

Université  
de Toulouse

# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

**Délivré par :**

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

**Discipline ou spécialité :**

Biologie structurale

---

**Présentée et soutenue par :**

Iordan IORDANOV

**le :** mardi 5 juin 2012

**Titre :**

Structure and dynamics of the outer membrane protein A from *Klebsiella pneumoniae*: a joint NMR - SMFS - proteolysis and MS approach

---

**Ecole doctorale :**

Biologie, Santé, Biotechnologies (BSB)

**Unité de recherche :**

"RMN et interactions Protéines-Membranes", IPBS

**Directeur(s) de Thèse :**

Alain MILON

**Rapporteurs :**

Jean-Louis BANÈRES, Erick DUFOURC, Dror WARSCHAWSKI

**Membre(s) du jury :**

Pierre-Emmanuel GLEIZES  
Jean-Louis BANÈRES  
Erick DUFOURC  
Dror WARSCHAWSKI  
Alain MILON

Professeur de l'Université Paul Sabatier, Président  
Directeur de recherche CNRS, Montpellier, Rapporteur  
Directeur de recherche CNRS, Bordeaux, Rapporteur  
Chargé de Recherche CNRS, Paris, Rapporteur  
Professeur de l'Université Paul Sabatier, Directeur de thèse



# Acknowledgements

Three years of time are not exactly an epoch, but on the human life scale they are not really negligible and, by the end of the day, I am glad to say that I fully enjoyed my stay in Toulouse. Having a good weather and good food are always benefits, but above all else I appreciated the calm, friendly and family-like environment in our NMR group. My special thanks go as follows:

To *Alain Milon*, my supervisor and group leader. Thank you for giving me the opportunity to learn so many things during these years as an early-stage researcher in the excellent SBMP network. Such kind of experience is not often accessible to every PhD student. I truly enjoyed your enthusiasm and optimism on the work and being part of your team was a great pleasure and honor for me. Also, I liked the barbecues at your house.

To *Pascal Demange*, a brilliant biochemist who always has an answer for your wetlab problems. Thank you for your invaluable know-how on protein expression and purification, as well as for all the experimental ideas (including the crazy ones). I wish we could have tried them all. Good luck with your garden.

To *Marie Renault* and *Masae Sugawara* - two former PhD students in the group, whose studies I used all the time as 'ground zero' for my own experiments. The work progress on KpOmpA presented here is heavily relying on their previous investigations on the protein.

To *Valérie Réat*, a dedicated ssNMR spectroscopist who got recently dedicated to two beautiful babies. This prevented our common work on the protein for quite some time (which was unfortunate for me and fortunate for her).

To *Olivier Saurel*, a remarkable spectroscopist who did the effort to explain me the basics of NMR, often more than one time (but never three times). The ssNMR chapter of this work wouldn't have been possible without him, for which of course I'm very grateful.

To *Pascal Ramos* who, sparing from his time dedicated to engineer's tasks and quality management duties, also did the effort to explain me some NMR concepts (even up to three times). He is also very proficient in preparing paella.

To *Olivier Cala*, a good scientist and a good friend, with whom I had lots of nice conversations during all those coffee breaks (the exact number of which I prefer to keep for myself).

To *Parth Rath*, who is another SBMP student in the group and still has some time left for the challenging task of purifying an oligomeric protein complex. Thank you for your advices and discussions throughout the years and I wish you a good luck for the rest of your project.

To *Corinne Hazan* and, subsequently, *Muriel Tauzin* - the European Project Managers of the SBMP - for their constant help in various administration-related issues involving lots of paperwork, in which a humble biologist is (apparently) easily lost.

To *Isabelle Muller*, for sharing the office space with me for three years without complaining, which is surely very brave from her side.

To all other members of the NMR group, present in one moment or another, coming or going, who contributed to this work to one extent or another. These are: *Isabelle Saves, Virginie Gervais, Jerzy Czaplicki, Deborah Gater, Nathalie Doncescu, Mansi Trivedi, Jade Durand, H el ene Eury, Gregory Menchon* and *Guillaume Nars*. Thank you for being here and for offering your help.

To *Carole Pichereaux* from the Proteomics department, for her help in the first steps of using a MALDI-TOF machine.

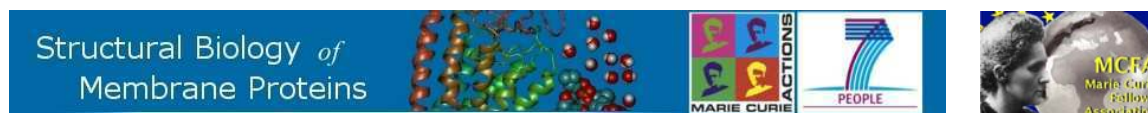


To the SBMP collaborators of our KpOmpA project in Basel, for their tremendous help in protein reconstitution and AFM experiments - *Andreas Engel, Daniel M uller, Patrick Bosshart* and *Shiho Kawamura*.



To my family and my friends back in the homeland, for their constant support and faith in me, even when “I didn’t call often enough”.

To the Marie Curie Fellows Association, for providing the bulk of the finances needed for this work and for giving me the opportunity to be part of the wonderful SBMP network.



These studies and the IPBS NMR equipment were financed by the French Research Ministry, CNRS, Universit  Paul Sabatier, the R gion Midi-Pyr n es and European structural funds. The research leading to these results has received funding from the European Community’s Seventh Framework Program FP7/2007-2013 under grant agreement n  211800, and from the European Drug Initiative on Channels and Transporters grant HEALTH-F4-2007-201924.

# Contents

<b>I. Introduction</b>	<b>1</b>
I.1. Enterobacteriaceae and <i>K. pneumoniae</i>	3
I.2. Cell envelope of the Gram-negative bacteria	5
I.3. Proteins of the bacterial outer membrane	7
I.4. C-terminal domain of Outer membrane protein A	11
I.5. N-terminal domain of Outer membrane protein A	14
I.6. Structural methods for membrane proteins	19
I.7. Aims of the present work	26
References	27
<b>1. Chapter 1 (Purification of KpOmpA)</b>	<b>33</b>
1.1. Construction of the full-length KpOmpA	35
1.2. Expression, refolding and purification of KpOmpA	41
1.3. Expression and purification of C-KpOmpA	47
1.4. Reconstitution of KpOmpA in lipid bilayers	53
References	61
<b>2. Chapter 2 (SMFS on KpOmpA)</b>	<b>63</b>
2.1. Introduction to AFM and SMFS	65
2.2. Basic concept of SMFS	68
2.3. Tip-sample interactions	70
2.4. Conclusion	70
References	71
The transmembrane protein KpOmpA anchoring the outer membrane of <i>Klebsiella pneumoniae</i> unfolds and refolds in response to tensile load [Structure 20, 121-127, January 2012]	
<b>3. Chapter 3 (ssNMR on KpOmpA)</b>	<b>73</b>
Abstract	76
3.1. Introduction	78
3.2. Materials and methods	79
3.2.1. Expression, purification and detergent exchange	79

3.2.2. Reconstitution .....	80
3.2.3. NMR sample preparation .....	81
3.2.4. Sucrose gradient centrifugation and colorimetric assays .....	81
3.2.5. Transmission electron microscopy .....	81
3.2.6. NMR spectroscopy .....	82
<b>3.3. Results .....</b>	<b>83</b>
3.3.1. Screening of experimental conditions for the reconstitution of KpOmpA TM domain in lipid bilayers .....	83
3.3.2. ssNMR characterization of KpOmpA TM domain in <i>E. coli</i> PLE vesicles .....	85
3.3.3. Dynamics of KpOmpA over different NMR time scales .....	87
<b>3.4. Discussion .....</b>	<b>89</b>
<b>Figures .....</b>	<b>93</b>
<b>Supplementary Information .....</b>	<b>99</b>
<b>References .....</b>	<b>101</b>
<b>4. Chapter 4 (Proteolysis and MS) .....</b>	<b>105</b>
4.1. Introduction .....	107
4.2. Material and methods .....	107
4.2.1. Lys-C and Trypsin proteases .....	107
4.2.2. Lys-C digestion of urea-denatured KpOmpA .....	108
4.2.3. Lys-C digestion of folded KpOmpA .....	109
4.2.4. Buffer exchange and digestion of KpOmpA in micelles .....	109
4.2.5. Buffer exchange and digestion of KpOmpA in bilayers .....	110
4.2.6. Digestion kinetics of N- and F-KpOmpA .....	111
4.2.7. Digestion kinetics of C-KpOmpA .....	111
4.2.8. Trypsin digestion of KpOmpA .....	111
4.2.9. MALDI-TOF measurements .....	112
4.3. Basic Concept .....	112
4.4. Complete enzyme digestion with Lys-C .....	122
4.4.1. Digestion of urea-denatured KpOmpA .....	122
4.4.2. Digestion of N- and F-KpOmpA in micelles .....	125
4.4.3. Digestion of N- and F-KpOmpA in liposomes .....	128
4.4.4. Influence of pH and ionic strength .....	132
4.5. Digestion kinetics in different environments .....	136
4.5.1. Digestion kinetics of N- and F-KpOmpA in micelles .....	136
4.5.2. Digestion kinetics of N- and F-KpOmpA in lipid bilayers .....	138
4.5.2. Digestion kinetics of C-KpOmpA .....	139

4.6. Trypsin digestion of KpOmpA .....	142
4.7. Conclusion .....	144
References .....	150
<b>Conclusions and perspectives .....</b>	<b>151</b>
References .....	156
<b>Annexes .....</b>	<b>157</b>
Annex 1 (F-KpOmpA construction) .....	159
Annex 2 (N- and F-KpOmpA purification) .....	163
Annex 3 (C-KpOmpA purification) .....	170
Annex 4 (N- and F-KpOmpA reconstitution) .....	173

# Abbreviations

OM	<b>O</b> uter <b>M</b> embrane
OmpA	<b>O</b> uter <b>M</b> embrane <b>P</b> rotein <b>A</b>
EcOmpA	<b>O</b> uter <b>M</b> embrane <b>P</b> rotein <b>A</b> of <i>Escherichia coli</i>
KpOmpA	<b>O</b> uter <b>M</b> embrane <b>P</b> rotein <b>A</b> of <i>Klebsiella pneumoniae</i>
F-KpOmpA	<b>F</b> ull-length polypeptide chain of <b>KpOmpA</b>
N-KpOmpA	<b>N</b> -terminal (membrane) domain of <b>KpOmpA</b>
C-KpOmpA	<b>C</b> -terminal (periplasmic) domain of <b>KpOmpA</b>
M.W.	<b>M</b> olecular <b>w</b> eight
OD	<b>O</b> ptical <b>d</b> ensity
SDS	<b>S</b> odium <b>d</b> odecyl <b>s</b> ulfate
IPTG	<b>I</b> sopropyl $\beta$ -D-1- <b>t</b> hiogalactopyranoside
CMC	<b>C</b> ritical <b>M</b> icelle <b>C</b> oncentration
OG	n- <b>o</b> ctyl- $\beta$ -D- <b>g</b> luco <b>p</b> yranoside
PLE	<b>P</b> olar <b>L</b> ipids <b>E</b> xtract
LPR	<b>L</b> ipid-to- <b>p</b> rotein ratio
ssNMR	<b>S</b> olid- <b>s</b> tate <b>N</b> uclear <b>M</b> agnetic <b>R</b> esonance
TEM	<b>T</b> ransmission <b>e</b> lectron <b>m</b> icroscopy
AFM	<b>A</b> tom <b>i</b> c <b>F</b> orce <b>M</b> icroscopy
SMFS	<b>S</b> ingle- <b>M</b> olecule <b>F</b> orce <b>S</b> pectroscopy
MALDI-TOF	<b>M</b> atrix- <b>a</b> ssisted laser <b>d</b> esorption/ionization - <b>t</b> ime- <b>o</b> f- <b>f</b> light
Lys-C	<b>L</b> ysine- <b>C</b> endoprotease
TFA	2,2,2- <b>T</b> rifluoro <b>a</b> cetic acid
AcN	<b>A</b> cet <b>o</b> nitrile (cyanomethane)
DHB	2,5- <b>d</b> ihydroxy <b>b</b> enzoic acid
CHCA	$\alpha$ - <b>c</b> yano-4- <b>h</b> ydroxycinnamic acid
SPA	<b>S</b> inapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid)



## Résumé



KpOmpA est une protéine de la membrane externe de *K. pneumoniae*. Elle fait partie de la famille des «outer membrane protein A», OmpA. KpOmpA est une protéine constituée de deux domaines: un domaine transmembranaire structuré en tonneau  $\beta$  et une partie soluble, périplasmique. Le domaine transmembranaire de KpOmpA présente une homologie importante avec celle d'OmpA d'*E. coli* dont la structure a été déterminée par cristallographie aux rayons X et spectroscopie RMN. OmpA d'*E. coli* est responsable lors de la formation de biofilm. Elle a un rôle d'adhésine et d'invasine. Elle est la cible préférentielle du système immunitaire et est le récepteur de bactériophages. Il est admis que la plupart de ces fonctions sont dues aux boucles extracellulaires de ces protéines. Les différences majeures entre les protéines KpOmpA et OmpA d'*E. coli* concernent les boucles extracellulaires de longueur plus importante dans le cas de KpOmpA. Elles jouent un rôle important au cours de l'activation des macrophages et des cellules dendritiques par la voie des récepteurs TLR2. Les boucles extracellulaires jouent un rôle essentiel au cours de l'activation du système immunitaire.

Mieux définir la structure et la dynamique de ces boucles est d'une importance essentielle afin de mieux appréhender la fonctionnalité des boucles extracellulaires de KpOmpA. Les informations structurales connues actuellement (structure RMN déterminée dans le groupe IPBS RMN en 2009) ont été obtenus jusqu'à présent avec des échantillons de protéines recombinantes purifiées et repliées dans des micelles de détergent. Dans le présent travail, nous avons d'abord établi un protocole de reconstitution de la protéine dans une membrane phospholipidique et caractérisé nos échantillons par microscopie électronique.

Des expériences de spectroscopie de force atomique sur molécule unique ont été réalisées pour caractériser le repliement de la protéine dans son environnement membranaire. Ces expériences suggèrent un nouveau rôle de KpOmpA au sein même de la membrane (collaboration D. Müller, ETH Zürich).

Le domaine soluble périplasmique de la protéine a été exprimé indépendamment du domaine membranaire. Les premières expériences HSQC réalisées montrent une structuration de ce domaine. La structure de ce domaine par spectroscopie RMN est en cours de réalisation.

Le comportement dynamique des boucles extracellulaires du domaine membranaire KpOmpA reconstitué dans des liposomes a été étudié par spectroscopie RMN à l'angle magique (MAS) et notamment par mesure des temps de relaxation. Nous avons montré que la dynamique intrinsèque de la protéine est indépendante de l'environnement (membrane vs micelle). Des expériences de protéolyse ménagée suivie par spectrométrie de masse (MALDI-TOF) ont été comparées avec les informations RMN afin d'évaluer plus précisément les niveaux de mobilité des différentes boucles extracellulaires.

La préservation au cours de l'évolution des boucles extracellulaires semble lier à leur dynamique, ce qui suggère l'importance de ces boucles extracellulaires, en termes de séquence, longueur mais aussi de dynamique lors de la réponse immunitaire.

## Summary



KpOmpA is a two-domain membrane protein from *Klebsiella pneumoniae* belonging to the outer membrane protein A (OmpA) family. It is composed of a transmembrane  $\beta$ -barrel with 8  $\beta$ -strands and a C-terminal, soluble periplasmic domain. The transmembrane domain presents a significant homology with *E. coli* OmpA whose three dimensional structure has been determined by X-ray crystallography and by NMR. The *E. coli* homologue can function as an adhesin and invasin, participate in biofilm formation, act as both an immune target and evasin, and serves as a receptor for several bacteriophages. It is assumed that most of these functions involve the four protein loops that emanate from the protein to the exterior of the cell. The difference between KpOmpA and *E. coli* OmpA is mostly concentrated in these extracellular loops which are larger in the case of KpOmpA. KpOmpA was shown to activate macrophages and dendritic cells through the TLR2 dependent pathway, and these larger loops are supposed to play a specific role in the interactions with the immune system.

Thus the structure and dynamics of these loops is of prime functional significance. The currently available information in this regard, including the NMR structure determined in the IPBS NMR group in 2009, have been obtained so far with recombinant protein samples purified and refolded in detergent micelles. In the present work we first established a reconstitution protocol that allowed the incorporation of the membrane protein in the more native environment of the lipid bilayer and characterised our samples by electron microscopy.

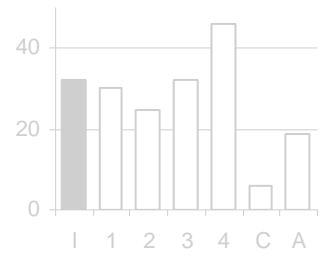
SMFS experiments were used to probe the reconstituted KpOmpA unfolding-refolding pathways, exploring the folding mechanisms for  $\beta$ -barrel proteins and suggesting a novel role for OmpA in the bacterial membrane (in collaboration with the group of D. Müller, ETH Zürich).

The C-terminal periplasmic domain of KpOmpA was expressed and purified as a separate product and the feasibility of its structure elucidation by NMR was demonstrated by obtaining a high quality HSQC spectrum.

The dynamic behaviour of the extracellular portion of the KpOmpA membrane domain reconstituted in liposomes has been investigated by solid state MAS NMR relaxation experiments. We confirmed that the previously observed gradient of dynamic along the molecule axis is an intrinsic property of the protein. Limited proteolysis and MALDI-TOF experiments were coupled with the NMR information in order to assess more precisely the different mobility levels in the loops.

Evolutional preservation of the different loops regions is related to their observed flexibility, pointing towards immunologically important, variable, dynamic and accessible loops sections.

# Introduction



## Summary

The Introduction section points out general aspects of the Gram-negative bacteria on the examples of *Escherichia sp.* and *Klebsiella sp.*, the organization of the outer membrane in these organisms and of their OmpA proteins in particular. The membrane (N-terminal) and the periplasmic (C-terminal) domains of the OmpA proteins are compared with homologous sequences and structures and their roles (known and putative) are described. A variety of structural methods used for investigating the structure-function relations in membrane proteins is briefly described. Finally, the aims of this work are summarized and explained in the light of this information.

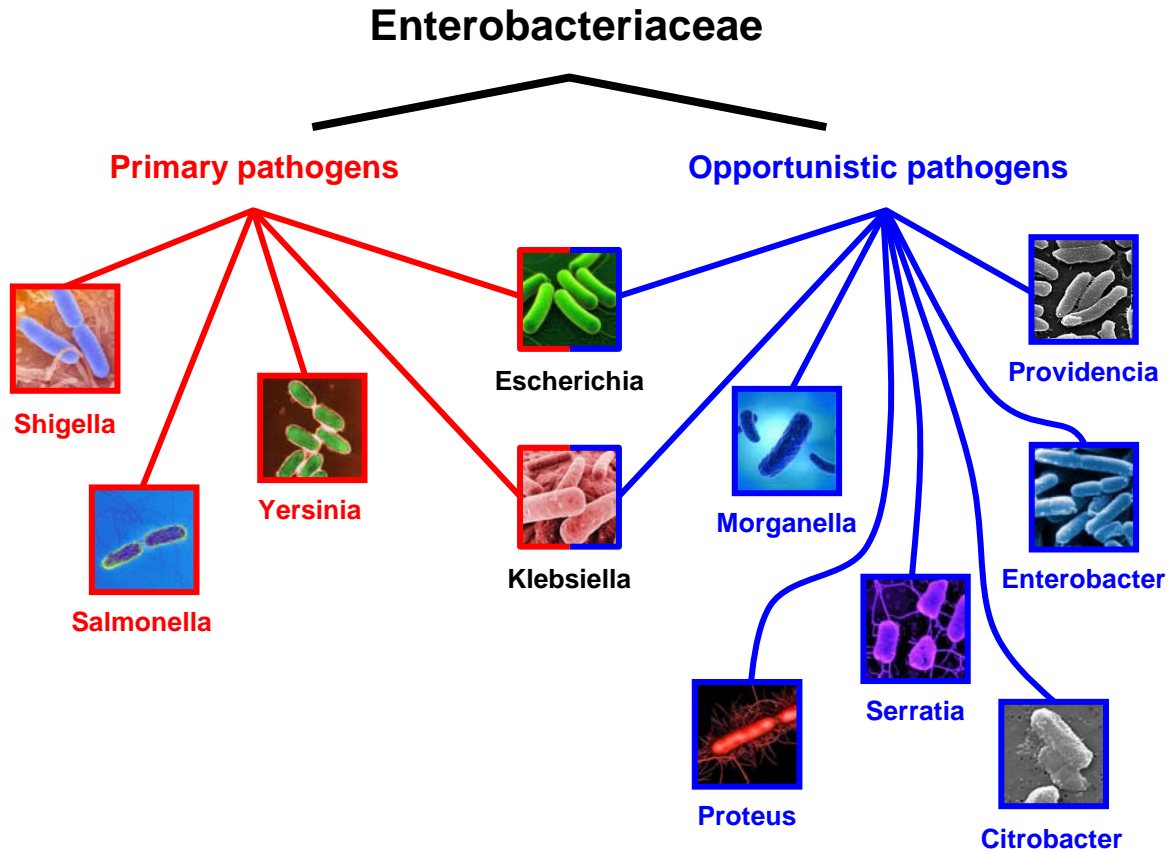
## I.1. Enterobacteriaceae and *K. pneumoniae*

The family *Enterobacteriaceae* belongs to the *Proteobacteria* phylum - a large bacterial taxonomy unit which is composed of Gram-negative species. The family itself contains around 50 different genera of rod-shaped, 1-5 µm long facultative anaerobes. They can be motile (possess flagella) or non-motile, pathogenic or symbionts. Natural habitats include soil, water, human and animals' intestines. The well-studied *Escherichia coli* belongs to this family.

The pathogens in *Enterobacteriaceae* are formally divided into 2 groups: the primary pathogens are capable of causing disease in different hosts or at a variety of conditions, while the opportunistic pathogens can do that under certain conditions and/or in certain hosts. The "famous" pathogens of the family include *Salmonella* (typhoid fever, food poisoning), *Shigella* (shigellosis, dysentery) and *Yersinia* (plague), while *Escherichia* (gastroenteritis, meningitis) and *Klebsiella* (pneumonia, septicemia) occupy an "intermediate" position of strain- and/or conditions-dependent pathogenicity (**Fig. I.1**).

The non-motile, encapsulated, opportunistic pathogen *Klebsiella pneumoniae* is the clinically most important representative of the *Klebsiella* genus, which (for the moment) contains 7 different species. They all express two main types of antigens on their surfaces - a lipopolysaccharide (O-antigen) and a capsular polysaccharide (K-antigen), with large variability of these in the different strains. The natural habitats include surface and sewage waters, soil, plants and mucosal, skin and intestine surfaces in mammals.

The *Klebsiella* genus is related to a variety of diseases such as atypical community-acquired pneumonia, urinary tract infections (UTI), atrophic rhinitis and rhinoscleroma. In particular, *K. pneumoniae* is responsible for UTI (second in rank after *E. coli*) and respiratory tract infections like pneumonia/bronchitis. In addition, new antibiotic-resistant strains are increasingly found in nosocomial (hospital-acquired) infections, particularly in immunocompromised patients that have already been weakened by a fight against another disease.

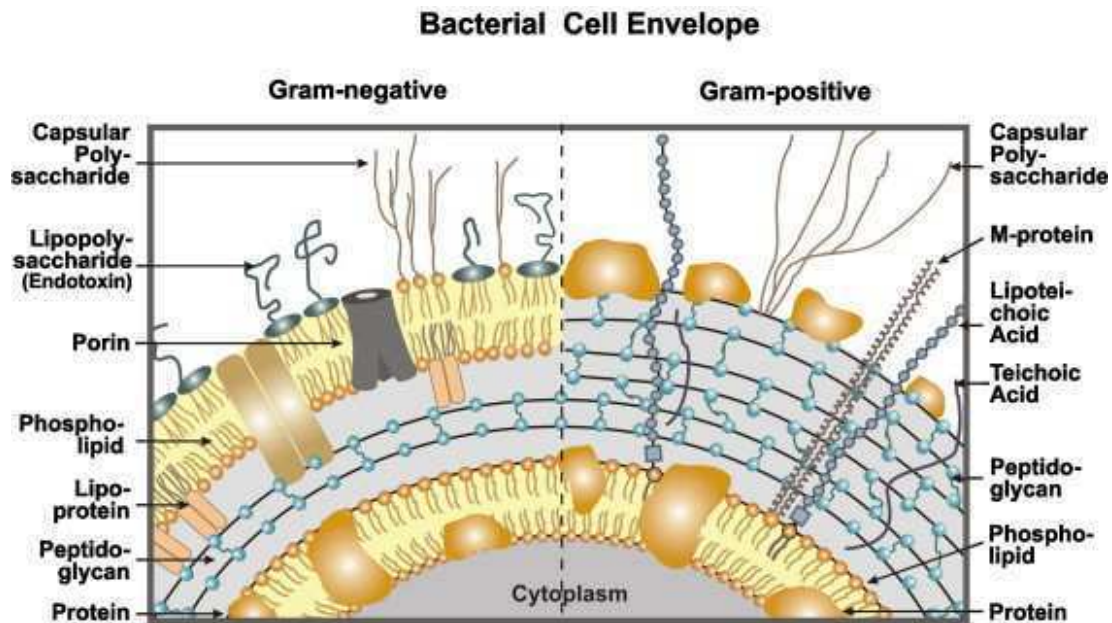


**Fig. I.1.** The *Klebsiella* genus (currently containing 7 different species) within the pathogens of the *Enterobacteriaceae* family. The primary pathogens are presented to the left (red color) and the opportunistic ones - to the right (blue color). The most important representative of *Klebsiella* is perhaps *K. pneumoniae*, causing urinary tract and respiratory infections.

According to the World Health Organization, pneumonia (although caused not only by *K. pneumoniae*) is the leading cause of death (18%) in children up to 5 years old worldwide with approximately 1.4 million casualties per year, which is more than the death toll from AIDS, malaria and tuberculosis combined.

## I.2. Cell envelope of the Gram-negative bacteria

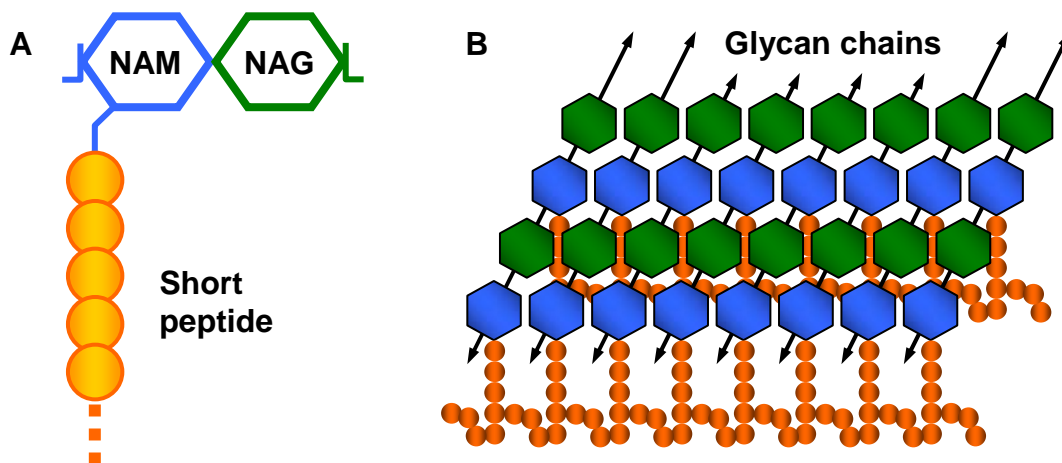
The bacterial cytoplasm is contained by the lipid bilayer of the plasma membrane, followed by a peptidoglycan network (part of the cell wall) of different thickness. In the case of Gram-positive bacteria, this cell wall is considerably thick and protrudes far into the extracellular environment, while the much thinner peptidoglycan lattice of the Gram-negative bacteria is trapped in the periplasmic space between the inner (i.e. the cytoplasmic) and outer membranes (Fig. I.2).



**Fig. I.2.** Schematic representation of the cell envelope of Gram-negative (left side) and Gram-positive (right side) bacteria. Gram-negative bacteria contain a unique feature - the outer membrane - which creates an additional compartment (the periplasmic space between the two membranes). The outer leaflet of the outer membrane is composed of lipopolysaccharides, whereas the inner leaflet is composed of phospholipids as those of the inner membrane [1, 2]. Image source: [3].

*Escherichia coli* is a well studied, typical representative of the rod-shaped Gram-negative bacteria. Its inner membrane is a symmetric distribution of three main types of lipids - phosphatidylethanolamine, phosphatidylglycerol and cardiolipin, in approximate weight ratio of 7:2:1, respectively. Various proteins, rich in  $\alpha$ -helical content, are embedded in this membrane and perform different roles, such as transport, enzymatic and signaling activities. The selective permeability properties of the cell envelope are attributed to this membrane, which is in contrast with the much more nonselective, rich in porins, outer membrane.

Beyond the inner membrane, the periplasm of the Gram-negative cell contains the 2-3 nm of peptidoglycan layer - a lattice of long sugar chains connected with each other via short peptide linkers (**Fig. I.3**). The sugar polymers are composed of glycosidic bond-connected N-acetylglucosamine (NAG) and N-acetylmuramate (NAM). These two alternating monomers form a repeating disaccharide unit which is the main longitudinal building block of the peptidoglycan network. Laterally, each NAM monomer is also linked to a short peptide of several amino acids whose composition varies among the different organisms. The peptides from two different NAM monomers (belonging to two different linear sugar polymers) can cross-link with each other, thus forming the 'mesh-like', stable and rigid structure of the peptidoglycan that still allows passage of fairly large (i.e. ~2 nm) molecules [4].

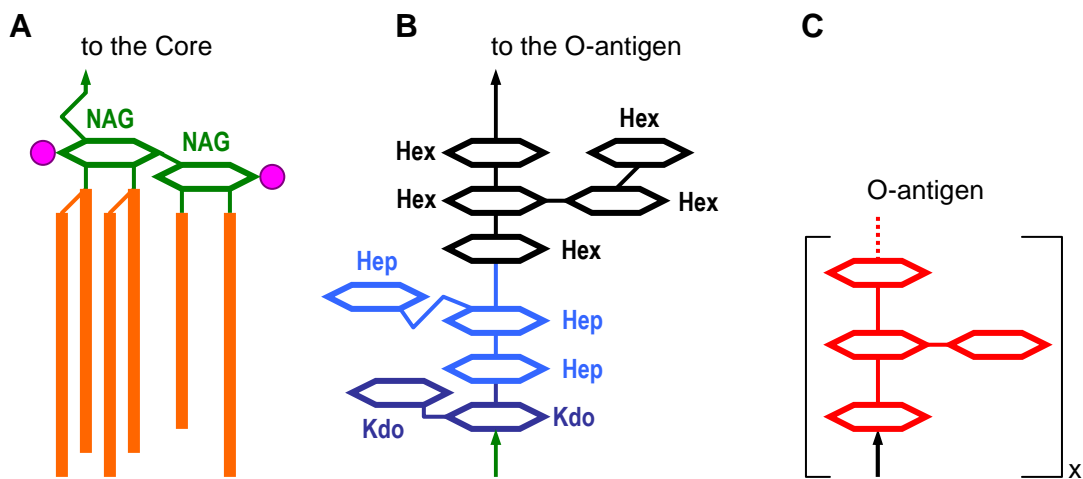


**Fig. I.3.** Cartoon representation of the peptidoglycan structure. (A) The disaccharide building block of the glycan chains with a short peptide attached to the NAM monomer. (B) Overview of the peptidoglycan network.

In a direction towards the extracellular environment, the peptidoglycan-containing periplasm of the Gram-negative bacteria is limited by the outer membrane (OM). Unlike the plasma membrane, the OM is not energized ([5]) and is highly asymmetric. Its inner (i.e. periplasmic) leaflet is virtually identical in lipid composition to the inner membrane, but its outer leaflet is formed exclusively by lipopolysaccharides (LPS). These molecules are built by a membrane-embedded lipid component (the Lipid A moiety) and a polysaccharide chain attached to the lipid head group and protruding outside of the cell. In the case of *E. coli*, Lipid A consists of six aliphatic chains (12-14 carbon atoms each) forming the hydrophobic core of the monolayer and a disaccharide hydrophilic head of two phosphorylated NAG monomers (**Fig. I.4**). The extracellular sugar chain (attached to one of the NAG monomers) possesses a



short core segment of several sugar monomers, followed by a long polysaccharide of several tens of monomers, often organized in repeating units and known as the O-antigen. The monosaccharides participating in the sugar component of LPS vary among the different species (and even strains, [6]) in terms of numbers and sequence arrangement. The presence of LPS obstructs the permeability of the OM for both hydrophilic and hydrophobic compounds, leaving the task of nutritional transport to the transmembrane proteins in the bilayer.



**Fig. I.4.** Schematic general structure (on the example of *E. coli*) of the LPS forming the outer leaflet of the outer membrane in Gram-negative bacteria. **(A)** The Lipid A moiety with a phosphorylated (pink) disaccharide NAG head (green) and 6 fatty chains (orange); **(B)** The core oligosaccharide is composed of inner core (a few mannooctulosonates (Kdo, purple) and heptose (Hep, blue) sugars), followed by the outer core of several hexose monomers (Hex, black); **(C)** The long O-antigen is composed of largely varying, species- and strain-specific sugar motives, often organized in repeating units.

Apart from the atypical, strong asymmetry of the lipid leaflets in the OM (in the past considered by some as “thermodynamically impossible”) and the host-response importance of LPS (one of the most immunologically reactive components that a bacteria can offer), most of the OM properties are maintained by another type of structures, largely present (~50%, [7]) in this bilayer and serving multiple purposes - the outer membrane proteins (OMPs).

### I.3. Proteins of the bacterial outer membrane

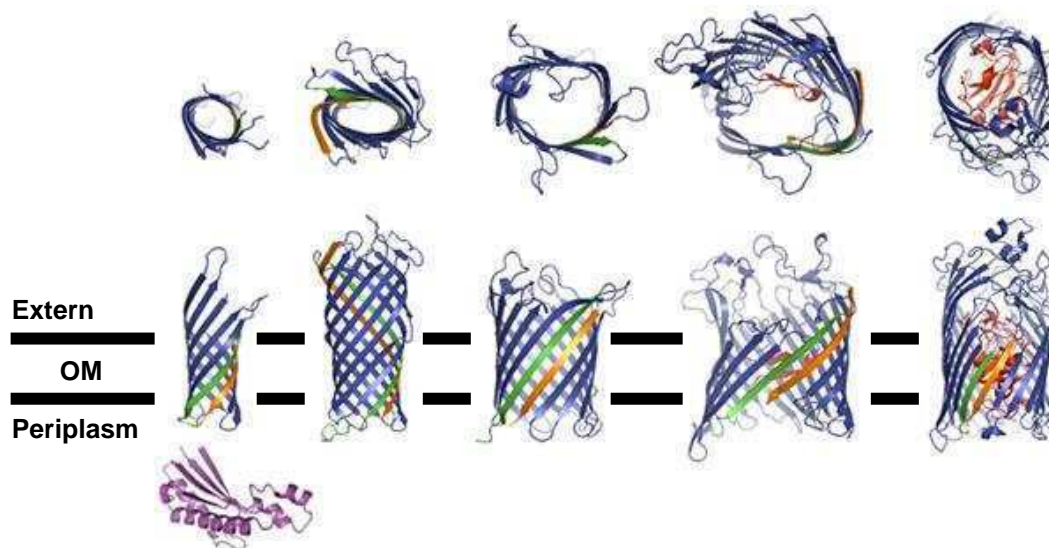
Up to date (2012), a total of 1568 3D structures of transmembrane proteins are known (~2% of all available structures, <http://pdbtm.enzim.hu>), with 320 of them considered unique

(excluding homologues, mutants, different structural methods used, etc., <http://blanco.biomol.ci.edu/mpstruc/listAll/list>). Some of those belong to proteins of the OM which perform several functions, such as structural (OmpA [8, 9]), enzymatic (OmpT [10], OmpLA [11]), active (FepA [12], FhuA [13]) and passive (LamB [14], FadL [15]) transport, non-specific transport (associated with the so called general porins, like OmpC [16], OmpF [17] and OmpG [18, 19] with a molecular 'cutoff' of 600 Da or less, or for smaller hydrophobic molecules as suggested for OmpW [20]), defensive capabilities (OmpX [21, 22]) and role in the cell adhesion and invasion (usually associated with the extracellular segments of the structural proteins, since they are present in large numbers across the bacterial OM). Detailed aspects and reviews on these subjects can be found, among others, in [23], [24], [25], [7], [1], [26] and [5].

Despite of the various roles of the different OMPs, they all share a common feature: structurally, their membrane-embedded sections consist of a cylindrical-like shaped  $\beta$ -barrels built by an even number of anti-parallel  $\beta$ -sheets, 'bended' around a central axis (i.e. the bilayer normal) and inclined (or tilted) at a certain angle with respect to that same axis. Few of the abovementioned OMPs (chosen to encompass different functions and sizes, i.e. number of  $\beta$ -sheets) are presented on **Fig. I.5** and shortly discussed below. It is noted that there was found (albeit only one for the moment) a  $\beta$ -barrel protein with odd number (19) of  $\beta$ -strands - the mitochondrial voltage-dependent anion channel (VDAC, [27]). This protein, however, belongs to the realm of *Eukaryota*. The exceptional case of its structure, if found in other examples or not, will have to be further explained in future, perhaps by taking into account co-evolutional processes of the first single-cell organisms with higher organization and their prokaryotic symbionts.

The OmpA family of proteins (**Fig. I.5A**, [8]) represents some of the simplest and smallest, 8-stranded  $\beta$ -barrels. Although the extracellular loops of the protein are thought to participate in various immunological processes ([28, 29]), the major function of these proteins is presumably structural - their membrane domains serve as anchors in the OM, linking this bilayer with the underlying peptidoglycan network via the periplasmic C-terminal domain. Unlike OmpA, the extracellular segment of the 10-stranded OmpT (**Fig. I.5B**, [10]) appears to be organized with high content of secondary structure, extending from the same  $\beta$ -sheets that lock the molecule in the OM. The active side of this protease is found in a groove of this 'extension', protruding into the extracellular space. The 14-stranded OmpG (**Fig. I.5C**, [19]) exists as a monomeric porin which appears to be pH-gated, with one of its extracellular loops closing the lumen at acidic pH. In contrast, the 18-stranded maltoporin (**Fig. I.5D**, [14]) forms trimers in the OM and uses part of its extracellular segments as a 'plug' for constraining the

lumen space, in the process of passive transport of maltose and other sugars. The 22-stranded monomeric FhuA (**Fig. I.5E**, [13]) acts as a TonB-dependent active transporter of ferric siderophores across the OM. An entire subdomain of the molecule (named ‘the cork’) is plugged into its lumen and LPS-binding site on the membrane-embedded part of the barrel was used to investigate LPS-protein interactions [30].

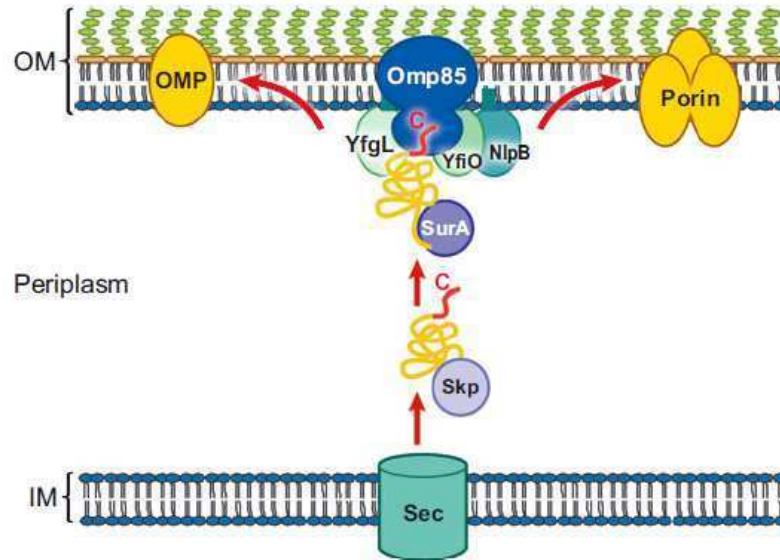


<u>Name:</u>	<b>OmpA</b>	<b>OmpT</b>	<b>OmpG</b>	<b>LamB</b>	<b>FhuA</b>
<u>PDB:</u>	1QJP	1I78	2X9K	1MPO	1QJQ
<u>β-sheets:</u>	8	10	14	18	22
<u>M.W.: kDa (a.a.)</u>	19 (171)	33 (297)	33 (280)	48 (421)	80 (725)
<u>Oligomeric state:</u>	Monomer	Monomer	Monomer	Trimer	Monomer
<u>Function:</u>	Structural	OmpTn	Porin	Passive tr.	Active tr.

