *Nat. Neurosci.* **15**, 558–564 (2012).

5. Schiaffino, M. V. *et al.* Ocular albinism: evidence for a defect in an intracellular signal transduction system. *Nat. Genet.* **23**, 108–112 (1999).

6. Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B. & Prossnitz, E. R. A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science* **307**, 1625–30 (2005).

7. Oksche, A. *et al.* Late endosomal/lysosomal targeting and lack of recycling of the ligand-occupied endothelin B receptor. *Mol. Pharmacol.* **57**, 1104–13 (2000).

8. Rozenfeld, R. & Devi, L. A. Regulation of CB1 cannabinoid receptor trafficking by the adaptor protein AP-3. *FASEB J.* **22**, 2311–22 (2008).

9. Sergin, I., Jong, Y.-J. I., Harmon, S. K., Kumar, V. & O'Malley, K. L. Sequences within the C Terminus of the Metabotropic Glutamate Receptor 5 (mGluR5) Are Responsible for Inner Nuclear Membrane Localization. *J. Biol. Chem.* **292**, 3637–3655 (2017).

10. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–72 (2003).

11. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

12. Tyndall, J. & Sandilya, R. GPCR Agonists and Antagonists in the Clinic. *Med. Chem. (Los. Angeles).* **1**, 405–421 (2005).

13. Lefkowitz, R. J. A Brief History of G-Protein Coupled Receptors (Nobel Lecture). *Angew. Chemie Int. Ed.* **52**, 6366–6378 (2013).

14. Lappano, R. & Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat. Rev. Drug Discov.* **10**, 47–60 (2011).

15. Allen, J. A. & Roth, B. L. Strategies to Discover Unexpected Targets for Drugs Active at G Protein–Coupled Receptors. *Annu. Rev. Pharmacol. Toxicol.* **51**, 117–144 (2011).

16. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).

17. Alexander, S. P. *et al.* The Concise Guide to PHARMACOLOGY 2015/16: G protein-coupled receptors. *Br. J. Pharmacol.* **172**, 5744–869 (2015).

18. Fredriksson, R. & Schiöth, H. B. The Repertoire of G-Protein-Coupled Receptors in Fully Sequenced Genomes. *Mol. Pharmacol.* **67**, 1414–1425 (2005).

19. Wheatley, M. *et al.* Lifting the lid on GPCRs: the role of extracellular loops. *Br. J. Pharmacol.* **165**, 1688–1703 (2012).

20. Simon, M. I., Strathmann, M. P. & Gautam, N. Diversity of G proteins in signal transduction. *Science* **252**, 802–8 (1991).

21. Downes, G. B. & Gautam, N. The G Protein Subunit Gene Families. *Genomics* **62**, 544–552 (1999).

22. Higashijima, T., Ferguson, K. M., Sternweis, P. C., Smigel, M. D. & G. & G., A. The Effect of GTP and Mg2’ on the GTPase Activity and the Fluorescent Properties of Go*. **20**, 41–45 (1988).

23. Ross, E. M. & Wilkie, T. M. GTPase-Activating Proteins for Heterotrimeric G Proteins: Regulators of G Protein Signaling (RGS) and RGS-Like Proteins. *Annu. Rev. Biochem.* **69**, 795–827 (2000).

24. Kimple, A. J., Bosch, D. E., Giguère, P. M. & Siderovski, D. P. Regulators of G-protein signaling and their Gα substrates: promises and challenges in their use as drug discovery targets. *Pharmacol. Rev.* **63**, 728–49 (2011).

25. Oesterhelt, D. & Stoeckenius, W. Rhodopsin-like Protein from the Purple Membrane of Halobacterium halobium. *Nat. New Biol.* **233**, 149–152 (1971).

26. Oesterhelt, D. & Stoeckenius, W. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2853–7 (1973).

27. Kayushin, L. P. & Skulachev, V. P. Bacteriorhodopsin as an electrogenic proton pump: Reconstitution of bacteriorhodopsin proteoliposomes generating Δψ and ΔpH. *FEBS Lett.* **39**, 39–42 (1974).

28. Haupts, U., Tittor, J. & Oesterhelt, D. Closing in on bacteriorhodopsin: Progress in Understanding the Molecule. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 367–399 (1999).

29. Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32 (1975).

30. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).

31. Lanyi, J. K. Mechanism of Ion Transport across Membranes. *J. Biol. Chem.* **272**, 31209–31212 (1997).

32. Wickstrand, C., Dods, R., Royant, A. & Neutze, R. Bacteriorhodopsin: Would the real structural intermediates please stand up? *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 536–553 (2015).

33. Hasegawa, N., Jonotsuka, H., Miki, K. & Takeda, K. X-ray structure analysis of bacteriorhodopsin at 1.3 Å resolution. *Sci. Rep.* **8**, 1–8 (2018).

34. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307**, 249–264 (2007).

35. Miroux, B. & Walker, J. E. 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 (1996).

36. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**, 203–206 (2004).

37. Burgess-Brown, N. A. *et al.* Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr. Purif.* **59**, 94–102 (2008).

38. Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3**, 303–313 (2006).

39. Neophytou, I. *et al.* Eukaryotic integral membrane protein expression utilizing the Escherichia coli glycerol-conducting channel protein (GlpF). *Appl. Microbiol. Biotechnol.* **77**, 375–381 (2007).

40. Mironova, R., Niwa, T., Handzhiyski, Y., Sredovska, A. & Ivanov, I. Evidence for non-enzymatic glycosylation of Escherichia coli chromosomal DNA. *Mol. Microbiol.* **55**, 1801–1811 (2005).

41. Mijakovic, I. *et al.* Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. *Nucleic Acids Res.* **34**, 1588–1596 (2006).

42. Strausberg, R. L. & Strausberg, S. L. Overview of Protein Expression in Saccharomyces cerevisiae. in *Current Protocols in Protein Science* **2**, 5.6.1-5.6.7 (John Wiley & Sons, Inc., 1995).

43. Bornert, O., Alkhalfioui, F., Logez, C. & Wagner, R. Overexpression of Membrane Proteins Using *Pichia pastoris*. in *Current Protocols in Protein Science* **29**, 29.2.1-29.2.24 (John Wiley & Sons, Inc., 2012).

44. Yang, Y. *et al.* High human GLUT1, GLUT2, and GLUT3 expression in Schizosaccharomyces pombe. *Biochem. (Mosc).* **74**, 75–80 (2009).

45. Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E. & Chisti, Y. Protein production using the baculovirus-insect cell expression system. *Biotechnol. Prog.* **30**, 1–18 (2014).

46. Ciccarone, V. C., Polayes, D. A. & Luckow, V. A. Generation of Recombinant Baculovirus DNA in E. coli Using a Baculovirus Shuttle Vector. in *Molecular Diagnosis of Infectious Diseases* 213–236 (Humana Press). doi:10.1385/0-89603-485-2:213

47. Louis, N., Evelegh, C. & Graham, F. L. Cloning and Sequencing of the Cellular–Viral Junctions from the Human Adenovirus Type 5 Transformed 293 Cell Line. *Virology* **233**, 423–429 (1997).

48. Zhu, J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* **30**, 1158–1170 (2012).

49. Khan, K. H., Hayat Khan, K. & Khan, K. H. Gene Expression in Mammalian Cells and its Applications. *Adv. Pharm. Bull.* **3**, 257–263 (2013).

50. Condreay, J. P., Witherspoon, S. M., Clay, W. C. & Kost, T. A. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 127–32 (1999).

51. Baldwin, S. *et al.* Transient and stable transfection of Chinese hamster ovary cells with the recombinant feline erythropoietin gene and expression, purification, and biological activity of feline erythropoietin protein. *Am J Vet Res.* **64**, 1465–71 (2012).

52. Thomas, P. & Smart, T. G. HEK293 cell line: A vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187–200 (2005).

53. Buchner, E. Alkoholische Gährung ohne Hefezellen. *Berichte der Dtsch. Chem. Gesellschaft* **30**, 117–124 (1897).

54. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1588–602 (1961).

55. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci.* **99**, 14652–14657 (2002).

56. Rosenblum, G. & Cooperman, B. S. Engine out of the chassis: Cell-free protein synthesis and its uses. *FEBS Lett.* **588**, 261–268 (2014).

57. Henrich, B., Lubitz, W. & Plapp, R. Lysis of Escherichia coli by induction of cloned phi X174 genes. *MGG Mol. Gen. Genet.* **185**, 493–497 (1982).

58. Goff, S. A. & Goldberg, A. L. An increased content of protease La, the lon gene product, increases protein degradation and blocks growth in Escherichia coli. *J. Biol. Chem.* **262**, 4508–4515 (1987).

59. Chrunyksp, B. a, Evans, J., Lillquistol, J., Youngti, P. & Ii, R. W. Formation and Protein Stability in Sequence Variants of Interleukin-lfl*. *Society* **268**, 18053–18061 (1993).

60. Schwarz, D. *et al.* Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat. Protoc.* **2**, 2945–2957 (2007).

61. Kurland, C. G. Translational accuracy in vitro. *Cell* **28**, 201–202 (1982).

62. Pavlov, M. Y. & Ehrenberg, M. Rate of Translation of Natural mRNAs in an Optimizedin VitroSystem. *Arch. Biochem. Biophys.* **328**, 9–16 (1996).

63. Klammt, C. *et al.* High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271**, 568–580 (2004).

64. Shimizu, M., Ikegami, T., Akiyama, K. & Morita, E. H. A Novel Way to Express Proline-Selectively Labeled Proteins with a Wheat Germ Cell–Free Protein Synthesis System. *J. Biochem.* **140**, 453–456 (2006).