**Fig. I.5.** Crystal structures of five OMPs with different sizes and functions. From left to right: OmpA, OmpT, OmpG, LamB and FhuA. The upper pictures represent the extracellular side of the molecules, while the lower pictures show their side view. The approximate boundaries of the outer membrane (OM) are schematically denoted with black lines. The first and the last β-sheet of each structure are colored in green and orange, respectively. The red color (when present) indicates a section of the molecule which is ‘plugged’ into the barrel lumen. The purple structure under the OmpA transmembrane β-barrel represents the soluble, C-terminal periplasmic domain in the native, full-length polypeptide chain of the protein. This structure was generated by Swiss-Model (<http://swissmodel.expasy.org>) based on the homologous structure of the C-terminal domain of RmpM from *Neisseria meningitidis* ([31], PDB: 1R1M). Short information for each protein is given below the respective picture (images created with PyMOL).

Natively, the OMPs are expressed with a short (~20 a.a.) signal peptide at their N-terminus, which is responsible for their transport (in unfolded state) across the inner membrane, via the Sec system [1]. Once in the periplasm, the unfolded OMP is assisted by chaperones, the most studied examples of which are Skp ([32, 33]) and SurA ([34]).

Mutations on the respective genes of these two chaperones exhibit reduced amount of OMPs and/or OMP assembly defects. The signal peptide of the OMP is cleaved in the periplasm (by Signal peptidase I) and the mature protein chain is integrated into the OM with the help of the  $\beta$ -Barrel Assembly Machinery (BAM). Its critical component is perhaps Omp85 [35], which natively forms complexes with the periplasmic, peptidoglycan-binding domain of RmpM ([36]), as well as with three lipoproteins (YfgL, YfiO and NlpB, **Fig. I.6**).



**Fig. I.6.** Model for the OMP biogenesis, as proposed by [5] (image source from the same reference). See text for details. Note that the red colored protein segment (designated with a red letter 'C') represents the so called 'sorting signal' and not the N-terminal signal peptide of the protein.

Interestingly, the BAM complex recognizes a C-terminal motif on the OMP chain which consists of a stretch of aromatic or hydrophobic residues alternating with any other residues [37]. This sequence is quite conserved among the OMPs and is also found in the barrel domain of OmpA, composing its last (i.e. 8<sup>th</sup>)  $\beta$ -sheet. Termed 'sorting signal', it was suggested that it represent the first segment of the OMP that penetrates into the bilayer. Also, it was proposed that one of two Serine residues in that segment is modified with a short amphipathic oligoester (oligo-(R)-3-hydroxybutyrate, or cOHB, [38]), believed to be important for the membrane targeting, assembly, and pore properties of OmpA [39]. It should be noted, however, that this signal is not necessarily essential for the OMP folding and membrane insertion *in vivo* [40]. *In vitro*, on the other hand, OmpA has been shown to refold and insert spontaneously into preformed lipid bilayers [41]. A multi-step mechanism for this insertion

was proposed ([42]), depicting all  $\beta$ -sheets as synchronized, simultaneous insertion units. This model was further investigated by monitoring the Tryptophan fluorescence emitted from the different  $\beta$ -strands during the process of membrane insertion [43], until the barrel structure is fully formed.

All known  $\beta$ -barrels share a common, cylindrical-like architecture, described by their number of  $\beta$ -sheets and the so called 'shear number' ([44]), reflecting the inclination angle ( $\alpha$ ) of the strands with respect to the bilayer normal. It appears that, for any given size (i.e. diameter and number of strands) of a  $\beta$ -barrel,  $\alpha$  is always in the range of 35-45° [45]. The crystal structure of OmpA from *E. coli* ([8, 46]) suggests a shear number of 10 for this 8-stranded  $\beta$ -barrel, with an average  $\alpha = 43^\circ$  and a radius of about 8Å. The solution state NMR structure of the same protein ([9]) is in good agreement with the crystal data and shares a number of similarities with the object of this study - KpOmpA: the OmpA protein from *Klebsiella pneumoniae*, whose  $\beta$ -barrel structure was recently determined by NMR in the IPBS NMR group ([47]). Due to the significant differences between the hydrophobic, membrane-embedded  $\beta$ -barrel and the soluble, periplasmic domain of this protein, the two sections of the molecule are described separately.

#### **I.4. C-terminal domain of Outer membrane protein A**

The C-terminal domain of OmpA is a soluble structure that natively resides in the bacterial periplasm. It is a representative of a largely spread structural motif (termed 'OmpA C-like super family'), responsible for binding to the peptidoglycan network of the cell. In contrast to the N-terminal membrane domain of OmpA (the  $\beta$ -barrel), this section of the molecule is less studied and lacks an available 3D structure. There are, however, homologous structures available, such as those of the C-terminal domains of *E. coli* YiaD (to be published by Ramelot et al.) and of Rv0899 from *M. tuberculosis* ([48]). The closest sequence match (38% identity) among these belongs to the C-terminal domain of RmpM from *N. meningitidis* ([31]). **Fig. I.7A** shows a sequence alignment between the C-terminal domains of *E. coli* OmpA and KpOmpA (named C-KpOmpA), while **Fig. I.7B** presents the alignment between C-KpOmpA and RmpM.

**A**

```

EcOmpA  VQTKHF1TLKSDVLFNFKATLKPEGQAALDQLYSQLSNLDPKDGSVVVLGYTDRI
KpOmpA  VATKHFTL1KSDVLFNFKATLKPEGQAALDQLYSQLSNMDPKDGSVVVLGYTDRI

EcOmpA  GSD2AYNQGLSERRAQSVVDYLISKGIPADKISARGMGESNPVTGNTCDNVKQRAA
KpOmpA  GSE2AYNQQLSEKRAQSVVDYLVAKGIPAGKISARGMGESNPVTGNTCDNVKARRA
                                                    *

EcOmpA  LIDCLAPDRRVEIEVKGIKD3VVTQPQA
KpOmpA  LIDCLAPDRRVEIEVKGYKE3VVTQPAA

```

**B**

```

KpOmpA  APAP4EVATKHFTLKSDVLFNFKATLKPEEGQAALDQLY5TQLSNMDPKDGS6AVV
RmpM    EQA4PQYVDE5TISLSAKT6LF7GF8DK9DSLRAEAQ10DN11LKVLAQR12LSRTNIQ--S13VRV

KpOmpA  LGY14TDRI15GSEAYNQQLSEKRAQSVVDYLVAKGIPAGKISARGMGESNPVTGNT
RmpM    EGH14TFMGS15DKYNQALSERRAYV16VANNLVSN17GV18PVSRISAVGLGESAQMTQV

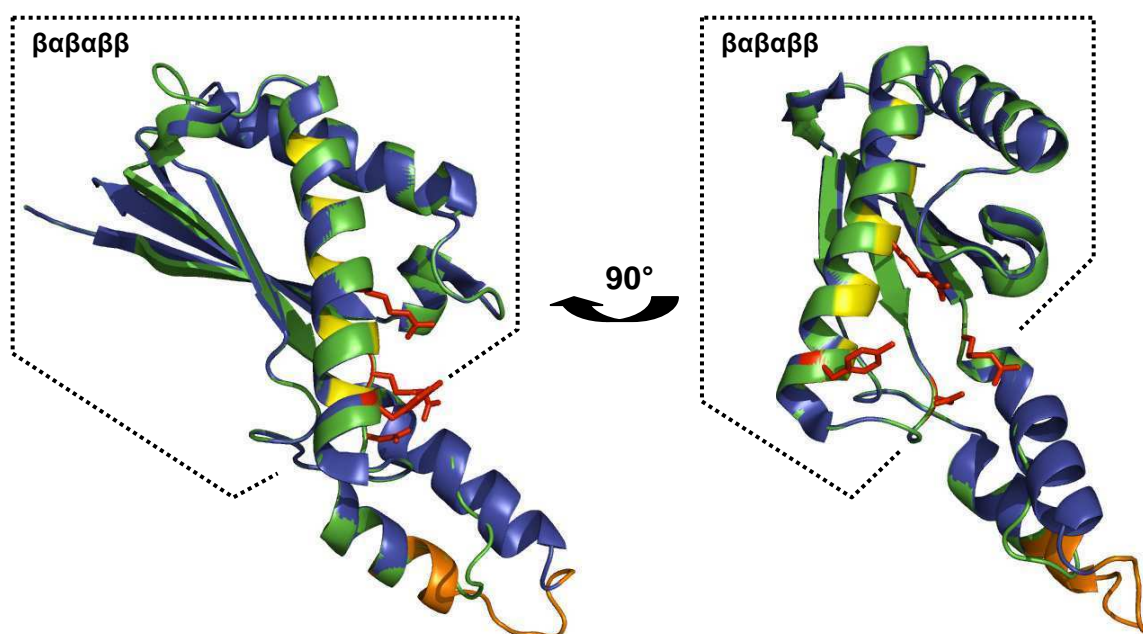
KpOmpA  CD-----NVKARAALIDCLAPDRRVEIEVKG19YKEVVTQPAA-----
RmpM    CEAEVAKLGAKVSKAKKREALIA19CEIPDRRVDVK20IRSI21VT22RQVV23PAH24NHHQHLE

```

**Fig. I.7.** Alignments (with ClustalW2) between different C-terminal OmpA sequences. The identities are indicated with black boxes and the homologues pairs D-E and I-V are in bold. **(A)** Alignment between OmpA from *E. coli* (EcOmpA) and from *K. pneumoniae* (KpOmpA). The asterisk (\*) denotes the 'standard' for this position Asparagine residue, while *K. pneumoniae* strain 52145 contains a Threonine at that position. **(B)** Alignment between the C-terminal domains of KpOmpA and RmpM. The first 3 (EQA) and the last 15 residues (TRQ...) of RmpM are denoted in italic. These disordered areas are not seen on the crystal structure of the protein (PDB: 1R1M). The black dots indicate the proposed peptidoglycan-binding residues in RmpM ([31]).

As expected, the similarities between the two OmpA homologues are much higher in comparison with these between KpOmpA and RmpM, although some key features of the molecules are preserved even in that case. For instance, the two Cysteine residues remain well aligned. These presumably form a disulfide bridge in KpOmpA, as was shown in the case of RmpM. Furthermore, one can distinguish several areas with higher preservation. The motif **NXXLSXXRAXXVXXXL** (with 'X' being any amino acid) is particularly interesting, since it was proposed that the conserved residues (marked in bold) are responsible for binding to the peptidoglycan ([49]). All these amino acids were predicted to stay on the same side of a putative  $\alpha$ -helix of the protein. Indeed, the structure elucidation of RmpM confirms this hypothesis. The authors of this work, however, propose a different arrangement of the peptidoglycan-binding (PGB) site emanating from four spatially close to each other residues, only one of which matches with the proposal of Koebnik ([49]). **Fig. I.8** displays a superimposition of the RmpM structure (PDB: 1R1M, blue color) and a modeled C-KpOmpA

structure (Swiss-Model, swissmodel.expasy.org, green color), based on the RmpM structure. The residues proposed by [49] for a PGB-site are painted in yellow and those proposed by [31] are in red sticks. Notably, both scenarios involve conserved residues only from the core structure of the OmpA C-like proteins, which is a  $\beta\alpha\beta\alpha\beta\beta$ -motif. The different homologues vary mostly in the presence and length of an elongation of the chain found between the last two  $\beta$ -strands of the core segment. In the case of RmpM, this section contains two  $\alpha$ -helices, while C-KpOmpA is slightly shorter in that area and the model constructed only one  $\alpha$ -helix. The longer chain of RmpM, as seen on the alignment above, is painted in orange on **Fig. I.8**. Modelling the C-terminal domain of *E. coli* OmpA (on the basis of RmpM) did not produce a significant difference on the observed structures alignment (data not shown).



**Fig. I.8.** Superimposition of the RmpM structure ([31], blue color) and a based on it, modeled C-KpOmpA structure (Swiss-Model, green color). The residues proposed by [49] for PGB-site are in yellow, and these proposed by [31] are in red sticks. The orange color denotes the mismatch on the alignment (**Fig. I.7**) between the longer RmpM and the shorter C-KpOmpA chains, which is found in the non-conservative area of the molecule (outside the  $\beta\alpha\beta\alpha\beta\beta$ -core motif, indicated with interrupted lines).

The role in peptidoglycan-binding was tested with RmpM and showed positive results. In addition, a PG-unit - NAG-NAM disaccharide with two amino acids (Ala-Glu) attached to it - was docked to the proposed PGB-site of the protein and exhibited well oriented arrangement with the sugars of the murodipeptide, interacting with three of the four putative PGB-residues

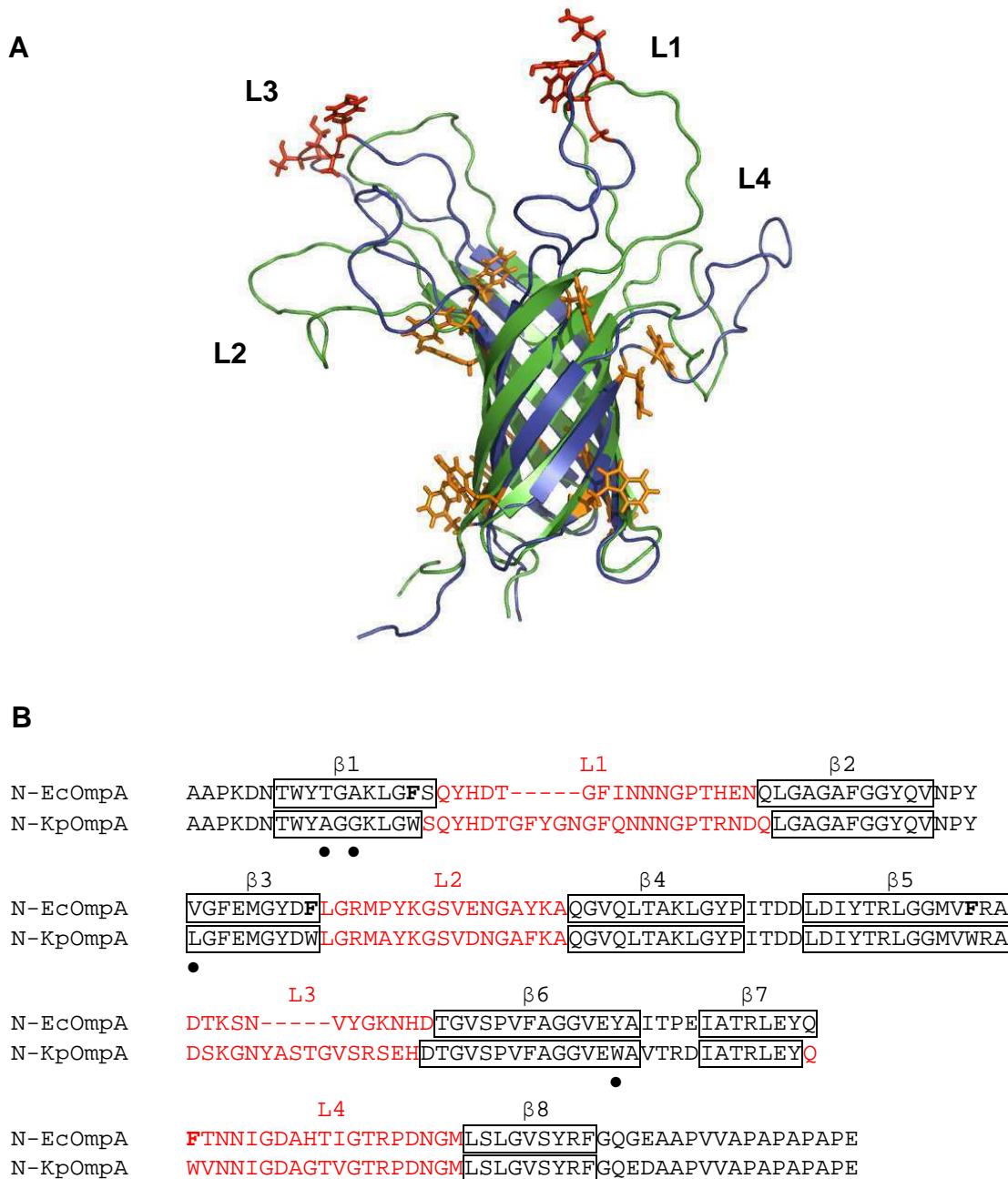
([31]). Furthermore, the authors of this work proposed a possible dimerization of two C-terminal domains (interacting with their first  $\beta$ -sheets of the core motif). Since the two putative PGB-sites remain accessible in this formation, this may further enhance the binding strength between RmpM and the glycan chains.

As expected, the C-terminal domain of OmpA can be expressed and folded independently from the transmembrane  $\beta$ -barrel, as is shown in this work as well. Interestingly, a chaperone role was proposed for this domain in *E. coli* OmpA, suggesting that the full-length polypeptide chain can exhibit 'self-support' properties during the moments between its synthesis and translocation across the inner membrane [50]. This concept may need further investigation. In the present work, the C-terminal domain of KpOmpA was designed on a gene level, expressed and purified (Chapter 1). Limited proteolysis experiments (Chapter 4) identified the most accessible (and presumably unstructured) area of the molecule, which correlated with the non-conservative loop emanating from the  $\beta\alpha\beta\alpha\beta\beta$ -core motif.

## **I.5. N-terminal domain of Outer membrane protein A**

The OmpA family of proteins is highly conserved among *Enterobacteria* and the most studied representative is the OmpA of *E. coli*. Crystallographic ([8, 46]) and solution state NMR ([9]) structures of the N-terminal domain of OmpA from *E. coli* (named 'N-EcOmpA'), as well as a solution state NMR structure ([47]) of the N-terminal domain of OmpA from *K. pneumoniae* (named 'N-KpOmpA') are available. The N-terminal domain of EcOmpA (**Fig. I.5A**) is a  $\beta$ -barrel that natively resides in the outer membrane. The structure of this domain possesses four long extracellular loops (L1-4) and three short periplasmic turns (T1-3) which connect the 8 antiparallel  $\beta$ -strands of the molecule. The structures of N-EcOmpA and N-KpOmpA are very similar (**Fig. I.9A**), with the major differences concentrated in the area of the extracellular protein loops, two of which (L1 and L3) are larger in the case of N-KpOmpA. The secondary structure elements of the two proteins, however, are nearly identical. Both sequences end with flexible Proline/Alanine-rich hinge region which natively links the N-terminal with the C-terminal domains of the molecules (**Fig. I.9B**).





**Fig. I.9.** Similarities between the NMR structures of N-EcOmpA ([9], PDB: 2GE4) and N-KpOmpA ([47], PDB: 2K0L). **(A)** Aligned structures (with PyMOL) of the  $\beta$ -barrel domains of EcOmpA (green) and KpOmpA (blue). The aromatic girdles at the membrane interface regions and the longer loop segments of N-KpOmpA (on L1 and L3) are in orange and red sticks, respectively. **(B)** Sequence alignment of N-EcOmpA and N-KpOmpA. The eight  $\beta$ -sheets ( $\beta$ 1-8) according to each structure are indicated with black boxes. The four extracellular loops (L1-4) are with red characters. Four Tryptophans in the N-EcOmpA chain have been mutated to Phenylalanines (denoted in bold), which may have affected the span of some of the  $\beta$ -sheets. Ignoring this artificially introduced difference leaves only 4 residues within the barrel core of the molecules which are not identical (marked with black dots).

Aside from the structural elucidations mentioned above, the folding and stability of N-EcOmpA was investigated with a variety of loops-truncated mutants [51]. It was shown that the formation of the  $\beta$ -barrel is independent from the presence or absence of the extracellular protein segments, i.e. all the structural information is contained in the 8  $\beta$ -strands. This may explain the strong evolutionary preservation of the hydrophobic protein core. The kinetics of EcOmpA folding when the urea-denatured protein spontaneously inserts into preformed lipid bilayers identified several membrane-bound intermediates in this process [43]. Interestingly, the final stage exhibited simultaneous, synchronized insertion of all  $\beta$ -sheets across the bilayer [52], thus differentiating the folding mechanisms for  $\beta$ -barrel and  $\alpha$ -helical proteins.

While the OmpA structural features are now well established, the question of the function(s) of this protein still remains open. Since many of the  $\beta$ -barrel structures belong to the major bacterial porins, one is tempted to ask whether OmpA itself is a pore. Early works with the protein in the field of channel conductance measurements suggested that EcOmpA is indeed capable of transporting non-specifically small molecules ([53, 54]). Later, two conductance states (50-80 and 260-320 pS) of EcOmpA were identified in DPhPC membranes [55]. The lower conductance state ( $\sim$ 75 pS) was also detected when C-terminal truncated EcOmpA (i.e. N-EcOmpA) was used [56]. Although the X-ray structure of N-EcOmpA ([8]) rules out a putative pore function, several water pockets can be observed in the barrel interior. Molecular dynamics studies further suggested that the pore formation in N-EcOmpA is feasible and even identified a putative salt-bridge mechanism for channel gating ([57, 58]). Interestingly, the latter work observed an interaction of the soluble C-terminal domain with the DMPC bilayer throughout the course of the simulation. Far before all of the abovementioned findings accumulated, an interesting hypothesis emerged: the full-length two-domain OmpA could, under certain conditions (i.e. physiological temperature, [59, 60]), reorganize its polypeptide chain so that it forms a single-domain, larger  $\beta$ -barrel that is embedded in the membrane and is a more efficient channel, in comparison with the N-terminal domain alone ([61]). Clearly this hypothesis and the putative pore function of OmpA in general need further investigation and unambiguous experimental evidences.

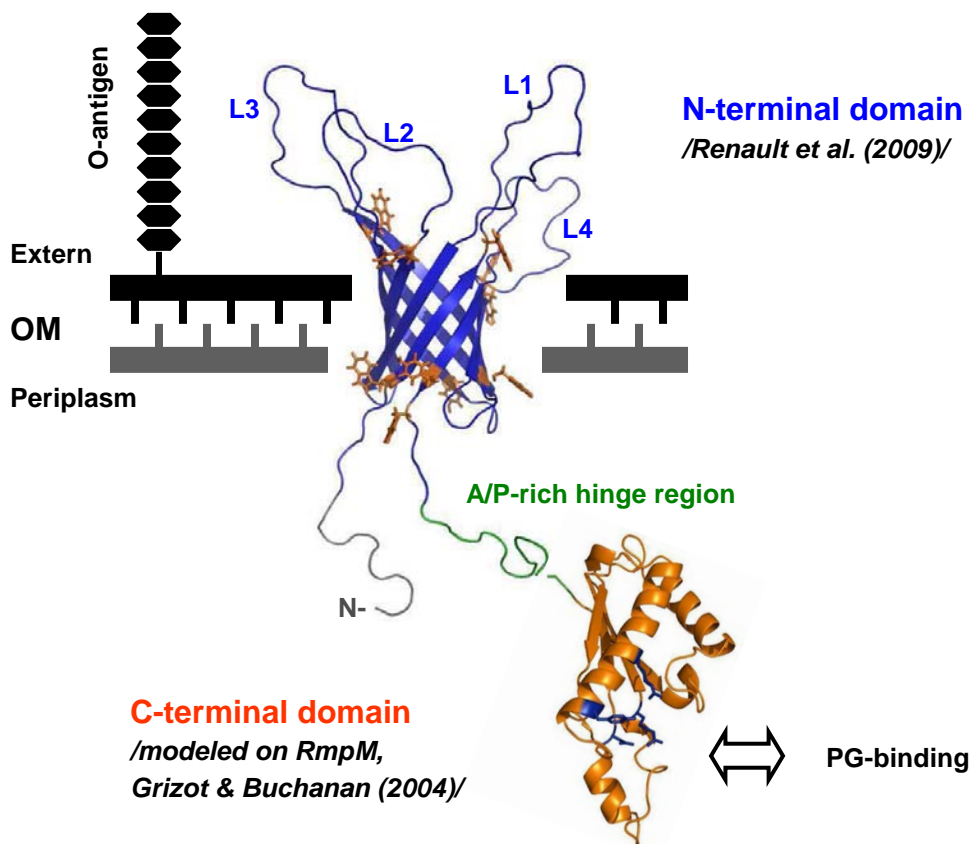
Apart from the structural and the suggested pore-forming properties attributed to the rigid barrel core of OmpA, another interesting portion of the N-terminal domain is the hydrophilic mass composed of the four extracellular loops. Those connect the eight  $\beta$ -strands and protrude out of the outer membrane, thus remaining exposed to the bacterial environment. Hence it is logical that the variety of OmpA-related immunological properties of *E. coli* is often associated with that region of the molecule. *E. coli* is a prominent cause of

neonatal meningitis, the successful treatment of which often suffers from the constantly emerging antibiotic-resistant strains [62]. This disease is associated with the OmpA-related invasion ([63]) of the bacteria through the blood-brain barrier - a tightly organized layer of brain microvascular endothelial cells. The four OmpA loops were shown to be crucial for binding to the host cells ([64]) and a study with several loops-mutated protein constructs provided a detailed mapping on the most important areas for intermolecular recognition ([65]). The OmpA receptor found on the brain endothelial cells surface, on the other hand, was also identified [66]. Named Ecgp, this glycoprotein is similar to gp96 and Hsp90. The binding between Ecgp and OmpA involves disaccharide epitopes on the receptor surface [29] and two putative regions in the OmpA loops region, one of which involves the most mobile sections of L1 and L2 [67]. Thus it is interesting to correlate the level of mobility in a certain protein segment with its putative functions, an attempt for which is made in the present work as well. The functions of OmpA described above categorize it as an adhesin and invasin. Other immunologically relevant properties of N-EcOmpA include evasion of the host defense by interfering with the complement activation, by inhibiting the cytokine synthesis and by allowing the bacteria to survive upon phagocytosis, i.e. to multiply within the macrophage [66, 68, 69]. Such a variety of invasion properties produces a rather gloomy picture for the interactions between this pathogen and its host, seemingly without a 'chance for survival' for the latter. However, possessing a multi-purpose surface-exposed molecule is always a double-edged knife. It was shown that EcOmpA is attacked by several bacteriophages ([70];[51]), recognizing areas in L1, L2 and L3. Interestingly, recently found alternative allele for OmpA (named *ompA2* [71] and further discussed in Chapter 4) that has altered L1 and L3 sections was somewhat resistant to phages. The innate immune system, on the other hand, is capable of targeting OmpA via its non-oxidative pathway, by using the neutrophil elastase [72]. In addition, the serum amyloid A protein that is produced at high level during inflammation binds to OmpA ([28]), thus opsonizing the bacteria.

In the particular case of KpOmpA, it was shown that the protein binds to a variety of immune cells ([73-75]) and that it activates macrophages and dendritic cells ([74, 76]). KpOmpA indirectly activates the Toll-like receptor 2 signaling pathway after initial binding to the scavenger receptors LOX-1 and SREC-I ([77]). In contrast with EcOmpA, specific extracellular regions of KpOmpA that might participate in these processes are not yet recognized, although some of the protein loops (L2 and L4) are quite similar to those of EcOmpA (**Fig. I.9B**). One possible approach in this direction is probing the flexibility of the protein loops by suitable methods such as NMR, while coupling these experiments with

biochemical data in the same context. *In vitro* screening for intermolecular interactions with specific partners can come at a later stage. Since dynamic data on the loops of N-KpOmpA in DHPC micelles is already available ([47]), one of the main goals of the present work will be a comparison of these in the more native environment of the lipid bilayer. Clearly the large extracellular portion of KpOmpA is immunologically very important, hence we will try to map the different mobility levels in the loops area and relate them, if possible, to their functional importance.

**Fig. I.10** summarizes the global fold of KpOmpA, its position/orientation in the bacterial OM and the main structural elements mentioned above for each domain of the protein.



**Fig. I.10.** Global fold and orientation of KpOmpA in the bacterial OM. The approximate boundaries of the OM are denoted by thick grey (for the inner leaflet) and black (the outer leaflet) lines. The N-terminal transmembrane domain (blue painted barrel) spans the OM with its 8  $\beta$ -strands, anchored in the lipids by hydrophobic residues and the aromatic girdles (orange side-chains) at the membrane-water interfaces. The four extracellular loops (L1-4) face the cell exterior and therefore are supposed to play a role in cell-cell recognition processes. At the C-terminus (found in the periplasmic space like the N-terminus), a flexible Proline/Alanine-rich hinge region (green color) links the  $\beta$ -barrel with the soluble domain of the protein (painted in orange). This domain connects the OM with the peptidoglycan layer in the periplasm, presumably via several conserved residues (blue side-chains).

## **I.6. Structural methods for membrane proteins**

As a PhD student recruited in the Marie Curie Initial Training Network of 'Structural Biology of Membrane Proteins' (SBMPs), I had the opportunity to 'get in touch' with a few structural methods, either commonly used in molecular biology or with growing importance and application in that field. A short overview of these (and other) methods is presented below, in the context of membrane proteins.

### **I.6.1. Molecular modelling**

This approach is now commonly used for investigating a number of membrane protein aspects, such as structure prediction, stability and dynamics, and modelling of protein-ligand and protein-protein complexes [78]. The simulations can provide dynamical view of the protein and build models of conformational states (like open and close states of channels, for instance). The protein-lipid interface, from its hand, may affect the topology, stability, oligomeric assembly, traffic and enzymatic activity [79]. With the increasing number of high resolution structures of membrane proteins that enables homologous modelling of more structures, a wide range of membrane proteins can now be simulated over time spans that allow observation of key biological processes [80]. Recent advances in the field were combined with experimental data to constrain modelling 3D structures of proteins ([81]) and their oligomers ([82]).

### **I.6.2. Atomic Force Microscopy**

Established in the groups of Andreas Engel and Daniel Müller, the Atomic Force Microscopy (AFM) and Single-Molecule Force Spectroscopy (SMFS) are novel biotechnological approaches for characterizing the structure-function relationships of reconstituted membrane proteins in their structurally intact, functionally active state in physiological buffer solutions at ambient temperatures ([83];[84]). Neither staining and/or fixing of the membranes are required by this technique, nor is the presence of a long-range protein order in the bilayer (although tight packing of the protein is advantageous). The high-resolution imaging of the method is capable of visualizing single membrane proteins at sub-nanometer resolution and, in some cases, even single secondary structure elements. The technique allowed observation of oligomeric protein states and structural assemblies, gap junctions from rat liver cells, the OmpF porin of *E. coli*, bacterial surfaces and others ([84];[85];[86];[87];[88];[89]).

The force spectroscopy mode of the apparatus enables detection of unfolding-refolding pathways of membrane proteins and their structural stability, probing their energy landscape and folding kinetics ([90]). The strength of such intra- and intermolecular interactions stabilizing the protein structure depends on environmental factors like oligomerization, temperature change, point mutations, electrolyte and pH variations ([91]). Recently, the strength and location of a single Na<sup>+</sup>-ion, bound as a ligand to individual antiporters, was mapped onto the protein structure at a precision of  $\pm 2$  amino acids. In addition, it was possible to observe inhibitor binding and deactivation of the same antiporter ([92];[93]). Thus the force-distance spectra reveal a ‘fingerprint’ of the protein, distinguishing between active and inhibited states. Similar work was done with the bovine rhodopsin ([93]). Establishing such critically important molecular interactions provides insights on their role in stabilizing the secondary structure elements of the membrane protein.

In the present work, SMFS was used for monitoring the unfolding-refolding pathways of the KpOmpA transmembrane domain and full-length proteins in DMPC bilayers, as described in Chapter 2.

### I.6.3. Electron crystallography

This method utilizes the propensity of membrane proteins to embed their hydrophobic segments into a lipid bilayer and to arrange themselves into long-range ordered lattices. The purified membrane protein must therefore be reconstituted into liposomes at a high protein-to-lipid ratio, increasing the possibility for 2D crystals formation [94]. This is achieved by removing the protein-solubilizing detergent molecules in the presence of mixed protein-detergent-lipids micelles ([95]), usually by means of dialysis or biobeads, or detergent-adsorbing molecules such as cyclodextrin ([96]). Upon reaching the critical micelle concentration of the detergent, aggregation of membrane proteins and lipids occur as a result of the hydrophobic effect. The method is particularly useful when the membrane protein is too fragile to form 3D crystals and in the same time allows it to be studied in the close-to-native environment of the membrane. Currently the electron crystallography is the method of choice for solving the structures of large protein complexes in membranes with low to medium resolution (typically 10-40 Å, although it is possible to go down to atomic resolution in favorable cases, as shown with bacteriorhodopsin [97]). A major advantage of the method is that it requires small amount of sample (in comparison with solid-state NMR, for instance) and that it may be applied on intact membranes [98], [99]. Recent technological

advancements in the field allowed the structure of Aquaporin-0 to be solved at 1.9 Å resolution [100].

In the present work, 2D crystals of KpOmpA were not observed, which prevented us from using the full power of electron crystallography (or electron cryomicroscopy). Transmission electron microscopy was used for KpOmpA-containing proteoliposomes visualization, as described in Chapter 3.

#### 1.6.4. X-ray crystallography

As already mentioned, the vast majority of structures deposited in the PDB belongs to soluble proteins, compared to their membrane counterparts. However, the dominant structural approach in each case falls to the X-ray crystallography ([101]). The technique relies on obtaining 3D crystals composed exclusively of the protein of interest, either by standard vapor diffusion methods starting from protein-detergent micelles, or by using lipid cubic phases ([102]). A bottleneck in the process emanates from the fact that the membrane proteins are often difficult to crystallize and subsequently the crystals are more fragile to handle. However, automation and robots usage allow screening of large panels of crystallization conditions with nanovolumes per trial, while irradiation with synchrotron-derived microfocused beams is capable of collecting complete data sets from crystals of just a few microns in size ([103]). A method recently developed at the IBS (Grenoble) allows automated screening of crystallization plates (and even data collection from the positive hits) without taking out the crystals from the growing medium ([104]). In X-ray crystallography, the collected data represents a diffraction pattern from scattered beams, appearing upon interaction with the gradually rotated (in the three dimensions) crystal. The data is then treated computationally to produce an electron density map. Obtaining an interpretable electron density map is one of the key steps, strongly dependent on acquiring information on both the amplitude and the phase of the scattered waves. While the amplitude is proportional to the intensity of the diffraction pattern's 'spots', the phase can not be measured directly. This problem is often dealt with directly (for small molecules with highly resolved data) or by molecular replacement (if a homologous structure is already available). Since the membrane proteins are in general large molecules and the database of unique structures is quite limited, this invokes phasing with more complex approaches, such as multiple isomorphous replacement (using introduction of heavy atoms) or multiple anomalous dispersion (using introduction of anomalous scattering atoms, such as selenium in seleno-methionine, [105]). Finally, the phased electron density is used for fitting-in the atomic model of the molecule,

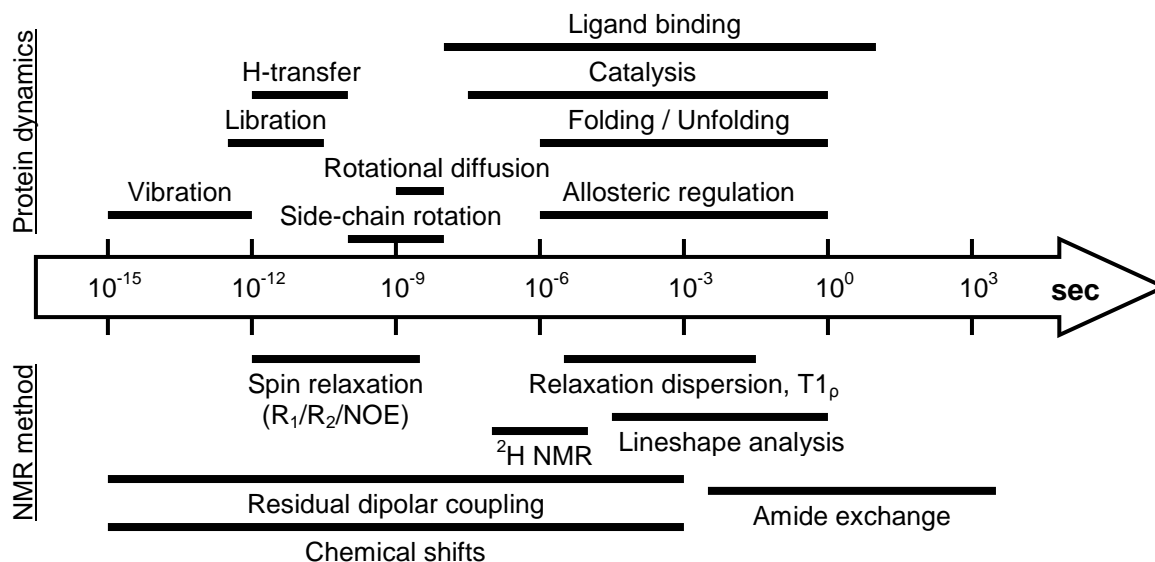
which is then refined. Even in the field of membrane proteins, and although these proteins present specific difficulties, X-ray diffraction remains the major source of high-resolution structural information on membrane proteins.

### I.6.5. Nuclear Magnetic Resonance

This spectroscopic method allows studying the molecular structure and dynamics of proteins in solution and/or in non-crystalline environment (solid-state NMR). The majority of NMR-derived structures have been obtained using aqueous solutions [106]. In the context of membrane proteins, this translates into the micellar environment of detergent-solubilized polypeptides and the structures of several  $\beta$ -barrel proteins and a pentameric  $\alpha$ -helical protein were determined [107]. The solid-state NMR approach, on the other hand, can investigate the lipid bilayer itself ([108]) and/or determine the structures and dynamics of membrane embedded proteins [109]. These studies can be conducted with microcrystalline or frozen samples, or extended to lipid bilayer preparations [110]. This will allow a comparison between particular protein features in the two environments, highlighting the influence of the surrounding lipid matrix on the protein structure. This was demonstrated, for instance, with a receptor-transducer complex of sensory rhodopsin [111]. In this way, NMR offers complementary spectroscopic means to study not only the protein structures, but also their dynamic behavior, ligand binding, formation of complexes and the influence of the protein environment. The macromolecules (such as proteins) are known to display a variety of dynamics on different timescales and orders of magnitude (**Fig. I.11**), which can be probed with various NMR methods [112].

The presence of protein crystals is not required for NMR due to the intrinsic short-range order of the observed interactions. This allows protein investigation in functional environment, which may be very powerful in combination with diffraction techniques. Recent progresses in solid-state NMR have used the dynamic nuclear polarization effect to increase NMR sensitivity by one or two orders of magnitude, opening new perspectives in the field [114]. One of the most exciting is the possibility to perform 'cellular NMR', that is to study membrane proteins in their natural environment without resorting to purifying recombinant proteins and reconstituting them in a membrane mimetic environment [115-117].





**Fig. I.11.** Timescales of the different protein dynamics and the respective NMR experiments for probing them. NOE - Nuclear Overhauser Effect. Adapted from [112] - [113].

In the present work, the transmembrane domain of KpOmpA was investigated by solid-state NMR after protein reconstitution at high protein-to-lipid ratio in the native-like environment of lipid bilayers. Several aspects of the mobility in different sections of the molecule are analyzed in relation with their putative functional roles (Chapter 3).

#### I.6.6. MALDI-TOF mass spectrometry

Although not strictly considered a structural technique, the mass spectrometry (MS) provides additional characterization of the object of interest. Developed since more than a century, MS variants are largely used nowadays for very different analytical purposes, ranging from paleontology to space exploration. In biology, MS perhaps finds its main application in the field of proteomics and protein identification [118]. A strong positive side of the technique is that it requires small amounts of sample (sub picomole amount).

The method relies on ionization of charged particles and their subsequent separation as a function of their mass-to-charge ratio. Since the conventional ionization sources are often destructive when large molecules (like proteins) are to be measured, the type of MS largely used in this work is the matrix-assisted laser desorption/ionization (MALDI, [119, 120]). This 'soft' ionization variant of MS allows the observation of large biopolymers after their ionization mediated by a crystalline, low molecular weight compound termed a matrix. The matrices used in this work are some of the most common ones - SPA (3,5-dimethoxy-4-

hydroxycinnamic acid, or sinapinic acid), CHCA ( $\alpha$ -cyano-4-hydroxycinnamic acid) and (exceptionally) DHB (2,5-dihydroxybenzoic acid). Upon mixing the sample with the matrix (on the specially designed metal MALDI-plate) and evaporation of the solute, the matrix molecules and the sample material (i.e. the protein) co-crystallize. According to the (still debated) mechanism for sample ionization, upon irradiation with a UV laser the matrix molecules absorb the bulk of the energy and transfer proton(s) to the protein, turning it into a charged molecule. The same type of ions (all that are  $1^+$ , or  $2^+$ , etc.) are then accelerated in an electric field and 'equalized' among themselves in terms of kinetic energy, i.e. the heavier objects will move slower and will need more time to reach the detector. This time (termed 'time-of-flight', or TOF) is measured and, for a given length of the flight-path and known acceleration voltage, is related to the mass-to-charge ratio of the particle. With MALDI-TOF, protein and protein fragments of very different sizes (up to  $\sim 100$  kDa) can be detected using the standard, lower resolution (or 'linear') measurement mode ([121]). Machines equipped with a reflectron 'ion mirror' device increase the flight-path and therefore the resolution significantly, but suffer from a limitation of the observed mass range (up to  $\sim 10$  kDa).

In this work, MALDI-TOF measurements were used for identification of protein fragments after limited proteolysis experiments, as well as for visualizing the intact polypeptides (Chapter 4).

#### I.6.7. Biochemistry methods and limited proteolysis

Last but not least, the traditional and modern techniques in molecular biology carry the task of supplying the structural methods with their objects of interest, as well as of analyzing these objects in parallel. Often the biochemistry role in the process is manifested at the early stages - cloning and constructing genes for recombinant proteins, expression and purification of these in substantial amounts, characterization and sample preparations for crystallography (2D and 3D crystals), NMR (different labeling schemes), etc. However, a combination of structural and biochemical methods during the phase of protein investigation can be particularly powerful, as it correlates the observed structural/dynamic features of the molecule with its functional aspects (folding, ligand binding, enzymatic or channel activity) under various treatment conditions. It is far beyond the scope of this work to encompass the vast arsenal of all molecular biology methods, neither it is necessary to describe largely used traditional techniques like polymerase chain reaction, electrophoresis of proteins and nucleic acids, centrifugation in density gradient, colorimetric titration of proteins and others. The construction of the recombinant full-length KpOmpA and of its C-terminal domain alone, as

well as the expression and purification of the three protein constructs used throughout the manuscript are described in details in Chapter 1. Therefore here we will provide a notion for just one particular biochemical approach - the limited proteolysis, - which is largely used in Chapter 4.

The limited proteolysis (for a nice review see [122]) relies on partial protein fragmentation by a proteolytic enzyme which depends on several parameters, such as the folding state of the protein, the enzyme-to-substrate ratio, the time and temperature of incubation [123]. The cornerstone of the approach is the postulation that folded protein segments (participating in secondary structure elements) will not be susceptible to cleavage, while exposed and unstructured, flexible portions of the molecule will be readily digested. Because the proteolytic events are said to be dictated by the accessibility (i.e. exposure and flexibility) at the cleavage site and not by the specificity of the used protease, the 'classic' version of the approach utilizes enzymes with broad target ranges (Subtilisin, Pepsin, etc). This was nicely exemplified with autolysis of Thermolysin ([124]), emphasizing the role of flexibility of protein surface areas in the process of protein-protein recognition. Alternatively, a sequence-specific enzyme (i.e. Trypsin which cleaves only after Lysine and Arginine residues, or Endo-Lys C which cleaves specifically after Lysine residues) may be used as well, since this will create a limited set of fragments and combinations of fragment, which can then be traced by other methods (such as SDS PAGE and mass spectrometry) and mapped on the protein structure (if available). This variant can be particularly beneficial when combining more than one enzyme, if the cleavable residues may sample various protein regions.

Thus the mobility in certain protein sections can be related to the cleavage pattern observed after treatment with a specific protease. If dynamic data (i.e. from NMR experiments) for the different protein regions is also available, the two datasets can be compared in order to pinpoint previously unobserved details or, by contrast, to confirm the expected outcome. Of note, in the case of soluble proteins too long reaction times can lead to complete digestion of the molecule. On the other hand, the folded membrane proteins (in detergent micelles or lipid bilayers), will have their transmembrane sections embedded in the hydrophobic environment (i.e. 'protected' from the enzyme). In such conditions they may avoid digestion even after exposure to the enzyme for long times and at higher temperatures. Examples of these situations will be seen and discussed in Chapter 4.

## I.7. Aims of the present work

In the light of the information presented above, the specific tasks of our work can be formulated as follows:

1. Cloning, expression and purification of the full-length polypeptide of KpOmpA (F-KpOmpA), which will serve for comparative experiments together with the already available  $\beta$ -barrel construct, N-KpOmpA (Chapter 1).

2. Design, expression and purification of the C-terminal periplasmic domain of KpOmpA (C-KpOmpA), opening the pathway to structure determination of that protein segment in future (Chapter 1).

3. Optimization of the reconstitution conditions for N- and F-KpOmpA in lipid bilayers with the help of electron microscopy and supplementary biochemistry techniques, providing opportunity for SMFS and solid-state NMR experiments with these preparations (Chapters 1 and 3).

4. Unfolding experiments (by SMFS) with N- and F-KpOmpA that will establish the rupture points in the  $\beta$ -barrel section, explore the unfolding-refolding pathways of the molecule and suggest a putative role of the extracellular protein segments in the bacterial membrane (Chapter 2).

5. Investigation (by solid-state NMR) of the dynamic behavior of the reconstituted N-KpOmpA and comparing it with the already available information for this protein construct in detergent micelles, thus exploring the influence of the protein environment on the loops mobility (Chapter 3).

6. Proteolytic digestion of the protein constructs and correlating the enzyme-accessibility of the different loop regions with their mobility levels (Chapter 4).

The combination of these experimental strategies will allow us to define precisely the functional dynamics of KpOmpA in various membrano-mimetic environments.

## References

1. Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited*. Microbiol Mol Biol Rev, 2003. **67**(4): p. 593-656.
2. Funahara, Y. and H. Nikaido, *Asymmetric localization of lipopolysaccharides on the outer membrane of Salmonella typhimurium*. J Bacteriol, 1980. **141**(3): p. 1463-5.
3. Muller-Loennies, S., L. Brade, and H. Brade, *Neutralizing and cross-reactive antibodies against enterobacterial lipopolysaccharide*. International Journal of Medical Microbiology, 2007. **297**(5): p. 321-340.
4. Demchick, P. and A.L. Koch, *The permeability of the wall fabric of Escherichia coli and Bacillus subtilis*. Journal of Bacteriology, 1996. **178**(3): p. 768-773.
5. Bos, M.P., V. Robert, and J. Tommassen, *Biogenesis of the gram-negative bacterial outer membrane*. Annual Review of Microbiology, 2007. **61**: p. 191-214.
6. Raetz, C.R.H. and C. Whitfield, *Lipopolysaccharide endotoxins*. Annual Review of Biochemistry, 2002. **71**: p. 635-700.
7. Koebnik, R., K.P. Locher, and P. Van Gelder, *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*. Molecular Microbiology, 2000. **37**(2): p. 239-253.
8. Pautsch, A. and G.E. Schulz, *Structure of the outer membrane protein A transmembrane domain*. Nature Structural Biology, 1998. **5**(11): p. 1013-1017.
9. Arora, A., et al., *Structure of outer membrane protein A transmembrane domain by NMR spectroscopy*. Nature Structural Biology, 2001. **8**(4): p. 334-338.
10. Vandeputte-Rutten, L., et al., *Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site*. Embo Journal, 2001. **20**(18): p. 5033-5039.
11. Snijder, H.J., et al., *Structural evidence for dimerization-regulated activation of an integral membrane phospholipase*. Nature, 1999. **401**(6754): p. 717-21.
12. Buchanan, S.K., et al., *Crystal structure of the outer membrane active transporter FepA from Escherichia coli*. Nature Structural Biology, 1999. **6**(1): p. 56-63.
13. Ferguson, A.D., et al., *Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide*. Science, 1998. **282**(5397): p. 2215-2220.
14. Schirmer, T., et al., *Structural Basis for Sugar Translocation through Maltoporin Channels at 3.1-Angstrom Resolution*. Science, 1995. **267**(5197): p. 512-514.
15. van den Berg, B., et al., *Crystal structure of the long-chain fatty acid transporter FadL*. Science, 2004. **304**(5676): p. 1506-1509.
16. Basle, A., et al., *Crystal structure of osmoporin OmpC from E-coli at 2.0 angstrom*. Journal of Molecular Biology, 2006. **362**(5): p. 933-942.
17. Cowan, S.W., et al., *Crystal structures explain functional properties of two E. coli porins*. Nature, 1992. **358**(6389): p. 727-33.
18. Subbarao, G.V. and B. van den Berg, *Crystal structure of the monomeric porin OmpG*. Journal of Molecular Biology, 2006. **360**(4): p. 750-759.
19. Yildiz, O., et al., *Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation*. Embo Journal, 2006. **25**(15): p. 3702-3713.
20. Hong, H.D., et al., *The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel*. Journal of Biological Chemistry, 2006. **281**(11): p. 7568-7577.
21. Vogt, J. and G.E. Schulz, *The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence*. Structure with Folding & Design, 1999. **7**(10): p. 1301-1309.
22. Fernandez, C., et al., *Solution NMR studies of the integral membrane proteins OmpX and OmpA from Escherichia coli*. Febs Letters, 2001. **504**(3): p. 173-178.

23. Buchanan, S.K., *beta-Barrel proteins from bacterial outer membranes: structure, function and refolding*. Current Opinion in Structural Biology, 1999. **9**(4): p. 455-461.
24. Murzin, A.G., A.M. Lesk, and C. Chothia, *Principles Determining the Structure of Beta-Sheet Barrels in Proteins .1. A Theoretical-Analysis*. Journal of Molecular Biology, 1994. **236**(5): p. 1369-1381.
25. Murzin, A.G., A.M. Lesk, and C. Chothia, *Principles Determining the Structure of Beta-Sheet Barrels in Proteins .2. The Observed Structures*. Journal of Molecular Biology, 1994. **236**(5): p. 1382-1400.
26. Tamm, L.K., H. Hong, and B.Y. Liang, *Folding and assembly of beta-barrel membrane proteins*. Biochimica Et Biophysica Acta-Biomembranes, 2004. **1666**(1-2): p. 250-263.
27. Hiller, S., et al., *Solution structure of the integral human membrane protein VDAC-1 in detergent micelles*. Science, 2008. **321**(5893): p. 1206-1210.
28. Hari-Dass, R., et al., *Serum amyloid A protein binds to outer membrane protein A of Gram-negative bacteria*. Journal of Biological Chemistry, 2005. **280**(19): p. 18562-18567.
29. Prasadarao, N.V., C.A. Wass, and K.S. Kim, *Endothelial cell GlcNAc beta 1-4GlcNAc epitopes for outer membrane protein A enhance traversal of Escherichia coli across the blood-brain barrier*. Infect Immun, 1996. **64**(1): p. 154-60.
30. Ferguson, A.D., et al., *A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins*. Structure with Folding & Design, 2000. **8**(6): p. 585-592.
31. Grizot, S. and S.K. Buchanan, *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. Molecular Microbiology, 2004. **51**(4): p. 1027-1037.
32. Chen, R. and U. Henning, *A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins*. Molecular Microbiology, 1996. **19**(6): p. 1287-1294.
33. Korndorfer, I.P., M.K. Dommel, and A. Skerra, *Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture*. Nature Structural & Molecular Biology, 2004. **11**(10): p. 1015-1020.
34. Lazar, S.W. and R. Kolter, *SurA assists the folding of Escherichia coli outer membrane proteins*. Journal of Bacteriology, 1996. **178**(6): p. 1770-1773.
35. Voulhoux, R., et al., *Role of a highly conserved bacterial protein in outer membrane protein assembly*. Science, 2003. **299**(5604): p. 262-265.
36. Prinz, T. and J. Tommassen, *Association of iron-regulated outer membrane proteins of Neisseria meningitidis with the RmpM (class 4) protein*. Fems Microbiology Letters, 2000. **183**(1): p. 49-53.
37. Struyve, M., M. Moons, and J. Tommassen, *Carboxy-Terminal Phenylalanine Is Essential for the Correct Assembly of a Bacterial Outer-Membrane Protein*. Journal of Molecular Biology, 1991. **218**(1): p. 141-148.
38. Xian, M., et al., *Sorting signal of Escherichia coli OmpA is modified by oligo-(R)-3-hydroxybutyrate*. Biochim Biophys Acta, 2007. **1768**(11): p. 2660-6.
39. Negoda, A., E. Negoda, and R.N. Reusch, *Oligo-(R)-3-hydroxybutyrate modification of sorting signal enables pore formation by Escherichia coli OmpA*. Biochimica Et Biophysica Acta-Biomembranes, 2010. **1798**(8): p. 1480-1484.
40. deCock, H., et al., *Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12*. Journal of Molecular Biology, 1997. **269**(4): p. 473-478.
41. Surrey, T. and F. Jahrig, *Refolding and Oriented Insertion of a Membrane-Protein into a Lipid Bilayer*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(16): p. 7457-7461.
42. Kleinschmidt, J.H. and L.K. Tamm, *Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism*. Biochemistry, 1996. **35**(40): p. 12993-13000.

43. Kleinschmidt, J.H. and L.K. Tamm, *Time-resolved distance determination by tryptophan fluorescence quenching: Probing intermediates in membrane protein folding*. Biochemistry, 1999. **38**(16): p. 4996-5005.
44. Schulz, G.E., *The structure of bacterial outer membrane proteins*. Biochimica Et Biophysica Acta-Biomembranes, 2002. **1565**(2): p. 308-317.
45. Abramson, J., et al., *The 3D structures of VDAC represent a native conformation*. Trends in Biochemical Sciences, 2010. **35**(9): p. 514-521.
46. Pautsch, A. and G.E. Schulz, *High-resolution structure of the OmpA membrane domain*. Journal of Molecular Biology, 2000. **298**(2): p. 273-282.
47. Renault, M., et al., *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, 2009. **385**(1): p. 117-30.
48. Yang, Y.S., et al., *Structure of the Mycobacterium tuberculosis OmpATb protein: A model of an oligomeric channel in the mycobacterial cell wall*. Proteins-Structure Function and Bioinformatics, 2011. **79**(2): p. 645-661.
49. Koebnik, R., *Proposal for a Peptidoglycan-Associating Alpha-Helical Motif in the C-Terminal Regions of Some Bacterial Cell-Surface Proteins*. Molecular Microbiology, 1995. **16**(6): p. 1269-1270.
50. Danoff, E.J. and K.G. Fleming, *The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane beta-barrel*. Biophys Chem, 2011. **159**(1): p. 194-204.
51. Koebnik, R., *Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from Escherichia coli*. Journal of Bacteriology, 1999. **181**(12): p. 3688-3694.
52. Kleinschmidt, J.H., et al., *Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism*. Biochemistry, 1999. **38**(16): p. 5006-5016.
53. Sugawara, E. and H. Nikaido, *Pore-Forming Activity of OmpA Protein of Escherichia-Coli*. Journal of Biological Chemistry, 1992. **267**(4): p. 2507-2511.
54. Sugawara, E. and H. Nikaido, *OmpA Protein of Escherichia-Coli Outer-Membrane Occurs in Open and Closed Channel Forms*. Journal of Biological Chemistry, 1994. **269**(27): p. 17981-17987.
55. Arora, A., et al., *Refolded outer membrane protein A of Escherichia coli forms ion channels with two conductance states in planar lipid bilayers*. Journal of Biological Chemistry, 2000. **275**(3): p. 1594-1600.
56. Saint, N., et al., *Ion channel formation by N-terminal domain: a common feature of OprFs of Pseudomonas and OmpA of Escherichia coli*. Fems Microbiology Letters, 2000. **190**(2): p. 261-265.
57. Bond, P.J., J.D. Faraldo-Gomez, and M.S.P. Sansom, *OmpA: A pore or not a pore? Simulation and modeling studies*. Biophysical Journal, 2002. **83**(2): p. 763-775.
58. Khalid, S., et al., *OmpA: Gating and dynamics via molecular dynamics simulations*. Biochimica Et Biophysica Acta-Biomembranes, 2008. **1778**(9): p. 1871-1880.
59. Zakharian, E. and R.N. Reusch, *Outer membrane protein A of Escherichia coli forms temperature-sensitive channels in planar lipid bilayers*. Febs Letters, 2003. **555**(2): p. 229-235.
60. Zakharian, E. and R.N. Reusch, *Kinetics of folding of Escherichia coli OmpA from narrow to large pore conformation in a planar bilayer*. Biochemistry, 2005. **44**(17): p. 6701-6707.
61. Stathopoulos, C., *An alternative topological model for Escherichia coli OmpA*. Protein Science, 1996. **5**(1): p. 170-173.
62. Kim, K.S., *Pathogenesis of bacterial meningitis: From bacteraemia to neuronal injury*. Nature Reviews Neuroscience, 2003. **4**(5): p. 376-385.
63. Prasadarao, N.V., et al., *Outer membrane protein A of Escherichia coli contributes to invasion of brain microvascular endothelial cells*. Infection and Immunity, 1996. **64**(1): p. 146-153.

64. Shin, S., et al., *Escherichia coli* outer membrane protein A adheres to human brain microvascular endothelial cells. *Biochem Biophys Res Commun*, 2005. **330**(4): p. 1199-204.
65. Prasadarao, N.V., et al., *Deciphering the Roles of Outer Membrane Protein A Extracellular Loops in the Pathogenesis of Escherichia coli K1 Meningitis*. *Journal of Biological Chemistry*, 2011. **286**(3): p. 2183-2193.
66. Prasadarao, N.V., *Identification of Escherichia coli outer membrane protein A receptor on human brain microvascular endothelial cells*. *Infection and Immunity*, 2002. **70**(8): p. 4556-4563.
67. Datta, D., et al., *Interaction of E. coli outer-membrane protein A with sugars on the receptors of the brain microvascular endothelial cells*. *Proteins*, 2003. **50**(2): p. 213-21.
68. Weiser, J.N. and E.C. Gotschlich, *Outer-Membrane Protein-a (Ompa) Contributes to Serum Resistance and Pathogenicity of Escherichia-Coli K-1*. *Infection and Immunity*, 1991. **59**(7): p. 2252-2258.
69. Sukumaran, S.K., H. Shimada, and N.V. Prasadarao, *Entry and intracellular replication of Escherichia coli K1 in macrophages require expression of outer membrane protein A*. *Infection and Immunity*, 2003. **71**(10): p. 5951-5961.
70. Morona, R., M. Klose, and U. Henning, *Escherichia-Coli K-12 Outer-Membrane Protein (Ompa) as a Bacteriophage Receptor - Analysis of Mutant-Genes Expressing Altered Proteins*. *Journal of Bacteriology*, 1984. **159**(2): p. 570-578.
71. Power, M.L., et al., *A naturally occurring novel allele of Escherichia coli outer membrane protein A reduces sensitivity to bacteriophage*. *Appl Environ Microbiol*, 2006. **72**(12): p. 7930-2.
72. Belaaouaj, A.A., K.S. Kim, and S.D. Shapiro, *Degradation of outer membrane protein A in Escherichia coli killing by neutrophil elastase*. *Science*, 2000. **289**(5482): p. 1185-1187.
73. Jeannin, P., et al., *OmpA targets dendritic cells, induces their maturation and delivers antigen into the MHC class I presentation pathway*. *Nature Immunology*, 2000. **1**(6): p. 502-509.
74. Soulas, C., et al., *Cutting edge: Outer membrane protein A (OmpA) binds to and activates human macrophages*. *Journal of Immunology*, 2000. **165**(5): p. 2335-2340.
75. Godefroy, S., et al., *Outer membrane protein A (OmpA) activates human epidermal Langerhans cells*. *European Journal of Cell Biology*, 2003. **82**(4): p. 193-200.
76. Jeannin, P., et al., *Outer membrane protein A (OmpA): a new pathogen-associated molecular pattern that interacts with antigen presenting cells-impact on vaccine strategies*. *Vaccine*, 2002. **20**: p. A23-A27.
77. Jeannin, P., et al., *Complexity and complementarity of outer membrane protein a recognition by cellular and humoral innate immunity receptors*. *Immunity*, 2005. **22**(5): p. 551-560.
78. Filipek, S., Modzelewska, A., *Molecular Modelling of Membrane Proteins*, in *Structural Genomics on Membrane Proteins*, K.H. Lundstrom, Editor. 2006, Taylor and Francis: New York. p. 331-48.
79. Schneiter, R. and A. Toulmay, *The role of lipids in the biogenesis of integral membrane proteins*. *Applied Microbiology and Biotechnology*, 2007. **73**(6): p. 1224-1232.
80. Ash, W.L., et al., *Computer simulations of membrane proteins*. *Biochimica Et Biophysica Acta-Biomembranes*, 2004. **1666**(1-2): p. 158-189.
81. Fleishman, S.J. and N. Ben-Tal, *Progress in structure prediction of alpha-helical membrane proteins*. *Current Opinion in Structural Biology*, 2006. **16**(4): p. 496-504.
82. Reggio, P.H., *Computational methods in drug design: Modeling G protein-coupled receptor monomers, dimers, and oligomers*. *Aaps Journal*, 2006. **8**(2): p. E322-E336.
83. Drake, B., et al., *Imaging Crystals, Polymers, and Processes in Water with the Atomic Force Microscope*. *Science*, 1989. **243**(4898): p. 1586-1589.
84. Engel, A. and D.J. Muller, *Observing single biomolecules at work with the atomic force microscope*. *Nature Structural Biology*, 2000. **7**(9): p. 715-718.
85. Seelert, H., et al., *Structural biology - Proton-powered turbine of a plant motor*. *Nature*, 2000. **405**(6785): p. 418-419.

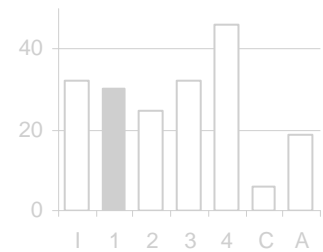


86. Fotiadis, D., et al., *Atomic-force microscopy: Rhodopsin dimers in native disc membranes*. Nature, 2003. **421**(6919): p. 127-128.
87. Fotiadis, D., et al., *Structural analysis of the reaction center light-harvesting complex I photosynthetic core complex of Rhodospirillum rubrum using atomic force microscopy*. Journal of Biological Chemistry, 2004. **279**(3): p. 2063-2068.
88. Scheuring, S. and J.N. Sturgis, *Chromatic adaptation of photosynthetic membranes*. Science, 2005. **309**(5733): p. 484-487.
89. Muller, D.J., et al., *Conformational changes in surface structures of isolated connexin 26 gap junctions*. Embo Journal, 2002. **21**(14): p. 3598-3607.
90. Kedrov, A., et al., *Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy*. Annual Review of Biophysics and Biomolecular Structure, 2007. **36**: p. 233-260.
91. Janovjak, H., et al., *Unfolding pathways of native bacteriorhodopsin depend on temperature*. Embo Journal, 2003. **22**(19): p. 5220-5229.
92. Kedrov, A., et al., *Locating ligand binding and activation of a single antiporter*. Embo Reports, 2005. **6**(7): p. 668-674.
93. Fotiadis, D., et al., *Structure of the rhodopsin dimer: a working model for G-protein-coupled receptors*. Current Opinion in Structural Biology, 2006. **16**(2): p. 252-259.
94. Braun, T., Engel, A., *Encyclopedia of Life Sciences*. 2005. pp. A000304.
95. Remigy, H.W., et al., *Membrane protein reconstitution and crystallization by controlled dilution*. Febs Letters, 2003. **555**(1): p. 160-169.
96. Signorell, G.A., et al., *Controlled 2D crystallization of membrane proteins using methyl-beta-cyclodextrin*. Journal of Structural Biology, 2007. **157**(2): p. 321-328.
97. Grigorieff, N., et al., *Electron-crystallographic refinement of the structure of bacteriorhodopsin*. J Mol Biol, 1996. **259**(3): p. 393-421.
98. Ubarretxena-Belandia, I. and D.L. Stokes, *Present and Future of Membrane Protein Structure Determination by Electron Crystallography*. Advances in Protein Chemistry and Structural Biology: Recent Advances in Electron Cryomicroscopy, Pt A, 2010. **81**: p. 33-60.
99. Wisedchaisri, G., S.L. Reichow, and T. Gonen, *Advances in Structural and Functional Analysis of Membrane Proteins by Electron Crystallography*. Structure, 2011. **19**(10): p. 1381-1393.
100. Gonen, T., et al., *Lipid-protein interactions in double-layered two-dimensional AQP0 crystals (vol 438, pg 633, 2005)*. Nature, 2006. **441**(7090): p. 248-248.
101. White, S.H., *The progress of membrane protein structure determination*. Protein Science, 2004. **13**(7): p. 1948-1949.
102. Landau, E.M. and J.P. Rosenbusch, *Lipidic cubic phases: A novel concept for the crystallization of membrane proteins*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(25): p. 14532-14535.
103. PebayPeyroula, E., et al., *X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases*. Science, 1997. **277**(5332): p. 1676-1681.
104. Jacquamet, L., et al., *Automated analysis of vapor diffusion crystallization drops with an X-ray beam*. Structure, 2004. **12**(7): p. 1219-1225.
105. Ealick, S.E., *Advances in multiple wavelength anomalous diffraction crystallography*. Current Opinion in Chemical Biology, 2000. **4**(5): p. 495-499.
106. Wüthrich, K., *NMR of proteins and nucleic acids*. 1986, New York: Wiley Interscience.
107. Tamm, L.K. and B.Y. Liang, *NMR of membrane proteins in solution*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2006. **48**(4): p. 201-210.
108. Seelig, J., *Deuterium Magnetic-Resonance - Theory and Application to Lipid-Membranes*. Quarterly Reviews of Biophysics, 1977. **10**(3): p. 353-418.
109. Opella, S.J. and F.M. Marassi, *Structure determination of membrane proteins by NMR spectroscopy*. Chemical Reviews, 2004. **104**(8): p. 3587-3606.

110. Baldus, M., *Molecular interactions investigated by multi-dimensional solid-state NMR*. Current Opinion in Structural Biology, 2006. **16**(5): p. 618-623.
111. Klare, J.P., et al., *Effects of solubilization on the structure and function of the sensory rhodopsin II/transducer complex*. Journal of Molecular Biology, 2006. **356**(5): p. 1207-1221.
112. Morin, S., *A practical guide to protein dynamics from 15N spin relaxation in solution*. Prog Nucl Magn Reson Spectrosc, 2011. **59**(3): p. 245-62.
113. Palmer, A.G., 3rd, *NMR characterization of the dynamics of biomacromolecules*. Chem Rev, 2004. **104**(8): p. 3623-40.
114. Debelouchina, G.T., et al., *Dynamic nuclear polarization-enhanced solid-state NMR spectroscopy of GNNQQNY nanocrystals and amyloid fibrils*. Physical Chemistry Chemical Physics, 2010. **12**(22): p. 5911-5919.
115. Jacso, T., et al., *Characterization of Membrane Proteins in Isolated Native Cellular Membranes by Dynamic Nuclear Polarization Solid-State NMR Spectroscopy without Purification and Reconstitution*. Angewandte Chemie-International Edition, 2012. **51**(2): p. 432-435.
116. Renault, M., et al., *Cellular solid-state nuclear magnetic resonance spectroscopy*. Proc Natl Acad Sci U S A, 2011.
117. Renault, M., et al., *Solid-State NMR Spectroscopy on Cellular Preparations Enhanced by Dynamic Nuclear Polarization*. Angew Chem Int Ed Engl, 2012, **51**: p. 2998-3001.
118. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
119. Karas, M., D. Bachmann, and F. Hillenkamp, *Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules*. Analytical Chemistry, 1985. **57**(14): p. 2935-2939.
120. Tanaka, K., et al., *Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry*. Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
121. Hortin, G.L., *The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome*. Clinical Chemistry, 2006. **52**(7): p. 1223-1237.
122. Fontana, A., et al., *Probing protein structure by limited proteolysis*. Acta Biochimica Polonica, 2004. **51**(2): p. 299-321.
123. Fontana, A., et al., *Probing the partly folded states of proteins by limited proteolysis*. Folding & Design, 1997. **2**(2): p. R17-R26.
124. Fontana, A., et al., *Correlation between Sites of Limited Proteolysis and Segmental Mobility in Thermolysin*. Biochemistry, 1986. **25**(8): p. 1847-1851.

# Chapter 1

## Expression, refolding, purification and reconstitution of the different KpOmpA constructs



## Summary

This chapter describes the expression/purification procedures of the different protein constructs used in this study. These include the KpOmpA membrane domain denoted as **N-KpOmpA** (expression, refolding, and purification), the KpOmpA full-length protein denoted as **F-KpOmpA** (cloning, expression, refolding, and purification) and the KpOmpA periplasmic domain denoted as **C-KpOmpA** (gene design, expression, and purification). In particular, the pilot reconstitution trials of N-KpOmpA in different buffers, lipids and lipid-to-protein ratios are discussed in more detail in Chapter 3, while here the basic reconstitution protocol is provided. Hence Chapter 1 represents detailed 'Material and methods' section on the biochemistry side of all other experiments, with further explanation and analysis of the different procedures. Precise step-by-step protocols of all the procedures (aiming faster reproducibility in future) can be found in Annexes.

## 1.1. Construction of the KpOmpA full-length protein (F-KpOmpA)

Natively, KpOmpA is a two-domain structure. The transmembrane N-terminal domain represents a  $\beta$ -barrel anchor in the membrane, while the C-terminal periplasmic domain is found between the inner and the outer membrane of Gram-negative bacteria. The N-KpOmpA construct was thoroughly investigated in terms of structure and dynamics by solution state NMR [1] and the respective information for its *E. coli* homologue is also substantial [2, 3]. However, the native protein possesses a C-terminal domain as well, which is responsible for additional functions of the molecule such as binding to the periplasmic peptidoglycan layer. Hence our interests in obtaining a protein construct closely corresponding to the full-length chain of KpOmpA, which can be further investigated (in Chapter 4).

Genomic DNA from *Klebsiella pneumoniae* strain 52145 was kindly provided by Dr. Tournebize (Institut Pasteur, Paris). The gene of F-KpOmpA in the bacterial genome is presented below:

```
[5']ATAATGAGGCGCAAAAAATGAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGCTTCGCTA
CCGTAGCGCAGGCCGCTCCGAAAGATAACACCTGGTATGCAGGTGGTAAACTGGGTTGGTCCAGTAT
CACGACACCGGTTTCTACGGTAACGGTTTTCCAGAACAACAACGGTCCGACCCGTAACGATCAGCTTGGT
GCTGGTGC GTTCGGTGGTTACCAGGTTAACCCGTACCTCGGTTTCGAAATGGGTTATGACTGGCTGGGC
CGTATGGCATATAAAGGCAGCGTTGACAACGGTGC'TTTCAAAGCTCAGGGCGTTCAGCTGACCGCTAAA
CTGGGTTACCCGATCACTGACGATCTGGACATCTACACCCGTCTGGGCGGCATGGTTTGGCGCGCTGAC
TCCAAAGGCAACTACGTTCTACCGCGTTTTCCCGTAGCGAACACGACACTGGCGTTTTCCCAGTATTT
GCTGGCGGCGTAGAGTGGGCTGTTACTCGTGACATCGCTACCCGTCTGGAATACCAGTGGGTAAACAAC
ATCGGCGACGCGGGCACTGTGGGTACCCGTCTGATAACGGCATGCTGAGCCTGGGCGTTTTCTACCGC
TTCGGTCAGGAAGATGCTGCACCGTTGTTGCTCCGGCTCCGGCTCCGGAAGTGGCTACCAAG
CACTTCACCCTGAAGTCTGACGTTCTGTTCAACTTCAACAAAGCTACCCTGAAACCGGAAGGTCAGCAG
GCTCTGGATCAGCTGTACTCAGCTGAGCAACATGGATCCGAAAGACGGTTCGGCTGTTGTTCTGGGC
TACACCGACCGCATCGGTTCCGAAGCTTACAACCAGCAGCTGTCTGAGAAACGTGCTCAGTCCGTTGTT
GACTACCTGGTTGCTAAAGGCATCCCGGCTGGCAAATCTCCGCTCGCGGCATGGGTGAATCCACCCG
GTTACTGGCAACACCTGTGACAACGTGAAAGCTCGCGCTGCCCTGATCGATTGCCCTGGCTCCGGATCGT
CGTGTAGAGATCGAAGTTAAAGGCTACAAAGAAGTTGTAACCTCAGCCGGCGGCTTAAGTTATAACCGAT
AAAAAAACCCGCTTC[3']
```

Starting after ATG for Methionine (colored red and in bold), the F-KpOmpA gene (bold letters) encodes a 20 amino acids (a.a.) signal peptide (KKTAIAIAVALAGFATVAQA, colored blue) responsible for the native transport of F-KpOmpA across the inner membrane. Upon cleavage of this signal peptide in the periplasmic space, the mature protein chain (starting with **GCT** coding for Alanine, black box) is inserted and folded in the outer membrane. The gene terminates with the “ochre” stop codon (TAA, colored red). In order to avoid transport across the inner membrane and to accumulate the product as inclusion bodies in the cytoplasm (similarly to the N-KpOmpA construct, see section 1.2. of this chapter), the signal peptide had to be omitted. For this purpose, the primers for PCR amplification of the F-KpOmpA gene (with positions underlined on the gene sequence above) were designed as follows and manufactured by Sigma-Aldrich:

5'- GTGGCACTGCATATG**AAAGCTACCGTAGCGCAGGCC**GCTCCGAAAGATAAC -3'

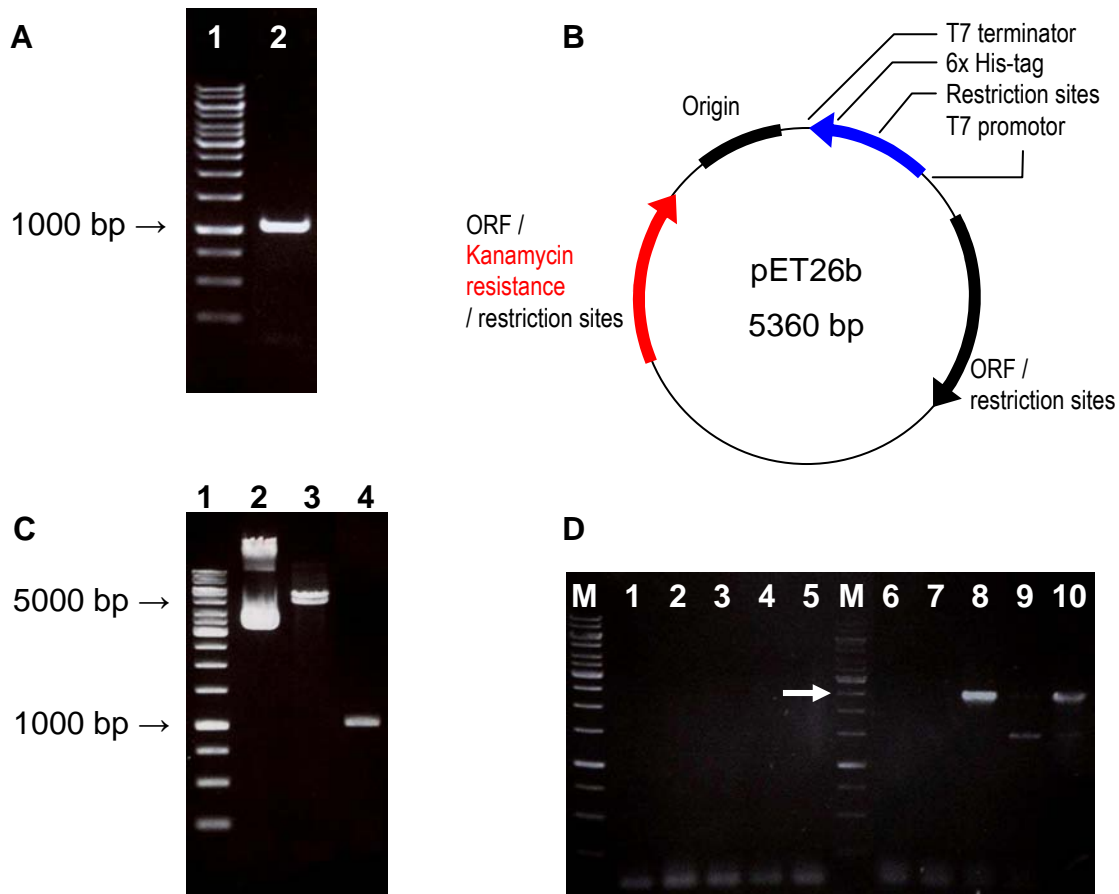
The 5' primer (*kpompa5'*), presented above, was used for the PCR reaction to introduce a unique NdeI restriction site containing the ATG start codon (5'-CATATG-3', underlined).

5'- GGTACCAGCTCGAGAGCCGCCGGCTGAGTTACAACTTCT -3'

The 3' primer (*kpompa3'*), presented above, was used to introduce a unique XhoI restriction site (5'-CTCGAG-3', underlined).

For PCR amplification of the target gene, 0.5 µl of *kpompa5'* (100 µM) and 0.5 µl of *kpompa3'* (100 µM) primers were mixed with 2 µl of *K. pneumoniae* genomic DNA (53 ng/µl), 1 µl dNTPs (10 mM mix, Promega), 10 µl 5x Phusion HF Buffer (Fynnzymes/Thermo Scientific), 35.5 µl Milli-Q water and 1 µl stock solution of Phusion Hot Start II High-Fidelity DNA Polymerase (2000 U/ml, Fynnzymes/Thermo Scientific). This 50-µl reaction mixture was then subjected to PCR (1000 Thermal Cycler, Bio-Rad) for amplification of the KpOmpA gene: 30 sec of initial denaturation step at 98°C, 40 cycles x 10 sec of denaturation (98°C) followed by 60 sec of combined annealing/extension step at 72°C, and a final extension step for 5 min at 72°C. The product of the PCR reaction is visualized on 1% agarose gel after staining with ethidium bromide (**Fig. 1.1A**), with the band around 1000 base pairs (bp) corresponding to the amplified KpOmpA gene-containing fragment (~350 a.a.).

40 µl of the PCR reaction was then subjected to centrifugation-based purification of the PCR-amplified gene using QIAquick PCR Purification kit (QIAGEN). The purified product at a concentration of 68 ng/µl (in 50 µl) was used for digestion with restriction enzymes and ligation in an empty vector, as described below.



**Fig. 1.1.** Representation of the major steps in the process of obtaining the F-KpOmpA construct. **(A)** PCR amplification of a KpOmpA gene-containing fragment from the *K. pneumoniae* genomic DNA: lane 1 - marker ladder (GeneRuler, Fermentas); lane 2 - PCR-amplified product. **(B)** Schematic image of the main components of the pET26b vector. **(C)** Purified and enzyme-digested pET26bII and KpOmpA amplified gene fragment: lane 1 - marker ladder; lane 2 - purified, non-digested pET26bII; lanes 3 and 4 - pET26bII and KpOmpA amplicon, respectively, both restricted with NdeI and XhoI enzymes. **(D)** PCR-based colony screen of empty pET26bII-transfected 'blank control' Top10 cells (lanes 1 to 5) and pET26bII-F-KpOmpA transfected Top10 cells (lanes 6 to 10); M - marker ladder. The arrow indicates the marker position at 2500 bp, which corresponds to the amplified [KpOmpA/kpompa5' - kanamycin resistance/3'-oligonucleotide] fragment from the vector.

The low-copy plasmid pET26bII was a kind gift from Wladimir Malaga (IPBS, Toulouse). It is derived from pET26b (**Fig. 1.1B**) and operates under the T7 promoter, contains multiple cloning sites (including such for restriction with NdeI and XhoI enzymes), a Thrombin-cleavable C-terminal hexahistidine tag for purification purposes and provides resistance to kanamycin. The modification of pET26bII is found as an introduced Thrombin cleavage site between the native protein chain and the C-terminal His-tag. *E. coli* Top10 cells, transfected

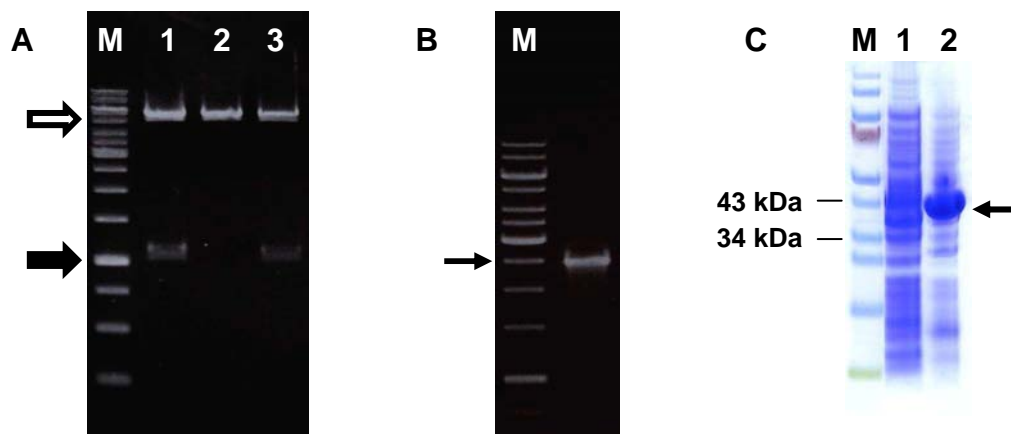
with pET26bII and deep-frozen, were thawed and grown in 50 ml Luria-Bertani (LB) medium supplemented with 30 mg/L kanamycin (Sigma Aldrich) at 37°C on a shaker (240 rpm). A small aliquot of this pre-culture was inoculated in 50 ml of fresh medium and left in the same conditions overnight. After that, the bacteria were collected (5000 g, 10 min, 4°C) and put on ice. Plasmid purification was performed according to the MidiPrep commercial protocol (Macherey-Nagel) with a NucleoBond AX 100 column. The purified plasmid at concentration of 285 ng/μl was stored at -20°C until further use.

The purified vector and KpOmpA gene-containing product were then digested with the restriction enzymes NdeI and XhoI (20 000 U/ml, New England Biolabs). 1 μl of each enzyme stock solution was mixed with 10 μl Milli-Q water, 2.5 μl 10x Buffer 4 (New England Biolabs) and 10 μl of either purified pET-26bII or KpOmpA amplified gene product. These 25-μl reactions were kept at 37°C for 30 min, after which another 1 μl from each enzyme stock was added for another 30 min. The reaction was then heat inactivated (10 min at 65°C) and a small aliquot from it was visualized on an agarose gel (**Fig. 1.1C**). Note the change in the migration of the non-digested vector (lane 2) exhibiting a variety of conformations, in comparison with the restricted pET26bII (lane 3) in which case certain populations (such as the coil and supercoil formations) are destroyed. As expected, the KpOmpA amplicon does not show a behavior different from that seen on **Fig. 1.1A**, since the restriction cleaves only small pieces from the termini of its linear DNA.