65. Tyler, R. C. *et al.* Comparison of cell-based and cell-free protocols for producing target proteins from the Arabidopsis thaliana genome for structural studies. *Proteins Struct. Funct. Bioinforma.* **59**, 633–643 (2005).

66. Vinarov, D. A. *et al.* Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat. Methods* **1**, 149–153 (2004).

67. Ryu, Y. & Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **3**, 263–265 (2006).

68. Zubay, G. In Vitro Synthesis of Protein in Microbial Systems. *Annu. Rev. Genet.* **7**, 267–287 (1973).

69. Kigawa, T. *et al.* Selenomethionine incorporation into a protein by cell-free synthesis. *J. Struct. Funct. Genomics* **2**, 29–35 (2002).

70. Kainosho, M. & Tsuji, T. Assignment of the three methionyl carbonyl carbon resonances in Streptomyces subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry* **21**, 6273–6279 (1982).

71. Guignard, L., Ozawa, K., Pursglove, S. E., Otting, G. & Dixon, N. E. NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. *FEBS Lett.* **524**, 159–162 (2002).

72. Shirokov, V. A., Kommer, A., Kolb, V. A. & Spirin, A. S. Continuous-Exchange Protein-Synthesizing Systems. in *In Vitro Transcription and Translation Protocols* 19–55 (Humana Press, 2007). doi:10.1007/978-1-59745-388-2_2

73. Kim, T.-W. *et al.* An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *J. Biotechnol.* **130**, 389–393 (2007).

74. Klammt, C., Schwarz, D., Dötsch, V. & Bernhard, F. Cell-Free Production of Integral Membrane Proteins on a Preparative Scale. in *In Vitro Transcription and Translation Protocols* 57–78 (Humana Press, 2007). doi:10.1007/978-1-59745-388-2_3

75. Berrier, C. *et al.* Cell-Free Synthesis of a Functional Ion Channel in the Absence of a Membrane and in the Presence of Detergent†. (2004). doi:10.1021/BI049049Y

76. Li, Y., Wang, E. & Wang, Y. A Modified Procedure for Fast Purification of T7 RNA Polymerase. *Protein Expr. Purif.* **16**, 355–358 (1999).

77. Nikola Trbovic *et al.* Efficient Strategy for the Rapid Backbone Assignment of Membrane Proteins. (2005). doi:10.1021/JA0540270

78. Elbaz, Y., Steiner-Mordoch, S., Danieli, T. & Schuldiner, S. In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1519–24 (2004).

79. Elhardt, D. & Böck, A. An in vitro polypeptide synthesizing system from methanogenic bacteria: Sensitivity to antibiotics. *MGG Mol. Gen. Genet.* **188**, 128–134 (1982).

80. Londei, P., Altamura, S., Cammarano, P. & Petrucci, L. Differential features of ribosomes and of poly(U)-programmed cell-free systems derived from sulphur-dependent archaebacterial species. *Eur. J. Biochem.* **157**, 455–462 (1986).

81. Ruggero, D., Creti, R. & Londei, P. In vitro translation of archaeal natural mRNAs at high temperature. *FEMS Microbiol. Lett.* **107**, 89–94 (1993).

82. Endoh, T., Kanai, T. & Imanaka, T. A highly productive system for cell-free protein synthesis using a lysate of the hyperthermophilic archaeon, Thermococcus kodakaraensis. *Appl. Microbiol. Biotechnol.* **74**, 1153–1161 (2007).

83. Chesters, J. K. Cell-Free Protein Synthesis by Rumen Protozoa. *J. Protozool.* **15**, 509–512 (1968).

84. Mager, J. & Lipmann, F. Amino acid incorporation and the reversion of its initial phase with cell-free Tetrahymena preparations. *Proc. Natl. Acad. Sci. U. S. A.* **44**, 305–9 (1958).

85. Reisner, A. H. & MacIndoe, H. Incorporation of Amino Acid into Protein by Utilizing a Cell-Free System from Paramecium. *J. Gen. Microbiol.* **47**, 1–15 (1967).

86. Carter, W. A., Levy, H. B. & Diamond, L. S. Protein Synthesis by Amoebal Ribosomes. *Nature* **213**, 722–724 (1967).

87. Basile, G. & Peticca, M. Recombinant Protein Expression in Leishmania tarentolae. *Mol. Biotechnol.* **43**, 273–278 (2009).

88. Kovtun, O., Mureev, S., Johnston, W. & Alexandrov, K. Towards the Construction of Expressed Proteomes Using a Leishmania tarentolae Based Cell-Free Expression System. *PLoS One* **5**, e14388 (2010).

89. Gallis, B. M., McDonnell, J. P., Hopper, J. E. & Young, E. T. Translation of poly(riboadenylic acid)-enriched messenger RNAs from the yeast, Saccharomyces cerevisiae, in heterologous cell-free systems. *Biochemistry* **14**, 1038–1046 (1975).

90. Sissons, C. H. Yeast protein synthesis. Preparation and analysis of a highly active cell-free system. *Biochem. J.* **144**, 131–140 (1974).

91. Algire, M. A. *et al.* Development and characterization of a reconstituted yeast translation initiation system. *RNA* **8**, S1355838202029527 (2002).

92. Mathews, M., Sonenberg, N. & Hershey, J. W. B. *Translational control in biology and medicine*. (Cold Spring Harbor Laboratory Press, 2007).

93. Iizuka, N., Najita, L., Franzusoff, A. & Sarnow, P. Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. *Mol. Cell. Biol.* **14**, 7322–30 (1994).

94. Rothblatt, J. A. & Meyer, D. I. Secretion in yeast: Reconstitution of the translocation and glycosylation of α-factor and invertase in a homologous cell-free system. *Cell* **44**, 619–628 (1986).

95. Roberts, B. E. & Paterson, B. M. 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. U. S. A.* **70**, 2330–4 (1973).

96. Hunter, A. R., Farell, P. J., Jackson, R. J. & Hunt, T. The Role of Polyamines in Cell-Free Protein Synthesis in the Wheat-Germ System. *Eur. J. Biochem.* **75**, 149–157 (1977).

97. Tse, T. P. H. & Taylor, J. M. Translation of albumin messenger RNA in a cell free protein synthesizing system derived from wheat germ. *J. Biol. Chem.* **252**, 1272–1278 (1977).

98. Madin, K., Sawasaki, T., Ogasawara, T. & Endo, Y. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 559–64 (2000).

99. Shields, D. & Blobel, G. Cell-free synthesis of fish preproinsulin, and processing by heterologous mammalian microsomal membranes. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2059–63 (1977).

100. Komoda, K., Naito, S. & Ishikawa, M. Replication of plant RNA virus genomes in a cell-free extract of evacuolated plant protoplasts. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1863–7 (2004).

101. Gursinsky, T., Schulz, B. & Behrens, S.-E. Replication of Tomato bushy stunt virus RNA in a plant in vitro system. *Virology* **390**, 250–260 (2009).

102. Buntru, M., Vogel, S., Spiegel, H. & Schillberg, S. Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnol.* **14**, 37 (2014).

103. Buntru, M., Vogel, S., Stoff, K., Spiegel, H. & Schillberg, S. A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates. *Biotechnol. Bioeng.* **112**, 867–878 (2015).

104. Smith, G. E., Summers, M. D. & Fraser, M. J. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* **3**, 2156–65 (1983).

105. Kubick, S. *et al.* In Vitro Translation in an Insect-Based Cell-Free System. in *Cell-Free Protein Expression* 209–217 (Springer Berlin Heidelberg, 2003). doi:10.1007/978-3-642-59337-6_25

106. Tarui, H. *et al.* Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment. *Appl. Microbiol. Biotechnol.* **55**, 446–453 (2001).

107. Sachse, R. *et al.* Synthesis of membrane proteins in eukaryotic cell-free systems. *Eng. Life Sci.* **13**, 39–48 (2013).

108. Fenz, S. F., Sachse, R., Schmidt, T. & Kubick, S. Cell-free synthesis of membrane proteins: Tailored cell models out of microsomes. *Biochim. Biophys. Acta - Biomembr.* **1838**, 1382–1388 (2014).

109. Shaklee, P. M. *et al.* Protein Incorporation in Giant Lipid Vesicles under Physiological Conditions. *ChemBioChem* **11**, 175–179 (2010).

110. Zampatis, D. E. *et al.* The protease-activated receptor 1 possesses a functional and cleavable signal peptide which is necessary for receptor expression. *FEBS Lett.* **586**, 2351–2359 (2012).

111. Merk, H., Gless, C., Maertens, B., Gerrits, M. & Stiege, W. Cell-free synthesis of functional and endotoxin-free antibody Fab fragments by translocation into microsomes. *Biotechniques* **53**, 153–160 (2012).

112. Stech, M. *et al.* Production of functional antibody fragments in a vesicle-based eukaryotic cell-free translation system. *J. Biotechnol.* **164**, 220–231 (2013).

113. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261 (2003).

114. Quast, R. B., Claussnitzer, I., Merk, H., Kubick, S. & Gerrits, M. Synthesis and site-directed fluorescence labeling of azido proteins using eukaryotic cell-free orthogonal translation systems. *Anal. Biochem.* **451**, 4–9 (2014).

115. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23**, 150–6 (2005).

116. Mikami, S., Kobayashi, T. & Imataka, H. Cell-Free Protein Synthesis Systems with Extracts from Cultured Human Cells. in 43–52 (Humana Press, 2010). doi:10.1007/978-1-60327-331-2_5

117. Brödel, A. K., Raymond, J. A., Duman, J. G., Bier, F. F. & Kubick, S. Functional evaluation of candidate ice structuring proteins using cell-free expression systems. *J. Biotechnol.* **163**, 301–310 (2013).