For ligation purposes, 10 μl of the KpOmpA gene restriction reaction were mixed with 5 μl of the plasmid restriction reaction, 2 μl Milli-Q water, 2 μl 10x T4 buffer (500 mM Tris (pH 7.5), 100 mM MgCl<sub>2</sub>, 100 mM DTT, 10 mM ATP, New England Biolabs) and 1 μl T4 DNA ligase stock solution (400 000 U/ml, New England Biolabs). Another such 20-μl reaction, but with 10 μl of water instead of KpOmpA restricted fragment was used as a 'blank control' for the cell transfection at the next stage. The ligation reaction was left for 30 min at 16°C, after which the enzyme was heat inactivated for 10 min at 65°C. The ligation reaction was then precipitated with ethanol and the dried pellet was covered with a 50 μl of electro-competent *E. coli* Top10 cells. The mixture of ligated plasmid and bacterial suspension was transferred to an electroporation cuvette with 2 mm space between the electrodes (Eurogentec) and subjected to a single pulse of 2.5 kV corresponding to 12.5 kV/cm (Micropulser, Bio-Rad). The cell suspension was then diluted with 900 μl LB medium and the cells were left for adaptation for 1h at 37°C, after which they were spread on petri dishes of LB medium with 1.5% agar and 30 mg/L kanamycin for selection of positive clones. The cells transformed with the 'blank control' were treated in exactly the same way. After 12h, colonies were found on



both plates, albeit in smaller number on the 'blank control' (data not shown). Five colonies from each plate were screened for the KpOmpA product using Dream Taq Polymerase (Fermentas), the forward *kpompa5'* primer and a commercial 3' primer for kanamycin (Sigma Aldrich, TGCGACAATCTATCGCTTGATGGGAA). Using the *kpompa3'* primer for colony screening purposes is not advisable, since the *E. coli* genome contains an *ompA* gene of its own. Due to the high level of identity between KpOmpA and *E. coli* OmpA not only on the protein (~85%) but also on the gene level (~84%), our KpOmpA-designed primers may happen to satisfy the native *ompA* gene as well, which will lead to false positive results in every screened colony and will prevent the separation of the true positive ones from the empty vector self-ligation artifacts. The results from this screening (**Fig. 1.1D**) show no positive hits in the 'blank control' cells, as expected, and positive hits in two (lanes 8 and 10) of the five screened colonies transfected with the pET26bII-F-KpOmpA ligation reaction. One other colony (lane 9) did not produce a convincing result, but nevertheless was further analyzed together with the other two. The three colonies (denoted as C8, C9 and C10) were grown in 5-ml LB cultures supplemented with 30 mg/L kanamycin, collected and the pET26bII-F-KpOmpA plasmid isolated according to the QIAprep Spin Miniprep kit (QIAGEN).



**Fig. 1.2.** Validation of the pET26bII-F-KpOmpA construct and first protein expression. **(A)** Enzyme restriction maps (with NdeI and XhoI) of the plasmids purified from colonies C8, C9 and C10 (lanes 1, 2 and 3, respectively; see text for details). Restrictions in C8 and C10 separated the KpOmpA gene fragment (~1000 bp, black arrow) from the rest of the linearized pET26bII (~5000 bp, white arrow). Note the missing KpOmpA gene in the case of C9. Compare with lanes 8, 9 and 10 on **Fig. 1.1D**. M - marker ladder. **(B)** Positive PCR screen (right lane) of the C8 BL21 (DE3) colony with the *kpompa5'* primer a 3' primer for the kanamycin-resistance gene. The arrow indicates the band at 2500 bp of the marker ladder (M, left lane). **(C)** Coomassie stained SDS PAGE gel of the same colony as in (B), grown in LB medium without induction (lane 1) and induced with 1 mM IPTG (lane 2). The marker ladder (M, PageRuler, Fermentas) denotes the bands with 43 and 34 kDa molecular weight and the arrow indicates the overexpressed F-KpOmpA (~38.7 kDa).

The isolated plasmids from the three colonies were then subjected to digestion with NdeI and XhoI and the corresponding restriction maps were visualized on 1% agarose gel (**Fig. 1.2A**). The purified product from colony C8 was then chosen for transfection in *E. coli* BL21 (DE3) cells, which was performed as with the Top10 strain. On the next day, five colonies from the kanamycin selective petri dish were screened as before for the presence of pET26bII-F-KpOmpA product, as a final precaution step before the pilot protein production. Surprisingly, not all colonies turned out to be positive, which is an example for the limited reliability of the PCR reaction as an ultimate visualization tool for the presence of a ligated product. Upon induction with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), all of the five screened BL21 (DE3) colonies (even those that appeared negative on the PCR colony screen) overexpressed the F-KpOmpA protein to a substantial level. An example of a PCR-positive BL21 (DE3) colony is shown on **Fig. 1.2B** and the F-KpOmpA produced by the same colony after culture growth and induction with IPTG is shown on **Fig. 1.2C**.

The composition of the pET26bII-F-KpOmpA construct was verified by sequencing (MilleGen). The expressed protein chain with a total of 359 a.a. (~38.7 kDa) thus corresponds to the following sequence:

MKATVAQAAPKDNTWYAGGKLGWSQYHDTGFYGNQFQNNNGPTRNDQLGAGAFGGYQVNPYLGFMGYD  
 WLGRMAYKGSVDNGAFKAQGVQLTAKLGYPIITDDLDIYTRLGGMVWRADSKGNASTGVSRSSEHDTGVS  
 PVFAGGVEWAVTRDIATRLEYQVWNNIGDAGTVGTRPDNGMLSLGVSYRFGQEDAAPVVAPAPAPAPE  
 VATKHFTLKSDVLFNFKATLKPEGQQALDQLYTLQLSNMDPKDGSVVVLGYTDTRIGSEAYNQQLSEKRA  
 QSVVDYLVAKGIPAGKISARGMGESTPVTGNTCDNVKARAALIDCLAPDRRVEIEVKGYKEVVTQPAAL  
**ELVPRGS***VEHHHHHH*

A short section from the native signal peptide is found after the first Methionine in the beginning of the chain (underlined). This fragment does not carry sufficient information for transportation across the inner membrane, therefore the protein accumulates in inclusion bodies (similarly to the N-KpOmpA construct). The black box in the middle of the sequence denotes the flexible, Alanine/Proline-rich region that connects the two domains of the molecule. Finally, there is a hexahistidine tag at the end of the sequence (in italic) and a thrombin cleavage site (LVPRGS, in bold) that allows eventual removal of the His-tag.

*Note: Annex 1 provides detailed protocols for this section of the chapter.*

## 1.2. Expression, refolding and purification of N-KpOmpA and F-KpOmpA

Since the N-KpOmpA and F-KpOmpA constructs do not express the signal peptide in the beginning of their protein sequences, both products accumulate in the bacterial cytoplasm in the form of inclusion bodies. In general, the purification procedure does not vary significantly between the two constructs or as a function of the cell culture medium used (rich or minimal, for details see Annex 2).

The construction of the pET26bII-F-KpOmpA plasmid (kanamycin resistance) was described in the previous section of this chapter. The construction of the N-KpOmpA (216 a.a.) fragment incorporated in pET21c vector was performed previously [1], resulting in a pET21c-N-KpOmpA plasmid (ampicillin resistance) containing short modification at the protein N-terminus (ARIMKAIFVLNA) and a hexahistidine tag at the C-terminus, immediately after the flexible hinge region of the molecule that connects its two domains.

Each plasmid is harbored by the *E. coli* BL21 (DE3) strain, frozen aliquots of which are kept at  $-80^{\circ}\text{C}$ . Upon cells collection from a small scale (50 ml) rich medium pre-culture supplemented with the respective antibiotic, a large scale culture (with the same antibiotic) is inoculated at optical density (OD) of around 0.1 (measured at 600 nm). The antibiotic concentrations are 50 mg/L ampicillin (Sigma Aldrich) for N-KpOmpA expression and 30 mg/L kanamycin (Sigma Aldrich) for F-KpOmpA. The expression of the respective protein product can then be induced with 1 mM IPTG during the logarithmic growth phase ( $\text{OD}_{600} = 0.6-0.8$ ) of the bacteria. **Fig. 1.3A** shows an example of a culture growth curve (in Terrific broth (TB) medium, Invitrogen) and the moment of induction of the protein synthesis. After that, the culture is further left for 4 hours or overnight (depending on the used culture medium) at  $37^{\circ}\text{C}$  with vigorous shaking. The cells are then harvested (5000 g, 10 min,  $4^{\circ}\text{C}$ ), resuspended (at a ratio of 10 ml per gram of cells) in lysis buffer (50 mM Tris (pH 8.5), 200 mM NaCl, 5 mM EDTA and freshly added 0.2% Tween 20, Sigma Aldrich and 0.5 mg/ml Lysozyme, Fluka Analytical) and left for 2 hours at room temperature with gentle mixing. The lysate is then sonicated (6 times x 1 min, 50% duty cycle, power 5, VWR) and the released inclusion bodies are collected (10 000 g, 30 min,  $4^{\circ}\text{C}$ ) and washed three times with the same lysis buffer containing 2M urea. The final inclusion bodies pellet is solubilized (at a ratio of 3-4 ml per gram of pellet) either in 7M urea with freshly added 10 mM DTT, or in 6M Guanidine-HCl buffer with 25 mM Tris (pH 8.3), 5 mM EDTA and freshly added 10 mM DTT. The pellet is allowed enough time to solubilize gently at room temperature and, after removal of the non-soluble fraction (10 000 g, 30 min,  $4^{\circ}\text{C}$ ), the cleared supernatant is mixed dropwise with 15

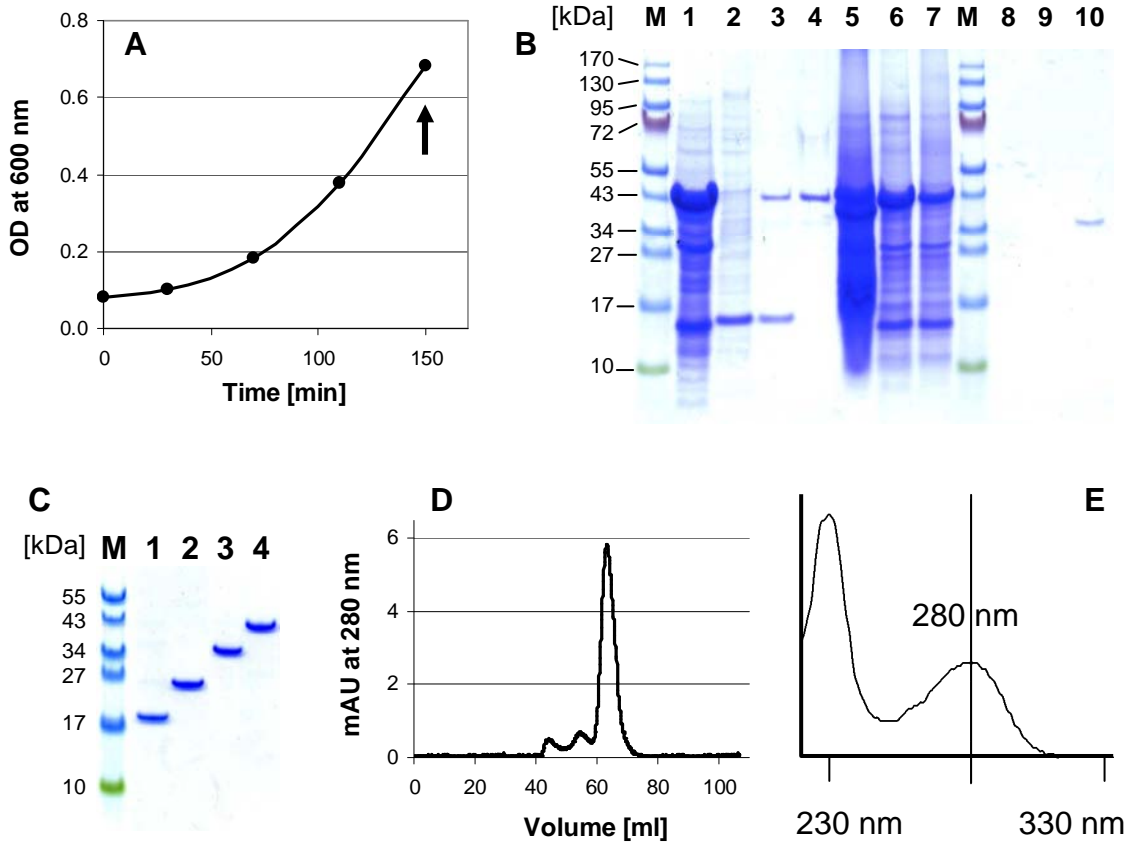
times larger volume of the refolding detergent-containing buffer (1% Zwittergent 3-14 (n-tetradecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, Anatrace) with critical micelle concentration (CMC) of 0.015%, 25 mM Tris (pH 8.5) and 150 mM NaCl). This mixture is left at 4°C for at least 4 hours (or overnight) of gentle mixing to allow for protein refolding. It is then centrifuged (10 000 g, 30 min, 4°C) in order to remove protein aggregates and it is dialyzed (Cellulose membranes with 12-14 kDa cutoff, 25 mm flat width, 16 mm diameter, 2 ml/cm, Spectra/Por) two times (2L each) for 12 hours against the same buffer with 10 times lower detergent concentration (0.1% Zwittergent 3-14) in order to remove the chaotropic agent and to decrease the detergent concentration. The higher detergent concentration (1%) facilitates the refolding, but it is unnecessary for further support of the folded protein, as seen on the SDS PAGE gel on **Fig. 1.3B,C** [4]. After the dialysis, the protein concentration is determined by measuring the absorption at 280 nm and using the theoretical molar extinction coefficient of each construct ( $50880 \text{ M}^{-1}\text{cm}^{-1}$  for N-KpOmpA and  $58455 \text{ M}^{-1}\text{cm}^{-1}$  for F-KpOmpA; coefficients calculated with *www.biomol.net*). The protein is then purified via affinity chromatography with discontinuous imidazole gradient. The dialyzed protein solution is mixed with water-washed nickel-coordinated resin (Ni-NTA Superflow, QIAGEN) at the ratio of ~5 mg protein per 1 ml of resin (equivalent to 2 ml of slurry) and left for 2-3 hours at 4°C with gentle mixing. The protein-bound resin was then separated from the supernatant (i.e. the flowthrough) and washed with at least 5 column volumes (CV) of 0.1% Zwittergent 3-14 buffer followed by further washing with 5 CV of the same buffer containing 20 mM imidazole (98.5% purity, Sigma-Aldrich) in order to remove eventual impurities and weakly bound misfolded and/or aggregated protein. Finally, the protein is detached from the resin with the same detergent buffer containing 400 mM imidazole. This eluate is then dialyzed two times (2L each) for 12 hours against the detergent buffer (0.1% Zwittergent 3-14, 25 mM Tris (pH 8.5), 150 mM NaCl) in order to remove the imidazole. The protein concentration is checked as before and the folded state of the protein is visualized on an SDS PAGE gel (**Fig. 1.3C**). The protein solution is then kept at 4°C until needed. The described above nickel-affinity purification steps (washings and elution) were performed either on a single-use column (Wizard Plus Midipreps, Promega) or with gentle centrifugations of the resin (3000 g, 10 min, 4°C), depending on the scale of the purification. In most of the cases, further concentration of the protein stock solution was not required. In the exceptional cases (such as after gel filtration, performed following the affinity chromatography) when the protein had to be concentrated, this was achieved by centrifuging (8000 g, 4°C) the protein/detergent solution in membrane falcons with 50 kDa cutoff (Vivaspin, Sartorius Group) until the desired

concentration was reached. Since this procedure inevitably leads to (at least partial) co-concentration of the detergent as well, the final solution was again dialyzed as described above in order to maintain the exact Zwittergent 3-14 concentration of 0.1%. Note that the membrane cutoff (50 kDa) is in fact larger than the size of each protein construct. However, the protein-detergent micelle size was previously found to be around 70 kDa for N-KpOmpA [1]. Indeed, a loss of neither N-KpOmpA (~23.4 kDa) nor F-KpOmpA (~38.7 kDa) was ever observed after concentration. The above described expression/purification procedure can be found as step-by-step protocol in Annex 2. The average protein yield after completion of this protocol is around 200 or 140 mg of protein per 1L of rich or minimal medium, respectively. The variation between the yields (in terms of protein mass) of the two constructs (N-KpOmpA and F-KpOmpA) obtained in the same culture medium is smaller than the variation between the different media used.

Representative aliquots from the major purification steps of N-KpOmpA were run on a 10% SDS PAGE gel and are shown on **Fig. 1.3B**. The OmpA proteins (and the  $\beta$ -barrel proteins in general) possess a specific feature called “heat modifiability”, which is a result from the high stability of their molecular structures. Since the barrel fold is not destroyed during the electrophoresis in denaturing conditions, but only if the gel aliquot was heated in advance (100°C, 5 min), this phenomenon is manifested as different migration positions of the folded and the unfolded protein bands on the gel [4]. Until present day, this Omp feature is one of the fastest, most commonly used assays for monitoring the folding of this class of proteins [5-7]. For example, when not boiled, the purified N-KpOmpA migrates at around 19 kDa (**Fig. 1.3C**), while the thermally denatured aliquot shifts to its expected molecular weight (~23 kDa). Respectively, native F-KpOmpA appears at around 34 kDa, while the true size position for this protein is seen only after boiling it (~39 kDa). Only in the case of F-KpOmpA, boiling the protein solution for longer times (> 5 min) and/or applying larger amount of boiled protein on the gel produces two additional weak bands, found around 24 and 15 kDa (not present on **Fig. 1.3C**). These two bands correspond to the presence of the DP-motif in the sequence of the periplasmic domain, which is significantly labile to acid-catalyzed hydrolysis ([8, 9]).

In addition to the visualization of the protein fold by gel electrophoresis, another suitable method for supplementing this aim is the size-exclusion chromatography. Furthermore, it can be used not only as analytical tool but also as a large scale purification step for elimination of protein aggregates, such as non-native dimer/trimer formations and unfolded molecules. **Fig. 1.3D** shows analytical run (1 ml/min, 60-cm column, S200, GE Healthcare) of 100  $\mu$ l purified

F-KpOmpA solution (1 mg/ml) in 0.1% Zwittergent 3-14, exhibiting a main symmetrical peak of the monomer population. The two smaller subpopulations of larger sizes correspond to non-native dimer and trimer formations. These subpopulations appear difficult to separate from the main peak if the column length is only 30 cm (data not shown). After gel filtration, the



**Fig. 1.3.** Summary of the major purification steps and preliminary protein analyses. **(A)** *E. coli* BL21 (DE3) culture growth curve in Terrific Broth medium, presented as optical density (OD) at 600 nm as a function of time. The arrow indicates the moment of protein overexpression induction with 1 mM IPTG. **(B)** SDS PAGE gel of the major purification steps (in this case - of F-KpOmpA). The induced cell culture (lane 1) is lysed as described in the text and the released inclusion bodies are washed three times (lanes 2-4; note the low molecular band of the lysozyme around 17 kDa and a certain loss of unfolded F-KpOmpA around 43 kDa). The small amount of pellet after inclusion bodies solubilization with chaotropic agent (lane 5), as well as the protein-aggregation pellets after protein refolding (lane 6) and dialysis (lane 7) exhibit a minor loss of F-KpOmpA. The flowthrough (lane 8) and the washing with imidazole-free buffer (lane 9) during the affinity chromatography do not contain any F-KpOmpA, while a small loss of folded protein is observed during the washing step with 20 mM imidazole (lane 10, around 34 kDa); M - marker ladder. **(C)** "Heat modifiability" (see text for details) of purified N-KpOmpA (lane 1: not boiled; lane 2: boiled) and F-KpOmpA (lane 3: not boiled; lane 4: boiled). M - marker ladder. **(D)** Gel filtration profile (1 ml/min, 60 cm column, S200, GE Healthcare) of 100  $\mu$ l purified F-KpOmpA (1 mg/ml) in 0.1% Zwittergent 3-14. **(E)** UV spectrum (220-330 nm) of purified N-KpOmpA with absorption value of 31.963 at 280 nm (measured with NanoDrop ND-1000 Spectrophotometer, Thermo Scientific), corresponding to protein concentration of 14.7 mg/ml.

fractions containing the separated in this way monomer are combined and concentrated in membrane falcon, followed by dialysis against 0.1% Zwittergent 3-14 buffer. The concentration of this protein stock solution is then determined again, after this final stage of the purification.

Since the  $\beta$ -barrel (present in both constructs) is rich in aromatic residues, the spectrophotometric determination of the protein concentration (at 280 nm) is particularly useful (**Fig. 1.3E**). The alternative of the colorimetry-based assays is not only more tedious and time consuming, but it also utilizes soluble proteins of known concentration as standards, which may introduce additional discrepancy when the target of interest is in fact a detergent-solubilized membrane protein which, in addition to that, does not unfold spontaneously in 1% SDS. For example, comparing the concentrations of three independently produced N-KpOmpA stock solutions (two uniformly- $^{13}\text{C}$ - $^{15}\text{N}$  labeled stocks and one non-labeled/natural abundance (NA) stock), measured photometrically and with the standard protein titration assay according to Lowry [10], revealed a significant discrepancy among the obtained concentration values with a tendency of the colorimetry to overestimate them up to 50%, as shown below:

N-KpOmpA concentration determination of three independent stocks:

<u>Protein stock</u>	<u>UV estimation</u>	<u>Colorimetric estimation</u>
[U- $^{13}\text{C}$ / $^{15}\text{N}$ ]	20 mg/ml	29 mg/ml
[U- $^{13}\text{C}$ / $^{15}\text{N}$ ]	17 mg/ml	24 mg/ml
Non-labeled (NA)	20 mg/ml	30 mg/ml

Colorimetric titration of 'blank samples' containing only the standard bovine serum albumin (BSA), BSA in 0.1% Zwittergent 3-14 and BSA in 0.1% Zwittergent 3-14 and 1% SDS did not show any significant difference (data not shown), which excluded a possible influence of the detergent. On the other hand, the molar extinction coefficients of N-KpOmpA and F-KpOmpA are theoretically derived and therefore not ideal. Furthermore, they do not provide a notion for the influence of the protein fold and of the different detergent environments on the measured absorption value. Nevertheless, the photometric approach appears to combine ease with enough precision and, as long as one follows a single method for protein concentration determination, the accuracy of the values should suffice. The future KpOmpA investigator is thus advised to take the above into account when determining the concentrations of N- and F-KpOmpA.

After the purification, the spectrophotometrically determined protein concentration is usually in the range of 5-10 mg/ml (depending on the fractions size and number), with higher values obtainable after concentration in membrane falcon (50 kDa cutoff). In order to limit the gradual aggregation of the protein and to extend the lifetime of the stock solution, it is advisable to preserve long-terms (~1 month) KpOmpA solutions at lower protein concentration (1-3 mg/ml) at 4°C, if not flash-frozen. The propensity for macroscopically observed aggregation appears to be higher in the case of the membrane domain alone (N-KpOmpA), presumably because a larger portion of the molecule of the full-length polypeptide is soluble. The long-term stability of the protein stock is also dependent on the detergent type. Zwittergent 3-14 (M.W. 363.6) is a zwitterionic detergent with low critical micelle concentration (CMC 0.012% = 0.33 mM), suitable for large-scale purification of KpOmpA which requires hundreds of milliliters from the different detergent-containing buffers. Assuming a concentration of N-KpOmpA in the range of 10 mg/ml (~0.43 mM) present in a buffer with 0.1% (2.75 mM) Zwittergent 3-14 (~10 times above its CMC), gives a ratio of about only 6 detergent molecules per protein molecule. It is clear that such amount of detergent can not support the  $\beta$ -barrel. On the other hand, the same solution can be concentrated by centrifugation without any loss through the 50 kDa cutoff of the membrane, even if the mass of N-KpOmpA is ~23.4 kDa. This indicates that the protein-detergent micelle is composed of, approximately, at least an order of magnitude higher number of detergent molecules, despite the 0.1% concentration of detergent in the surrounding solution. This difference is the result of the refolding and dialysis steps of the purification procedure, during which the protein-detergent micelles form in the presence of a large excess of zwittergent molecules. It is thus possible to maintain the  $\beta$ -barrel fold in such rather low (apparent) detergent concentration at a later stage. Since the detergent molecules are not 'bound' to the protein, but are in dynamic exchange between the protein micelle and its surroundings, then the ability of the detergent solution to support the monomeric protein for a certain time, without allowing the protein molecules themselves to collide into one another, will strongly depend on the protein concentration itself. The hence described problem of protein precipitation related to the protein-to-detergent ratio can be easily avoided by using alternative, high-CMC detergents for further analysis of KpOmpA. For instance, n-octyl- $\beta$ -D-glucopyranoside (M.W. 292.4, CMC 0.7% = 25 mM) was successfully used with *E. coli* OmpA ([11]) and here it was utilized for the reconstitution of KpOmpA in lipid bilayers, as described in section 1.4 of this chapter.

*Note: Annex 2 provides detailed protocols for this section of the chapter.*



### 1.3. Gene design, expression and purification of C-KpOmpA

Natively, after cleavage of its signal peptide (20 a.a.) and integration in the outer membrane, KpOmpA forms a two-domain structure with a transmembrane  $\beta$ -barrel domain (N-KpOmpA, 180 a.a.), a flexible periplasmic Proline/Alanine-rich hinge region (GQEDAAPVVAPAPAPAPE, 18 a.a.) and a C-terminal periplasmic domain (C-KpOmpA, 137 a.a.). Hence a fraction of about 41% of the protein chain belongs to C-KpOmpA. This domain is 44% homologous (and ~35% identical) to the C-terminal domain of RmpM ([12], crystal structure available with PDB: 1R1M) - a putative peptidoglycan binding protein from *Neisseria meningitidis* that adopts a  $\beta\alpha\beta\alpha\beta$ -fold, as in the case of the chorismate mutase from *Bacillus subtilis*. In addition to the expected properties of these domains to bind non-covalently to the peptidoglycan layer in the periplasm, a role of a chaperone was proposed for C-KpOmpA [13]. Furthermore, a dimerization process was suggested for the C-terminal domain of RmpM for more efficient binding to peptidoglycan - a possibility that exists also in the case of C-KpOmpA, despite that the  $\beta$ -barrel domain of the native chain is exclusively found as a monomer in the membrane. Construction and expression of the C-terminal domain of KpOmpA could therefore be used in the future for structure determination of this soluble segment of the molecule and for investigating its properties, either alone or as a part of F-KpOmpA.

The hinge region of KpOmpA (immediately after the C-terminus of the  $\beta$ -barrel) is expected to be unstructured and to provide a flexible link between the two domains of the protein. Therefore this segment could be excluded from the N-terminal part of the C-KpOmpA construct. Instead, this terminus was designed to contain a hexahistidine tag, followed by a TEV protease recognition site (EXXYXQ–G/S, with X being any amino acid and the dash representing the actual cleavage point) for eventual His-tag removal after the protein expression. The rest of the C-terminal part of the construct was preserved as found in the native gene. The thus designed construct is shown below. The sequence starts with ATG for Methionine (red color) and three additional amino acids (GSS, blue color), followed by a hexahistidine tag (black box), another three linking amino acids (SSG, green color) and a TEV protease restriction site (ENLYFQGS, underlined). The native protein sequence (bold letters) starts after that and terminates with the “ochre” stop codon at the 3' end (TAA, red color). The nucleotide motif CCATGG at the very beginning represents an NcoI restriction site which was introduced for potential restriction and ligation in another vector.

M G S S H H H H H H S S G

[5']ACCATGGGCAGCAGCCATCATCATCATCATCAAGCAGCGGCGAGAATCTTTATTTTCAGGGATCC  
**GAAGTGGCTACCAAGCACTTCACCCCTGAAGTCTGACGTTCTGTTCAACTTCAACAAAGCTACCCCTGAAA**  
**CCGGAAGGTCAGCAGGCTCTGGATCAGCTGTACACTCAGCTGAGCAACATGGATCCGAAAGACGGTTCC**  
**GCTGTTGTTCTGGGCTACACCGACCGCATCGGTTCCGAAGCTTACAACCAGCAGCTGTCTGAGAAACGT**  
**GCTCAGTCCGTTGTTGACTACCTGGTTGCTAAAGGCATCCCGGCTGGCAAATCTCCGCTCGCGGCATG**  
**GGTGAATCCAACCCGGTTACTGGCAACACCTGTGACAACGTGAAAGCTCGCGCTGCCCTGATCGATTGC**  
**CTGGCTCCGGATCGTCGTGTAGAGATCGAAGTTAAAGGCTACAAAGAAGTTGTAACCTCAGCCGGCGGCT**  
TAA[3']

The designed gene was then sent to DNA2.0 ([www.dna20.com](http://www.dna20.com)) for sequence optimization, gene synthesis, plasmid integration and sequence verification. The C-KpOmpA gene was optimized for *E. coli* preferred codons and synthesized with a consensus ribosome-binding site (AGGAGGT) and a short spacer of five adenines at the 5' end. The fragment was then integrated in the vector pJexpress411 (DNA2.0, **Fig. 1.4A**). This vector offers a T7 promoter (for expression in strains such as *E. coli* BL21 (DE3) or T7 express) and resistance to kanamycin. The final gene sequence (500 bp) in the plasmid pJexpress411-C-KpOmpA is thus:

[5']AGGAGGTAAAAACCATGGGCAGCTCTCACCACCATCATCACCACAGCAGCGGCGAGAACTTGTATT  
TTCAGGGATCCGAGGTTGCGACGAAGCATTT**CACGCTGAAGAGCGACGTGCTGTTTA**ACTT**CAATAAAG**  
**CAACCCTGAAACCGGAAGGTCAACAGGCGCTGGATCAGCTGTATACGCAATTGAGCAATATGGACCCGA**  
**AGGATGGTTCGGCCGTTGTTCTGGGTTACACCGATCGTATCGGCAGCGAAGCTTATAACCAACAGCTGA**  
**GCGAGAAACGCGCGCAATCCGTCGTTGATTACCTGGTCGCGAAAGGTATTCCGGCTGGCAAGATCAGCG**  
**CCCGTGGTATGGGTGAGAGCAATCCGGTGACCGGCAACACCTGTGACAATGTCAAAGCACGTGCGGCGC**  
**TGATTGACTGCCTGGCACCGGACCGCGTGTGGAAATCGAAGTGAAGGGTTACAAAGAGGTTGTCACTC**  
**AGCCAGCCGCA**TAACTCGAG[3']

The ribosome-binding site is indicated with a blue color at the 5' end, while the start (ATG) and stop (TAA) codons are in red color. The 6xHis-tag (black box) and the TEV protease cleavage site (underlined) are found in the beginning of the sequence. The native C-KpOmpA chain (~15 kDa, 138 a.a.) is denoted in bold letters. The corresponding expressed polypeptide sequence from this construct (159 a.a., 17 300 Da) is presented below:

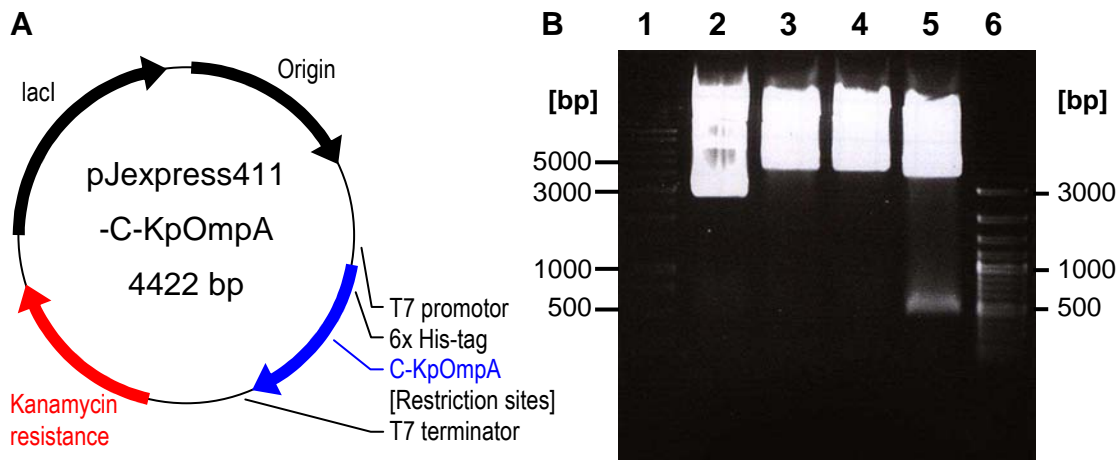
```

1  MGSS[HHHHHH] SSGENLYFQG SEVATKHFTL KSDVLFNFNK ATLKPEGQQA
51  LDQLYTQLSN MDPKDGSVVV LGYTDRIGSE AYNQQLSEKR AQSVDYDLVA
101 KGIPAGKISA RGMGESNPVT GNTCDNVKAR AALIDCLAPD RRVEIEVKGY
151 KEVVTQPAA

```

The N-terminal 6xHis-tag is indicated with a black box and the removable piece of the TEV protease site is underlined. If digested with this enzyme, the polypeptide will be shortened down to 140 a.a. and only two non-native amino acids (GS, denoted with bold letters) will be left at the N-terminus of the protein.

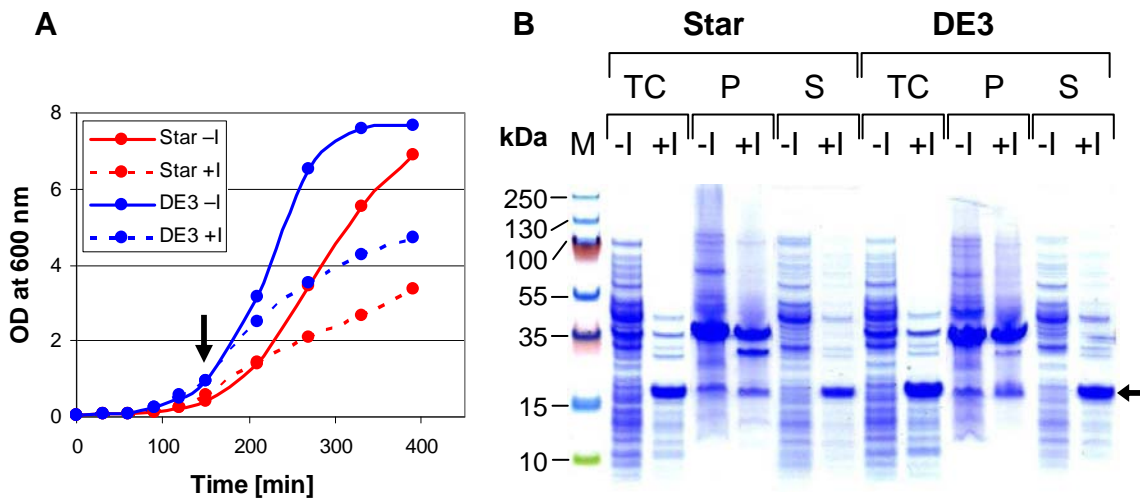
The product pJexpress411-C-KpOmpA was received as a lyophilized pellet and solubilized in sterile Milli-Q water. The plasmid concentration was spectrophotometrically estimated to be 0.8 µg/µl. Restriction map of the vector was obtained with digestion of 2 µl pJexpress411-C-KpOmpA stock solution either with 1 µl NcoI alone, or with 1 µl XhoI alone (New England Biolabs), or with both enzymes (**Fig. 1.4B**). Note the C-KpOmpA gene (~500 bp) appearing in lane 5.



**Fig. 1.4.** (A) Schematic picture of the pJexpress411-C-KpOmpA plasmid (DNA2.0). (B) Restriction map of the product obtained with NcoI and/or XhoI enzyme digestion: lane 1 - marker ladder (1 kbp DNA Ladder, Promega); lane 2 - not digested pJexpress411-C-KpOmpA plasmid; lanes 3 and 4 - the plasmid digested with NcoI or XhoI, respectively; lane 5 - the plasmid digested with both enzymes; lane 6 - marker ladder (GeneRuler 100 bp Plus DNA Ladder, Fermentas).

The plasmid pJexpress411-C-KpOmpA was then used for transfection via electroporation (12.5 kV/cm, with 2 µl of plasmid solution at 0.018 µg/µl) in two electro-

competent *E. coli* strains - BL21 Star and BL21 (DE3), hereby abbreviated as 'Star' and 'DE3', respectively. Each strain was spread on a petri dish with LB/agar medium supplemented with 30 mg/L kanamycin and left at 37°C overnight. Single colonies of each strain were picked, inoculated in 25-ml pre-cultures of LB/kanamycin medium and grown overnight at 37°C with vigorous shaking (280 rpm). Four 50-ml cultures of fresh LB/kanamycin medium were then prepared (two for each strain) and inoculated at OD (at 600 nm) of 0.05. The four flasks were incubated at 37°C (280 rpm) and the OD was regularly monitored. The DE3 strain exhibited faster growth rate and after a total culture time of 150 min the OD of its two flasks was in the range of 0.90, while the Star strain was just reaching an OD of 0.50 (Fig. 1.5A).



**Fig. 1.5.** Cultures growth and C-KpOmpA overexpression. **(A)** OD at 600 nm as a function of time for the non-induced (-I, solid lines) and induced (+I, punctured lines) cell cultures of Star (red color) and DE3 (blue color) strains. Note the reduced growth rate for the induced cultures. The black arrow indicates the moment of IPTG induction. **(B)** SDS PAGE image of non-induced (-I) and induced (+I) Star (left side) and DE3 (right side) *E. coli* strains as 'total cells' (TC), pellet fractions (P) and supernatant fractions (S) after the cell lysis and sonication (see text for the details). M - marker ladder. The arrow indicates the position of C-KpOmpA.

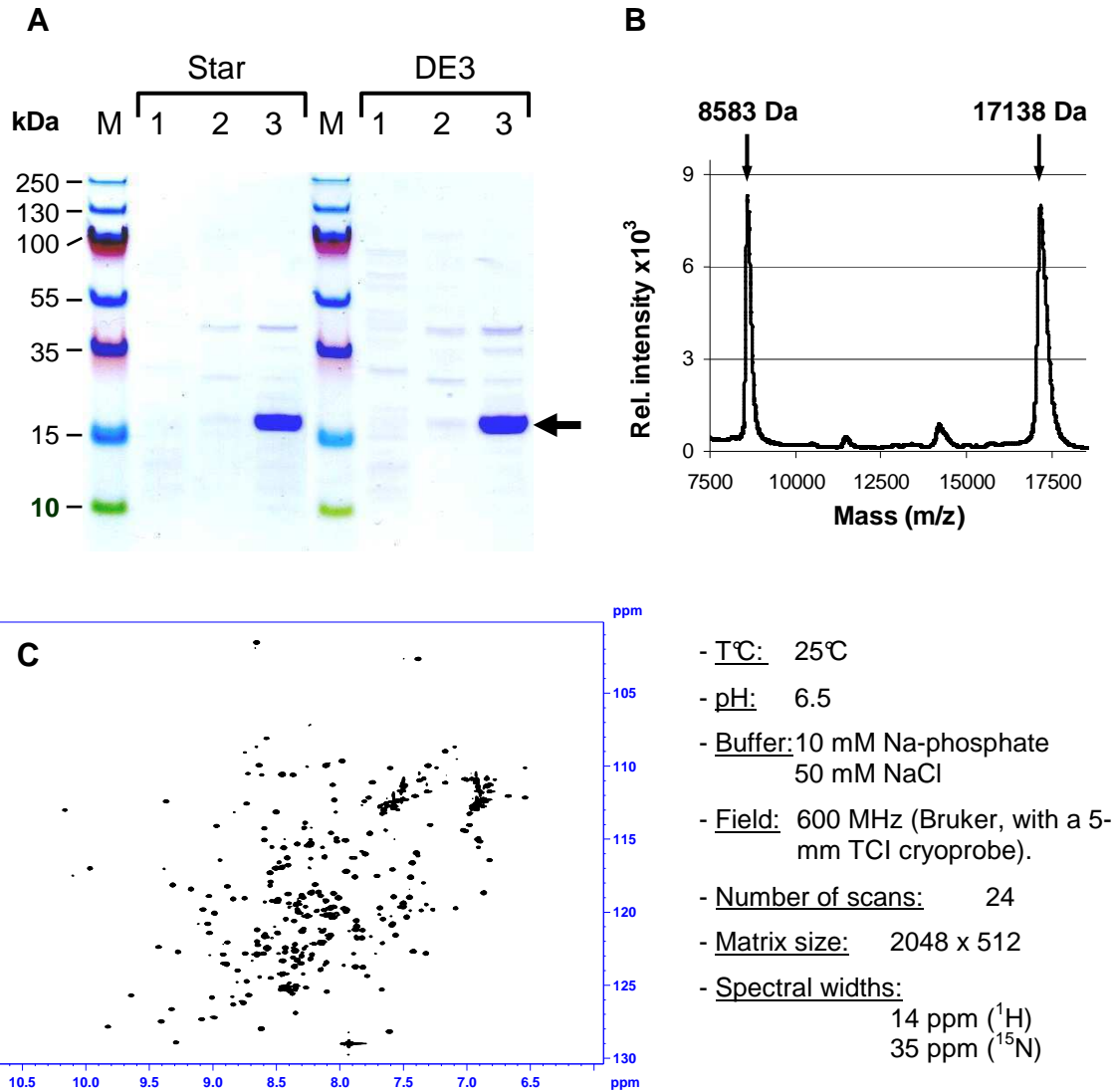
One 50-ml flask from each strain type was then induced with 1 mM IPTG and all flasks were further incubated for another 4 hours. Finally, different small-volume aliquots (of about 500 µl) from each of the four cultures were collected. These aliquots, representing 'total cell' fractions, were centrifuged (5000 g, 10 min 4°C) and prepared for direct loading on an SDS PAGE gel. The rest (~50 ml) of the four cell cultures were collected in the same way, the cell

pellets were resuspended in 10 ml lysis buffer (20 mM sodium phosphate (pH 8), 100 mM NaCl, 0.1% Tween 20) and subsequently sonicated (3 times x 20 sec, 50% active cycle, power 5, VWR). Upon centrifugation (10 000 g, 30 min, 4°C), the four supernatant were collected, while the four pellets were washed once with 12 ml of 20 mM Na-phosphate buffer (pH 8), 100 mM NaCl and centrifuged again. The four final pellets (together with the four supernatants and the four 'total cell' aliquots) were visualized on an SDS PAGE gel (**Fig. 1.5B**). The C-KpOmpA (~17.3 kDa) appears to be well overexpressed in each strain. Due to its slightly faster growth rate, the DE3 strain is presumably a better candidate for a large scale expression system. Note that the protein, as expected, is exclusively found in the soluble fractions from both strains, with no (or minimal) amount in the cell pellet.

The two supernatants from the induced Star and DE3 cultures (**Fig.1.5B**) were each mixed with 2 ml washed nickel-coordinated resin (Ni-NTA Superflow, QIAGEN) and left for 3 hours at 4°C with gentle mixing. After the batch, the resin was transferred to a single-use chromatography column and the flowthrough was collected. The protein-bound resin was then washed with about 10 ml 20 mM Na-phosphate buffer (pH 8), 100 mM NaCl, followed by 6 ml washing with the same buffer containing 20 mM imidazole and elution step with 3 ml 400 mM imidazole in the same buffer. Small aliquots from the preparation were then visualized on an SDS PAGE gel. As seen on **Fig. 1.6A**, the protein (~17.3 kDa) with a good purity is released during the elution step. The expected mass of the protein was confirmed by mass spectrometry (MALDI-TOF in linear measurement mode with  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) as matrix), as shown on **Fig. 1.6B**. Uniformly  $^{13}\text{C}/^{15}\text{N}$ -labeled C-KpOmpA (0.8 mM, following the same labeling protocol as for N-KpOmpA) produced a high-quality HSQC spectrum (**Fig. 1.6C**), confirming the protein fold and establishing a starting point for its structure determination.

The hereby established preliminary protocol (found in step-by-step fashion in Annex 3) for the C-KpOmpA purification can be used in future when overexpressing this protein construct. Furthermore, the protocol can be improved in several ways, such as by utilization of nucleases and protease inhibitors during the cell lysis step and afterwards, as well as eventual dialysis step of the cell lysate against clean phosphate/salt buffer prior to the purification stage. Two Cysteine residues are found on the chain of C-KpOmpA (C124 and C136. according to the numeration of the expressed construct). Natively, these residues are expected to form a disulfide bridge (similarly to the case of RmpM [12]) in the periplasmic space, after the unfolded polypeptide chain is transferred through the inner membrane. Since the artificially expressed protein is not exported from the bacterial cytoplasm, it is unclear

whether this bridge is actually formed. On the other hand, the importance of this putative bridge for the protein structure and function is also not known. These aspects of C-KpOmpA should be taken into account during the future work with the protein.



**Fig.1.6.** Visualization of the purified C-KpOmpA: **(A)** Purification of C-KpOmpA from supernatants of ruptured cells of *E. coli* strains BL21 Star (left side) and BL21 (DE3) (right side). Lanes 1 and 4: flowthroughs; lanes 2 and 5: washing steps with imidazole-free buffer; lanes 3 and 6: elution steps with 400 mM imidazole. M - marker ladder. The arrow indicates the position of C-KpOmpA (~17.3 kDa). **(B)** MALDI-TOF linear mode spectrum (matrix: CHCA), showing the main C-KpOmpA signal at ~17 138 Da (theoretical mass of 17 169 Da) and its 2<sup>+</sup> ion at ~8583 Da (theoretical mass of 8584 Da). **(C)** <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of 0.8 mM uniformly [<sup>13</sup>C/<sup>15</sup>N]-labeled C-KpOmpA (acquisition parameters are present to the right).

*Note: Annex 3 provides detailed protocols for this section of the chapter.*

#### 1.4. Reconstitution of N-KpOmpA and F-KpOmpA in lipid bilayers

Providing a hydrophobic environment is a necessary step in the process of membrane proteins purification since it allows the newly expressed molecules to adopt their native fold. In the majority of the cases this task is fulfilled by suitable types of detergent or, more recently, by amphipols ([14]). In other, more exceptional cases, the protein can be detergent-solubilized or co-inserted in preformed liposomes simultaneously with its *in vitro* expression ([15]). Studies on the kinetics of protein folding, on the other hand, rely on mixing the denatured protein with lipid vesicles and monitor the process of folding in time ([16]). Nevertheless, most of the membrane proteins isolations and characterizations, in particular those aimed at structural studies, still utilize a detergent-based folding of the protein of interest. The protein-detergent micelle is thus the basic form in which the protein is often found at the end of a purification/refolding procedure. This form is suitable for electrophoresis analyses and dynamic light scattering, for UV spectrometry and protein functional assays (provided the detergent does not interfere with the latter). Furthermore, due to its relatively small size and fast tumbling rate (in comparison with liposomes), the micellar form is usable for structural and dynamical studies by solution state NMR. This form is also the usual starting point for 3D crystallization of membrane proteins, although recent advancements in the field of the Lipid Cubic Phases offer an alternative ([17]).

However, natively the membrane proteins exist in the lipid bilayer of the cellular membrane. This environment differs from the detergent micelle in terms of several parameters, such as the membrane thickness, curvature, lipid composition and organization. In particular, the  $\beta$ -barrel proteins of Gram-negative bacteria (such as KpOmpA) are situated in oriented fashion in the highly asymmetric leaflets of the outer membrane. It is thus important to investigate, when possible, the features of these molecules in the native or close-to-native environment of the lipid bilayer. Often this is achieved by reconstitution of the protein of interest (upon detergent removal) in homogeneously distributed lipids that form the membranes of unilamellar vesicles. Although the genuine asymmetry of the bacterial outer membrane can not be recreated in this way, these bilayers are already a step closer in simulating the native environment of the protein, in comparison to the starting point of the detergent micelles. Furthermore, the results from an investigation of the protein in bilayers can later be compared with data from the micellar samples, thus revealing that key features of the polypeptide are either present in both environments, or lost in one of them, or gained. The  $\beta$ -barrel proteins are composed of rigid membrane-embedded barrel core and a set of

mobile extracellular loops. As shown previously with KpOmpA [1], these immunologically important long loops tend to be very dynamic when the protein is present in detergent micelles. One possible direction for investigation is thus monitoring the loops dynamics after protein reconstitution in lipid bilayers, which may provide insights on the structural basis for recognition of host-cell receptors in the process of KpOmpA-mediated *Klebsiella* invasion.

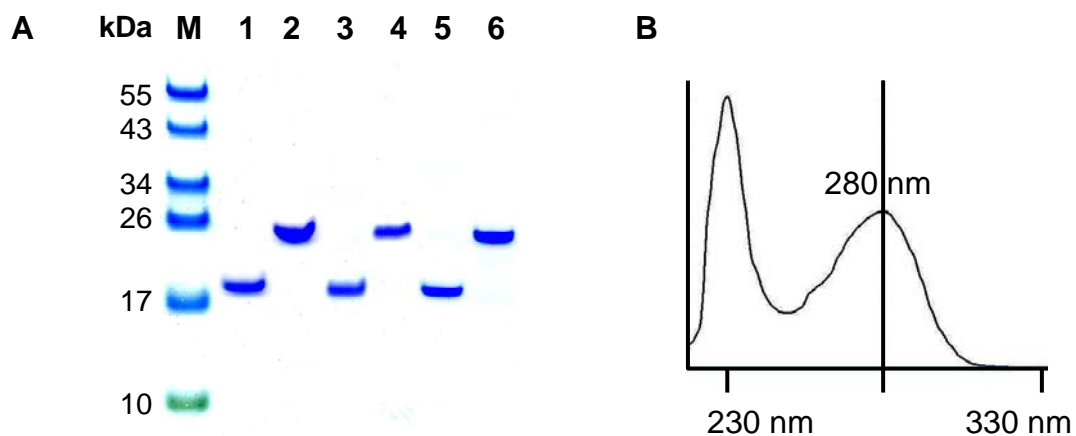
The abovementioned concept is dealt with more thoroughly in Chapter 3. Similarly, detailed description of the variety of small-scale reconstitution trials of N-KpOmpA in different lipids, lipid-to-protein ratios and dialysis buffers can also be found there. They do not differ essentially from the protocol described in the present section of this chapter. This section focuses on the procedure for KpOmpA reconstitution as such (in particular during the sample preparation for solid state NMR experiments), as well as on analyzing the prepared sample with several biochemical methods. Detailed protocol for the N- and F-KpOmpA reconstitution can be found in Annex 4.

The purified protein (either N- or F-KpOmpA), prepared according to section 1.2 of this chapter and kept at 4°C, is found in the stock solution of 0.1% Zwittergent 3-14, 25 mM Tris (pH 8.5) and 150 mM NaCl. Exchange of the protein environment from detergent micelle to lipid bilayer will create essentially different sample which can be analyzed with several methods, such as dynamic light scattering, centrifugation in sucrose gradient and solid state NMR (ssNMR). Due to its lower sensitivity, the ssNMR is particularly demanding in terms of substantial protein content. Since the total sample volume is limited by the size of the NMR rotor, the ratio between the lipids and the protein must be shifted towards the latter to the highest possible extent. However, in order to reconstitute the protein at such low lipid-to-protein ratios, one can not rely on the low efficiency of the spontaneous insertion of unfolded protein into preformed liposomes. By contrast, the protein reconstitution via detergent removal, coupled with the simultaneous formation of the bilayers, is an alternative approach that allows the incorporation of large amount of protein into the newly formed unilamellar vesicles.

This type of reconstitution is dependent on the effective, as complete as possible removal of the detergent (either by biobeads or by dialysis), which excludes the usability of low-CMC detergents. Therefore the first step in this direction is to exchange the 0.1% Zwittergent 3-14 of the stock solution with a high-CMC detergent, such as the abovementioned octylglucoside (OG, CMC 0.7% = 25 mM). For this purpose, certain amount of KpOmpA (from few to 22 mg, depending on the sample purposes and desired final protein content) in 0.1% Zwittergent 3-14 buffer is mixed with washed nickel-charged resin (Ni-NTA



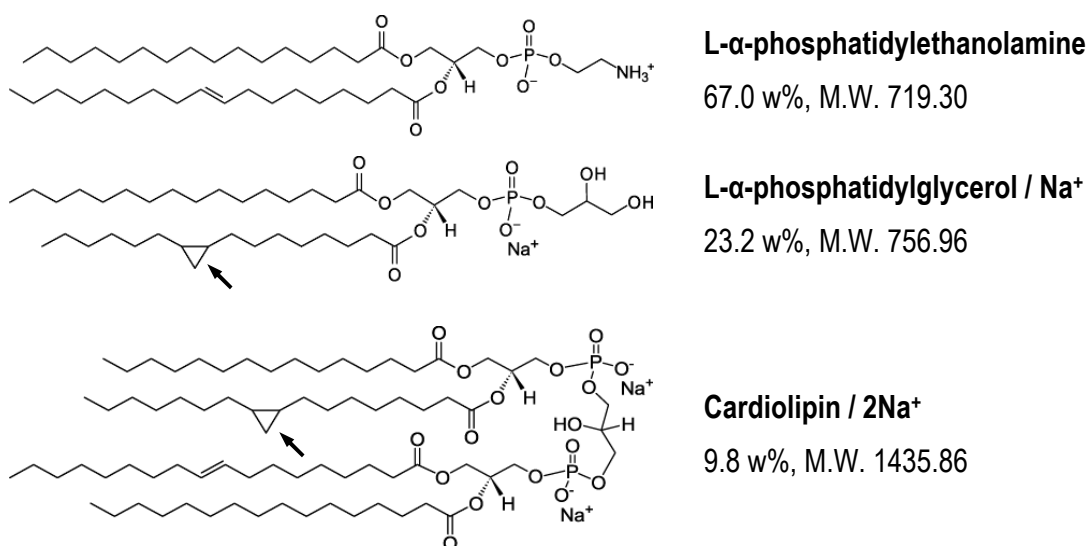
Superflow, QIAGEN) at a ratio of about 5 mg protein per 1 ml resin (equivalent to 2 ml slurry). After gentle mixing for 3-4 hours at 4°C, the protein-bound resin is transferred to a 50-ml centrifuge tube and covered with 45-50 ml of 2% OG (99.8% purity, Carbosynth) buffer containing 25 mM Tris (pH 8.5) and 150 mM NaCl. The resin-bound protein is then allowed to exchange its detergent micelle from Zwittergent 3-14 to OG by gentle mixing for 15 min at room temperature, after which the resin is collected by centrifugation (3000 g, 10 min, 4°C). This detergent-exchange step is repeated two more times and finally the resin is loaded onto single-use chromatography column (Wizard Plus Midipreps, Promega). After the collection of the flowthrough, the resin is washed with 5 column volumes of 2% OG buffer and then the protein is eluted with the same buffer containing 400 mM imidazole. The protein folding state and concentration are checked by SDS PAGE (**Fig. 1.7A**, lanes 1-4) and photometrically (**Fig. 1.7B**). The protein solution in 2% OG is kept at 4°C, without imidazole removal.



**Fig. 1.7.** Monitoring the protein folding and concentration during the detergent exchange step and the final stage of the procedure. **(A)** SDS PAGE gel of the N-KpOmpA stock solution in 0.1% Zwittergent 3-14 (lanes 1 and 2), after exchange to 2% OG (lanes 3 and 4) and after reconstitution in PLE-liposomes (lanes 5 and 6). The aliquots on lanes 2, 4 and 6 were heat denatured (5 min, 100°C); M - marker ladder. **(B)** UV spectrum (220-330 nm) of N-KpOmpA eluted in 2% OG (with 400 mM imidazole) with absorption value of 6.430 at 280 nm (measured with NanoDrop ND-1000 Spectrophotometer, Thermo Scientific), corresponding to protein concentration of about 3 mg/ml.

Prior to the protein preparation for reconstitution, a stock solution of the desired lipid or lipids mixture can be prepared. The *E. coli* Polar Lipids Extract (PLE,  $M.W._{AVERAGE} \approx 760$ , Avanti Polar Lipids, Inc.) is a powder mixture of the three main components in the bacterial inner membrane and the periplasmic leaflet of the outer membrane, thus providing a close-to-native environment for the membrane proteins of Gram-negative bacteria. It is composed of 67% (w/w) L- $\alpha$ -phosphatidylethanolamine, 23.2% (w/w) L- $\alpha$ -phosphatidylglycerol and 9.8%

(w/w) Cardiolipin, presented on **Fig. 1.8**. The gel-to-liquid crystalline phase transition temperature of this lipid mixture was found to be rather low [18], which ensures a liquid crystalline phase when working with it at room temperature and throughout the ssNMR measurements (also at room temperature). The PLE powder was transferred to a borosilicate glass vial and dissolved in few milliliters of chloroform, up to an approximate concentration of 20 mg lipids/ml. This solution was evaporated under nitrogen stream and then in a vacuum chamber overnight. The exact mass of the dry lipid film was detected with analytical balance. The film was then covered with the same 2% OG buffer used for the protein, at a concentration of 10 mg lipids/ml. Several cycles of heating (5 min, 42°C) and sonication in water bath (5 min) ensured the dissolving of the lipids and the formation of homogenous mixture of micelles, composed (in molar terms) of about 5 detergent molecules per lipid molecule. This detergent/lipid stock solution can be kept frozen at -20°C for several months and, before each use, should be thawed at room temperature, warmed (42°C) and sonicated in water bath.



**Fig. 1.8.** Composition of the Polar Lipids Extract (PLE) from *E. coli* B (ATCC 11303) grown in Kornberg Minimal media at 37°C (Avanti Polar Lipids). The arrows denote cyclopropane rings on the acyl chains.

The protein and lipid stock solution, prepared as described above, can then be mixed at the desired lipid-to-protein ratio (LPR). The LPR can be expressed either in mass (w/w), or in molar (mol/mol) terms when taking into account the respective molecular weights of the participating compounds:

$$LPR(w/w) = \frac{mg_{LIPID}}{mg_{PROTEIN}}$$

$$LPR(mol/mol) = \frac{mg_{LIPID} / MW_{LIPID}}{mg_{PROTEIN} / MW_{PROTEIN}}$$

Several comparative examples for different LPRs (used in this work) of *E. coli* PLE and the two protein constructs are shown below:

<b><u>N-KpOmpA / PLE</u></b>		<b><u>F-KpOmpA / PLE</u></b>	
<u>LPR (w/w)</u>	<u>LPR (mol/mol)</u>	<u>LPR (w/w)</u>	<u>LPR (mol/mol)</u>
0.1	3:1	0.1	5:1
0.5	15:1	0.5	25:1
1.0	30:1	1.0	50:1

For the reconstitution, the protein stock solution (usually at 1-3 mg/ml in 2% OG buffer and 400 mM imidazole) was mixed with the detergent/lipids (10 mg lipids/ml) stock solution at the desired LPR. In this case, the detergent and salts concentrations of the two solutions were the same. However, this approach does not exclude the possibility to have a small difference in one of the ingredients, since the protein solution is usually applied in considerably larger volume and therefore becomes 'dominant' in determining the composition of the final mixture. The ternary mixture of protein-detergent-lipids was then allowed 30 min of gentle mixing, after which it was transferred to a dialysis tube (12-14 kDa cutoff) and dialyzed for 12 hours at 37°C against 2L of detergent-free buffer. After exchanging the dialysis buffer with fresh 2L for another 12 hours, the white powder-like liposomes replaced the previously cleared micellar solution. The dialysis steps can be repeated one or two more times, depending on the initial sample amount, in order to ensure maximal removal of the detergent and imidazole.

For ssNMR experiments, 22 mg of N-KpOmpA were mixed with 11 mg PLE (LPR 0.5 w/w) and liposomes were prepared in the described way. In this case, the dialysis was extended by several more steps with gradual decrease in the NaCl concentration (in 25-mM steps, 12 hours each), in order to avoid presence of NaCl during the NMR measurement. Furthermore, macroscopically the vesicles exhibited lower propensity for aggregation at low-salt conditions which was previously shown for the PLE [18], although it is probable that at pH 8.5 the negatively charged protein (with theoretical pI 6.24), apart from the negatively

charged lipids, contributes to this effect. The folding state of the reconstituted protein was checked with gel electrophoresis, as before (**Fig. 1.7A**, lanes 5 and 6).

The reconstituted uniformly- $^{13}\text{C}/^{15}\text{N}$  labeled N-KpOmpA could then be used for ssNMR spectroscopy. For this purpose, the liposomes were collected (15 000 g, 15 min, 4°C) and then further stacked with ultracentrifugation (200 000 g, 90 min, 10°C, TLA 100.4 rotor, Optima TLX Ultracentrifuge, Beckman Coulter). After discarding the supernatant, the pellet was dehydrated down to a certain level under nitrogen stream, while the total weight of the sample was monitored until it dropped to about 50 mg. Initially composed of 22 mg protein and 11 mg PLE lipids (LPR of 0.5), the total of 33 mg of sample material is thus mixed with about 17 mg of 25 mM Tris buffer (pH 8.5). The approximate 30% (w/w) of water content ensures the close-to-native environment for the bilayers, while in the same time a maximal amount of labeled protein is introduced in the 4-mm NMR rotor (with about 50  $\mu\text{l}$  volume). A sample prepared in this way is further analyzed in Chapter 3.

Aside from the ssNMR sample preparation at high protein content, the sample resulting from the reconstitution procedure described above can be also characterized with other methods and/or at different LPR, without partial dehydration or a need of labeled protein. The centrifugation in sucrose gradient is one such approach which offers information on the reconstitution efficiency, since protein-rich liposomes will exhibit substantially different density than protein-free ones. Therefore vesicles containing only lipids will stay 'on top' of the tube, while the proteoliposomes will migrate deeper into the gradient. Sucrose powder (99.5% purity, Sigma-Aldrich) was dissolved in Milli-Q water in two stock solutions with concentrations of 20 and 60%, respectively. Starting with the higher concentration, 5-ml volumes from the two solutions were gradually mixed from the bottom to the top of a centrifuge tube (Beckman Coulter), forming a continuous gradient. The proteoliposomes (a total of 24-30 mg of protein and lipid material) were collected (15 000 g, 15 min, 4°C), resuspended in 0.5 ml of 20 mM Tris (pH 8.5) and a small aliquot was separated for dynamic light scattering measurements. Half of the remaining material (~250  $\mu\text{l}$ ) was subjected to 5 freeze-thaw cycles in liquid nitrogen, in order to determine if this affects the sample homogeneity, and was then applied on the top of a sucrose gradient tube. The other half was directly applied on the top of another tube. The samples were then centrifuged at 4°C for around 18 hours at 28 000 rpm in SW41 rotor (~100 000 g, Beckman Coulter). In the case of both N- and F-KpOmpA, regardless of the use of freeze-thaw cycles or not, the centrifugation produced a single band at mid-point position in the tube. Fractions of about 1-ml size were collected starting from the top of each tube and the concentration of sucrose was directly

measured with refractometer. **Fig. 1.9A** shows these values with indication of the band-containing fraction (~47% and ~45% for N- and F-KpOmpA, respectively).

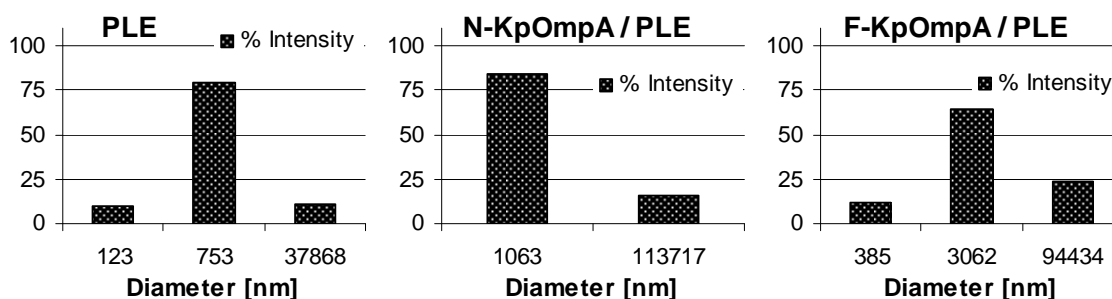
(A) Sucrose concentration in several fractions from each sample.

Sample	N-kpOmpA / PLE (LPR 0.5 w/w)					F-KpOmpA / PLE (LPR 0.5 w/w)				
Fraction	6	7	<b>8</b>	9	10	5	6	<b>7</b>	8	9
[%] sucrose, <i>Not treated</i>	38.5	41.5	<b>47</b>	54	57	33.5	37.5	<b>45.0</b>	52	56
[%] sucrose, <i>Freeze-thaw</i>	37.5	41.5	<b>47</b>	54	57	33	37.5	<b>44.5</b>	50	53.5

(B) Colorimetric estimation of the LPR in each band-containing fraction.

Sample	LPR (w/w), <i>Not treated</i>	LPR (w/w), <i>Freeze-thaw</i>
N-kpOmpA / PLE, Fraction 8	0.31	0.32
F-KpOmpA / PLE, Fraction 7	0.31	0.25

(C) Dynamic light scattering of N- and F-KpOmpA reconstituted in PLE.



**Fig.1.9.** Sucrose gradient centrifugation, colorimetric and dynamic light scattering analyses of the reconstituted protein constructs. (A) Estimation of the sucrose concentration in and around the fraction containing the proteoliposomes band (indicated in bold) for both protein constructs and sample treatment (with freeze-thaw cycles applied or not). (B) Colorimetric calculations of the LPR in the isolated bands from each sample, showing overestimation of the protein content. (C) Size distribution of protein-free PLE liposomes (left), or vesicles containing N-KpOmpA (middle) or F-KpOmpA (right) at LPR 0.5 (w/w). All preparations were made by dialysis-driven removal of the detergent (2% OG).

The isolated bands from each sample were then subjected to colorimetric analyses for estimation of the protein [10] and phospholipids [19, 20] contents. Due to the high viscosity of the sucrose solution in the bands-containing fractions, a large amount of the material was

gradually lost with each step of the subsequent treatment (band collection and sucrose clearing with several centrifugations in water). This prevented a realistic quantification of the total amount of protein/lipid material in the band fractions, but a relative estimation (protein-to-lipid ratio) was still possible. **Fig. 1.9B** shows the LPR in the isolated band fraction from each sample, calculated by colorimetry. As previously mentioned, the Lowry titration tends to overestimate the protein concentration by up to 50%, when compared with the spectrophotometric detection. However, the starting protein amount for each reconstitution experiment is estimated from the UV-spectrum of the respective 2% OG solution, which explains the observed discrepancies. Taking this into account, the actual LPR values of the isolated bands then appears to be close to 0.4-0.5 (w/w), as initially adjusted in the protein-detergent-lipid ternary complex before its dialysis. This, in addition to the absence of additional bands and the lack of influence of the freeze-thaw cycles on the band(s) distribution, suggested that the utilized dialysis-driven detergent removal is a suitable method for homogenous reconstitution of the two protein constructs in PLE at the desired LPR.

The proteoliposomes aliquots (taken prior to performing centrifugation in sucrose gradient with the rest of the samples) were then diluted 1000 times in 20 mM Tris (pH 8.5), subjected to dynamic light scattering measurements (Wyatt) and analyzed with Dynamics V6 (Protein solutions) for checking the size distribution of the vesicles (**Fig. 1.9C**). As expected, the dialysis-driven detergent removal did not produce uniform-size vesicles, but several populations with different sizes. However, one major population was detected in each case, with slightly larger diameter in the proteoliposomes in comparison to the protein-free vesicles. The populations with diameters at around 120 nm (PLE) and 385 nm (F-KpOmpA / PLE) appear to be minor in numbers and, furthermore, a negligible percentage of the material is concentrated there. Occasionally, such populations were also found in the case of N-KpOmpA / PLE (data not shown). In all cases, the major population of the proteoliposomes was observed in the range of 1-3  $\mu\text{m}$  in diameter, for both N- and F-KpOmpA reconstitutions in PLE. The smaller numbers of extremely large sizes ( $\sim 40$ -110  $\mu\text{m}$ ), found in each case, are outside of the instrumentation range and are probably a result from the noise in the correlation curve at long delay times, as suggested by the manufacturer's instructions. This leads to the conclusion that the majority of the sample material is found in large (or even giant) vesicles of sizes between 1 and 3  $\mu\text{m}$ . Despite of their variation in size, these liposomes possess the same LPR, as verified by the centrifugation in sucrose gradient. For ssNMR purposes this is satisfactory, as all vesicles with larger sizes will behave similarly. Smaller vesicles might exhibit rotational diffusion of the membrane components which will

interfere with solid state NMR experiments ([21]). Such vesicles, even if present in the described preparation, will contain negligible amount of the sample material. If needed, homogenization of the size distribution is usually achieved by extrusion of the sample. However, the high protein content (LPR 0.5 w/w) in the bilayers tends to 'rigidify' the vesicles up to a level which makes the extrusion difficult. Furthermore, PLE as such was found to be less prone to extrusion, particularly at higher salt concentrations ([18]). Neither extrusion nor sonication was used in any of the N- and F-KpOmpA reconstitution-related experiments described here.

*Note: Annex 4 provides detailed protocols for this section of the chapter.*

## References

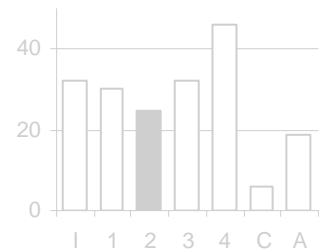
1. Renault, M., et al., *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, 2009. **385**(1): p. 117-30.
2. Arora, A., et al., *Structure of outer membrane protein A transmembrane domain by NMR spectroscopy*. Nature Structural Biology, 2001. **8**(4): p. 334-338.
3. Pautsch, A. and G.E. Schulz, *Structure of the outer membrane protein A transmembrane domain*. Nature Structural Biology, 1998. **5**(11): p. 1013-1017.
4. Nakamura, K. and S. Mizushima, *Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12*. J Biochem, 1976. **80**(6): p. 1411-22.
5. Kleinschmidt, J.H. and L.K. Tamm, *Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism*. Biochemistry, 1996. **35**(40): p. 12993-13000.
6. Kleinschmidt, J.H., M.C. Wiener, and L.K. Tamm, *Outer membrane protein A of E-coli folds into detergent micelles, but not in the presence of monomeric detergent*. Protein Science, 1999. **8**(10): p. 2065-2071.
7. Kleinschmidt, J.H. and L.K. Tamm, *Secondary and tertiary structure formation of the beta-barrel membrane protein OmpA is synchronized and depends on membrane thickness*. Journal of Molecular Biology, 2002. **324**(2): p. 319-330.
8. Davagnino, J., et al., *Acid-Hydrolysis of Monoclonal-Antibodies*. Journal of Immunological Methods, 1995. **185**(2): p. 177-180.
9. Skribanek, Z., et al., *Mass spectrometric and chemical stability of the Asp-Pro bond in herpes simplex virus epitope peptides compared with X-Pro bonds of related sequences*. Journal of Peptide Science, 2002. **8**(8): p. 398-406.
10. Lowry, O.H., et al., *Protein measurement with the Folin phenol reagent*. J Biol Chem, 1951. **193**(1): p. 265-75.
11. Ohnishi, S., K. Kameyama, and T. Takagi, *Characterization of a heat modifiable protein, Escherichia coli outer membrane protein OmpA in binary surfactant system of sodium dodecyl sulfate and octylglucoside*. Biochimica Et Biophysica Acta-Biomembranes, 1998. **1375**(1-2): p. 101-109.

12. Grizot, S. and S.K. Buchanan, *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. *Molecular Microbiology*, 2004. **51**(4): p. 1027-1037.
13. Danoff, E.J. and K.G. Fleming, *The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane beta-barrel*. *Biophys Chem*, 2011. **159**(1): p. 194-204.
14. Popot, J.L., et al., *Amphipols: polymeric surfactants for membrane biology research*. *Cellular and Molecular Life Sciences*, 2003. **60**(8): p. 1559-1574.
15. Schneider, B., et al., *Membrane protein expression in cell-free systems*. *Methods Mol Biol*, 2010. **601**: p. 165-86.
16. Kleinschmidt, J.H., et al., *Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism*. *Biochemistry*, 1999. **38**(16): p. 5006-5016.
17. Caffrey, M. and V. Cherezov, *Crystallizing membrane proteins using lipidic mesophases*. *Nature Protocols*, 2009. **4**(5): p. 706-731.
18. White, G.F., et al., *Physical properties of liposomes and proteoliposomes prepared from Escherichia coli polar lipids*. *Biochimica Et Biophysica Acta-Biomembranes*, 2000. **1468**(1-2): p. 175-186.
19. Murphy, J. and J.P. Riley, *A single-solution method for the determination of soluble phosphate in sea water*. *Journal of the Marine Biological Association of the United Kingdom*, 1958. **37**(01): p. 9-14.
20. Murphy, J. and J.P. Riley, *Citation-Classic - a Modified Single Solution Method for the Determination of Phosphate in Natural-Waters*. *Current Contents/Agriculture Biology & Environmental Sciences*, 1986(12): p. 16-16.
21. Traikia, M., et al., *High-resolution spectra of liposomes using MAS NMR. The case of intermediate-size vesicles*. *Journal of Magnetic Resonance*, 1997. **125**(1): p. 140-144.



# Chapter 2

## Single-Molecule Force Spectroscopy of KpOmpA in lipid bilayers



## Summary

Single-Molecule Force Spectroscopy has been used to investigate the reaction of KpOmpA to mechanical stress. The molecule was found to unfold stepwise as a function of the tip distance and to refold spontaneously in the bilayer upon relieve of the tensile load. This work suggests that membrane proteins, apart from their other functions in the native environment, can act as 'stress-relieving anchors' in the bacterial membrane. In addition, it proposes the possibility that the two-state folding and insertion model adopted for  $\alpha$ -helical membrane proteins can be extended to the  $\beta$ -barrel proteins as well.

This work has been published as:  
**Patrick D. Bosshart, Iordan Iordanov, Carlos Garzon-Coral, Pascal Demange, Andreas Engel, Alain Milon, and Daniel J. Müller,**  
The transmembrane protein KpOmpA anchoring the outer membrane of *Klebsiella pneumoniae* unfolds and refolds in response to tensile load.  
*Structure* 20, 121-127, January 2012.

## 2.1. Introduction to AFM and SMFS

The Atomic Force Microscopy (AFM) is a novel type of microscopy with a resolution limit in the nano- and subnanometer scale [1]. Commercially available for just around 20 years now, the AFM has already proved to be an excellent tool for imaging in biology and other disciplines, as well as for mechanical manipulation of the sample with forces in the range of piconewtons ([2]; see also different AFM manufacturers such as JPK, Novascan, Angstrom Advanced Inc., NanoScience Instruments, Park Systems, Bruker AXS and others).

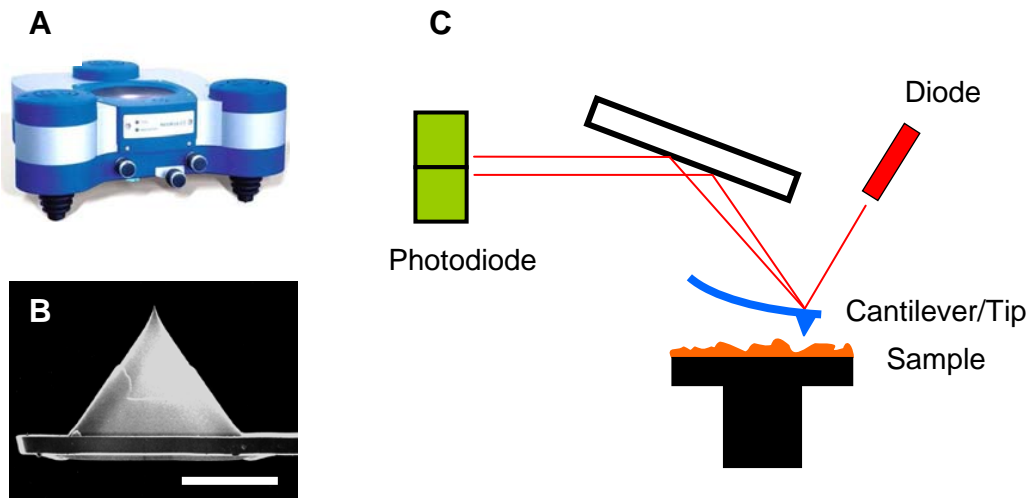
In the field of biological sciences, AFM finds application in the rational drug design and biosensor design, molecular electronics and biomaterials development. Specifically, investigating the intermolecular forces in biology with AFM was first shown for ligand-receptor interactions and studies on DNA [3, 4], opening the path for direct observation of the structure-function relations in biomacromolecules. Nowadays AFM is used in biology to analyze dynamic events on the minute's timescale (such as cell movement, [5]), measure cell-surface interactions during muscle contraction [6], measure forces between single molecules [7], characterize biological assemblies [8] and DNA-protein complexes in particular [9]. 'Time-laps AFM' allows the following of biological processes in solution, like transcription in *E. coli* [10] or DNA degradation [11]. Recent advances in AFM lead to its increased resolution and only part of them belong to instrumentation development. Another important factor is the sample preparation itself, including stable immobilization of the sample, pH and ionic strength of the buffer, etc. [12, 13]. At the instrument level, novel sub-types of AFM offer specific benefits for the biological application of the technique, such as analyzing frozen, highly rigid samples with Cryo-AFM, or increasing the resolution by using new generations of carbon nanotubes tips. Structural conformations and functional states of molecular complexes can be correlated by 'real-time AFM' [14] and 'fast AFM' was used to increase the imaging speed for these and other purposes by an order of magnitude [15].

The high precision of the AFM measurements is ensured by the use of piezoelectric elements, which change their size in dependence of an applied voltage. The sample is spread onto atomically flat support surface, such as the hydrophilic mica or the hydrophobic pyrolytic graphite. The choice of this surface and of a variety of other parameters (buffer conditions, cantilever type, etc.) can greatly influence the success of an AFM experiment [16, 17]. The AFM microscopes (**Fig. 2.1A**) utilize a very sharp probe (called a tip, **Fig. 2.1B**) attached to a cantilever. A diode beam strikes the upper surface of the cantilever, while the sharp end of the tip faces the sample (such as lipid bilayer patch with reconstituted proteins).

The reflected beam falls into photodiode detectors and subsequent amplifier collects their signal. These photodiodes form position-sensitive detector which reacts to a displacement of the cantilever. Upon movement of the cantilever in a close proximity to the sample surface, the tip bends (or deflects) under the influence of various forces (electrostatic, capillary, van der Waals, etc.) according to the Hooke's law of elasticity:

$$F = -kx,$$

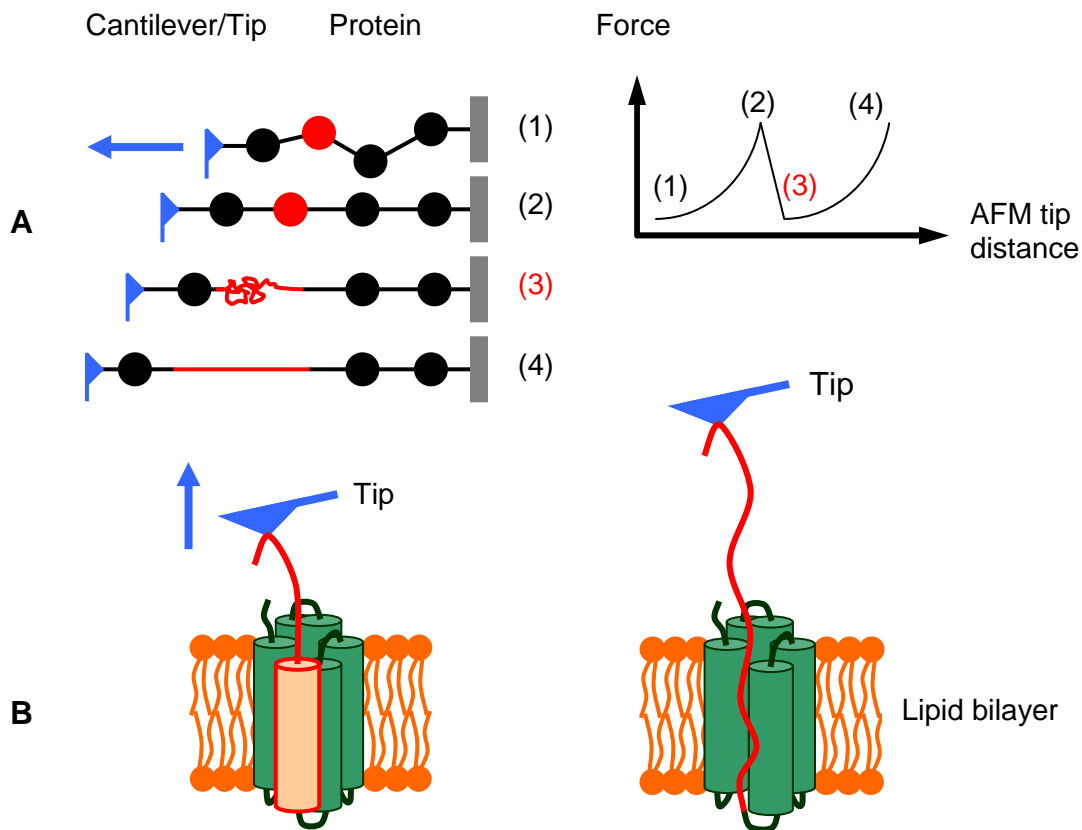
where the exerted force  $F$  (N) is proportional to the distance (or displacement)  $x$  (m) of the AFM tip and  $k$  (N/m) is the cantilever's spring constant which relates the other two parameters. In this way the AFM microscope receives information on the "height" of the sample. When moving the AFM tip with constantly applied force (the so called 'constant force contact mode') on the two dimensions ( $x$  and  $y$ ) of the surface, the scanning collects data on the third dimension ( $z$ ), thus producing a topological image of the sample (**Fig. 2.1C**).



**Fig. 2.1.** The basics of an AFM microscope. **(A)** Global view of an AFM microscope (NanoWizard® II, JPK Instruments, Germany). The cantilever is attached to the bottom surface of the instrument and is facing downwards. After positioning of a biological sample below the microscope - in the center of the tripod formation - the cantilever will be submerged in the aqueous solution and brought in contact with the sample surface. Image adapted from [www.jpk.com](http://www.jpk.com). **(B)** Electron micrograph of a sharpened, silicon nitride, square pyramidal tip (Olympus, Japan) of 2.9  $\mu\text{m}$  height. The scale bar is 2.4  $\mu\text{m}$ . Adapted from [probe.olympus-global.com](http://probe.olympus-global.com). **(C)** Schematic representation of the AFM microscope working principle (see text for details).

The resolution of this technique can be remarkable, since it was successfully used to visualize details on the bilayer surface even as small as conformational change of OmpG at

different pH [18] and of bacteriorhodopsin [19]. A critical point in this regard is that the AFM data acquisition allows the sample to stay in close-to-native environment. The main drawback is that shear forces resulting from the tip movement may damage soft samples (like soft biological and polymer surfaces) and distort the generated images.