118. Schweet, R., Lamfrom, H. & Allen, E. The synthesis of hemoglobin in a cell-free system. *Proc. Natl. Acad. Sci. U. S. A.* **44**, 1029–35 (1958).

119. Pelham, H. R. B. & Jackson, R. J. An Efficient mRNA-Dependent Translation System from Reticulocyte Lysates. *Eur. J. Biochem.* **67**, 247–256 (1976).

120. D, Mac Donald, M. R., Mccourts, D. W. & Krauseq, E. Posttranslational Processing of CY- ,. 15176–15183 (1988).

121. Bailey, C. A., Gerber, L., Howard, A. D. & Udenfriend, S. Processing at the carboxyl terminus of nascent placental alkaline phosphatase in a cell-free system: evidence for specific cleavage of a signal peptide. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 22–6 (1989).

122. Dalley, J. A. & Bulleid, N. J. The endoplasmic reticulum (ER) translocon can differentiate between hydrophobic sequences allowing signals for glycosylphosphatidylinositol anchor addition to be fully translocated into the ER lumen. *J. Biol. Chem.* **278**, 51749–57 (2003).

123. Brödel, A. K., Sonnabend, A. & Kubick, S. Cell-free protein expression based on extracts from CHO cells. *Biotechnol. Bioeng.* **111**, 25–36 (2014).

124. Brödel, A. K., Wüstenhagen, D. A. & Kubick, S. Cell-Free Protein Synthesis Systems Derived from Cultured Mammalian Cells. in 129–140 (2015). doi:10.1007/978-1-4939-2230-7_7

125. Kim, J. Y., Kim, Y.-G. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* **93**, 917–930 (2012).

126. Brödel, A. K. *et al.* IRES-Mediated Translation of Membrane Proteins and Glycoproteins in Eukaryotic Cell-Free Systems. *PLoS One* **8**, e82234 (2013).

127. Picanco-Castro, V., Tage Biaggio, R., Tadeu Cova, D. & Swiech, K. Production of Recombinant Therapeutic Proteins in Human Cells: Current Achievements and Future Perspectives.

128. Bradrick, S. S., Nagyal, S. & Novatt, H. A miRNA-responsive cell-free translation system facilitates isolation of hepatitis C virus miRNP complexes. *RNA* **19**, 1159–69 (2013).

129. Weber, L. A., Feman, E. R. & Baglioni, C. Cell free system from HeLa cells active in initiation of protein synthesis. *Biochemistry* **14**, 5315–5321 (1975).

130. Mikami, S., Kobayashi, T., Masutani, M., Yokoyama, S. & Imataka, H. A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. *Protein Expr. Purif.* **62**, 190–198 (2008).

131. Mikami, S., Masutani, M., Sonenberg, N., Yokoyama, S. & Imataka, H. An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr. Purif.* **46**, 348–357 (2006).

132. Novoa, I., Zeng, H., Harding, H. P. & Ron, D. Feedback inhibition of the unfolded protein response by GADD34-mediated dephosphorylation of eIF2alpha. *J. Cell Biol.* **153**, 1011–22 (2001).

133. Mikami, S. *et al.* N-terminally truncated GADD34 proteins are convenient translation enhancers in a human cell-derived in vitro protein synthesis system. *Biotechnol. Lett.* **32**, 897–902 (2010).

134. Kim, D.-M. & Swartz, J. R. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol. Bioeng.* **66**, 180–188 (1999).

135. Kim, T.-W., Kim, D.-M. & Choi, C.-Y. Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. *J. Biotechnol.* **124**, 373–380 (2006).

136. Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. & Alakhov, Y. B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**, 1162–4 (1988).

137. Kim, D.-M. & Choi, C.-Y. A Semicontinuous Prokaryotic Coupled Transcription/Translation System Using a Dialysis Membrane. *Biotechnol. Prog.* **12**, 645–649 (1996).

138. Stech, M. *et al.* A Continuous-Exchange Cell-Free Protein Synthesis System Based on Extracts from Cultured Insect Cells. *PLoS One* **9**, e96635 (2014).

139. Alakhov, J. B. *et al.* Method of preparing polypeptides in cell-free translation system. (1992).

140. Kigawa, T. & Yokoyama, S. A Continuous Cell-Free Protein Synthesis System for Coupled Transcription-Translation1. *J. Biochem.* **110**, 166–168 (1991).

141. Spirin, A. S. High-throughput cell-free systems for synthesis of functionally active proteins. *Trends Biotechnol.* **22**, 538–545 (2004).

142. Spirin, A. S. Cell-Free Protein Synthesis. in *Cell-Free Translation Systems* 3–20 (Springer Berlin Heidelberg, 2002). doi:10.1007/978-3-642-59379-6_1

143. Kim, R. G. & Choi, C. Y. Expression-independent consumption of substrates in cell-free expression system from Escherichia coli. *J. Biotechnol.* **84**, 27–32 (2000).

144. Kigawa, T., Muto, Y. & Yokoyama, S. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* **6**, 129–134 (1995).

145. Jewett, M. C. & Swartz, J. R. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* **86**, 19–26 (2004).

146. Schwarz, D. *et al.* 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 (2007).

147. Kawasaki, T., Gouda, M. D., Sawasaki, T., Takai, K. & Endo, Y. Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *Eur. J. Biochem.* **270**, 4780–4786 (2003)

.

148. Ryabova, L. A., Desplancq, D., Spirin, A. S. & Plückthun, A. Functional antibody production using cell-free translation: Effects of protein disulfide isomerase and chaperones. *Nat. Biotechnol.* **15**, 79–84 (1997).

149. Yin, G. & Swartz, J. R. Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnol. Bioeng.* **86**, 188–195 (2004).

150. Jiang, X., Ookubo, Y., Fujii, I., Nakano, H. & Yamane, T. Expression of Fab fragment of catalytic antibody 6D9 in an Escherichia coli in vitro coupled transcription/translation system. *FEBS Lett.* **514**, 290–294 (2002).

151. Kolb, V. A., Kommer, A. & Spirin, A. S. Co-Translational Protein Folding in Prokaryotic and Eukaryotic Cell-Free Translation Systems. in *Cell-Free Translation Systems* 131–140 (Springer Berlin Heidelberg, 2002). doi:10.1007/978-3-642-59379-6_11

152. Volkman, J. Sterols in microorganisms. *Appl. Microbiol. Biotechnol.* **60**, 495–506 (2003).

153. Devaux, P. F. & Morris, R. Transmembrane Asymmetry and Lateral Domains in Biological Membranes. *Traffic* **5**, 241–246 (2004).

154. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).

155. Lundbæk, J. A., Birn, P., Girshman, J., Hansen, A. J. & Andersen, O. S. Membrane Stiffness and Channel Function. *Biochemistry* **35**, 3825–3830 (1996).

156. Wang, X., Bogdanov, M. & Dowhan, W. Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. *EMBO J.* **21**, 5673–81 (2002).

157. Abi-Char, J. *et al.* Membrane cholesterol modulates Kv1.5 potassium channel distribution and function in rat cardiomyocytes. *J. Physiol.* **582**, 1205–1217 (2007).

158. Krueger-Koplin, R. D. *et al.* An evaluation of detergents for NMR structural studies of membrane proteins. *J. Biomol. NMR* **28**, 43–57 (2004).

159. Vinogradova, O., Sönnichsen, F. & Sanders, II, C. R. On choosing a detergent for solution NMR studies of membrane proteins. *J. Biomol. NMR* **11**, 381–386 (1998).

160. Sanders, C. R., Kuhn Hoffmann, A., Gray, D. N., Keyes, M. H. & Ellis, C. D. French Swimwear for Membrane Proteins. *ChemBioChem* **5**, 423–426 (2004).

161. Sanders, C. R. & Sönnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn. Reson. Chem.* **44**, S24–S40 (2006).

162. Page, R. C. *et al.* Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J. Struct. Funct. Genomics* **7**, 51–64 (2006).

163. Poget, S. F. & Girvin, M. E. Solution NMR of membrane proteins in bilayer mimics: Small is beautiful, but sometimes bigger is better. *Biochim. Biophys. Acta - Biomembr.* **1768**, 3098–3106 (2007).

164. Raschle, T., Hiller, S., Etzkorn, M. & Wagner, G. Nonmicellar systems for solution NMR spectroscopy of membrane proteins. *Curr. Opin. Struct. Biol.* **20**, 471–479 (2010).

165. Naito, A. Structure elucidation of membrane-associated peptides and proteins in oriented bilayers by solid-state NMR spectroscopy. *Solid State Nucl. Magn. Reson.* **36**, 67–76 (2009).

166. McDermott, A. Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annu. Rev. Biophys.* **38**, 385–403 (2009).

167. Tanford, C. & Reynolds, J. A. Characterization of membrane proteins in detergent solutions. *Biochim. Biophys. Acta - Rev. Biomembr.* **457**, 133–170 (1976).

168. Helenius, A. & Simons, K. Solubilization of membranes by detergents. *Biochim. Biophys. Acta - Rev. Biomembr.* **415**, 29–79 (1975).

169. Sanders, C. R. & Landis, G. C. Reconstitution of Membrane Proteins into Lipid-Rich Bilayered Mixed Micelles for NMR Studies. *Biochemistry* **34**, 4030–4040 (1995).

170. Lee, D. *et al.* Bilayer in Small Bicelles Revealed by Lipid−Protein Interactions Using NMR Spectroscopy. *J. Am. Chem. Soc.* **130**, 13822–13823 (2008).