**Fig. 2.2.** Schematic representation of SMFS experiments. **(A)** SMFS of a soluble, multidomain protein with the “weakest point” (in red color) being the domain which unfolds first. The force-distance curve (to the right) indicates the initial retraction of the AFM tip (1), the stretching of the polypeptide chain that leads to bending of the tip and hence a force increase (2), the unfolding of one domain which leads to temporary “release” of the tip (3) and the subsequent stretch of the chain (4). **(B)** In the case of membrane proteins, the bilayer environment provides a strong constraint due to the hydrophobic interactions with the protein. Therefore the segment(s) that are most closely attached to the AFM tip are pulled out first, followed by the next secondary structure elements along the chain length. In this example, the left picture would correspond to point (2) in (A) and the right one - to point (3).

Apart from providing a high-resolution imaging technique, the AFM microscopes can also be used for Single-Molecule Force Spectroscopy (SMFS). During the SMFS experiments one does not acquire an image of the sample surface. Instead, the instrument measures the

interaction forces between the AFM tip and the sample, as a function of the distance between them [20]. Hence the acquired data is called 'force curves'. The tip is periodically forced in contact with the surface for a given time and at a given speed (extension), after which it is separated from the sample and brought up to its initial position (retraction). In the case of biological samples of reconstituted membrane proteins in lipid bilayers, there is a certain (albeit rather small, i.e. around once over 1000 trials) chance that the exterior segments of the protein chain can be attached to the tip after its extension. Upon retraction of the tip, the attached protein is gradually unfolded and extracted out of the bilayer in a stepwise fashion, with characteristic force-distance peak positions and force values. The force peak positions can be subsequently mapped on the protein sequence in order to reveal the rupture points, thus providing information on areas crucial for the stability of the molecule (**Fig. 2.2**). This was nicely shown, for example, on wild-type and mutated bacteriorhodopsin [21] and on the *E. coli*  $\beta$ -barrel protein OmpG which was shown to unfold its  $\beta$ -sheets in pairs, as hairpins [22]. It was suggested, however, that under different conditions (i.e. pulling speed, ionic strength, etc., [23]) some of the hairpins can cluster and therefore unfold together, providing insights into the mechanisms of  $\beta$ -sheets aggregation. Additionally, AFM-imaging on OmpG revealed the pH-induced pore closure by one of the extracellular protein loops, demonstrating the versatility of AFM in monitoring proteins conformational change ([18]).

## 2.2. Basic concept of SMFS

One of the key points in an SMFS experiment is the proper notion of the distance between the tip and the sample. The directly controlled parameter in this regard is the displacement of the piezo element, which however does not take into account the tip-sample interactions (leading to deflection in the cantilever's extremity) and the sample propensity for deformation. In close proximity to the sample surface, this often results in larger apparent (i.e. measured) distances than the real ones [24]. The term 'force-displacement curves' is therefore often used to distinguish between these and the actual 'force-distance curves'. The latter are derived by taking into account additional parameters, such as the elastic constants of the cantilever and the sample surface. However, such information is not always available or easily derived. On the other hand, when stretching long polymers (such as proteins of hundreds amino acids) the retraction distances are often long enough in order to decrease this 'error' down to a negligible value.

The acquired force-distance curves (as in **Fig. 2.2A**) then represent the force needed for rupturing a given section of the folded protein as a function of the tip-sample separation distance. Calculating the cantilever's spring constant before each experiment represents a calibration procedure that allows the expression of this separation in actual distance units (usually nanometers). In the context of SMFS of proteins, the polypeptide is considered a continuously flexible polymer which, when unfolded and stretched, has an average distance of  $\sim 0.35$  nm per monomer (i.e. amino acid [25]). Of note, this value also depends on the speed of the tip retraction (for instance, [21] uses  $\sim 0.32$  nm per amino acid at 200 pN pulling force). The peak positions on the spectra represent the moments when a rupture event occurred, i.e. part of the protein has unfolded and the tip is now relaxed, while still retracting and further stretching the chain towards the next event. These peaks are fitted using the 'worm-like chain model' (WLC) and the lengths (in amino acids) of the protein segments participating in one rupture event or another are established [26]. The precision of such a calculation strongly depends on two factors: (i) the approach (i.e. extension) of the tip to the sample surface must be performed gently, so that the tip is indeed in contact with the surface (and not hanging above it), but without deforming it; This is to determine the 'zero level', from which the stretching of the protein (i.e. the 'counting' of the amino acids at a later point) is initiated; (ii) the second assumption is that the attachment of the polymer chain to the tip happens very close (if not exactly) at one of the protein termini. Usually, the drastically different scenarios are distinguishable. These include capturing and pulling an extracellular loop of the (membrane) protein, or extending from the other (non-symmetrical) terminus, etc. The curve patterns in these cases are substantially different and are filtered out. Since the probability for attachment of the molecule to the tip is in general quite low ( $\sim 0.1\%$ ), the accumulation of (at least)  $\sim 100$  'good' curves requires thousands of extensions-retractions. The larger number of curves is needed during the statistical analysis of the data, because the same rupture events in two different spectra may differ from one another in the range of a few nanometers. Averaging out of many curves, however, increases the accuracy of the fit and highlights the consistently repeating events. In the case of unfolding GFP, for instance, the error from 87 measurements was shown to be 0.33 nm, i.e.  $\pm 1$  amino acid [25]. The bilayer environment of reconstituted membrane proteins adds additional constraints during the pulling of the molecule and the rupture forces are no longer a function of only the protein structure, but also of the interactions with the surrounding lipids. Sample inhomogeneity, non-native adhesive forces and contamination of the tip surface with the 'sticky' lipids accidentally scratched from the sample are only a few of the additional obstacles found in the realm of

SMFS of membrane proteins. The accuracy of the fitted rupture positions is thus often reduced down to several amino acids.

### 2.3. Tip-sample interactions

The OMCL-RC800PSA tips (Olympus) used in the present work were coated with silicon nitride which, according to the manufacturer (<http://probe.olympus-global.com>), reduces the wear of the tip in comparison with a standard silicon tip. Although it is possible to render this kind of tips predominantly hydrophilic (with a plasma reactor) or hydrophobic (with a silane coating), such treatments were not applied in our case. Apart from silicon nitride, the tip coating also contains smaller amounts of carbon and oxygen (~10% for each). The strongest interactions at short tip-sample distances are the Van der Waals forces, while other, long-range interactions (capillary, electrostatic, etc.) become dominant when the tip retracts away from the sample surface.

Unsharpened tips (with ~50 nm radius of curvature), even if much larger than the average dimension of about 5 nm for a typical protein, are preferred for SMFS of proteins due to the higher probability of binding the macromolecule in the “trial and error” acquisition process. The so called “stiff cantilevers” are usually used for these experiments, since those ensure that the main elastic component of the system will be the protein (and not the tip itself) which simplifies the analysis. During a pulling experiment, the proteins attach (or adsorb) nonspecifically to the tip surface [27].

### 2.4. Conclusion

The work presented in this chapter exploits the behavior of the full-length KpOmpA protein (with N-terminal membrane-embedded domain and C-terminal peptidoglycan-binding domain, situated outside of the DMPC bilayer) upon application of mechanical force. This force tends to simulate a number of *in vivo* found conditions in which a bacterium may experience similar mechanical stress. KpOmpA was found to reversibly unfold and fold back in the membrane, suggesting its possible function as stress-relieving anchor in the bacterial outer membrane.

This hypothesis is further supported by the fact that the last unfolded structural segment (the N-terminal  $\beta$ -strand) exhibits unusually high anchoring strength in the membrane. The reader will note that, with the exception of the first rupture point (around residue 191), the other rupture points are found on top of the extracellular loops of the molecule. This is in



contrast with what was observed in the case of OmpG [22]. From SMFS point of view, the classical bacterial porins (such as OmpF [28] and OmpG [29]) are differing from OmpA/KpOmpA in two distinct ways: 1) they are much larger in size, which leaves an empty lumen in the core of the molecule; 2) their barrel-to-extracellular loops ratio is considerably higher, that is to say the rigidly structured portions of the molecules are larger in these cases. Of note, the KpOmpA transmembrane domain represents a narrow barrel with many side chains in close proximity in its lumen, while the loops area of the protein consists of nearly 50% of all residues [30]. It is known that strong, non-specific attachment of the biological sample to the AFM support surface could alter the native physical state of the lipids and the protein, thus influencing the measurement [17]. On the one hand, the side chains interactions in the lumen of the barrel could add to the mechanical stability of the  $\beta$ -sheets during their unfolding while, on the other, the long extracellular loops may contribute to the same effect by strongly adhering to the hydrophilic mica surface. Nevertheless, it is notable that even small  $\beta$ -barrel protein like KpOmpA can unfold stepwise, rather than being pulled out of the membrane in a single step, even if the unfolded segments do not match the behavior of those from the larger porins. Furthermore, this finding supports the suggestion that  $\beta$ -barrel proteins can fold in the membrane following the two-step insertion/folding mechanism, previously proposed for  $\alpha$ -helical proteins [31].

## References

1. Jalili, N. and K. Laxminarayana, *A review of atomic force microscopy imaging systems: application to molecular metrology and biological sciences*. Mechatronics, 2004. **14**(8): p. 907-945.
2. Tamayo, J., et al., *High-Q dynamic force microscopy in liquid and its application to living cells*. Biophysical Journal, 2001. **81**(1): p. 526-537.
3. Florin, E.L., V.T. Moy, and H.E. Gaub, *Adhesion Forces between Individual Ligand-Receptor Pairs*. Science, 1994. **264**(5157): p. 415-417.
4. Lee, G.U., L.A. Chrisey, and R.J. Colton, *Direct Measurement of the Forces between Complementary Strands of DNA*. Science, 1994. **266**(5186): p. 771-773.
5. Hoh, J.H. and C.A. Schoenenberger, *Surface-Morphology and Mechanical-Properties of Mdc Monolayers by Atomic-Force Microscopy*. Journal of Cell Science, 1994. **107**: p. 1105-1114.
6. Domke, J., et al., *Mapping the mechanical pulse of single cardiomyocytes with the atomic force microscope*. European Biophysics Journal with Biophysics Letters, 1999. **28**(3): p. 179-186.
7. Clausen-Schaumann, H., et al., *Force spectroscopy with single bio-molecules*. Current Opinion in Chemical Biology, 2000. **4**(5): p. 524-530.
8. Yang, Y., H. Wang, and D.A. Erie, *Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy*. Methods, 2003. **29**(2): p. 175-187.
9. Erie, D.A., et al., *DNA Bending by Cro Protein in Specific and Nonspecific Complexes - Implications for Protein Site Recognition and Specificity*. Science, 1994. **266**(5190): p. 1562-1566.

10. Argaman, M., et al., *Phase imaging of moving DNA molecules and DNA molecules replicated in the atomic force microscope*. Nucleic Acids Research, 1997. **25**(21): p. 4379-4384.
11. Bezanilla, M., et al., *Motion and Enzymatic Degradation of DNA in the Atomic-Force Microscope*. Biophysical Journal, 1994. **67**(6): p. 2454-2459.
12. Muller, D.J. and A. Engel, *The height of biomolecules measured with the atomic force microscope depends on electrostatic interactions*. Biophys J, 1997. **73**(3): p. 1633-44.
13. Muller, D.J., et al., *Electrostatically balanced subnanometer imaging of biological specimens by atomic force microscope*. Biophys J, 1999. **76**(2): p. 1101-11.
14. Stolz, M., et al., *Monitoring biomolecular interactions by time-lapse atomic force microscopy*. Journal of Structural Biology, 2000. **131**(3): p. 171-180.
15. Schitter, G., Stark, R.W., Stemmer, A., *Fast AFM-imaging of biological specimen*. European Cells and Materials, 2003. **6**(Suppl. 1): p. 73.
16. Muller, D.J. and A. Engel, *Atomic force microscopy and spectroscopy of native membrane proteins*. Nature Protocols, 2007. **2**(9): p. 2191-2197.
17. Muller, D.J. and A. Engel, *Strategies to prepare and characterize native membrane proteins and protein membranes by AFM*. Current Opinion in Colloid & Interface Science, 2008. **13**(5): p. 338-350.
18. Mari, S.A., et al., *pH-Induced Conformational Change of the beta-Barrel-Forming Protein OmpG Reconstituted into Native E. coli Lipids*. Journal of Molecular Biology, 2010. **396**(3): p. 610-616.
19. Engel, A., C.A. Schoenenberger, and D.J. Muller, *High resolution imaging of native biological sample surfaces using scanning probe microscopy*. Current Opinion in Structural Biology, 1997. **7**(2): p. 279-284.
20. Butt, H.J., B. Cappella, and M. Kappl, *Force measurements with the atomic force microscope: Technique, interpretation and applications*. Surface Science Reports, 2005. **59**(1-6): p. 1-152.
21. Muller, D.J., et al., *Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy*. Biophysical Journal, 2002. **83**(6): p. 3578-3588.
22. Sapa, K.T., et al., *One beta Hairpin after the Other: Exploring Mechanical Unfolding Pathways of the Transmembrane beta-Barrel Protein OmpG*. Angewandte Chemie-International Edition, 2009. **48**(44): p. 8306-8308.
23. Kedrov, A., et al., *Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy*. Annual Review of Biophysics and Biomolecular Structure, 2007. **36**: p. 233-260.
24. Cappella, B. and G. Dietler, *Force-distance curves by atomic force microscopy*. Surface Science Reports, 1999. **34**(1-3): p. 1-104.
25. Dietz, H. and M. Rief, *Exploring the energy landscape of GFP by single-molecule mechanical experiments*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(46): p. 16192-16197.
26. Bustamante, C., et al., *Entropic Elasticity of Lambda-Phage DNA*. Science, 1994. **265**(5178): p. 1599-1600.
27. Arrondo, J.L.R. and A. Alonso, eds. *Advanced Techniques in Biophysics*. 1 ed. Springer Series in Biophysics. Vol. 10. 2006, Springer. 280.
28. Yildiz, O., et al., *Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation*. Embo Journal, 2006. **25**(15): p. 3702-3713.
29. Subbarao, G.V. and B. van den Berg, *Crystal structure of the monomeric porin OmpG*. Journal of Molecular Biology, 2006. **360**(4): p. 750-759.
30. Renault, M., et al., *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, 2009. **385**(1): p. 117-30.
31. Popot, J.L. and D.M. Engelman, *Helical membrane protein folding, stability, and evolution*. Annual Review of Biochemistry, 2000. **69**: p. 881-922.

# The Transmembrane Protein KpOmpA Anchoring the Outer Membrane of *Klebsiella pneumoniae* Unfolds and Refolds in Response to Tensile Load

Patrick D. Bosshart,<sup>1</sup> Jordan Jordanov,<sup>2</sup> Carlos Garzon-Coral,<sup>1</sup> Pascal Demange,<sup>2</sup> Andreas Engel,<sup>3,\*</sup> Alain Milon,<sup>2,\*</sup> and Daniel J. Müller<sup>1,\*</sup>

<sup>1</sup>Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland

<sup>2</sup>Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, Centre National de la Recherche Scientifique, F-31077 Toulouse, France

<sup>3</sup>Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106, USA

\*Correspondence: andreas.engel@case.edu (A.E.), alain.milon@ipbs.fr (A.M.), daniel.mueller@bsse.ethz.ch (D.J.M.)

DOI 10.1016/j.str.2011.11.002

## SUMMARY

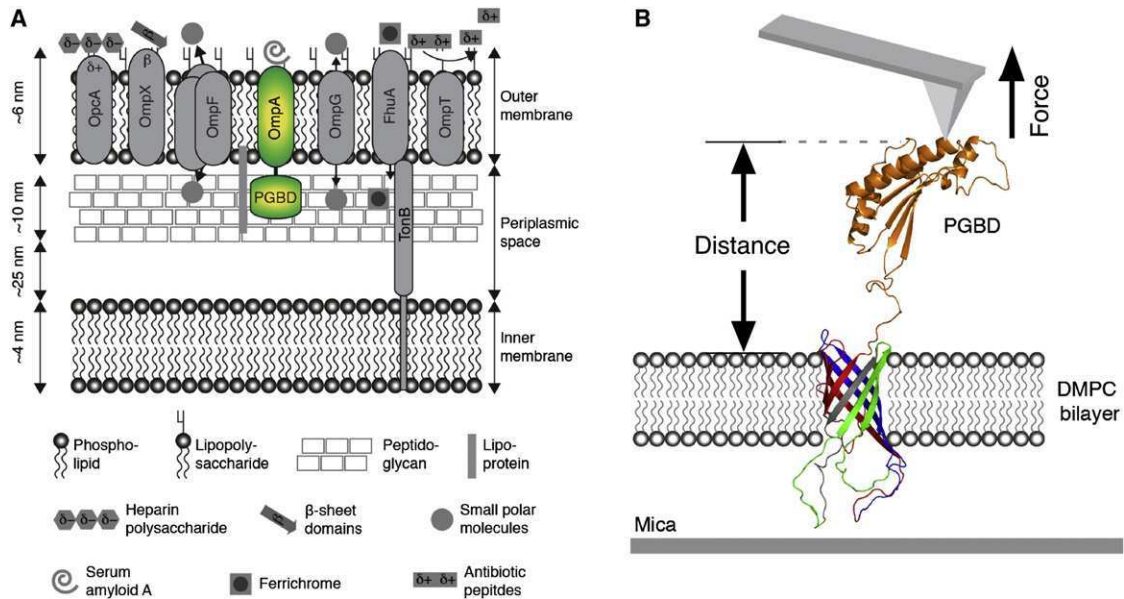
In *Klebsiella pneumoniae* the transmembrane  $\beta$ -barrel forming outer membrane protein KpOmpA mediates adhesion to a wide range of immune effector cells, thereby promoting respiratory tract and urinary infections. As major transmembrane protein OmpA stabilizes Gram-negative bacteria by anchoring their outer membrane to the peptidoglycan layer. Adhesion, osmotic pressure, hydrodynamic flow, and structural deformation apply mechanical stress to the bacterium. This stress can generate tensile load to the peptidoglycan-binding domain (PGBD) of KpOmpA. To investigate how KpOmpA reacts to mechanical stress, we applied a tensile load to the PGBD and observed a detailed unfolding pathway of the transmembrane  $\beta$ -barrel. Each step of the unfolding pathway extended the polypeptide connecting the bacterial outer membrane to the peptidoglycan layer and absorbed mechanical energy. After relieving the tensile load, KpOmpA reversibly refolded back into the membrane. These results suggest that bacteria may reversibly unfold transmembrane proteins in response to mechanical stress.

## INTRODUCTION

Outer membrane protein A (OmpA) is highly conserved among *Enterobacteriaceae* and represents the most abundant protein of the outer membranes. Due to its multiple essential functions, OmpA is one of the best-studied membrane proteins in Gram-negative bacteria (Faraldo-Gómez and Sansom, 2003; Koebnik et al., 2000; Smith et al., 2007). OmpA provides mechanical stability by anchoring the outer membrane to the murein layer via a C-terminal peptidoglycan-binding domain (PGBD). OmpA acts as a receptor for bacteriophages and bacteriocins and as a target for host cell defense systems (Smith et al., 2007; Wang, 2002). Moreover, OmpA has been described as a gating

transmembrane pore (Hong et al., 2006). Importantly, OmpA is a key protein facilitating bacterial adhesion to mammalian and plant cells. OmpA-mediated adhesion is involved in meningitis, enterohemorrhagic infections, immune invasion, and pneumonia (Jeannin et al., 2000; Shin et al., 2005; Soulas et al., 2000). Among all outer membrane proteins, OmpA forms the smallest  $\beta$ -barrel, with eight antiparallel  $\beta$ -strands. The residues inside this  $\beta$ -barrel are tightly packed so that much of the lumen inside the barrel is filled with polar side chains that interact through a network of hydrogen bonds and electrostatic interactions (Pautsch and Schulz, 2000). In *Klebsiella pneumoniae*, OmpA (KpOmpA) mediates adhesion to a wide range of immune effector cells and is responsible for respiratory tract and urinary infections (Jeannin et al., 2000; Pichavant et al., 2003; Soulas et al., 2000). Like OmpA, KpOmpA comprises an eight  $\beta$ -stranded transmembrane  $\beta$ -barrel with extracellular loops that are significantly larger than those of *Escherichia coli* OmpA (Renault et al., 2009). It is thought that *K. pneumoniae* infection is mediated via adhesion of these loops, whereas the C-terminal PGBD of the  $\beta$ -barrel protein remains strongly anchored to the peptidoglycan layer in the periplasmic space (Figure 1A) (Hizukuri et al., 2009; Wang, 2002).

Several factors can mechanically stress the outer membrane of Gram-negative bacteria. Among these are osmotic pressure, hydrodynamic flow, mechanical deformation, and cell adhesion (Hizukuri et al., 2009; Smith et al., 2007; Wang, 2002). OmpA, as the major transmembrane protein ( $\approx 100,000$  copies per bacterium [Koebnik et al., 2000]) of the outer membrane, mechanically stabilizes the Gram-negative bacterium by anchoring the outer membrane to the peptidoglycan layer (Endermann and Henning, 1979; Endermann et al., 1978; Koebnik, 1995; Sonntag et al., 1978) and protects the bacterium from environmental stress (Wang, 2002). In absence of other major mechanisms anchoring the outer membrane to the peptidoglycan layer, one must assume that mechanical stress applied to the outer membrane of the bacterium will apply tensile load to the PGBD that anchors KpOmpA of the outer membrane to the peptidoglycan layer. To characterize how KpOmpA may react to mechanical stress applied to the outer membrane of *K. pneumoniae*, we applied a tensile load to the C-terminal anchor of KpOmpA and characterized the response of the  $\beta$ -barrel protein that was anchored in the lipid membrane



**Figure 1. Omps of Gram-Negative Bacteria and SMFS Experiment**

(A) Omps (indicated) support diverse functions that are essential for bacterial survival. These include the binding and transport of solutes and molecules (bottom) anchoring the outer membrane to the peptidoglycan and mediating bacterial adhesion to hosts.

(B) SMFS experiment mimicking tensile load applied between the transmembrane  $\beta$ -barrel and the PGDB of KpOmpA. The tip of an AFM cantilever is used to mechanically stress the nonspecifically attached PGDB domain. While retracting the cantilever, its deflection (force) and distance between tip and membrane is measured in a force-distance curve (Figure 2A). The PGDB structure is a homology model generated using SWISS-MODEL based on the structure of the OmpA-like domain RmpM (PDB entry code 1R1M.pdb) (Bordoli et al., 2009).

(Figure 1B). We found that upon exposure to sufficient forces corresponding to that stabilizing single cell adhesion bonds ( $\approx 60$ – $200$  pN) (Helenius et al., 2008; Müller et al., 2009), individual  $\beta$ -strands of KpOmpA start unfolding in distinct steps, with the last  $\beta$ -strand showing an unusually high anchoring strength in the membrane. This stepwise unfolding absorbs mechanical energy and extends the polypeptide connecting peptidoglycan and the outer membrane up to  $\approx 80$  nm. As soon as it is relieved from the tensile load, the unfolded KpOmpA polypeptide reversibly folds back into the membrane.

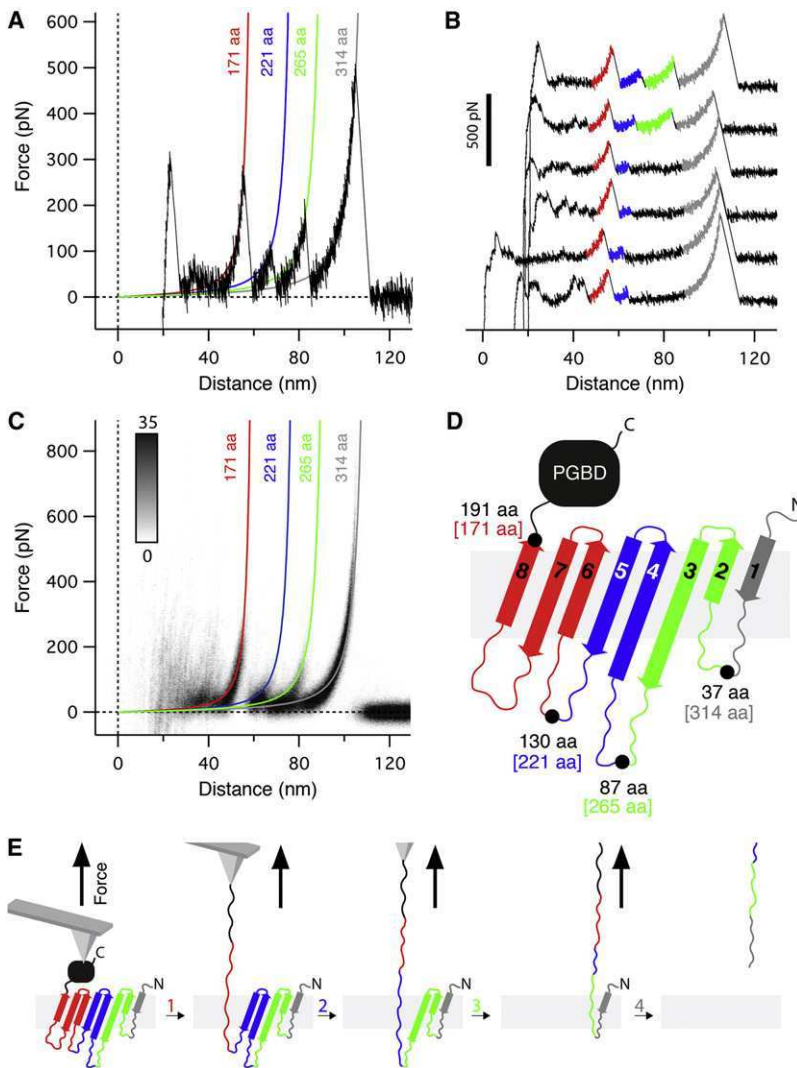
## RESULTS AND DISCUSSION

### Applied to Tensile Load KpOmpA Unfolds Stepwise

To apply a tensile load to the PGDB of full-length KpOmpA reconstituted in lipid membranes (Supplemental Experimental Procedures available online), we used single-molecule force spectroscopy (SMFS) (Figure 1B). To facilitate nonspecific attachment of the PGDB, the atomic force microscopy (AFM) tip was pushed onto the KpOmpA surface. Withdrawing the AFM tip stretched the PGDB and induced the stepwise unfolding and extraction of KpOmpA from the membrane. Force-distance curves recorded the interactions that occurred upon unfolding of a single KpOmpA (Figures 2A and 2B). We analyzed only force-distance curves that corresponded to the fully stretched length of a KpOmpA polypeptide unfolded from the PGDB (Experimental Procedures and Supplemental Experimental Procedures). This selection criterion ensured that KpOmpA was mechanically unfolded by stretching its PGDB (Kedrov et al., 2007).

Individual force-distance curves showed a series of peaks that varied in force and occurrence (Figure 2B). Every force peak of a force-distance curve reflected an interaction that was established by an unfolding intermediate with all intermediates describing the unfolding pathway of KpOmpA. Superimposing all force-distance curves showed a clear pattern of predominant force peaks (Figure 2C). Unfolding a C-terminally truncated version of KpOmpA lacking the PGDB showed a similar pattern of prevalent force peaks as observed for the full-length KpOmpA (Figure S1). The difference between both force peak patterns was that the pattern of C-terminally truncated KpOmpA shifted toward shorter distances (i.e., contour lengths). This shift highlighted that full-length KpOmpA was unfolded from its C-terminal end (Bosshart et al., 2008; Kedrov et al., 2007; Kedrov et al., 2004). Furthermore, this experiment demonstrated that the prevalent force peaks observed for full-length KpOmpA (Figure 2C) detected the unfolding intermediates and pathways of the transmembrane  $\beta$ -barrel.

Reproducibly occurring force peaks were fitted using the worm-like chain (WLC) model to reveal the contour lengths of unfolded polypeptide stretches (Figures 2A and 2C). The average contour lengths of these unfolded polypeptide stretches (Figure S2) allowed assigning the structural segments that, upon unfolding, transformed one unfolding intermediate to the next (Figure 2D). In contrast to the mechanical unfolding of water-soluble proteins, transmembrane proteins unfold sequentially (Engel and Gaub, 2008; Kedrov et al., 2007; Oesterhelt et al., 2000). The reason for this sequential unfolding behavior is that transmembrane proteins are embedded in and anchored by the highly anisotropic lipid membrane. Thus, when pulling



the terminal end, the first structural segment that unfolds is directly linked to the terminal end from which the AFM tip applies the tensile load. After this, the forthcoming structural segment that unfolds is the next adjacent one. Therefore, the sequence of the force peaks reflects the sequential unfolding steps of the structural segments of the transmembrane protein (Engel and Gaub, 2008; Kedrov et al., 2007). Consequently, the first structural segment that unfolded from the transmembrane  $\beta$ -barrel was created by three  $\beta$ -strands connected to the C-terminal end. The second and third structural segments that unfolded were established by two  $\beta$ -strands, and the last unfolding intermediate was established by a single N-terminal  $\beta$ -strand. Figure 2E demonstrates the predominant unfolding pathway of KpOmpA.

#### KpOmpA Shows a Unique Unfolding Pathway with $\beta$ Strands Forming Unfolding Intermediates

Previous mechanical unfolding experiments showed that main unfolding steps of the transmembrane  $\beta$ -barrel protein OmpG were established by single  $\beta$ -hairpins of two  $\beta$ -strands (Sapra

#### Figure 2. Mechanical Unfolding of KpOmpA Embedded in Lipid Membranes

(A) Force-distance curve recorded during unfolding of a single KpOmpA shows force peaks that detect unfolding intermediates of the  $\beta$ -barrel protein. Individual force peaks have been fitted using the WLC model to obtain the contour lengths (given in amino acids [aa]) of the unfolded polypeptide chain.

(B) Selection of force-distance curves each recorded unfolding a single KpOmpA.

(C) Superposition of force-distance curves ( $n = 183$ ) shows the reproducible unfolding pattern of KpOmpA. Colored lines are WLC fits as shown in (A). The gray-scale bar ranging from 0 to 35 allows assigning the density levels of the superposition.

(D) Mapping the main unfolding intermediates of the transmembrane  $\beta$ -barrel. Equally colored  $\beta$ -strands unfold in cooperative events. The N-terminal  $\beta$ -strand unfolds in a single step. Numbers in brackets indicate contour lengths revealed from WLC fits, while other numbers locate the interaction detected at the corresponding amino acid position in KpOmpA (PDB entry code 2K0L.pdb). In case the interaction had to be assumed to lie opposite of the membrane, 11 aa ( $\approx 4$  nm) were added to the contour length to structurally locate the interaction. This procedure is called “membrane compensation” (Supplemental Experimental Procedures) (Müller et al., 2002).

(E) Predominant unfolding pathway of KpOmpA. The secondary structure schemes show that upon mechanical pulling of the PGBD, the  $\beta$ -strands 8, 7, and 6 (red) unfold in a single step. After this, the  $\beta$ -strands 5 and 4 (blue) unfold, followed by the  $\beta$ -strands 3 and 2 (green). In a final step, the N-terminal  $\beta$ -strand (gray) is extracted from the membrane. See also Figures S1 and S2.

et al., 2009). Upon changing the functional state of OmpG,  $\beta$ -strands 8, 9, and 10 grouped to unfold together, and  $\beta$ -strand 11 unfolded individually (Damaghi et al., 2010). Thus, it is not surprising that up to three  $\beta$ -strands of KpOmpA can unfold individually.

The observation that a transmembrane protein mediating bacterial adhesion and providing mechanical stability to the outer membrane can unfold in several steps lines up interesting mechanistic insight on structural and functional design principles. Based on the extended hydrogen-bonding network of charged and polar residues in the lumen of the OmpA  $\beta$ -barrel, it may be assumed that the transmembrane  $\beta$ -barrel acts as one entity (Bowie, 2004; Pautsch and Schulz, 1998, 2000). In addition, chemically and thermally induced unfolding and refolding experiments suggest that OmpA and a similar transmembrane  $\beta$ -barrel protein, PagP, insert into the lipid bilayer in one major step (Huysmans et al., 2010; Kleinschmidt, 2006; Tamm et al., 2004). Therefore, a priori, one may assume that when applying a sufficiently high tensile load, KpOmpA is extracted in one step from the membrane. Instead, the  $\beta$ -barrel protein unfolds in consecutive steps, with each unfolding step requiring the input of a tensile load. This stepwise unfolding of KpOmpA destroys mechanical energy stressing the PGBD. Thus, one may speculate that too high tensile load applied to the outer membrane of the bacteria could be relieved by multiple

unfolding events. It may be further speculated that the stepwise unfolding of KpOmpA brings along another advantage that extracellular loops, which facilitate bacterial adhesion (Pichavant et al., 2003; Smith et al., 2007; Wang, 2002), are extracted from potential binding sites as soon as the tensile load applied becomes too high.

Our experiments show that the unfolding intermediates of KpOmpA are established by single and grouped  $\beta$ -strands. Their detailed unfolding pathways are different from the almost spontaneous force-induced unfolding of the water-soluble  $\beta$ -barrel forming green fluorescent protein (GFP) (Dietz and Rief, 2004). In case of GFP unfolding, the exposure of the hydrophobic core to the hydrophilic aqueous solution is one of the driving unfolding forces. In contrast, the anisotropic environment of the lipid membrane significantly contributes to the structural stability of the embedded protein so that unfolding forces must be repetitively applied for a sequence of unfolding intermediates until the entire membrane protein has been unfolded. The stepwise unfolding behavior of KpOmpA is similar to that observed for the larger transmembrane  $\beta$ -barrel protein OmpG comprising 14  $\beta$ -strands (Damaghi et al., 2010; Sapra et al., 2009). This finding is surprising since it is thought that the interaction of  $\beta$ -barrel proteins with the membrane also depends on the size of the  $\beta$ -barrel (Kleinschmidt, 2006; White and Wimley, 1999).

### The N-terminal $\beta$ Strand Is Unusually Strongly Anchored in the Membrane

Our experiment shows that the first unfolding events of grouped  $\beta$ -strands of KpOmpA require forces of  $\approx 100$ – $200$  pN, with the last unfolding event of the N-terminal  $\beta$ -strand requiring the highest forces of  $\approx 388 \pm 17$  pN ( $n = 183$ ) (Figure S2). These forces, which reflect the interaction strengths stabilizing the  $\beta$ -strands, are higher compared to those ( $\approx 100$ – $150$  pN) required to unfold  $\beta$ -strands from the larger  $\beta$ -barrel membrane protein OmpG (Sapra et al., 2009) and to those ( $\approx 150$ – $200$  pN) required to unfold the entire  $\beta$ -barrel protein GFP (Dietz and Rief, 2004) at similar conditions. Thus, our experiments suggest that KpOmpA is an unusually stable protein.

It is surprising that the last N-terminal  $\beta$ -strand of KpOmpA requires an average pulling force of  $\approx 388$  pN to be unfolded and extracted from the membrane bilayer. Alternatively, one may assume that the strong force peak at  $\approx 314$  amino acids (aa) that detects the N-terminal  $\beta$ -strand (Figure 2) could also reflect the rupture of the attachment between the AFM tip and the polypeptide. In such a case, the detected force would underestimate the force required to unfold and extract the N-terminal  $\beta$ -strand from the membrane bilayer. Nevertheless, the N-terminal  $\beta$ -strand shows at least an  $\approx 2$ – $4$  times higher stability than that determined for the unfolding of two or three  $\beta$ -strands (Figure S2). This finding suggests that the N-terminal  $\beta$ -strand has been designed to anchor much more stably in the membrane than all other  $\beta$ -strands of the  $\beta$ -barrel.

### In Absence of Externally Applied Tension KpOmpA Refolds into the Membrane

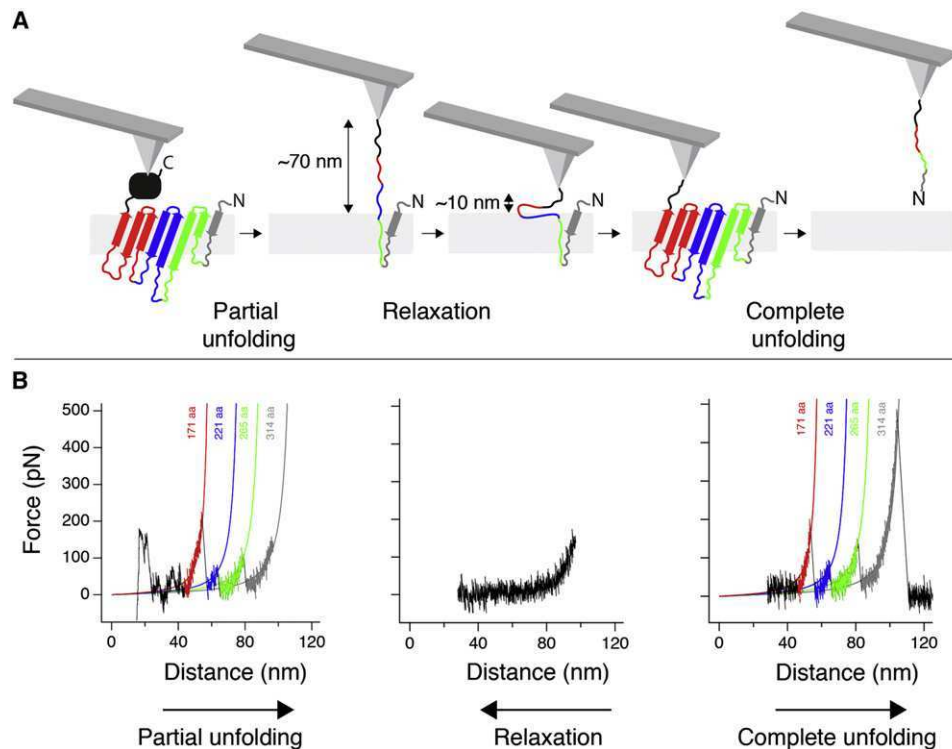
We investigated what happens once the KpOmpA polypeptide has been unfolded. Therefore, we unfolded single KpOmpA molecules leaving the last N-terminal  $\beta$ -strand anchored in the

membrane (Figure 3; Figure S3). Then, we relaxed the polypeptide and provided sufficient time ( $\approx 2$  s) for refolding. After this, we stretched the polypeptide again to detect which secondary structures refolded. In  $\approx 2.5\%$  of 200 refolding experiments performed, we observed that the unfolded polypeptide refolded. In these cases, the unfolding force peaks of the refolded polypeptide reoccurred at the same positions as detected upon initial unfolding of KpOmpA (Figure 3B). This suggests that the secondary structures of KpOmpA refolded. If the  $\beta$ -strands would have folded without inserting into the membrane or simply adsorbed to the membrane surface, the force peaks would have been detected at shifted positions (Supplemental Experimental Procedures). Similarly, because force peaks are characteristic for the membrane protein fold, they would have changed position in the presence of misfolding events (Kedrov et al., 2007; Kedrov et al., 2006). Moreover, the average unfolding forces were similar upon initial unfolding of KpOmpA and upon unfolding the refolded KpOmpA. This experiment showed that when partially unfolded and relaxed, KpOmpA refolds into the membrane to establish secondary structures and interactions similarly to those observed of native KpOmpA.

### Possible Extension of the Two-State Folding and Insertion Model to $\beta$ -Barrel Membrane Proteins

In the case of transmembrane  $\alpha$ -helical proteins, it was shown that the mechanically stressed polypeptide not only induces the stepwise unfolding of the membrane protein but also folds back stepwise into the membrane bilayer (Kedrov et al., 2006; Kedrov et al., 2004; Kessler et al., 2006). Similarly, mechanically loaded OmpG unfolds stepwise and refolds stepwise into the membrane (Damaghi et al., 2011). Here we show that when mechanically stressing the PGBD of KpOmpA, transmembrane  $\beta$ -strands form stable unfolding intermediates and, as soon as the tensile load disappears, fold back into the lipid membrane.

Bulk unfolding experiments suggest that OmpA from *E. coli* unfolds and folds reversibly (Kleinschmidt, 2006; Tamm et al., 2004). The folding process is described to occur via three membrane-bound folding intermediates. In the first stage, the OmpA polypeptide adsorbs to the water-membrane interface. In the second stage, the polypeptide adopts a partially membrane-inserted folding intermediate, and in the third stage, the fully inserted polypeptide forms the tertiary structure of OmpA. Differences between SMFS and conventional denaturation and renaturation experiments (Kleinschmidt, 2006; Tamm et al., 2004) of OmpA may be due to different experimental conditions. In our experiments, KpOmpA was embedded in a lipid membrane and investigated in buffer solution at room temperature. Conventional unfolding and refolding experiments using thermal or chemical denaturants (e.g., 4–8 M urea) are thought to induce very different unfolding scenarios of, in most cases solubilized, OmpA. However, in vivo and in vitro experiments have shown that OmpA fragments can insert into the membrane and fold into biologically active OmpA (Debnath et al., 2010; Koebnik, 1996). This suggests that single and grouped  $\beta$ -strands can form stable structures in the membrane and that  $\beta$ -stranded fragments can assemble into functional OmpA complexes. More generally, our results suggest that the two-state folding and insertion model in which transmembrane  $\alpha$  helices insert and fold into the membrane as independently



**Figure 3. Refolding a Single KpOmpA**

(A) Schematic of refolding experiment. After attaching the PGBD to the AFM tip, the cantilever is withdrawn to initiate unfolding of KpOmpA. After all domains have been unfolded except for the last  $\beta$ -strand ( $>70$  nm), the unfolded polypeptide is relaxed, bringing the AFM tip in close proximity ( $\approx 10$  nm) to the membrane. Then, a certain time is left for the polypeptide to refold ( $\approx 2$  s) and the AFM tip is withdrawn to detect the refolded structures.

(B) Force-distance curves recorded according to (A). First, KpOmpA is partially unfolded until the N-terminal  $\beta$ -strand embedded in the membrane bilayer is detected. Then, the unfolded polypeptide is relaxed. After a refolding time of 2 s, KpOmpA is completely unfolded. See also Figure S3.

stable units that then assemble into the functional membrane protein (Engelman et al., 2003) may be expandable to  $\beta$ -stranded membrane proteins.

### Relevance to Outer Membrane of *K. pneumoniae* Being Exposed to Mechanical Stress

OmpA that mediates bacterial adhesion is embedded in the outer membrane and attached via the C terminus to the peptidoglycan layer of the periplasmic space (Faraldo-Gómez and Sansom, 2003; Koebnik et al., 2000; Smith et al., 2007). Force spectroscopy studies show that single receptor-ligand bonds facilitating cell adhesion rupture at forces ranging between  $\approx 20$ – $200$  pN when mechanically stressed at similar experimental conditions as used in this work (Helenius et al., 2008; Müller et al., 2009). The magnitude of these rupture forces depends on the bond formed, on the reaction trajectory the receptor and ligand are forced to separate, and on the cell state and experimental environment (Evans and Calderwood, 2007). Thus, the force values exerted by single adhesive bonds suggest that they could become sufficiently large to induce unfolding of a single KpOmpA. In addition to KpOmpA, other molecules (e.g., proteins or carbohydrates) can significantly contribute to bacterial adhesion and, thus, to the mechanical stress applied to the bacterial outer membrane (Alsteens et al., 2009; Aprikian et al., 2011; Verbelen et al., 2008; Yakovenko et al., 2008). It

has been reported that the adhesion strength mediated by single bacterial proteins can reach sufficient values to induce the complete or partial unfolding of water-soluble proteins (Alsteens et al., 2009; Aprikian et al., 2011; Yakovenko et al., 2008). Most importantly, however, the adhesion of the entire bacterial cell to, for example, a host is much stronger than that facilitated by single receptor-ligand bonds and can easily reach several tens of nanonewtons (Razatos et al., 1998). Such large adhesion forces are sufficient to deform the membrane and thus apply tensile load to the PGBD of KpOmpA. Moreover, osmotic pressure, hydrodynamic flow, and mechanical deformation can add significantly to the tensile load between PGBD and transmembrane  $\beta$ -barrel of KpOmpA (Hizukuri et al., 2009; Smith et al., 2007; Wang, 2002). Thus, we can assume that the tensile load applied in this work to unfold KpOmpA indeed can occur in vivo. In absence of the anchorage to the peptidoglycan, even little mechanical stress significantly deforms the relatively soft outer membrane (Sonntag et al., 1978). In comparable situations, small adhesive forces of 60–80 pN start deforming the cellular membrane until membrane tethers (or nanotubes) are extracted (Müller et al., 2009; Sheetz, 2001). To prevent this effect, the bacterial transmembrane protein OmpA anchors the outer membrane to the peptidoglycan layer. Our SMFS experiments applied tensile load to the PGBD of KpOmpA. Upon reaching a critical force, KpOmpA starts stepwise

unfolding. This stepwise unfolding behavior reduces mechanical stress applied to bacterial membrane, membrane protein, and peptidoglycan anchor. Taking this escape route, KpOmpA employs stepwise unfolding pathways to dissipate too large tensile loads, being of mechanical origin, and to escape from unfavorable mechanically stressing situations. As soon as the tensile load disappeared the unfolded KpOmpA polypeptide refolds back into the membrane to reanchor membrane and peptidoglycan layer so that native conformations can be re-established. These results indicate that cells may employ unfolding and refolding pathways of transmembrane proteins to fulfill dedicated functional tasks.

## EXPERIMENTAL PROCEDURES

### SMFS

Full-length KpOmpA was prepared as described in detail in the [Supplemental Experimental Procedures](#). Briefly, KpOmpA was expressed into *E. coli* inclusion bodies, refolded in the presence of Octyl-POE (Alexis Biochemicals, Switzerland), purified by metal affinity and size exclusion chromatography. Purified protein was reconstituted into a 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) (Avanti Polar Lipids Inc., USA) lipid membrane by dialysis-driven detergent removal.

Lipid membranes containing densely packed KpOmpA were adsorbed to freshly cleaved mica in buffer solution (pH 8, 20 mM Tris-HCl, 150 mM NaCl). After an adsorption time of  $\approx 30$  min nonadsorbed membranes were removed by washing the sample with adsorption buffer. After this, the membranes were localized by AFM imaging in adsorption buffer at room temperature (Müller and Engel, 2007). For SMFS, the AFM tip (OMCL-RC800PSA, Olympus, Japan) was pushed onto the membrane until reaching a force of  $\approx 1$  nN for  $\approx 500$  ms. This forced contact led to the attachment of the PGBD from KpOmpA to the tip in  $\approx 0.1\%$  of all cases. Then the cantilever to which the AFM tip was attached was retracted at 600 nm/s to induce protein unfolding. A force-distance curve recorded the forces required to overcome the interactions that stabilized the unfolding intermediates of the membrane protein. For refolding, the protein was attached to the AFM tip as described before. Subsequently, the tip was retracted by  $\approx 90$  nm leading to partial KpOmpA unfolding. After that, the piezo was extended so that the cantilever rested at a height of  $\approx 10$  nm above the membrane surface. This position was kept constant for  $\approx 2$  s allowing refolding of KpOmpA into the lipid bilayer. In the final step the cantilever was retracted again with a speed of 600 nm/s leading to complete KpOmpA unfolding. Only force-distance curves where the cantilever was not in contact with the membrane surface during refolding were accepted for analysis.

### Data Selection and Analysis

For analysis we selected only force-distance curves that were sufficiently long to ensure that KpOmpA was unfolded from its PGBD. The fully stretched KpOmpA polypeptide shows 359 aa and assuming a contour length of 0.36 nm per amino acid this corresponds to a length of  $\approx 129$  nm. The C-terminal domain, to which the AFM tip preferentially attached, was  $\approx 140$  aa long ( $\approx 50$  nm), whereas the N-terminal end was  $\approx 14$  aa long ( $\approx 5$  nm). Since the exact attachment point of the AFM tip to the C-terminal PGBD was not known ([Supplemental Experimental Procedures](#)), we selected force-distance curves which were  $\geq 70$  nm long. This selection criterion together with the controls ([Figure S1](#)) ensured that we selected force-distance curves that were recorded when pulling the C-terminal domain of KpOmpA. Force-distance curves were manually superimposed and every force peak of every curve was fitted using the WLC model and a persistence length of 0.4 nm. These WLC fits provided for every force peak the contour length of the unfolded and stretched polypeptide segment (Kedrov et al., 2007; Müller and Engel, 2007). The metric contour length was converted into numbers of amino acids by division with 0.36 nm, which is the average distance between the C $\alpha$ -atoms of two adjacent amino acids (Pauling et al., 1951). A histogram of all contour lengths fitted from all force-distance curves analyzed was generated ([Figure S2](#)). A Gaussian was fitted to the main peaks of the histogram

in order to extract the average contour length values. These contour lengths were taken to assign the interactions that established structurally stable segments on the transmembrane  $\beta$ -barrel of KpOmpA (Protein Data Bank [PDB] entry code 2K0L.pdb). The first peak in the contour length histogram was assigned to the periplasmic end of the C-terminal  $\beta$ -strand. In some cases the contour length suggested that the interaction anchoring the unfolded polypeptide had to be located at the extracellular membrane surface opposite to the AFM tip pulling from the periplasmic surface. To locate this interaction we applied a 'membrane compensation' called procedure ([Supplemental Experimental Procedure](#)).

## SUPPLEMENTAL INFORMATION

Supplemental information includes three figures and Supplemental Experimental Procedures and can be found with this article online at [doi:10.1016/j.str.2011.11.002](https://doi.org/10.1016/j.str.2011.11.002).

## ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (grant MU 1791) and the European Union Seventh Framework Programme (FP7/2007-2013, grant agreement no. 211800). We thank C. Bippes for assistance and critical reading of the manuscript and J.-L. Popot for helpful discussions.

Received: July 28, 2011

Revised: October 20, 2011

Accepted: November 4, 2011

Published: January 10, 2012

## REFERENCES

- Alsteens, D., Dupres, V., Klotz, S.A., Gaur, N.K., Lipke, P.N., and Dufrène, Y.F. (2009). Unfolding individual Als5p adhesion proteins on live cells. *ACS Nano* 3, 1677–1682.
- Aprikan, P., Interlandi, G., Kidd, B.A., Le Trong, I., Tcheshnokova, V., Yakovenko, O., Whitfield, M.J., Bullitt, E., Stenkamp, R.E., Thomas, W.E., and Sokurenko, E.V. (2011). The bacterial fimbrial tip acts as a mechanical force sensor. *PLoS Biol.* 9, e1000617.
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4, 1–13.
- Bosshart, P.D., Casagrande, F., Frederix, P.L., Ratera, M., Bippes, C.A., Müller, D.J., Palacin, M., Engel, A., and Fotiadis, D. (2008). High-throughput single-molecule force spectroscopy for membrane proteins. *Nanotechnology* 19, 384014.
- Bowie, J.U. (2004). Membrane proteins: a new method enters the fold. *Proc. Natl. Acad. Sci. USA* 101, 3995–3996.
- Damaghi, M., Bippes, C., Köster, S., Yildiz, O., Mari, S.A., Kühlbrandt, W., and Müller, D.J. (2010). pH-dependent interactions guide the folding and gate the transmembrane pore of the beta-barrel membrane protein OmpG. *J. Mol. Biol.* 397, 878–882.
- Damaghi, M., Köster, S., Bippes, C.A., Yildiz, O., and Müller, D.J. (2011). One  $\beta$  hairpin follows the other: exploring refolding pathways and kinetics of the transmembrane  $\beta$ -barrel protein OmpG. *Angew. Chem. Int. Ed. Engl.* 50, 7422–7424.
- Debnath, D., Nielsen, K.L., and Otzen, D.E. (2010). *In vitro* association of fragments of a beta-sheet membrane protein. *Biophys. Chem.* 148, 112–120.
- Dietz, H., and Rief, M. (2004). Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. USA* 101, 16192–16197.
- Endermann, R., Krämer, C., and Henning, U. (1978). Major outer membrane proteins of *Escherichia coli* K-12: evidence for protein II being a transmembrane protein. *FEBS Lett.* 86, 21–24.
- Endermann, R., and Henning, U. (1979). Nearest neighbors of major proteins in the outer membrane of *Escherichia coli* K12. *FEBS Lett.* 97, 339–342.



- Engel, A., and Gaub, H.E. (2008). Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* 77, 127–148.
- Engelman, D.M., Chen, Y., Chin, C.N., Curran, A.R., Dixon, A.M., Dupuy, A.D., Lee, A.S., Lehnert, U., Matthews, E.E., Reshetnyak, Y.K., et al. (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett.* 555, 122–125.
- Evans, E.A., and Calderwood, D.A. (2007). Forces and bond dynamics in cell adhesion. *Science* 316, 1148–1153.
- Faraldo-Gómez, J.D., and Sansom, M.S. (2003). Acquisition of siderophores in gram-negative bacteria. *Nat. Rev. Mol. Cell Biol.* 4, 105–116.
- Helenius, J., Heisenberg, C.P., Gaub, H.E., and Muller, D.J. (2008). Single-cell force spectroscopy. *J. Cell Sci.* 121, 1785–1791.
- Hizukuri, Y., Morton, J.F., Yakushi, T., Kojima, S., and Homma, M. (2009). The peptidoglycan-binding (PGB) domain of the *Escherichia coli* pal protein can also function as the PGB domain in *E. coli* flagellar motor protein MotB. *J. Biochem.* 146, 219–229.
- Hong, H., Szabo, G., and Tamm, L.K. (2006). Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nat. Chem. Biol.* 2, 627–635.
- Huysmans, G.H., Baldwin, S.A., Brockwell, D.J., and Radford, S.E. (2010). The transition state for folding of an outer membrane protein. *Proc. Natl. Acad. Sci. USA* 107, 4099–4104.
- Jeannin, P., Renno, T., Goetsch, L., Miconnet, I., Aubry, J.P., Delneste, Y., Herbault, N., Baussant, T., Magistrelli, G., Soulas, C., et al. (2000). OmpA targets dendritic cells, induces their maturation and delivers antigen into the MHC class I presentation pathway. *Nat. Immunol.* 1, 502–509.
- Kedrov, A., Ziegler, C., Janovjak, H., Kühlbrandt, W., and Müller, D.J. (2004). Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J. Mol. Biol.* 340, 1143–1152.
- Kedrov, A., Janovjak, H., Ziegler, C., Kuhlbrandt, W., and Muller, D.J. (2006). Observing folding pathways and kinetics of a single sodium-proton antiporter from *Escherichia coli*. *J. Mol. Biol.* 355, 2–8.
- Kedrov, A., Janovjak, H., Sapra, K.T., and Müller, D.J. (2007). Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* 36, 233–260.
- Kessler, M., Gottschalk, K.E., Janovjak, H., Muller, D.J., and Gaub, H.E. (2006). Bacteriorhodopsin folds into the membrane against an external force. *J. Mol. Biol.* 357, 644–654.
- Kleinschmidt, J.H. (2006). Folding kinetics of the outer membrane proteins OmpA and FomA into phospholipid bilayers. *Chem. Phys. Lipids* 141, 30–47.
- Koebnik, R. (1995). Proposal for a peptidoglycan-associating  $\alpha$ -helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* 16, 1269–1270.
- Koebnik, R. (1996). In vivo membrane assembly of split variants of the *E. coli* outer membrane protein OmpA. *EMBO J.* 15, 3529–3537.
- Koebnik, R., Locher, K.P., and Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37, 239–253.
- Müller, D.J., and Engel, A. (2007). Atomic force microscopy and spectroscopy of native membrane proteins. *Nat. Protoc.* 2, 2191–2197.
- Müller, D.J., Kessler, M., Oesterhelt, F., Möller, C., Oesterhelt, D., and Gaub, H. (2002). Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J.* 83, 3578–3588.
- Müller, D.J., Helenius, J., Alsteens, D., and Dufrene, Y.F. (2009). Force probing surfaces of living cells to molecular resolution. *Nat. Chem. Biol.* 5, 383–390.
- Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E., and Müller, D.J. (2000). Unfolding pathways of individual bacteriorhodopsins. *Science* 288, 143–146.
- Pauling, L., Corey, R.B., and Branson, H.R. (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 37, 205–211.
- Pautsch, A., and Schulz, G.E. (1998). Structure of the outer membrane protein A transmembrane domain. *Nat. Struct. Biol.* 5, 1013–1017.
- Pautsch, A., and Schulz, G.E. (2000). High-resolution structure of the OmpA membrane domain. *J. Mol. Biol.* 298, 273–282.
- Pichavant, M., Delneste, Y., Jeannin, P., Fourneau, C., Bricchet, A., Tonnel, A.B., and Gosset, P. (2003). Outer membrane protein A from *Klebsiella pneumoniae* activates bronchial epithelial cells: implication in neutrophil recruitment. *J. Immunol.* 171, 6697–6705.
- Razatos, A., Ong, Y.L., Sharma, M.M., and Georgiou, G. (1998). Molecular determinants of bacterial adhesion monitored by atomic force microscopy. *Proc. Natl. Acad. Sci. USA* 95, 11059–11064.
- Renault, M., Saurel, O., Czaplicki, J., Demange, P., Gervais, V., Löhr, F., Réat, V., Piotto, M., and Milon, A. (2009). Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications. *J. Mol. Biol.* 385, 117–130.
- Sapra, K.T., Damaghi, M., Köster, S., Yildiz, O., Kühlbrandt, W., and Muller, D.J. (2009). One beta hairpin after the other: exploring mechanical unfolding pathways of the transmembrane beta-barrel protein OmpG. *Angew. Chem. Int. Ed. Engl.* 48, 8306–8308.
- Sheetz, M.P. (2001). Cell control by membrane-cytoskeleton adhesion. *Nat. Rev. Mol. Cell Biol.* 2, 392–396.
- Shin, S., Lu, G., Cai, M., and Kim, K.S. (2005). *Escherichia coli* outer membrane protein A adheres to human brain microvascular endothelial cells. *Biochem. Biophys. Res. Commun.* 330, 1199–1204.
- Smith, S.G., Mahon, V., Lambert, M.A., and Fagan, R.P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol. Lett.* 273, 1–11.
- Sonntag, I., Schwarz, H., Hirota, Y., and Henning, U. (1978). Cell envelope and shape of *Escherichia coli*: multiple mutants missing the outer membrane lipoprotein and other major outer membrane proteins. *J. Bacteriol.* 136, 280–285.
- Soulas, C., Baussant, T., Aubry, J.P., Delneste, Y., Barillat, N., Caron, G., Renno, T., Bonnefoy, J.Y., and Jeannin, P. (2000). Outer membrane protein A (OmpA) binds to and activates human macrophages. *J. Immunol.* 165, 2335–2340.
- Tamm, L.K., Hong, H., and Liang, B. (2004). Folding and assembly of beta-barrel membrane proteins. *Biochim. Biophys. Acta* 1666, 250–263.
- Verbelen, C., Dupres, V., Raze, D., Bompard, C., Locht, C., and Dufrene, Y.F. (2008). Interaction of the mycobacterial heparin-binding hemagglutinin with actin, as evidenced by single-molecule force spectroscopy. *J. Bacteriol.* 190, 7614–7620.
- Wang, Y. (2002). The function of OmpA in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 292, 396–401.
- White, S.H., and Wimley, W.C. (1999). Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28, 319–365.
- Yakovenko, O., Sharma, S., Forero, M., Tcheshnokova, V., Aprikian, P., Kidd, B., Mach, A., Vogel, V., Sokurenko, E., and Thomas, W.E. (2008). FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J. Biol. Chem.* 283, 11596–11605.

## Supplemental Information

### The Transmembrane Protein KpOmpA Anchoring the Outer Membrane of *Klebsiella pneumoniae*

#### Unfolds and Refolds in Response to Tensile Load

Patrick D. Bosshart, Iordan Iordanov, Carlos Garzon-Coral, Pascal Demange,  
Andreas Engel, Alain Milon, and Daniel J. Müller

#### Inventory of Supplemental Information

**Supplemental Figure S1.** This figure shows SMFS experiments unfolding C-terminally truncated KpOmpA. The experimental results are used to identify from which terminal end full-length KpOmpA was unfolded and at the same time to assign the unfolding pattern of the transmembrane  $\beta$ -barrel domain (related to Figure 2A-D).

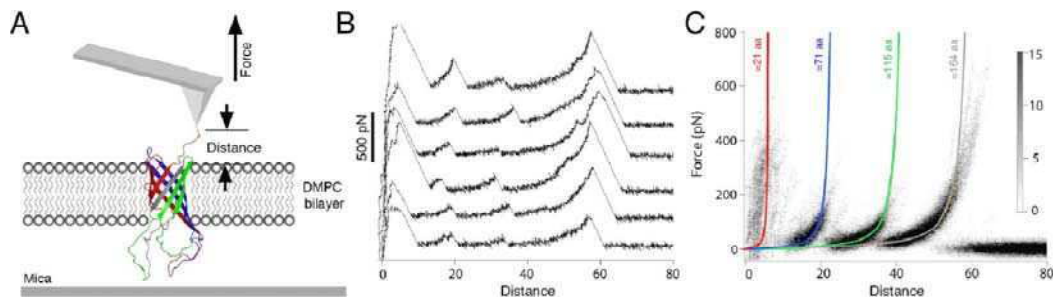
**Supplemental Figure S2.** This figure shows the histograms of the average contour lengths and the average forces detected upon unfolding of full-length KpOmpA. The average contour lengths are used to assign the unfolding intermediates of the KpOmpA unfolding pathway (related to Figure 2D,E).

**Supplemental Figure S3.** This figure shows further details (time trace) of the experiment refolding single KpOmpA (related to Figure 3).

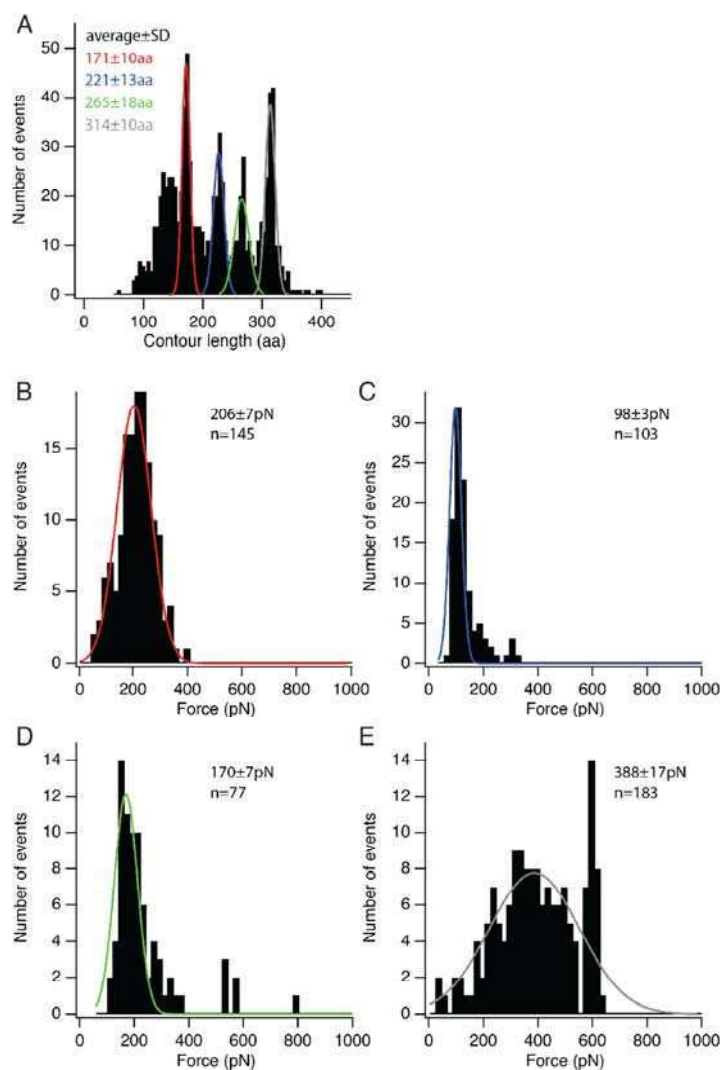
**Supplemental Experimental Procedures.** Here we describe the cloning, expression, refolding, purification and reconstitution of full-length KpOmpA. Furthermore we provide details of data analysis and interpretation these include:

- i) How we selected force-distance curves that were recorded with the peptidoglycan-binding domain (PGBD) attached to the AFM tip.
- ii) How the membrane compensation was applied to locate interactions within the membrane or at the membrane surface opposite the pulling AFM tip.
- iii) Why the unfolding spectra recorded by SMFS is specific for the folding of the protein.

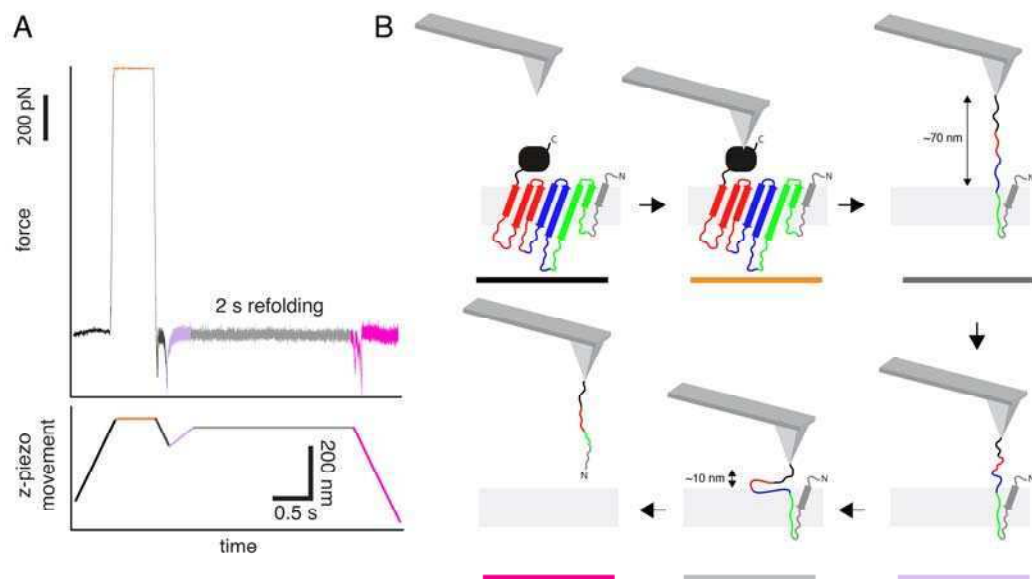
## Supplemental Data



**Supplemental Figure S1, related to Figure 2. Mechanical unfolding of KpOmpA lacking the C-terminal domain ( $\Delta$ CKpOmpA).** (A) SMFS experiment showing the AFM tip applying tensile load to the truncated C-terminal end of the transmembrane  $\beta$ -barrel of  $\Delta$ CKpOmpA. (B) Selection of force-distance curves each recorded upon unfolding of single  $\Delta$ CKpOmpAs. (C) Superimposition of force-distance curves ( $n=130$ ) shows the reproducible unfolding pattern of  $\Delta$ CKpOmpA. The gray scale bar ranging from 0 to 15 allows assigning the density levels of the superimposition. The truncated version of KpOmpA lacking the PGBD having a length of 216 aa and a shortened C-terminal end of 25 aa (19 aa from the protein sequence plus a hexahistidine-tag) was expressed, refolded and purified as described (Renault et al., 2009).



**Supplemental Figure S2, related to Figure 2. Histograms of the average contour lengths and average forces of all force peaks detected for full-length KpOmpA.** (A) Histogram showing the average contour lengths of all force peaks detected. The ‘first peak’ detected before the force peak at a contour length of 171 amino acids (aa) roughly presents where the PGDB domain was picked up with the AFM tip (**Supplemental Experimental Procedures**). Averages and standard deviations are given for every fit. (B-E) Distribution of average forces detected upon mechanically unfolding full-length KpOmpA. Shown are average forces ( $\pm$  standard error) of the force peaks detected at contour lengths of (B) 171 aa, (C) 221 aa, (D) 265 aa, and (E) 314 aa. All 183 force-distance curves superimposed in **Figure 2** were analyzed. To reveal the average values every peak of the histogram was fitted with a Gaussian (colored curves). The color coding of the Gaussian fits corresponds to that used for **Figures 2, 3, S1, and S3**. Because the individual force peaks had a certain probability to occur the number of force peaks ( $n$ ) contributing to the averages differed. This variable probability of unfolding force peaks (unfolding events) is frequently observed when unfolding  $\alpha$ -helical and  $\beta$ -barrel membrane proteins (Janovjak et al., 2003; Kedrov et al., 2007; Sapra et al., 2009).



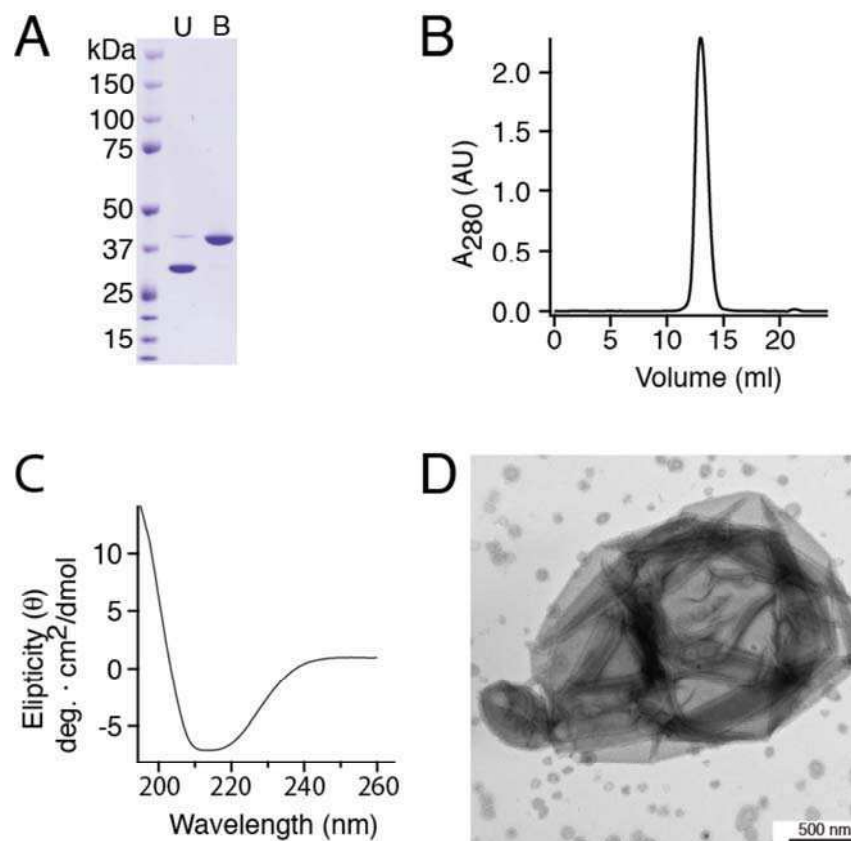
**Supplemental Figure S3, related to Figure 3. Time trace showing the refolding experiment of full-length KpOmpA.** Time trace (A) and schematic drawing (B) showing the segments of a typical membrane protein refolding experiment as described in **Figure 3** and in the *Experimental procedures*. In the first step the piezoelectric element of the AFM was extended to bring the cantilever tip into contact with KpOmpA (black segment). Pushing the tip on the KpOmpA at  $\approx 1$  nN for  $\approx 0.5$  s (orange segment). Retraction of the piezo separates the cantilever tip and stresses the protein. This separation if continued until the KpOmpA unfolded up to its last N-terminal anchor (dark gray segment). Then, the unfolded polypeptide was relaxed again (violet segment) by extending the z-piezo and bringing the cantilever tip close ( $\approx 10$  nm) to the membrane surface. After a certain time (e.g.  $\approx 2$  s) left for re-insertion and refolding of the polypeptide into the lipid bilayer (light gray segment) the cantilever tip is retracted to record the unfolding spectrum of the refolded full-length KpOmpA (magenta segment). (B) Schematic representation of the steps shown in (A).

## Supplemental Experimental Procedures

### Cloning, expression, refolding, purification and reconstitution of full-length KpOmpA

Wild type full-length KpOmpA comprising the N-terminal membrane domain and the soluble C-terminal domain was amplified by polymerase chain reaction (PCR) from genomic DNA of *Klebsiella pneumoniae* (ATCC 700721) introducing a 5'-end Nde1 and a 3'-end Xho1 restriction site. The 5'-end signal sequence for insertion into the outer membrane was removed to promote the formation of inclusion bodies during expression. The PCR product was subcloned into a pET22b vector between its Nde1 and Xho1 cleavage sites yielding full-length KpOmpA-pET22b. The construct codes for a 359 aa long protein with a C-terminal hexahistidine-tag and a predicted mass of  $\approx 39$  kDa. The sequence of the final construct was verified by DNA sequencing.

Recombinant full-length KpOmpA was expressed in *E. coli* BL21(DE3) grown in low-salt LB-medium supplemented with 100  $\mu\text{g/ml}$  ampicillin at 37°C. The expression was induced with 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside at  $\text{OD}_{600} \approx 0.6$  leading to the formation of inclusion bodies. Inclusion bodies were harvested, washed as described (Renault et al., 2010) and dissolved in 25 mM Tris-HCl, 6 M guanidine hydrochloride (Gdn-HCl), 5 mM  $\beta$ -mercaptoethanol, pH 8.5 (0.2 g/ml inclusion bodies) by vigorous stirring for 2 h at 37°C. Full-length KpOmpA was refolded by rapid dilution in 25 volumes of refolding buffer (150 mM NaCl, 600 mM L-arginine, 5 mM  $\beta$ -mercaptoethanol, 5% n-octylpolyoxyethylene (Octyl-POE) (w/v), 25 mM Tris-HCl, pH 8.5) at room temperature and incubated overnight at 4°C under gentle stirring. Refolded full-length KpOmpA (**Figure S4**) was ultracentrifuged (100,000 g, 30 min, 4°C), filtered (Millex® GV Filter Unit 0.22  $\mu\text{m}$ , Millipore, Ireland) and dialyzed against 40 volumes of 150 mM NaCl, 1% Octyl-POE (w/v), 25 mM Tris-HCl, pH 8.5 for 48 h at 4°C. The protein was purified using nickel nitrilotriacetic acid (NiNTA) (Qiagen, Germany) (in 150 mM NaCl, 1% Octyl-POE (w/v), 500mM imidazole, 25 mM Tris-HCl, pH 8.5) and gel filtration chromatography on a Superose 6 10/300 GL column (GE Healthcare Life Sciences, Switzerland) (**Figure S4B**). The protein was eluted with 150 mM NaCl, 1% Octyl-POE (w/v), 25 mM Tris-HCl, pH 8.5. Peak fractions were pooled and the concentration was adjusted to 1 mg/ml using the absorbance at 280 nm ( $\epsilon = 59945\text{M}^{-1}\text{cm}^{-1}$ ). Refolding efficiency was checked by linear (4-12%) SDS-PAGE assay (Schweizer et al., 1978). For protein analysis the band of a colloidal Coomassie-blue stained SDS-PAGE was excised, the protein was digested and subjected to mass spectrometry. Refolded full-length KpOmpA was reconstituted into a 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) (Avanti Polar Lipids, USA) lipid bilayer by dialysis-driven detergent removal. Full-length KpOmpA was mixed with 1% Octyl-POE-solubilized lipids at a lipid-to-protein ratio of 0.5 (w/w) and dialyzed against detergent-free buffer (100 mM NaCl, 20 mM Tris-HCl, 0.01% (w/v)  $\text{NaN}_3$ , pH 8) for 5 days at 37°C. After reconstitution, the SDS-PAGE assay (Schweizer et al., 1978) was repeated to verify that full-length KpOmpA remained folded during reconstitution. All buffers were prepared from nanopure water (resistivity  $> 18\text{M}\Omega\cdot\text{cm}$ ). Reconstitution of full-length KpOmpA into DMPC bilayers was checked by transmission electron microscopy of uranyl-acetate stained samples (**Figure S4D**).



**Supplemental Figure S4. Purification and reconstitution of full-length KpOmpA into DMPC lipid bilayers.** (A) Coomassie-blue stained linear SDS-PAGE (4-12%) of refolded full-length KpOmpA. The unboiled sample (**U**) migrates to a lower molecular weight than the boiled (**B**) sample. This gel-shift is typical for properly folded  $\beta$ -barrel proteins (Schweizer et al., 1978). (B) Gel filtration profile of purified full-length KpOmpA run on a Superose 6 10/300 GL column. The symmetrical peak indicates the monodispersity of the refolded full-length KpOmpA. (C) Average ( $n=10$ ) far-UV circular dichroism (CD) spectrum of refolded and purified full-length KpOmpA recorded in PBS (pH 8) supplemented with 1% Octyl-POE. The average CD spectrum indicates that the protein contained secondary structure elements with a high  $\beta$ -sheet content, which agrees with the structure of the membrane domain. (D) Transmission electron microscopy micrograph of an uranyl-acetate stained DMPC vesicle that contains densely packed full-length KpOmpA.

#### Attachment of the peptidoglycan-binding domain (PGBD) to the AFM tip

Full-length KpOmpA has 359 residues, which would result in a contour length of  $\approx 129$  nm if the polypeptide was fully unfolded and stretched. However, in most cases (>95%) the force-distance curves had a length of  $\approx 80$  nm indicating that KpOmpA was not attached to the AFM tip at its last amino acid. This assumption is corroborated by the structure of an OmpA-like PGBD (RmpM, PDB entry code 1R1M.pdb) where the N- and C-termini are very close to each other. This structure indicates that the C-terminus of the KpOmpA PGBD points towards the membrane part (N-terminus) and is not accessible for the non-specific attachment to the AFM tip. The PGBD can be approximated as a globular domain, which non-specifically attaches to the AFM tip at various positions leading to shorter force-distance curves than expected. For this reason we have selected only force-distance curves showing a length of  $\geq 70$  nm for further analysis (see **Experimental Procedures**).

**Applying a membrane compensation to locate interactions within the membrane or at the membrane surface opposite to the pulling AFM tip**

In some cases the contour length suggested that the interaction anchoring the unfolded polypeptide was located at the membrane surface (extracellular) opposite to the pulling AFM tip. To locate this interaction, the thickness of the membrane ( $\approx 4$  nm) was added to the measured contour length of the unfolded polypeptide. This 'membrane compensation' called procedure (Kedrov et al., 2007; Muller et al., 2002) adds  $\approx 11$  amino acids (aa) to the contour length of the unfolded polypeptide locates the interaction inside the membrane (**Figure 2D**). In other cases the anchor of the polypeptide had to be assumed to locate in the membrane. Depending on the location we added ' $n$ ' aa to the contour length with  $n \cdot 0.36$  nm equals the vertical position of the anchor in the membrane.

**Specificity of the unfolding spectra for protein folding**

Force-distance curves such as shown in **Figure 2** are specific for the unfolding of single transmembrane proteins from lipid membranes (Cisneros et al., 2008; Kedrov et al., 2007; Sapra et al., 2009). Taking the examples of various different membrane proteins, others and we have shown that force-distance curves recorded upon unfolding of a membrane protein are specific for the folding of a membrane protein (Engel and Gaub, 2008; Kedrov et al., 2007). It has been also demonstrated that the force-distance spectrum is sensitive to misfolding events of a membrane protein (Kedrov et al., 2004; Kessler et al., 2006). Smallest misfolding events detected so far were  $\alpha$ -helical fragments. Moreover, it has been shown that force-distance curves can detect the functional state of a membrane protein and sensitively react upon changing their functional state (Kedrov et al., 2005; Kedrov et al., 2006). It has also been demonstrated that the force-distance curves of OmpG changes with the functional state of the membrane protein (Damaghi et al., 2010a; Damaghi et al., 2010b). This is because the pH-dependent gating of the transmembrane OmpG pore established interactions that shift individual unfolding peaks. The force-distance curves of membrane proteins can also change when inserting a point mutation (Sapra et al., 2008), replacing a polypeptide loop (Kienberger et al., 2005), or changing the membrane protein assembly (Sapra et al., 2006). Thus, if the unfolded KpOmpA polypeptide would adsorb onto the membrane instead of inserting into the membrane one would expect significant changes of the force-distance curves. Even if the KpOmpA polypeptide would fold into a  $\beta$ -barrel on the surface of the protein membrane all peaks of the force-distance curves would be shifted by at least half the thickness of the KpOmpA membrane ( $\approx 6$  nm). In contrast, we observe that within the accuracy of our method ( $\pm 2-4$  aa) the unfolding force peaks of refolded KpOmpA remain at the same position as observed upon initial unfolding. Thus, we can conclude that KpOmpA has inserted in the membrane and refolded into the same conformation as observed for native KpOmpA.



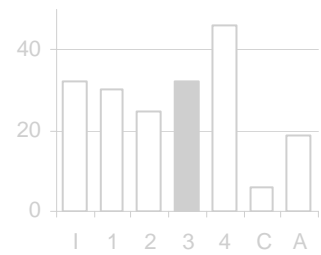
## Supplemental references

- Cisneros, D.A., Oberbarnscheidt, L., Pannier, A., Klare, J.P., Helenius, J., Engelhard, M., Oesterhelt, F., and Muller, D.J. (2008). Transducer binding establishes localized interactions to tune sensory rhodopsin II. *Structure* *16*, 1206-1213.
- Damaghi, M., Bippes, C., Koster, S., Yildiz, O., Mari, S.A., Kuhlbrandt, W., and Muller, D.J. (2010a). pH-dependent interactions guide the folding and gate the transmembrane pore of the beta-barrel membrane protein OmpG. *J. Mol. Biol.* *397*, 878-882.
- Damaghi, M., Sapra, K.T., Koster, S., Yildiz, O., Kuhlbrandt, W., and Muller, D.J. (2010b). Dual energy landscape: the functional state of the beta-barrel outer membrane protein G molds its unfolding energy landscape. *Proteomics* *10*, 4151-4162.
- Engel, A., and Gaub, H.E. (2008). Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* *77*, 127-148.
- Janovjak, H., Kessler, M., Gaub, H.E., Oesterhelt, D., and Muller, D.J. (2003). Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J.* *22*, 5220-5229.
- Kedrov, A., Janovjak, H., Sapra, K.T., and Muller, D.J. (2007). Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* *36*, 233-260.
- Kedrov, A., Krieg, M., Ziegler, C., Kuhlbrandt, W., and Muller, D.J. (2005). Locating ligand binding and activation of a single antiporter. *EMBO Rep.* *6*, 668-674.
- Kedrov, A., Ziegler, C., Janovjak, H., Kuhlbrandt, W., and Müller, D.J. (2004). Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J. Mol. Biol.* *340*, 1143-1152.
- Kedrov, A., Ziegler, C., and Muller, D.J. (2006). Differentiating ligand and inhibitor interactions of a single antiporter. *J. Mol. Biol.* *362*, 925-932.
- Kessler, M., Gottschalk, K.E., Janovjak, H., Muller, D.J., and Gaub, H.E. (2006). Bacteriorhodopsin folds into the membrane against an external force. *J. Mol. Biol.* *357*, 644-654.
- Kienberger, F., Kada, G., Mueller, H., and Hinterdorfer, P. (2005). Single molecule studies of antibody-antigen interaction strength versus intra-molecular antigen stability. *J. Mol. Biol.* *347*, 597-606.
- Muller, D.J., Kessler, M., Oesterhelt, F., Moeller, C., Oesterhelt, D., and Gaub, H. (2002). Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J.* *83*, 3578-3588.
- Renault, M., Saurel, O., Czaplicki, J., Demange, P., Gervais, V., Lohr, F., Reat, V., Piotto, M., and Milon, A. (2009). Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications. *J. Mol. Biol.* *385*, 117-130.
- Renault, M., Saurel, O., Demange, P., Reat, V., and Milon, A. (2010). Solution-state NMR spectroscopy of membrane proteins in detergent micelles: structure of the *Klebsiella pneumoniae* outer membrane protein A, KpOmpA. *Methods Mol. Biol.* *654*, 321-339.
- Sapra, K.T., Balasubramanian, G.P., Labudde, D., Bowie, J.U., and Muller, D.J. (2008). Point mutations in membrane proteins reshape energy landscape and populate different unfolding pathways. *J. Mol. Biol.* *376*, 1076-1090.
- Sapra, K.T., Besir, H., Oesterhelt, D., and Muller, D.J. (2006). Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. *J. Mol. Biol.* *355*, 640-650.
- Sapra, K.T., Damaghi, M., Koester, S., Yildiz, O., Kuhlbrandt, W., and Muller, D.J. (2009). One beta hairpin after the other: Exploring mechanical unfolding pathways of the transmembrane beta-barrel protein OmpG. *Angew. Chem. Int. Ed.* *48*, 8306-8308.
- Schweizer, M., Hindennach, I., Garten, W., and Henning, U. (1978). Major proteins of the *Escherichia coli* outer cell envelope membrane. Interaction of protein II with lipopolysaccharide. *Eur. J. Biochem.* *82*, 211-217.