171. Tribet, C., Audebert, R. & Popot, J.-L. L. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15047–50 (1996).

172. Chae, P. S. *et al.* A New Class of Amphiphiles Bearing Rigid Hydrophobic Groups for Solubilization and Stabilization of Membrane Proteins. *Chem. - A Eur. J.* **18**, 9485–9490 (2012).

173. Bayburt, T. H., Carlson, J. W. & Sligar, S. G. Reconstitution and Imaging of a Membrane Protein in a Nanometer-Size Phospholipid Bilayer. *J. Struct. Biol.* **123**, 37–44 (1998).

174. Glück, J. M. *et al.* Integral Membrane Proteins in Nanodiscs Can Be Studied by Solution NMR Spectroscopy. *J. Am. Chem. Soc.* **131**, 12060–12061 (2009).

175. Ekaterina N. Lyukmanova *et al.* Lipid−Protein Nanoscale Bilayers: A Versatile Medium for NMR Investigations of Membrane Proteins and Membrane-Active Peptides. (2008). doi:10.1021/JA0777988

176. Linke, D. Chapter 34 Detergents: An Overview. *Methods Enzymol.* **463**, 603–617 (2009).

177. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. (2007). doi:10.1021/JP073016L

178. Zhou, H.-X. & Cross, T. A. Influences of Membrane Mimetic Environments on Membrane Protein Structures. *Annu. Rev. Biophys.* **42**, 361–392 (2013).

179. Hwang, P. M., Bishop, R. E. & Kay, L. E. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Natl. Acad. Sci.* **101**, 9618–9623 (2004).

180. Catoire, L. J., Warnet, X. L. & Warschawski, D. E. Micelles, Bicelles, Amphipols, Nanodiscs, Liposomes, or Intact Cells: The Hitchhiker's Guide to the Study of Membrane Proteins by NMR. in *Membrane Proteins Production for Structural Analysis* 315–345 (Springer New York, 2014). doi:10.1007/978-1-4939-0662-8_12

181. Ding, Y., Fujimoto, L. M., Yao, Y., Plano, G. V. & Marassi, F. M. Influence of the lipid membrane environment on structure and activity of the outer membrane protein Ail from Yersinia pestis. *Biochim. Biophys. Acta - Biomembr.* **1848**, 712–720 (2015).

182. Andersen, O. S. & Koeppe, R. E. Bilayer Thickness and Membrane Protein Function: An Energetic Perspective. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 107–130 (2007).

183. Hong, H. & Tamm, L. K. Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4065–70 (2004).

184. Negron, C. & Keating, A. E. *Multistate protein design using CLEVER and CLASSY*. *Methods in Enzymology* **523**, (Elsevier Inc., 2013).

185. Cullis, P. R. & Hope, M. J. Chapter 1 Physical properties and functional roles of lipids in membranes. *New Compr. Biochem.* **20**, 1–41 (1991).

186. Mouritsen, O. G. *Life — As a Matter of Fat*. (Springer Berlin Heidelberg, 2005). doi:10.1007/b138577

187. Park, S. H., De Angelis, A. A., Nevzorov, A. A., Wu, C. H. & Opella, S. J. Three-Dimensional Structure of the Transmembrane Domain of Vpu from HIV-1 in Aligned Phospholipid Bicelles. *Biophys.J.* **91**, 3032–3042 (2006).

188. Strandberg, E. *et al.* Tilt angles of transmembrane model peptides in oriented and non-oriented lipid bilayers as determined by 2H solid-state NMR. *Biophys. J.* **86**, 3709–21 (2004).

189. Egorova-Zachernyuk, T. A. *et al.* Heteronuclear 2D-correlations in a uniformly [13C, 15N] labeled membrane-protein complex at ultra-high magnetic fields. *J. Biomol. NMR* **19**, 243–253 (2001).

190. Abdine, A. *et al.* Structural study of the membrane protein MscL using cell-free expression and solid-state NMR. *J. Magn. Reson.* **204**, 155–159 (2010).

191. Hu, J. *et al.* Structural biology of transmembrane domains: efficient production and characterization of transmembrane peptides by NMR. *Protein Sci.* **16**, 2153–65 (2007).

192. Sanders, C. R., Prestegard, J. H. & Prestegard, J. H. Magnetically orientable phospholipid bilayers containing small amounts of a bile salt analogue, CHAPSO. *Biophys. J.* **58**, 447–60 (1990).

193. Sanders, C. R. & Schwonek, J. P. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry* **31**, 8898–8905 (1992).

194. Vold, R. R., Prosser, R. S. & Deese, A. J. Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides. *J. Biomol. NMR* **9**, 329–35 (1997).

195. Loudet, C. *et al.* Bicelle membranes and their use for hydrophobic peptide studies by circular dichroism and solid state NMR. *Biochim. Biophys. Acta - Gen. Subj.* **1724**, 315–323 (2005).

196. Triba, M. N., Zoonens, M., Popot, J.-L., Devaux, P. F. & Warschawski, D. E. Reconstitution and alignment by a magnetic field of a β-barrel membrane protein in bicelles. *Eur. Biophys. J.* **35**, 268–275 (2006).

197. McKibbin, C. *et al.* Opsin Stability and Folding: Modulation by Phospholipid Bicelles. *J. Mol. Biol.* **374**, 1319–1332 (2007).

198. Biverståhl, H., Lind, J., Bodor, A. & Mäler, L. Biophysical studies of the membrane location of the voltage-gated sensors in the HsapBK and KvAP K+ channels. *Biochim. Biophys. Acta - Biomembr.* **1788**, 1976–1986 (2009).

199. Chartrand, E., Arnold, A. A., Gravel, A., Jenna, S. & Marcotte, I. Potential role of the membrane in hERG channel functioning and drug-induced long QT syndrome. *Biochim. Biophys. Acta* **1798**, 1651–62 (2010).

200. Marcotte, I. & Auger, M. Bicelles as model membranes for solid- and solution-state NMR studies of membrane peptides and proteins. *Concepts Magn. Reson. Part A* **24A**, 17–37 (2005).

201. Ram, P. & Prestegard, J. H. Magnetic field induced ordering of bile salt/phospholipid micelles: new media for NMR structural investigations. *Biochim. Biophys. Acta - Biomembr.* **940**, 289–294 (1988).

202. Park, S. H. & Opella, S. J. Triton X-100 as the "short-chain lipid" improves the magnetic alignment and stability of membrane proteins in phosphatidylcholine bilayers for oriented-sample solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 12552–3 (2010).

203. Girvin, M. E., Rastogi, V. K., Abildgaard, F., John L. Markley & Fillingame, R. H. Solution Structure of the Transmembrane H+-Transporting Subunit c of the F1Fo ATP Synthase. (1998). doi:10.1021/BI980511M

204. Schwaiger, M. *et al.* NMR investigation of the multidrug transporter EmrE, an integral membrane protein. *Eur. J. Biochem.* **254**, 610–619 (1998).

205. Dmitriev, O. Y., Altendorf, K. & Fillingame, R. H. Subunit a of the E. coli ATP synthase: reconstitution and high resolution NMR with protein purified in a mixed polarity solvent. *FEBS Lett.* **556**, 35–38 (2004).

206. Bayburt, T. H. *et al.* Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. (2002). doi:10.1021/NL025623K

207. Nath, A., William, M. A. & Sligar, S. G. Applications of Phospholipid Bilayer Nanodiscs in the Study of Membranes and Membrane Proteins. (2007). doi:10.1021/BI602371N

208. Borch, J. & Hamann, T. The nanodisc: a novel tool for membrane protein studies. *Biol. Chem.* **390**, 805–814 (2009).

209. Bayburt, T. H. & Sligar, S. G. Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6725–30 (2002).

210. Jonas, A. [32] Reconstitution of high-density lipoproteins. *Methods Enzymol.* **128**, 553–582 (1986).

211. Katzen, F. *et al.* Insertion of Membrane Proteins into Discoidal Membranes Using a Cell-Free Protein Expression Approach. *J. Proteome Res.* **7**, 3535–3542 (2008).

212. Popot, J.-L. Amphipols, Nanodiscs, and Fluorinated Surfactants: Three Nonconventional Approaches to Studying Membrane Proteins in Aqueous Solutions. *Annu. Rev. Biochem.* **79**, 737–775 (2010).

213. Popot, J.-L. *et al.* Amphipols From A to Z*. *Annu. Rev. Biophys.* **40**, 379–408 (2011).

214. Elter, S. *et al.* The Use of Amphipols for NMR Structural Characterization of 7-TM Proteins. *J. Membr. Biol.* **247**, 957–964 (2014).

215. Gohon, Y. & Popot, J.-L. Membrane protein–surfactant complexes. *Curr. Opin. Colloid Interface Sci.* **8**, 15–22 (2003).

216. Breyton, C., Tribet, C., Olive, J., Dubacq, J. P. & Popot, J. L. Dimer to monomer conversion of the cytochrome b6 f complex. Causes and consequences. *J. Biol. Chem.* **272**, 21892–900 (1997).

217. Popot, J.-L. L. *et al. Amphipols: Polymeric surfactants for membrane biology research*. *Cellular and Molecular Life Sciences* **60**, 1559–1574 (Birkhäuser-Verlag, 2003).

218. Kigawa, T. *et al.* Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **5**, 63–68 (2004).

219. Zawada, J. F. Preparation and Testing of E. coli S30 In Vitro Transcription Translation Extracts. in 31–41 (2012). doi:10.1007/978-1-61779-379-0_2

220. Pletnev, P., Osterman, I., Sergiev, P., Bogdanov, A. & Dontsova, O. Survival guide: Escherichia coli in the stationary phase. *Acta Naturae* **7**, 22–33 (2015).

221. Wada, A. Growth phase coupled modulation of Escherichia coli ribosomes. *Genes to Cells* **3**, 203–208 (1998).

222. Sarubbi, E. *et al.* Characterization of the spoT gene of Escherichia coli. *J. Biol. Chem.* **264**, 22–31 (1989).

223. Metzger, S., Sarubbi, E., Glaser, G. & Cashell, M. Protein Sequences Encoded by the relA and spoT Genes of Escherichia coli are interrelated. *J. Biol. Chem.* **264**, 9122–9125 (1989).

224. Gentry, D. R. & Cashel, M. Mutational analysis of the Escherichia coli spoT gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. *Mol. Microbiol.* **19**, 1373–1384 (1996).

225. Etzkorn, M. *et al.* Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: Biophysical properties and NMR accessibility. *Structure* **21**, 394–401 (2013).

226. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic regulation of archaeal proteasome gate opening as studied by trosy NMR. *Science (80-.).* **328**, 98–102 (2010).

227. Gans, P. *et al.* Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chemie - Int. Ed.* **49**, 1958–1962 (2010).

228. Arbogast, L. W., Brinson, R. G. & Marino, J. P. Mapping Monoclonal Antibody Structure by 2D 13C NMR at Natural Abundance. *Anal. Chem.* **87**, 3556–3561 (2015).

229. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H-13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).

230. Schuetz, A. K. & Kay, L. E. A dynamic molecular basis for malfunction in disease mutants of p97/VCP. *Elife* **5**, (2016).

231. Macek, P. *et al.* Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci. Adv.* **3**, e1601601 (2017).

232. Kitevski-Leblanc, J. L. *et al.* Investigating the Dynamics of Destabilized Nucleosomes Using Methyl-TROSY NMR. *J. Am. Chem. Soc.* **140**, 4774–4777 (2018).

233. Tugarinov, V. & Kay, L. E. An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* **28**, 165–172 (2004).

234. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **1**, 749–754 (2006).

235. Tugarinov, V. & Kay, L. E. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **125**, 13868–13878 (2003).

236. Kainosho, M. & Gntert, P. SAIL-Stereo-array isotope labeling. *Q. Rev. Biophys.* **42**, 247–300 (2009).

237. Lazarova, M. *et al.* Precursor-Based Selective Methyl Labeling of Cell-Free Synthesized Proteins. *ACS Chem. Biol.* **13**, 2170–2178 (2018).

238. Linser, R. *et al.* Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. *J. Am. Chem. Soc.* **136**, 11308–11310 (2014).

239. Henrich, E. *et al.* Synthetic Biology-Based Solution NMR Studies on Membrane Proteins in Lipid Environments. in *Methods in Enzymology* **614**, 143–185 (Academic Press Inc., 2019).

240. Nygaard, R. *et al.* The dynamic process of β2-adrenergic receptor activation. *Cell* **152**, 532–542 (2013).

241. Kofuku, Y. *et al.* Functional dynamics of deuterated β2-adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angew. Chemie - Int. Ed.* **53**, 13376–13379 (2014).

242. Okude, J. *et al.* Identification of a Conformational Equilibrium That Determines the Efficacy and Functional Selectivity of the μ-Opioid Receptor. *Angew. Chemie - Int. Ed.* **54**, 15771–15776 (2015).

243. Kofuku, Y. *et al.* Deuteration and selective labeling of alanine methyl groups of β2-adrenergic receptor expressed in a baculovirus-insect cell expression system. *J. Biomol. NMR* **71**, 185–192 (2018).

244. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112 (2010).

245. Sounier, R. *et al.* Propagation of conformational changes during μ-opioid receptor activation. *Nature* **524**, 375–378 (2015).

246. Eddy, M. T., Didenko, T., Stevens, R. C. & Wüthrich, K. β2-Adrenergic Receptor Conformational Response to Fusion Protein in the Third Intracellular Loop. *Structure* **24**, 2190–2197 (2016).

247. Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P. & Scott Prosser, R. Activation of the A 2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016).

248. Cardillo, R., Fuganti, C., Ghiringhelli, D., Grasselli, P. & Gatti, G. Pattern of incorporation of leucine samples asymmetrically labelled with 13C in the isopropyl unit into the C5-isoprenoid units of echinuline and flavoglaucine. *J. Chem. Soc. Chem. Commun.* 474–476 (1977). doi:10.1039/C39770000474

249. Oba, M., Terauchi, T., Miyakawa, A., Kamo, H. & Nishiyama, K. Stereoselective deuterium-labelling of diastereotopic methyl and methylene protons of L-leucine. *Tetrahedron Lett.* **39**, 1595–1598 (1998).

250. August, R. A., Khan, J. A., Moody, C. M. & Young, D. W. Stereospecific synthesis of (2S,4R)-[5,5,5-2H3]-leucine. *Tetrahedron Lett.* **33**, 4617–4620 (1992).

251. Kelly, N. M., Gordon Reid, R., Willis, C. L. & Winton, P. L. Methods for the synthesis of L-leucine selectively labelled with carbon-13 or deuterium in either diastereotopic methyl group. *Tetrahedron Lett.* **36**, 8315–8318 (1995).

252. Fletcher, M. D. *et al.* Three approaches to the synthesis of L-leucine selectively labelled with carbon-13 or deuterium in either diastereotopic methyl group. *J. Chem. Soc. Perkin Trans. 1* 43–52 (2001). doi:10.1039/a907598d

253. Oba, M., Ohkuma, K., Hitokawa, H., Shirai, A. & Nishiyama, K. Convenient synthesis of deuterated glutamic acid, proline and leucine via catalytic deuteration of unsaturated pyroglutamate derivatives. *J. Label. Compd. Radiopharm.* **49**, 229–235 (2006).

254. Shattuck, J. C. & Meinwald, J. The preparation of L-(2S,3S)-4,4,4-[2H3] valine. *Tetrahedron Lett.* **38**, 8461–8464 (1997).

255. Humphrey, J. M., Hart, J. A. & Richard Chamberlin, A. Efficient syntheses of diastereoselectively labeled (3S)-[4-13C]-l-valine, and regioselectively labeled [3-13CH3]-l-isoleucine hydrochlorides. *Bioorganic Med. Chem. Lett.* **5**, 1315–1320 (1995).

256. Kainosho, M. *et al.* Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57 (2006).

257. Miyanoiri, Y. *et al.* Differential isotope-labeling for Leu and Val residues in a protein by E. coli cellular expression using stereo-specifically methyl labeled amino acids. *J. Biomol. NMR* **57**, 237–249 (2013).

258. Siebum, A. H. G., Woo, W. S. & Lugtenburg, J. Preparation and Characterization of [5-13C]-(2S,4R)-Leucine and [4-13C]-(2S,3S)-Valine− Establishing Synthetic Schemes to Prepare Any Site-Directed Isotopomer ofL-Leucine,L-Isoleucine andL-Valine. *European J. Org. Chem.* **2003**, 4664–4678 (2003).

259. Kitamura, M., Shirakawa, S. & Maruoka, K. Powerful chiral phase-transfer catalysts for the asymmetric synthesis of α-alkyl- and α,α-dialkyl-α-amino acids. *Angew. Chemie - Int. Ed.* **44**, 1549–1551 (2005).

260. Mizukoshi, Y. *et al.* Improvement of Ligand Affinity and Thermodynamic Properties by NMR-Based Evaluation of Local Dynamics and Surface Complementarity in the Receptor-Bound State. *Angew. Chemie - Int. Ed.* **55**, 14606–14609 (2016).

261. Bax, A., Mehlkopf, A. F. & Smidt, J. Absorption spectra from phase-modulated spin echoes. *J. Magn. Reson.* **35**, 373–377 (1979).

262. Minato, Y. *et al.* Conductance of P2X4 purinergic receptor is determined by conformational equilibrium in the transmembrane region. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4741–4746 (2016).

263. Sitarska, A. *et al.* Affordable uniform isotope labeling with 2H, 13C and 15N in insect cells. *J. Biomol. NMR* **62**, 191–197 (2015).

264. Pritišanac, I. *et al.* Automatic Assignment of Methyl-NMR Spectra of Supramolecular Machines Using Graph Theory. *J. Am. Chem. Soc.* **139**, 9523–9533 (2017).

265. Pritišanac, I., Würz, J. M., Alderson, T. R. & Güntert, P. Automatic structure-based NMR methyl resonance assignment in large proteins. *Nat. Commun.* **10**, 1–12 (2019).

266. Kurauskas, V. *et al.* Sensitive proton-detected solid-state NMR spectroscopy of large proteins with selective CH3 labelling: Application to the 50S ribosome subunit. *Chem. Commun.* **52**, 9558–9561 (2016).

267. Robson, S., Arthanari, H., Hyberts, S. G. & Wagner, G. Nonuniform Sampling for NMR Spectroscopy. in *Methods in Enzymology* **614**, 263–291 (Academic Press Inc., 2019).

268. Roychoudhury, A., Haussinger, D. & Oesterhelt, F. Effect of the Compatible Solute Ectoine on the Stability of the Membrane Proteins. *Protein Pept. Lett.* **19**, 791–794 (2012).

269. Roychoudhury, A., Bieker, A., Häussinger, D. & Oesterhelt, F. Membrane protein stability depends on the concentration of compatible solutes – a single molecule force spectroscopic study. *Biol. Chem.* **394**, 1465–1474 (2013).

270. Zoonens, M. & Popot, J.-L. L. Amphipols for Each Season. *J. Membr. Biol.* **247**, 759–796 (2014).

271. Dahmane, T. *et al.* Amphipol-assisted in vitro folding of G protein-coupled receptors. *Biochemistry* **48**, 6516–6521 (2009).

272. Pocanschi, C. L. *et al.* Amphipathic polymers: Tools to fold integral membrane proteins to their active form. *Biochemistry* **45**, 13954–13961 (2006).

273. Banères, J.-L., Popot, J.-L. & Mouillac, B. New advances in production and functional folding of G-protein-coupled receptors. *Trends Biotechnol.* **29**, 314–322 (2011).

274. Mary, S. *et al.* Amphipols in G Protein-Coupled Receptor Pharmacology: What Are They Good For? *J. Membr. Biol.* **247**, 853–860 (2014).

275. Zoonens, M., Catoire, L. J., Giusti, F. & Popot, J.-L. NMR study of a membrane protein in detergent-free aqueous solution. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8893–8 (2005).

276. Planchard, N. *et al.* The Use of Amphipols for Solution NMR Studies of Membrane Proteins: Advantages and Constraints as Compared to Other Solubilizing Media. *J. Membr. Biol.* **247**, 827–842 (2014).

277. Catoire, L. J. *et al.* Structure of a GPCR Ligand in Its Receptor-Bound State: Leukotriene B4 Adopts a Highly Constrained Conformation When Associated to Human BLT2. *J. Am. Chem. Soc.* **132**, 9049–9057 (2010).

278. Breit, A. *et al.* Alternative G protein coupling and biased agonism: New insights into melanocortin-4 receptor signalling. *Mol. Cell. Endocrinol.* **331**, 232–240 (2011).

279. Tao, Y.-X. The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology. *Endocr. Rev.* **31**, 506–43 (2010).

280. Fani, L., Bak, S., Delhanty, P., van Rossum, E. F. C. & van den Akker, E. L. T. The melanocortin-4 receptor as target for obesity treatment: a systematic review of emerging pharmacological therapeutic options. *Int. J. Obes.* **38**, 163–169 (2014).

281. Liang, L., Angleson, J. K. & Dores, R. M. Using the human melanocortin-2 receptor as a model for analyzing hormone/receptor interactions between a mammalian MC2 receptor and ACTH(1-24). *Gen. Comp. Endocrinol.* **181**, 203–210 (2013).

282. Picard, M. *et al.* Protective and Inhibitory Effects of Various Types of Amphipols on the Ca2+-ATPase from Sarcoplasmic Reticulum: A Comparative Study. (2006). doi:10.1021/BI051954A

283. Gohon, Y. *et al.* Partial specific volume and solvent interactions of amphipol A8-35. *Anal. Biochem.* **334**, 318–334 (2004).

284. Tribet, C. *et al.* Thermodynamic Characterization of the Exchange of Detergents and Amphipols at the Surfaces of Integral Membrane Proteins. *Langmuir* **25**, 12623–12634 (2009).

285. Klammt, C. *et al.* Facile backbone structure determination of human membrane proteins by NMR spectroscopy. *Nat. Methods* **9**, 834–9 (2012).

286. Schneider, B. *et al.* Membrane Protein Expression in Cell-Free Systems. in *Methods in molecular biology (Clifton, N.J.)* **601**, 165–186 (Humana Press, 2010).

287. Charvolin, D., Picard, M., Huang, L.-S., Berry, E. A. & Popot, J.-L. Solution Behavior and Crystallization of Cytochrome bc 1 in the Presence of Amphipols. *J. Membr. Biol.* **247**, 981–996 (2014).

288. Gohon, Y. *et al.* Bacteriorhodopsin/Amphipol complexes: Structural and functional properties. *Biophys. J.* **94**, 3523–3537 (2008).

289. Manuela Zoonens *et al.* Dynamics of Membrane Protein / Amphipol Association Studied by Fo ¨ rster Resonance Energy Transfer : Implications for in Vitro Studies of Amphipol-Stabilized Membrane Proteins †. *Biochemistry* **46**, 10392–10404 (2007).

290. Arunmanee, W., Harris, J. R. & Lakey, J. H. Outer membrane protein F stabilised with minimal amphipol forms linear arrays and LPS-dependent 2D crystals. *J. Membr. Biol.* **247**, 949–56 (2014).

291. Dahmane, T., Rappaport, F. & Popot, J.-L. L. Amphipol-assisted folding of bacteriorhodopsin in the presence or absence of lipids: Functional consequences. *Eur. Biophys. J.* **42**, 85–101 (2013).

292. Liu, J. & Rost, B. Comparing function and structure between entire proteomes. *Protein Sci.* **10**, 1970–9 (2001).

293. Aridor, M. & Hannan, L. A. Traffic Jam: A Compendium of Human Diseases that Affect Intracellular Transport Processes. *Traffic* **1**, 836–851 (2000).

294. Aridor, M. & Hannan, L. A. Traffic Jams II: An Update of Diseases of Intracellular Transport. *Traffic* **3**, 781–790 (2002).

295. Cotton, R. G. H. & Horaitis, O. The HUGO Mutation Database Initiative. Human Genome Organization. *Pharmacogenomics J.* **2**, 16–9 (2002).

296. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD ®): 2003 update. *Hum. Mutat.* **21**, 577–581 (2003).

297. Heyse, S., Stora, T., Schmid, E., Lakey, J. H. & Vogel, H. Emerging techniques for investigating molecular interactions at lipid membranes. *Biochim. Biophys. Acta - Rev. Biomembr.* **1376**, 319–338 (1998).

298. Tanaka, M. & Sackmann, E. Polymer-supported membranes as models of the cell surface. *Nature* **437**, 656–663 (2005).

299. Köper, I. Insulating tethered bilayer lipid membranes to study membrane proteins. *Mol. Biosyst.* **3**, 651 (2007).

300. Vockenroth, I. K. *et al.* Stable insulating tethered bilayer lipid membranes. *Biointerphases* **3**, FA68–FA73 (2008).

301. Deverall, M. A. *et al.* Membrane Lateral Mobility Obstructed by Polymer-Tethered Lipids Studied at the Single Molecule Level. *Biophys. J.* **88**, 1875–1886 (2005).

302. Giess, F., Friedrich, M. G., Heberle, J., Naumann, R. L. & Knoll, W. The protein-tethered lipid bilayer: a novel mimic of the biological membrane. *Biophys. J.* **87**, 3213–20 (2004).

303. Rigaud, J.-L. & Lévy, D. Reconstitution of Membrane Proteins into Liposomes. *Methods Enzymol.* **372**, 65–86 (2003).

304. Zocher, M. *et al.* Single-Molecule Force Spectroscopy from Nanodiscs: An Assay to Quantify Folding, Stability, and Interactions of Native Membrane Proteins. *ACS Nano* **6**, 961–971 (2012).

305. Petia P. Atanasova, Atanasov, V. & Köper, I. Anchor-Lipid Monolayers at the Air−Water Interface; Prearranging of Model Membrane Systems. (2007). doi:10.1021/LA7002854

306. Sinner, E.-K. & Knoll, W. Functional tethered membranes. *Curr. Opin. Chem. Biol.* **5**, 705–711 (2001).

307. Pfeiffer, I., Petronis, S., Köper, I., Kasemo, B. & Zäch, M. Vesicle Adsorption and Phospholipid Bilayer Formation on Topographically and Chemically Nanostructured Surfaces. *J. Phys. Chem. B* **114**, 4623–4631 (2010).

308. Roder, F. *et al.* Reconstitution of Membrane Proteins into Polymer-Supported Membranes for Probing Diffusion and Interactions by Single Molecule Techniques. *Anal. Chem.* **83**, 6792–6799 (2011).

309. Coutable, A. *et al.* Preparation of Tethered-Lipid Bilayers on Gold Surfaces for the Incorporation of Integral Membrane Proteins Synthesized by Cell-Free Expression. *Langmuir* **30**, 3132–3141 (2014).

310. Oberbarnscheidt, L., Janissen, R. & Oesterhelt, F. Direct and model free calculation of force-dependent dissociation rates from force spectroscopic data. *Biophys. J.* **97**, L19-21 (2009).

311. Engel, A. & Gaub, H. E. Structure and Mechanics of Membrane Proteins. *Annu. Rev. Biochem.* **77**, 127–148 (2008).

312. Harris, N. C., Song, Y. & Kiang, C.-H. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. *Phys. Rev. Lett.* **99**, 068101 (2007).

313. Lee, G. U., Kidwell, D. A. & Colton, R. J. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir* **10**, 354–357 (1994).

314. Oberbarnscheidt, L., Janissen, R., Martell, S., Engelhard, M. & Oesterhelt, F. Single-Molecule Force Spectroscopy Measures Structural Changes Induced by Light Activation and Transducer Binding in Sensory Rhodopsin II. *J. Mol. Biol.* **394**, 383–390 (2009).

315. Pelaseyed, T. *et al.* Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell-protective device. *FEBS J.* **280**, 1491–1501 (2013).

316. Baumann, R.-P., Schranz, M. & Hampp, N. Bending of purple membranes in dependence on the pH analyzed by AFM and single molecule force spectroscopy. *Phys. Chem. Chem. Phys.* **12**, 4329 (2010).

317. Müller, D. J., Schabert, F. A., Büldt, G. & Engel, A. Imaging purple membranes in aqueous solutions at sub-nanometer resolution by atomic force microscopy. *Biophys. J.* **68**, 1681–6 (1995).

318. Kessler, M. & Gaub, H. E. Unfolding Barriers in Bacteriorhodopsin Probed from the Cytoplasmic and the Extracellular Side by AFM. *Structure* **14**, 521–527 (2006).

319. Klammt, C. *et al.* Cell-free production of G protein-coupled receptors for functional and structural studies. *J. Struct. Biol.* **158**, 482–493 (2007).

320. Lata, S. & Piehler*, J. Stable and Functional Immobilization of Histidine-Tagged Proteins via Multivalent Chelator Headgroups on a Molecular Poly(ethylene glycol) Brush. (2005). doi:10.1021/AC048813J

321. Butt, H.-J. & Jaschke, M. Calculation of thermal noise in atomic force microscopy. *Nanotechnology* **6**, 1–7 (1995).

322. Janissen, R., Oberbarnscheidt, L. & Oesterhelt, F. Optimized straight forward procedure for covalent surface immobilization of different biomolecules for single molecule applications. *Colloids Surfaces B Biointerfaces* **71**, 200–207 (2009).

323. Sullan, R. M. A., Li, J. K. & Zou, S. Direct Correlation of Structures and Nanomechanical Properties of Multicomponent Lipid Bilayers. *Langmuir* **25**, 7471–7477 (2009).

324. Giocondi, M.-C. *et al.* Surface topography of membrane domains. *Biochim. Biophys. Acta - Biomembr.* **1798**, 703–718 (2010).

325. Giocondi, M.-C., Vié, V., Lesniewska, E., Goudonnet, J.-P. & Le Grimellec, C. In Situ Imaging of Detergent-Resistant Membranes by Atomic Force Microscopy. *J. Struct. Biol.* **131**, 38–43 (2000).

326. Choucair, A., Chakrapani, M., Chakravarthy, B., Katsaras, J. & Johnston, L. J. Preferential accumulation of Aβ(1−42) on gel phase domains of lipid bilayers: An AFM and fluorescence study. *Biochim. Biophys. Acta - Biomembr.* **1768**, 146–154 (2007).

327. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. Biophys. Acta - Biomembr.* **1666**, 105–117 (2004).

328. Vasudevan, S. V, Schulz, J., Zhou, C. & Cocco, M. J. Protein folding at the membrane interface, the structure of Nogo-66 requires interactions with a phosphocholine surface. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6847–51 (2010).

329. Lin, H., Bhatia, R. & Lal, R. Amyloid β protein forms ion channels: implications for Alzheimer's disease pathophysiology. *FASEB J.* **15**, 2433–2444 (2001).

330. Redondo-Morata, L., Giannotti, M. I. & Sanz, F. Influence of Cholesterol on the Phase Transition of Lipid Bilayers: A Temperature-Controlled Force Spectroscopy Study. *Langmuir* **28**, 12851–12860 (2012).

331. Tang, J. *et al.* Detection of metal binding sites on functional S-layer nanoarrays using single molecule force spectroscopy. *J. Struct. Biol.* **168**, 217–222 (2009).

332. Heymann, B. & Grubmüller, H. Elastic properties of poly(ethylene-glycol) studied by molecular dynamics stretching simulations. *Chem. Phys. Lett.* **307**, 425–432 (1999).

333. Oesterhelt, F., Rief, M. & Gaub, H. E. Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water. *New J. Phys.* **1**, 6–6 (1999).

334. Parker, M. J., Aulton-Jones, M., Hounslow, A. M. & Craven, C. J. A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J. Am. Chem. Soc.* **126**, 5020–1 (2004).

335. Lohr, F. *et al.* Combinatorial triple-selective labeling as a tool to assist membrane protein backbone resonance assignment. *J. Biomol. NMR* **52**, 197–210 (2012).

336. Lohr, F., Tumulka, F., Bock, C., Abele, R. & Dötsch, V. An extended combinatorial 15N, 13Calpha, and 13C' labeling approach to protein backbone resonance assignment. *J. Biomol. NMR* **62**, 263–279 (2015).

337. Cappuccio, J. A. *et al.* Cell-free co-expression of functional membrane proteins and apolipoprotein, forming soluble nanolipoprotein particles. *Mol. Cell. Proteomics* **7**, 2246–2253 (2008).

338. Katzen, F., Peterson, T. C. & Kudlicki, W. Membrane protein expression: no cells required. *Trends Biotechnol.* **27**, 455–460 (2009).

339. Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).

340. Hefke, F. *et al.* Optimization of amino acid type-specific 13C and 15N labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. *J. Biomol. NMR* **49**, 75–84 (2011).

341. Kooijman, L. *et al.* Backbone and methyl assignment of bacteriorhodopsin incorporated into nanodiscs. *J. Biomol. NMR* **74**, 45–60 (2020).

342. Pervushin, K. V., Arseniev, A. S., Kozhich, A. T. & Ivanov, V. T. Two-dimensional NMR study of the conformation of (34-65)bacterioopsin polypeptide in SDS micelles. *J. Biomol. NMR* **1**, 313–322 (1991).

343. Pervushin, K. V., Orekhov, V. Y., Popov, A. I., Musina, L. Y. & Arseniev, A. S. Three-dimensional structure of (1–71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by 15N-1H heteronuclear NMR spectroscopy. *Eur. J. Biochem.* **219**, 571–583 (1994).

344. Krishnamani, V., Hegde, B. G., Langen, R. & Lanyi, J. K. Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. *Biochemistry* **51**, 1051–1060 (2012).

345. Krishnamani, V. & Lanyi, J. K. Molecular dynamics simulation of the unfolding of individual bacteriorhodopsin helices in sodium dodecyl sulfate micelles. *Biochemistry* **51**, 1061–1069 (2012).

346. Huang, K. S., Bayley, H. & Khorana, H. G. Delipidation of bacteriorhodopsin and reconstitution with exogenous phospholipid. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 323–327 (1980).

347. Huang, K., Bayley, H., Liao, M., London, E. & Khorana, H. G. Refolding of an Integral Membrane Protein. *J. Biol. Chem.* **256**, 3802–3809 (1981).

348. Riley, M. L., Wallace, B. A., Flitsch, S. L. & Booth, P. J. Slow alpha helix formation during folding of a membrane protein. *Biochemistry* **36**, 192–196 (1997).

349. Pan, Y., Brown, L. & Konermann, L. Site-directed mutagenesis combined with oxidative methionine labeling for probing structural transitions of a membrane protein by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **21**, 1947–1956 (2010).

350. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science (80-.).* **277**, 1676–1681 (1997).

351. Grigorieff, N., Ceska, T. A. A., Downing, K. H. H., Baldwin, J. M. M. & Henderson, R. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).

352. London, E. & Khorana, H. G. Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. *J. Biol. Chem.* **257**, 7003–7011 (1982).

353. Popott, J. *et al.* Refolding of Bacteriorhodopsin in Lipid Bilayers A Thermodynamically Controlled Two-stage Process. *J. Mol. Biol.* **198**, 655–676 (1987).

354. Curran, A. R. *et al.* Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. *Biochemistry* **38**, 9328–9336 (1999).

355. Popot, J. L. & Engelman, D. M. Membrane Protein Folding and Oligomerization: The Two-Stage Model. *Biochemistry* **29**, 4031–4037 (1990).

356. Booth, P. J. *et al.* bacteriorhodopsi n. **2**, 139–143 (1995).

357. Hunt, J. F. *et al.* A biophysical study of integral membrane protein folding. *Biochemistry* **36**, 15156–15176 (1997).

358. Kahn, T. W., Sturtevant, J. M. & Engelman, D. M. Thermodynamic measurements of the contributions of helix-connecting loops and of retinal to the stability of bacteriorhodopsin. *Biochemistry* **31**, 8829–8839 (1992).

359. Booth, P. J. Unravelling the folding of bacteriorhodopsin. *Biochim. Biophys. Acta - Bioenerg.* **1460**, 4–14 (2000).

360. Booth, P. J. *et al.* Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nat. Struct. Biol.* **2**, 139–143 (1995).

361. Rehorek, M. & Heyn, M. P. Binding of all-trans-Retinal to the Purple Membrane. Evidence for Cooperativity and Determination of the Extinction Coefficient. *Biochemistry* **18**, 4977–4983 (1979).

362. Oesterhelt, D., Meentzen, M. & Schuhmann, L. Reversible Dissociation of the Purple Complex in Bacteriorhodopsin and Identification of 13-cis and all-trans-Retinal as its Chromophores. *Eur. J. Biochem.* **40**, 453–463 (1973).

363. Schreckenbach, T., Walckhoff, B. & Oesterhelt, D. Studies on the Retinal-Protein Interaction in Bacteriorhodopsin. *Eur. J. Biochem.* **76**, 499–511 (1977).

364. Balbach, J. *et al.* Protein folding monitored at individual residues during a two-dimensional NMR experiment. *Science (80-.).* **274**, 1161–1163 (1996).

365. Etzkorn, M., Böckmann, A., Penin, F., Riedel, D. & Baldus, M. Characterization of folding intermediates of a domain-swapped protein by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 169–175 (2007).

366. Etzkorn, M., Böckmann, A. & Baldus, M. Kinetic analysis of protein aggregation monitored by real-time 2D solid-state NMR spectroscopy. *J. Biomol. NMR* **49**, 121–129 (2011).

367. Kim, P. S. & Baldwin, R. L. Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding. *Annu. Rev. Biochem.* **51**, 459–489 (1982).

368. Schreckenbach, T., Walckhoff, B. & Oesterhelt, D. Specificity of the Retinal Binding Site of Bacteriorhodopsin: Chemical and Stereochemical Requirements for the Binding of Retinol and Retinal. *Biochemistry* **17**, 5353–5359 (1978).

369. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1**, (2011).

370. Planelles, V. *et al.* Vpr-induced cell cycle arrest is conserved among primate lentiviruses. *J. Virol.* **70**, 2516–24 (1996).

371. Guenzel, C. A., Hérate, C. & Benichou, S. HIV-1 Vpr-a still 'enigmatic multitasker'. *Frontiers in Microbiology* **5**, 127 (2014).

372. Li, G., Bukrinsky, M. & Zhao, R. Y. HIV-1 viral protein R (Vpr) and its interactions with host cell. *Curr. HIV Res.* **7**, 178–83 (2009).

373. Kogan, M. & Rappaport, J. HIV-1 Accessory Protein Vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* **8**, 1–20 (2011).

374. Schwefel, D. *et al.* Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234–238 (2014).

375. Pierce, J. V. & Kumamoto, C. A. Variation in Candida albicans EFG1 expression enables host-dependent changes in colonizing fungal populations. *MBio* **3**, (2012).

376. Odds, F. C. Candida infections: An overview. *Crit. Rev. Microbiol.* **15**, 1–5 (1987).

377. Calderone, R. A. & Fonzi, W. A. Virulence factors of Candida albicans. *Trends in Microbiology* **9**, 327–335 (2001).

378. Cole, G. T., Halawa, A. A. & Anaissie, E. J. The Role of the Gastrointestinal Tract in Hematogenous Candidiasis: From the Laboratory to the Bedside. *Clin. Infect. Dis.* **22**, S73–S88 (1996).

379. Brown, A. J. P. & Gow, N. A. R. Regulatory networks controlling Candida albicans morphogenesis. *Trends in Microbiology* **7**, 333–338 (1999).

380. Lo, H. J. *et al.* Nonfilamentous C. albicans mutants are avirulent. *Cell* **90**, 939–949 (1997).

381. Ernst, J. F. Transcription factors in Candida albicans-environmental control of morphogenesis. *Microbiology* **146**, 1763–1774 (2000).

382. Stoldt, V. R. Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. *EMBO J.* **16**, 1982–1991 (1997).

383. Massari, M. E. & Murre, C. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Mol. Cell. Biol.* **20**, 429–440 (2000).

384. Robinson, K. A. Survey and summary: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. *Nucleic Acids Res.* **28**, 1499–1505 (2000).

385. Lassak, T. *et al.* Target specificity of the Candida albicans Efg1 regulator. *Mol. Microbiol.* **82**, 602–618 (2011).

386. Maruyama, T. *et al.* Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* **298**, 714–719 (2002).

387. Foord, S. M. *et al.* International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacological Reviews* **57**, 279–288 (2005).

388. Keitel, V. *et al.* The G-protein coupled bile salt receptor TGR5 is expressed in liver sinusoidal endothelial cells. *Hepatology* **45**, 695–704 (2007).

389. Tiwari, A. & Maiti, P. TGR5: an emerging bile acid G-protein-coupled receptor target for the potential treatment of metabolic disorders. *Drug Discovery Today* **14**, 523–530 (2009).

390. Kumar, D. P. *et al.* Activation of transmembrane bile acid receptor TGR5 stimulates insulin secretion in pancreatic β cells. *Biochem. Biophys. Res. Commun.* **427**, 600–605 (2012).

391. Broeders, E. P. M. *et al.* The bile acid chenodeoxycholic acid increases human brown adipose tissue activity. *Cell Metab.* **22**, 418–426 (2015).

392. van Nierop, F. S. *et al.* Clinical relevance of the bile acid receptor TGR5 in metabolism. *The Lancet Diabetes and Endocrinology* **5**, 224–233 (2017).

393. Stölting, G., Fischer, M. & Fahlke, C. CLC channel function and dysfunction in health and disease. *Front. Physiol.* **5**, 378 (2014).

394. Fahlke, C. & Fischer, M. Physiology and pathophysiology of ClC-K/barttin channels. *Front. Physiol.* **1**, 155 (2010).

395. Landau, D., Shalev, H., Ohaly, M. & Carmi, R. Infantile variant of Bartter syndrome and sensorineural deafness: A new autosomal recessive disorder. *Am. J. Med. Genet.* **59**, 454–459 (1995).

396. Rickheit, G. *et al.* Endocochlear potential depends on Cl- channels: mechanism underlying deafness in Bartter syndrome IV. *EMBO J.* **27**, 2907–17 (2008).

397. Riazuddin, S. *et al.* Molecular basis of DFNB73: mutations of BSND can cause nonsyndromic deafness or Bartter syndrome. *Am. J. Hum. Genet.* **85**, 273–80 (2009).

398. Tan, H., Bungert-Plümke, S., Fahlke, C. & Stölting, G. Reduced Membrane Insertion of CLC-K by V33L Barttin Results in Loss of Hearing, but Leaves Kidney Function Intact. *Front. Physiol.* **8**, 269 (2017).

399. Estévez, R. *et al.* Barttin is a Cl- channel β-subunit crucial for renal Cl- reabsorption and inner ear K+ secretion. *Nature* **414**, 558–561 (2001).

400. Hayama, A., Rai, T., Sasaki, S. & Uchida, S. Molecular mechanisms of Bartter syndrome caused by mutations in the BSND gene. *Histochem. Cell Biol.* **119**, 485–493 (2003).

401. Scholl, U. *et al.* Barttin modulates trafficking and function of ClC-K channels. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.0601631103

402. Fischer, M., Janssen, A. G. H. & Fahlke, C. Barttin activates ClC-K channel function by modulating gating. *J. Am. Soc. Nephrol.* **21**, 1281–9 (2010).

403. Wojciechowski, D. *et al.* Activation of renal ClC-K chloride channels depends on an intact N terminus of their accessory subunit barttin. *J. Biol. Chem.* **293**, 8626–8637 (2018).

404. Steinke, K. V. *et al.* Human CLC-K Channels Require Palmitoylation of Their Accessory Subunit Barttin to Be Functional. *J. Biol. Chem.* **290**, 17390–400 (2015).

405. Tajima, M., Hayama, A., Rai, T., Sasaki, S. & Uchida, S. Barttin binds to the outer lateral surface of the ClC-K2 chloride channel. *Biochem. Biophys. Res. Commun.* **362**, 858–864 (2007).

406. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett.* **584**, 1721–7 (2010).

407. Nasr, M. L. *et al.* Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nat. Methods* **14**, 49–52 (2017).

408. Viennet, T. *et al.* Structural insights from lipid-bilayer nanodiscs link α-Synuclein membrane-binding modes to amyloid fibril formation. *Commun. Biol.* **1**, 44 (2018).

409. Viegas, A., Viennet, T. & Etzkorn, M. The power, pitfalls and potential of the nanodisc system for NMR-based studies. *Biol. Chem.* **397**, 1335–1354 (2016).

410. Fox, D. A. *et al.* Structure of the Neisserial outer membrane protein Opa_{60}: loop flexibility essential to receptor recognition and bacterial engulfment. *J. Am. Chem. Soc.* **136**, 9938–46 (2014).

411. Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J. Am. Chem. Soc.* **135**, 1919–1925 (2013).

412. Mineev, K. S., Goncharuk, S. A., Kuzmichev, P. K., Vilar, M. & Arseniev, A. S. NMR Dynamics of Transmembrane and Intracellular Domains of p75NTR in Lipid-Protein Nanodiscs. *Biophys. J.* **109**, 772–82 (2015).

413. Ritchie, T. K. *et al.* Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol.* **464**, 211–31 (2009).

414. Mitchell, T. W., Ekroos, K., Blanksby, S. J., Hulbert, A. J. & Else, P. L. Differences in membrane acyl phospholipid composition between an endothermic mammal and an ectothermic reptile are not limited to any phospholipid class. *J. Exp. Biol.* **210**, 3440–3450 (2007).

415. Kim, O. Y. *et al.* Altered heart and kidney phospholipid fatty acid composition are associated with cardiac hypertrophy in hypertensive rats. *Clin. Biochem.* **46**, 1111–1117 (2013).

416. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & S. G. Sligar. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. (2004). doi:10.1021/JA0393574

417. Schägger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368–379 (1987).

418. Pervushin, K., Riek, R., Wider, G., Wuthrich, K. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94**, 12366–12371 (1997).

419. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Kurt Wüthrich. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. (1999). doi:10.1021/JA9834226

420. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*. (2004).

421. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59**, 687–696 (2005).

422. Theillet, F.-X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).

423. Viennet, T. *et al.* Selective Protein Hyperpolarization in Cell Lysates Using Targeted Dynamic Nuclear Polarization. *Angew. Chemie Int. Ed.* **55**, 10746–10750 (2016).

424. Katayama, H. *et al.* Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3453–7 (2010).

425. Tsukamoto, H., Sinha, A., DeWitt, M. & Farrens, D. L. Monomeric rhodopsin is the minimal functional unit required for arrestin binding. *J. Mol. Biol.* **399**, 501–11 (2010).

426. Rues, R.-B., Dötsch, V. & Bernhard, F. Co-translational formation and pharmacological characterization of beta1-adrenergic receptor/nanodisc complexes with different lipid environments. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1306–1316 (2016).

427. Raschle, T. *et al.* Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J. Am. Chem. Soc.* **131**, 17777–9 (2009).

428. Viegas, A. *et al.* UTOPIA NMR: activating unexploited magnetization using interleaved low-gamma detection. *J. Biomol. NMR* **64**, 9–15 (2016).

429. Morgado, L., Zeth, K., Burmann, B. M., Maier, T. & Hiller, S. Characterization of the insertase BamA in three different membrane mimetics by solution NMR spectroscopy. *J. Biomol. NMR* **61**, 333–345 (2015).

430. Shenkarev, Z. O. *et al.* Lipid-protein nanodiscs: Possible application in high-resolution NMR investigations of membrane proteins and membrane-active peptides. *Biochem.* **74**, 756–765 (2009)

.

431. Hagn, F. & Wagner, G. Structure refinement and membrane positioning of selectively labeled OmpX in phospholipid nanodiscs. *J. Biomol. NMR* **61**, 249–60 (2015).