# Chapter 3

## Solid-state NMR spectroscopy of KpOmpA in lipid bilayers



## Summary

Solid-state NMR spectroscopy has been used to investigate the dynamic behavior of the extracellular loops of the KpOmpA transmembrane domain, after protein reconstitution in *E. coli* lipids at low lipid-to-protein ratio. It is shown that the gradient of dynamics along the barrel axis, previously observed in detergent micelles, is also present in lipid bilayers. Thus the loops mobility is an intrinsic property of the protein and not a function of its environment. The different mobility levels of the loops regions of the molecule are associated with their evolutionary preservation, suggesting that specific immunological properties of KpOmpA may be related with the most flexible loop areas.

This work has been accepted for publication in *BBA Biomembranes* as:

**Iordan Iordanov, Marie Renault, Valérie Réat,**

**Patrick D. Bosshart, Andreas Engel, Olivier Saurel, Alain Milon,**

Dynamics of *Klebsiella pneumoniae* OmpA transmembrane domain: The four extracellular loops display restricted motion behavior in micelles and in lipid bilayers.

doi:10.1016/j.bbamem.2012.05.004

## Title

Dynamics of *Klebsiella pneumoniae* OmpA transmembrane domain: The four extracellular loops display restricted motion behavior in micelles and in lipid bilayers.

Iordan Iordanov<sup>a,b</sup>, Marie Renault<sup>c</sup>, Valérie Réat<sup>a,b</sup>, Patrick D. Bosshart<sup>d</sup>, Andreas Engel<sup>e</sup>, Olivier Saurel<sup>a,b</sup>, Alain Milon<sup>a,b\*</sup>

- a) Institute of Pharmacology and Structural Biology, Université de Toulouse - UPS, 205 route de Narbonne, 31077 Toulouse, France
- b) IPBS – UMR 5089, CNRS, 205 route de Narbonne, BP 64182, 31077 Toulouse, France
- c) UPCAM iSm2 service 512, Campus Scientifique de St Jérôme, 13397 Marseille cedex 20, France
- d) Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland
- e) Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106, USA

### E-mail addresses of authors

iordan.iordanov@ipbs.fr

marie31.renault@gmail.com

valerie.reat@ipbs.fr

patrick.bosshart@bsse.ethz.ch

andreas.engel@case.edu

olivier.saurel@ipbs.fr

alain.milon@ipbs.fr

### Corresponding author

\* Alain Milon

Address: Institute of Pharmacology and Structural Biology, IPBS – UMR 5089, Université de Toulouse – UPS and CNRS, 205 route de Narbonne, BP 64182, 31077 Toulouse Cedex 04, France.

Tel.: +33 (0)5 61 17 54 23

Fax: +33 (0)5 61 17 59 94

E-mail: alain.milon@ipbs.fr

## Abstract

The transmembrane domain of *Klebsiella pneumoniae* OmpA (KpOmpA) possesses four long extracellular loops that exhibit substantial sequence variability throughout OmpA homologues in *Enterobacteria*, in comparison with the highly conserved membrane-embedded  $\beta$ -barrel core. These loops are responsible for the immunological properties of the protein, including cellular and humoral recognition. In addition to key features revealed by structural elucidation of the KpOmpA transmembrane domain in detergent micelles, studies of protein dynamics provide insight into its function and/or mechanism of action. We have investigated the dynamics of KpOmpA in a lipid bilayer, using magic angle spinning solid-state NMR. The dynamics of the  $\beta$ -barrel and loop regions were probed by the spin-lattice relaxation times of the  $C^\alpha$  and  $C^\beta$  atoms of the serine and threonine residues, and by cross-polarization dynamics. The  $\beta$ -barrel core of the protein is rigid; The C-terminal halves of two of the four extracellular loops (L1 and L3), which are particularly long in KpOmpA, are highly mobile. The other two loops (L2 and L4), which are very similar to their homologues in *E. coli* OmpA, and the N-terminal halves of L1 and L3 exhibit more restricted motions. We suggest a correlation between the sequence variability and the dynamics of certain loop regions, which accounts for their respective contributions to the structural and immunological properties of the protein.

## Keywords

outer membrane protein A  
protein reconstitution  
protein dynamics  
spin-lattice relaxation  
solid-state NMR  
electron microscopy

## Abbreviations

KpOmpA:	<i>Klebsiella pneumoniae</i> Outer membrane protein A
TM:	transmembrane
LPR:	lipid-to-protein ratio
CMC:	critical micelle concentration
PLE:	<i>E. coli</i> polar lipids extract
OG:	<i>n</i> -octyl- $\beta$ -D-glucopyranoside
DHPC:	1,2-dihexanoyl- <i>sn</i> -glycero-3-phosphocholine
DOPC:	1,2-dioleoyl- <i>sn</i> -glycero-3-phosphocholine
DMPC:	1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphocholine
ssNMR:	solid-state Nuclear Magnetic Resonance
PDSD:	proton-driven spin diffusion
CP:	cross-polarization

## Highlights

- KpOmpA loop dynamics were assessed in lipid bilayers.
- $T_1$ ,  $T_{1\rho}$ , CP and J-based excitations revealed a variety of motional regimes.
- The L2 and L4 loops display a restricted motional regime.
- The C-terminal halves of L1 and L3 are highly mobile.

### 3.1. Introduction

Membrane proteins represent approximately one-third of all proteins in living organisms [1], are involved in many key physiological processes, and are favored targets for therapeutic drug development [2]. Detailed investigation of membrane protein structure, dynamics and molecular interactions in functional environments is, therefore, a major challenge for structural biology.

Recent advances in NMR spectroscopy have provided unique opportunities to probe the structure of integral membrane proteins at atomic resolution, as well as their associated molecular motions over a large time scale, under different conditions [3, 4]. A combination of refined isotope-labeling schemes and multidimensional transverse-relaxation optimized spectroscopy (TROSY)-based solution-state NMR experiments can be used to determine the entire 3D molecular structures of small to medium-size  $\beta$ -barrel and  $\alpha$ -helical integral membrane proteins in membrane mimetic environments, provided that molecular entities tumble rapidly [5, 6]. Solid-state NMR (ssNMR) spectroscopy offers a complementary spectroscopic tool to monitor the molecular structure and dynamics of larger integral membrane proteins at atomic resolution and in complex settings, such as membrane bilayers [7, 8].

We have previously determined the 3D structure the 210-residue TM domain of the outer membrane protein A from *Klebsiella pneumoniae* (KpOmpA) in DHPC detergent micelles [9]. *Klebsiella pneumoniae* is a Gram-negative bacterium that is responsible for respiratory tract and urinary infections. KpOmpA belongs to the OmpA family of proteins that is well conserved among *Enterobacteria* and has diverse roles in bacterial cellular processes [10-12]. For example, OmpA can function as an adhesin and invasin, participates in biofilm formation, acts as both an immune target and evasin, and serves as a receptor for several bacteriophages [13]. Pore-forming activity of KpOmpA has been reported [14] and subsequently debated [13, 15]. Although sequence conservation within the OmpA family is particularly high in the TM domain (~80%), the extracellular loops exhibit larger variability in terms of length and amino-acid sequence. Interestingly, the extracellular loops of KpOmpA appear to play important roles in the immunological properties of the protein, such as cellular recognition mediated by the scavenger receptors LOX-1 and SREC-I, and humoral recognition mediated by the long pentraxin PTX3 [11, 16]. When solubilized in DHPC micelles, the structure of the TM domain of KpOmpA (Met3-Glu207) consists in an antiparallel 8-stranded  $\beta$ -barrel with four extracellular loops of different lengths and three short periplasmic turns that closely mirror available 3D structures of *E. coli* OmpA TM domains [17-



20]. We reported that the KpOmpA TM domain in DHPC micelles exhibits a mobility gradient, with mobility increasing from the  $\beta$ -barrel core toward the extracellular loops [9]. Although the centre of the  $\beta$ -barrel is highly rigid, the protein segments that surround the two aromatic girdles at the interface exhibit sizable motions on the millisecond time scale. Finally, the extracellular loops are disordered and characterized by nanosecond time-scale motions of different amplitudes.

Key to further characterization of KpOmpA is the investigation of its structure and dynamics at atomic resolution in native-like environments, such as lipid bilayers, where lateral pressure [21], membrane curvature [22], lateral organization, domain formation [23] and protein oligomerization [24, 25] occur and may play important roles in the protein structure, dynamic and function. In order to achieve this aim, we reconstituted the uniformly [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled KpOmpA TM domain in lipid bilayers from a well-folded and soluble form in detergent micelles, and have investigated the folding and dynamics of the protein using ssNMR spectroscopy under magic angle spinning (MAS) conditions. First, the reconstitution of KpOmpA in lipid bilayers was optimized. Successful reconstitutions with a lipid-to-protein ratio (LPR) equal to or below 0.5 (w/w) were analyzed by electron microscopy (EM) to monitor sample homogeneity and the presence of protein patches. Then, KpOmpA-containing proteoliposomes were analyzed by ssNMR spectroscopy to probe the folding and local dynamics of the protein. Using 1- and 2D ssNMR experiments employing either scalar- or dipolar-based magnetization transfer steps associated with  $^{13}\text{C}$  spin-lattice relaxation measurements and cross-polarization dynamics, we assessed the dynamics of the KpOmpA membrane domain in bilayers composed of *E. coli* polar lipids extracts. A new and refined picture of the loops dynamics emerges from the comparison of ssNMR data in lipid bilayers and heteronuclear NOE data previously obtained in a micellar environment [9].

## 3.2. Materials and methods

### 3.2.1. Expression, purification and detergent exchange

The cloning, expression and purification of *Klebsiella pneumoniae* outer membrane protein A (KpOmpA) TM domain was performed as described previously [9]. For the small-scale reconstitution trials and sucrose gradient centrifugation, the unlabeled protein was purified from *E. coli* BL21 (DE3) culture grown in Terrific Broth (Invitrogen) medium. For NMR experiments, the bacteria were grown in M9 minimal medium supplemented with 2 g/L  $^{15}\text{NH}_4\text{Cl}$  and 2 g/L U- $^{13}\text{C}_6$  D-glucose (Cambridge Isotope Laboratories, Inc.) as the only

nitrogen and carbon sources, respectively, producing uniformly [<sup>13</sup>C, <sup>15</sup>N]-labeled KpOmpA TM domain.

For detergent exchange, purified KpOmpA/Zwittergent 3-14 complexes were incubated with nickel-chelating resin (Ni-NTA Superflow, QIAGEN) for 3 hours at 4°C under gentle mixing. The resin-bound protein was washed three times with 50 ml of 25 mM Tris/HCl (pH 8.5), 150 mM NaCl, 2% *n*-octyl-β-D-glucopyranoside (OG, Anatrace) and isolated by centrifugation (4000 g, 10 min, 4°C). The KpOmpA/OG complexes were then eluted in one step with 25 mM Tris/HCl (pH 8.5), 150 mM NaCl, 2% OG and 400 mM imidazole. The concentration of the purified protein (either in 0.1% Zwittergent 3-14 or in 2% OG micelles) was determined spectrophotometrically by measuring absorbance at 280 nm and using the theoretical protein molar extinction coefficient of 50880 M<sup>-1</sup>cm<sup>-1</sup>. The successful refolding of the protein in both detergents was checked with SDS PAGE [26].

### 3.2.2. Reconstitution

Protein reconstitution was examined in three types of lipid: DOPC, DMPC and *E. coli* polar lipids extract (PLE) (all from Avanti Polar Lipids). Proteoliposomes containing the TM domain of KpOmpA were prepared using a detergent dilution method adapted from procedures described by Rigaud et al. [27]. Briefly, each lipid type was dissolved in chloroform, and dried under a stream of nitrogen gas and then in a vacuum chamber overnight. Mixed micelles were prepared by hydration of the lipid film with 25 mM Tris/HCl (pH 8.5), 150 mM NaCl and 2% OG with a final lipid concentration of 6 or 10 mg/ml. Purified and unlabeled monomeric KpOmpA TM domain (at a concentration of 1-3 mg/ml, depending on the different preparations) was added to the mixed micelles solution at the desired lipid-to-protein ratio (LPR) and incubated for 10 min at 4°C. The solution of KpOmpA/OG/lipid ternary complexes was then dialyzed twice (12 h each) against 2 L of the desired detergent-free buffer at 37°C to achieve the complete removal of detergent and imidazole and the formation of large unilamellar vesicles (LUVs). In order to screen a large panel of experimental conditions, the reconstitution trials were conducted in dialysis buttons of ~70 µl volume. The LPRs tested were between 0.1 and 1 (in 0.1 steps), 10 and 20 (w/w). The different dialysis buffers used in these small-scale reconstitution trials were: Buffer A (140 mM NaCl, 20 mM HEPES, pH 7.0), Buffer B (300 mM NaCl, 20 mM HEPES, pH 7.0), Buffer C (100 mM NaCl, 20 mM Tris, pH 8.5), Buffer D (140 mM NaCl, 20 mM Tris, pH 8.5) and Buffer E (300 mM NaCl, 20 mM Tris, pH 8.5). All buffers contained 0.01% NaN<sub>3</sub> to prevent the growth of contaminating bacteria.

### 3.2.3 NMR sample preparation

For solid-state NMR (ssNMR) experiments, 22 mg of purified, uniformly [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled KpOmpA in 25 mM Tris buffer (pH 8.5) containing 150 mM NaCl and 2% OG was reconstituted in 11 mg of *E. coli* PLE as described above, resulting in an LPR of 0.5 (w/w), equal to a molar ratio of ~15 lipid molecules per protein molecule. The mixture was transferred to a dialysis tube with a 12-14 kDa cutoff (SpectraPor) and dialyzed twice against 2 L of 25 mM Tris/HCl (pH 8.5), 150 mM NaCl. Following the removal of detergent and imidazole, and the formation of LUVs during the first two dialysis steps, the dialysis buffers were exchanged every 12 hours with a gradual decrease in the concentration of NaCl in 25-mM steps until the salt was completely removed. The size of the resulting proteoliposomes was verified by dynamic light scattering (DynaPro NanoStar, Wyatt Technology Corporation) and had a major population of particles that were ~1-3  $\mu\text{m}$  in diameter (data not shown). The LUVs were collected by centrifugation (200 000g, 90 min, 10°C) and partially dehydrated under a stream of nitrogen gas until the total sample weight was reduced to ~50 mg. This value corresponded to approximately 33 mg of protein/lipid material with an LPR of 0.5 (w/w) and 17 mg water (30 w/v% of the entire sample). The proteoliposomes were then transferred to a 4-mm magic angle spinning (MAS) rotor by low-speed centrifugation (3000 g, 10 min, 10°C) and subsequently analyzed by ssNMR spectroscopy.

### 3.2.4 Sucrose gradient centrifugation and colorimetric assays

Unlabeled KpOmpA-containing liposomes prepared in exactly the same way as the ssNMR sample (but not dehydrated under a stream of nitrogen gas) were collected (15 000 g, 20 min, 4°C) and subjected to centrifugation at 4°C in a 20-60% continuous sucrose gradient for 18 hours at 100 000 g in SW41 rotor (Beckman) for an assessment of homogeneity and verification of the LPR. The isolated single band was subsequently analyzed for protein content by Lowry titration [28] and for lipid content by phosphoric acid titration [29].

### 3.2.5 Transmission electron microscopy

The reconstitution of KpOmpA into lipid bilayers was checked by transmission electron microscopy. For that purpose the sample was adsorbed to a carbon-coated copper grid, which was rendered hydrophilic by glow discharge at low pressure. The grids were then washed in nanopure water (resistivity > 18 M $\Omega$ •cm), stained with 2% (w/v) uranyl acetate, blotted and air-dried. Electron micrographs were recorded on a charge-coupled device (CCD)

with a CM100 transmission electron microscope (FEI, Eindhoven, Netherlands) operated at 80 kV acceleration voltage.

### 3.2.6 NMR spectroscopy

ssNMR experiments were carried out on a Bruker AVANCE standard-bore NMR spectrometer (Bruker Biospin) operating at a  $^1\text{H}$  Larmor frequency of 700 MHz and equipped with a 4 mm-double resonance  $^1\text{H}\{-^{15}\text{N}\text{-}^{13}\text{P}\}$  cross-polarization (CP)-MAS probe. The MAS frequency was set to 12 kHz and the sample temperature was regulated at 20°C. The typical  $\pi/2$  pulse length for  $^1\text{H}$  was 3.1  $\mu\text{s}$ . Unless otherwise indicated, the  $^1\text{H}\text{-}^{13}\text{C}$  CP magnetization transfer step employed a linear ramp (50 to 100% field strength) on the  $^1\text{H}$  channel. 1D CP-MAS spectra were obtained using a CP contact time of 400  $\mu\text{s}$  and a  $^{13}\text{C}$  radio frequency field of 60 kHz. High-power proton decoupling was obtained using TPPM [30] schemes and a decoupling field of 80 kHz. 2D  $^{13}\text{C}\text{-}^{13}\text{C}$  proton-driven spin diffusion (PDSD) [31] experiments were performed using similar acquisition parameters and a PDSD mixing time of 40 ms. The  $^{13}\text{C}$  spin-lattice relaxation measurements were achieved using 2D  $T_1$ -edited ( $^{13}\text{C}\text{-}^{13}\text{C}$ ) PDSD correlation experiments (*i.e.*, with a  $90^\circ\text{-}\tau\text{-}90^\circ$  element between the  $^1\text{H}\text{-}^{13}\text{C}$  CP and  $t_1$  evolution period). Spectra were obtained for nine spin-lattice relaxation times (0.02, 10, 100, 250, 500, 700, 1000, 2000 and 4000 ms), using 48 scans for each of the 364  $t_1$  increments. Characteristic  $\text{C}^\alpha\text{-C}^\beta$  cross-peaks of Thr and Ser residues were integrated. Cross-peak integrals were plotted as a function of the  $T_1$  relaxation delay ( $\tau$ ) and fitted using a mono-exponential decay function to determine  $T_1$  relaxation time. The CP dynamics measurement relied on 13 2D  $^{13}\text{C}\text{-}^{13}\text{C}$  PDSD correlation experiments using 76 scans for each 256  $t_1$  increments and employing CP contact times of 0.025, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 1, 2, 4 and 8 s. Cross-peaks intensities of Thr and Ser  $\text{C}^\alpha\text{-C}^\beta$  were plotted as a function of the CP contact time and fitted with the equation of the two-stage model to determine the  $T_{\text{CP}}$  and  $T_{1\rho}$  values [32]. Uncertainty in values was determined from the standard errors of the fitting. 1D ( $^1\text{H}\text{-}^{13}\text{C}$ ) refocused-INEPT (with  $J_{\text{CH}}$  of 142 Hz) and  $^{13}\text{C}$  direct excitation experiments employed  $\pi/2$  pulse lengths of 3.1  $\mu\text{s}$  ( $^1\text{H}$ ) and 4  $\mu\text{s}$  ( $^{13}\text{C}$ ), corresponding to  $^1\text{H}$  and  $^{13}\text{C}$  RF fields of 80 and 62 kHz, respectively. Similarly to CP-based ssNMR experiments, proton decoupling was achieved using SPINAL64 schemes [33] and a decoupling field of 80 kHz. For  $^{13}\text{C}$  direct and CP measurements, 2K scans were acquired, and 12K scans were acquired for the refocused-INEPT experiment.  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts were referenced with respect to DSS. Spectra were processed using topsin1.3 (Bruker Biospin) and  $T_1$  and CP dynamic curves were fitted using the GOSA program [34].

### 3.3. Results

#### 3.3.1. Screening of experimental conditions for the reconstitution of KpOmpA<sup>TM</sup> domain in lipid bilayers

Prior to the overexpression, purification and reconstitution of labeled *Klebsiella pneumoniae* outer membrane protein A (KpOmpA) for solid-state NMR (ssNMR) experiments, several reconstitution trials were conducted with non-labeled protein in order to optimize the conditions for the formation of proteoliposomes in terms of lipid composition, lipid-to-protein ratio (LPR) and dialysis buffer. The examination of negatively stained preparations by electron microscopy represents a suitable tool for this purpose because it enables rapid visualization and diffraction pattern-based analysis of the different preparations in order to identify regions of long-range ordered molecules in the bilayer. The goal of these pilot experiments was to establish conditions suitable for the arrangement of the protein in 2D crystals, which is beneficial for the ssNMR spectroscopic sensitivity and spectral resolution.

The reconstitution of KpOmpA into liposomes was achieved by slow dialysis-driven detergent removal. This is particularly important in the case of samples with low LPRs, in which the relatively large amount of protein must receive sufficient time for incorporation into the newly formed bilayers. Because the detergent used for the protein purification (Zwittergent 3-14) has a low critical micellar concentration (CMC ~0.012%), it is unsuitable for removal in this manner and therefore was exchanged for the high-CMC detergent *n*-octyl- $\beta$ -D-glucopyranoside (OG, CMC ~0.7%), which is commonly used for the solubilization of *E. coli* OmpA [35]. The homogeneity of the micelle size was checked with gel filtration (data not shown) and the protein was found to be monomeric, as expected. **Fig. 1E** (lanes 1 to 4) shows an SDS PAGE gel of purified KpOmpA<sup>TM</sup> domain in 0.1% Zwittergent 3-14 (lanes 1 and 2) and in 2% OG (lanes 3 and 4), with the folded (lanes 1 and 3) and unfolded (heat-denatured, lanes 2 and 4) states visualized by the different migration of the bands [26]. The protein in 2% OG was then mixed with a particular lipid type and dialyzed following the procedure described in the materials and methods section to obtain proteoliposomes. The different trials were checked macroscopically, and then by electron microscopy.

We tested three types of lipids: DOPC, DMPC and *E. coli* polar lipids extract (PLE). These lipid types were chosen to comprise a variety of chain lengths and acyl chain saturation levels: the DOPC has long and unsaturated chains (18:1/18:1); the DMPC is shorter and saturated (14:0/14:0); and the “native-like” lipid mixture of the *E. coli* PLE contains different levels of lengths and saturation in its composition of L- $\alpha$ -

phosphatidylethanolamine (67.0%), L- $\alpha$ -phosphatidylglycerol (23.2%) and cardiolipin (9.8%). The dialysis buffers tested encompassed different pH and ionic strengths. The range of LPRs investigated varied from 0.1 to 20 (w/w), corresponding to ~1 protein for between 3 and 600 lipid molecules. The results of these trials to determine appropriate conditions for the formation of proteoliposomes are reported in **Fig. 1A**.

We initially observed that samples contained high levels of macroscopically visible aggregated protein and/or vesicles, usually in conditions of lower pH and higher salt concentrations. The negative effect of low pH on sample quality is illustrated by the DOPC sample, which produced small and aggregated vesicles (**Fig. 1B**, pH 7.0). Based on the reconstitution trials, we concluded that, independently of the chosen LPR and ionic strength, the lower pH (HEPES at pH 7.0) does not appear to be as suitable for the formation of proteoliposomes as the higher pH (Tris at pH 8.5). This result agrees with previous observations that OmpA (as well as several other OmPs) exhibits improved folding in slightly basic conditions [36, 37]. We therefore excluded the samples at pH 7.0 from further experiments.

Because NMR is a demanding technique in terms of sensitivity, it is imperative to keep the LPR to a minimum in order to maximize the quantity of protein in the sample. We systematically observed aggregated samples at very low LPRs (i.e., 0.1-0.3 w/w) for any buffer condition, and to a lesser extent for smaller LPRs (i.e., 0.3 and 0.4) at higher pH and salt concentrations. As an illustration of the lack of sample homogeneity at low LPR values, **Fig. 1C** shows an aggregated DMPC sample with an LPR of 0.2 and a high pH (8.5). Such a low LPR (corresponding to ~6 lipid molecules per protein molecule), in the absence of 2D protein crystallization, threatens overall sample homogeneity and might lead to aggregation of at least part of the protein population. All of the lipids tested produced usable samples in conditions where neither the pH nor the LPR were too low. An LPR of 0.5 (w/w, corresponding to about 15 lipids per protein molecule) is acceptable for our purposes, as this condition significantly reduces the probability of protein aggregation and still enables the preparation of a sample for NMR containing more than 20 mg of protein. For example, **Fig. 1D** demonstrates the homogeneity of a typical sample at this LPR value in *E. coli* PLE. Control 1D and 2D PDS  $^{13}\text{C}$  NMR spectra were performed at LPR = 1 and no difference was observed with those at LPR = 0.5. Therefore, samples with an LPR of 0.5 were selected for use in ssNMR experiments, and samples with an LPR above this value and a high salt concentration were discarded (see gray squares in **Fig. 1A**).

Further examination of the different pilot samples led to the conclusion that the most suitable lipid type for our purposes was the *E. coli* PLE, although we concluded that all three lipids tested were appropriate for the dialysis-based reconstitution of KpOmpA in these conditions. However, the substantially longer chains of DOPC (18:1/18:1) form bilayers with a thickness (~5.5 nm at room temperature [38]) that does not correlate closely with the section of the protein known to be embedded in the membrane (~2 nm), and this hydrophobic mismatch could result in altered dynamics at the interface between the  $\beta$ -barrel and the loops. Conversely, DMPC has the disadvantage of a higher gel to fluid phase transition temperature (~23°C). However, the *E. coli* PLE is usable at room temperature and should mimic the protein's natural environment in the outer membrane of Gram-negative bacteria in terms of polar-head composition.

Finally, in the range of 100-300 mM NaCl, no significant influence of the ionic strength on protein reconstitution was observed by electron microscopy. The electron diffraction experiments provided no evidence, in any sample preparations, of the existence of 2D-crystals, such as those observed with OmpG [39]. The monomeric nature of KpOmpA and the geometrical symmetry of its  $\beta$ -barrel, combined with the lack of structural features provided by the loops [40] presumably contribute to this lack of long-range 2D order. Based on the screening of experimental conditions, we finally decided to reconstitute the KpOmpA TM domain in *E. coli* PLE at an LPR of 0.5 (w/w) and in buffer C (i.e. pH 8.5) without salts, in order to prevent sample heating during the NMR experiments. The large-scale proteoliposome preparation for NMR spectroscopy was first performed with unlabeled protein and the full-scale sample was further characterized by sucrose gradient centrifugation. The gradient centrifugation produced a single, well-defined band at ~45-47% sucrose that was subsequently analyzed with protein and phospholipid colorimetric assays. The LPR of the isolated band was confirmed to be ~0.5 (w/w). The proteoliposomes generated, regardless of their variation in size and aggregation state, possess the same LPR, as demonstrated by the single band in the sucrose gradient centrifugation. The folded and unfolded states of the reconstituted protein were also verified with SDS PAGE (**Fig. 1E**, lanes 5 and 6).

### 3.3.2. ssNMR characterization of KpOmpA TM domain in *E. coli* PLE vesicles.

**Fig. 2A** shows the 2D ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) proton-driven spin diffusion (PDSD) correlation spectrum of (U- $^{13}\text{C}$ ,  $^{15}\text{N}$ )-KpOmpA/PLE proteoliposomes (LPR of 0.5, w/w) recorded at 20°C with 40 ms mixing time. Under these conditions, the lipid bilayer is fluid and the ssNMR spectrum is dominated by the intraresidue correlation network of the protein. Characteristic

one- to three-bond intraresidue CC connectivities were identified for several amino-acid types, including Thr, Ser, Pro, Ala and Ile (**Fig. 2A**, boxes). As an example, the complete correlation network between the  $\alpha$ ,  $\beta$  and  $\gamma_2$  carbons of two Thr residues is shown (solid and dashed lines). According to standard amino acid-peak positions [41] and the overall correlation pattern, the PDS spectrum of KpOmpA reflects predominantly  $\beta$ -sheet and random coil protein segments, in accordance with the topological profile of the protein in detergent micelles. Spectral resolution is given by  $^{13}\text{C}$  line-widths of 0.6-0.9 ppm for isolated resonances (**Fig. 2B**). This is in agreement with the literature, where the typical  $^{13}\text{C}$  line-widths observed by ssNMR vary from 0.5 to 0.7 ppm for small (<100 residues) uniformly [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled proteins in micro-crystalline form [42-44] and seven-helix receptors (NpSRII) in native membranes [45], respectively. Although the spectral resolution is high, the CC correlation spectrum is characterized by a significant resonance overlap resulting from the high frequency of occurrence of particular residue types in  $\beta$ -sheet TM segments (e.g., Leu and Val) and loop regions (e.g., Ser and Arg), which significantly hampered the resonance assignment procedure and residue-specific analysis. We thus compared our results with data obtained using the perdeuterated KpOmpA TM domain in DHPC micelles (BMRB accession code 15651). Prior to chemical-shift analysis,  $^{13}\text{C}^{\alpha}\text{C}^{\beta}$  chemical shifts were corrected for the  $^2\text{H}$ -isotope effect as described by Venters et al. [46]. The intraresidue correlation network was then calculated from the corrected solution NMR chemical shifts and the prediction was overlaid with the 2D PDS spectrum of KpOmpA (**Fig. S1**). A remarkable agreement between the solution NMR chemical shifts and the overall CC correlation pattern of KpOmpA was observed, suggesting that the global fold of the protein in DHPC micelles is well preserved in PLE bilayers. Notably, isolated backbone resonances corresponding to Thr, Ser, Val, Ala and Leu residues within KpOmpA TM segments were readily observed in the vicinity of peak positions predicted from solution NMR chemical shifts (**Fig. S1**, solid squares). In contrast, significant deviations between ssNMR and solution NMR data were observed for a set of residues located in the unstructured N-terminal extremity (Ile3, Ile7) and in the first (Asn63 and Pro64) and second (Pro103 and Ile104) periplasmic turns, suggesting that conformational and/or dynamical changes may occur around these residues in the lipid environment. Finally, correlations corresponding to residues within extracellular loops were also visible, suggesting a reduced mobility within unstructured protein segments in the lipid environment (**Fig. S1**, dashed squares).

The contribution of the unstructured extracellular loops to the KpOmpA PDS spectrum was investigated in more detail by examining the spectral region of the  $\text{C}^{\alpha}\text{-C}^{\beta}$  resonances of



the Thr and Ser residues (**Fig. 3A**), which are well-distributed within the extracellular loops (from L1 to L4) and  $\beta$ -sheet TM segments (TM1, TM3-TM8) of the protein (**Fig. 3C and D**). Taking into account the number of residues forming the cross-peaks, the intensity is systematically smaller when the residue is located in the loop than in the  $\beta$ -sheet (**Fig. 3B**). Considering the Thr residues, the intensity of the  $C^\beta$ - $C^\alpha$  cross-peak is much smaller for the five residues constituting the loops (T-33,47,130,174,177) compared with the seven residues of the  $\beta$ -barrel (T-18,97,105,112,139,154,159). Similarly, the signals of Ser residues within the loops is smaller compared with Ser signals in the  $\beta$ -sheet protein segments, and only three of eight assigned Ser residues are located into the  $\beta$ -barrel. This difference in the intensities is obviously induced by different dynamic regimes in two distinct moieties of the protein: the TM domain and the loops. Thus the CP + PDS transfer schemes select a subset of the loops' resonances, attenuating but not completely suppressing the most mobile signals arising from the loop region. Further characterization of these remaining resonances is presented in the following section.

### 3.3.3. Dynamics of KpOmpA over different NMR time scales

The rotational diffusion correlation times of membrane embedded proteins are much slower than for small soluble proteins (typically above microseconds [47]) and NMR relaxation rates are thus sensitive to dynamics up to microsecond timescales. Various ssNMR methods have been developed to measure  $R_1$ ,  $R_{1\rho}$ , and heteronuclear NOEs [48-50]. Using extended model free formalism, spectral density functions could be expressed in terms of order parameters and correlations times for slow and fast motions and dynamics time scale could be determined [51, 52].

To investigate the different degrees of molecular mobility within the KpOmpA TM domain, we recorded a series of three  $^{13}\text{C}$ -detected 1D spectra using different excitation schemes (*i.e.*, direct  $^{13}\text{C}$  excitation,  $^1\text{H}$ - $^{13}\text{C}$  refocused-INEPT and  $^1\text{H}$ - $^{13}\text{C}$  CP, **Fig. 4**). The direct  $^{13}\text{C}$  excitation spectrum (**Fig. 4A**) constitutes the reference spectrum, in which signal intensities are most directly related to the number of carbons. Accordingly, Thr  $C^\beta$  resonances revealed similar peak intensities for residues within the  $\beta$ -sheet (B peak at 71.8 ppm, 7 Thr) and random coil segments (L peak at 70 ppm, 5 Thr). In contrast, the  $^1\text{H}$ - $^{13}\text{C}$  INEPT spectrum (**Fig. 4C**) displays signal intensity only at 70 ppm, which arises from the loop regions. Indeed, the  $^1\text{H}$ - $^{13}\text{C}$  INEPT type experiments employ a magnetization transfer step mediated by  $^1\text{H}$ - $^{13}\text{C}$  scalar (through-bond) coupling. This coupling takes a few milliseconds to establish, during

which time the magnetization transfer mediated by strong dipolar couplings vanishes due to fast  $R_2$  relaxation for rigid segments. Therefore, only protein segments undergoing fast (ns) isotropic motions persist in the INEPT spectrum. In contrast, CP-based experiments filter out magnetization from spins in fast motional regimes, and thus signals from rigid protein regions dominate the spectrum. As expected, in the CP spectrum an intense peak was observed at 71.8 ppm, characteristic of rigid protein TM segments (**Fig. 4B**). Interestingly, a signal of lower intensity is still observed at 70 ppm in the CP spectrum, indicating the presence of restricted protein motions in the unstructured extracellular loops. These results suggest the presence of multiple motional regimes within the extracellular loops of KpOmpA.

Next, we measured the  $^{13}\text{C}$  spin-lattice relaxation time ( $T_1$ ) of the  $\text{C}^\alpha$  and  $\text{C}^\beta$  of the Ser and Thr residues from 2D PDS  $^{13}\text{C}$ - $^{13}\text{C}$  spectra by incorporating a longitudinal relaxation delay between the CP and the first  $^{13}\text{C}$  time evolution.  $^{13}\text{C}$   $T_1$  relaxation is dominated by the C-H dipolar interaction mechanism and depends on the motional regime. The results of this experiment can only be discussed qualitatively, because they are averaged over several overlapping residues and because of the contribution of proton driven carbon-carbon spin diffusion [53]. We systematically observed shorter  $T_1$  values for the residues constituting the loops compared with those in the  $\beta$ -sheets (**Fig. 5A**). For example, for the Thr  $\text{C}_\alpha$  carbons:  $T_{1-\text{C}\alpha} = 0.46 \pm 0.16$  s in the loops, and  $T_{1-\text{C}\alpha} = 1.14 \pm 0.07$  s in the  $\beta$ -barrel. For the Ser  $\text{C}_\alpha$  carbons:  $T_{1-\text{C}\alpha} = 0.32 \pm 0.18$  s in the loops and  $T_{1-\text{C}\alpha} = 1.63 \pm 0.30$  s in the  $\beta$ -barrel. This result reveals that there are increased motions in the loops around the  $^{13}\text{C}$  Larmor frequency, because the  $R_1$  longitudinal relaxation rate ( $1/T_1$ ) is enhanced by these motions, with characteristic times in the nanosecond and sub-nanosecond range.

The dynamics of the CP were examined by using contact times from 50  $\mu\text{s}$  to 8 ms on the 2D  $^{13}\text{C}$ - $^{13}\text{C}$  spectrum. The CP dynamics are described most simply by a two-stage model. The “build-up” phase starts at very short contact times and is caused by the initial CP transfer from  $^1\text{H}$  to the closest  $^{13}\text{C}$  via  $^1\text{H}$ - $^{13}\text{C}$  dipolar interactions. This CP transfer is usually characterized by a constant time, called  $T_{\text{CP}}$ . This initial phase is followed by a “decay” phase caused by the relaxation of the spin-locked  $^1\text{H}$ , with a characteristic time  $T_{1p}$  [32]. We observed that  $T_{\text{CP}}$  values for each residue and carbon type are systematically longer when the Ser and Thr residues are located in a loop rather than in the  $\beta$ -barrel (**Fig. 5B**). For example,  $T_{\text{CP}}$  for the Thr  $\text{C}^\beta$  is equal to  $51 \pm 10$   $\mu\text{s}$  in the  $\beta$ -barrel and increases to  $108 \pm 14$   $\mu\text{s}$  in the loops. Similarly,  $T_{\text{CP}}$  for the Ser  $\text{C}^\beta$  is  $26 \pm 10$   $\mu\text{s}$  in the  $\beta$ -barrel and  $82 \pm 14$   $\mu\text{s}$  in the loops. This slower polarization transfer in the loop regions is caused by partial averaging of

the  $^1\text{H}$ - $^{13}\text{C}$  dipolar interaction by the greater mobility of the loops in the sub-microsecond range.

The  $T_{1\rho}$  values extracted from the 2D CP dynamic experiment analysis vary from 1.6 to 3.6 ms depending on the carbon and residue types (**Fig. 5C**). These values are in agreement with values reported for proteorhodopsin in a lipid environment [54]. For the analysis of the spin-lattice relaxation in the laboratory frame ( $T_1$ ) analysis, the comparison was restricted to  $T_{1\rho}$  values for the same carbon in two distinct secondary structure elements. In these conditions, both the  $\text{C}^\alpha$  and  $\text{C}^\beta$  of the Thr residues exhibited larger  $T_{1\rho}$  values for residues in the  $\beta$ -barrel compared with the loops (i.e., Thr  $\text{C}^\alpha$   $T_{1\rho} = 3.72 \pm 0.48$  ms and  $1.75 \pm 0.29$  ms, respectively). The faster average  $R_{1\rho}$  relaxation rate in the loops is probably related to the Thr localization: three of five Thr residues constituting the loops (T33, T174 and T177, in orange in **Fig. 6**) are located in the region of the loops that undergo restricted mobility with motions in the ms to  $\mu\text{s}$  range, which are known to be important for  $R_{1\rho}$  relaxation. In contrast with the Thr residues, the  $\text{C}^\alpha$  of the Ser residues present in the loops exhibited less efficient  $R_{1\rho}$  relaxation and the majority of the Ser residues participating in the loop cross-peak are located in the C-terminal region of L3 (S129, S133 and S135), which is highly flexible. These Ser residues are expected to be dominated by fast isotropic motions, outside the millisecond to microsecond range, and therefore to give less efficient  $R_{1\rho}$  relaxation.

### 3.4. Discussion

The structure and dynamics of KpOmpA have been studied by NMR in two distinct environments: in DHPC micelles [9]; and in *E. coli* polar lipids extract (PLE) proteoliposomes (present work). Careful optimization of the reconstitution protocol has permitted the preparation of homogeneous liposomes with a large lipid to protein ratio (LPR = 0.5, typically 22 mg protein for 11 mg lipid, molar ratio of 1 to 15) that is compatible with a detailed ssNMR characterization, which is quite demanding in terms of sensitivity. Electron microscopy demonstrated the samples to be homogeneous under these conditions, with no indication of 2D crystallization of the protein in the lipid bilayer planes. 2D PDSD showed good resolution of the  $^{13}\text{C}$  resonances and significant similarity between  $^{13}\text{C}$  chemical shifts in bilayers and in DHPC micelles, revealing that the overall fold of the protein is identical in both environments. 2D PDSD acquired at a LPR of 1 and 0.5 were identical, suggesting that the molecular crowding did not affect much the protein dynamics in this range of protein concentration. The lipid to protein ratio in Gram negative bacteria outer membranes is estimated to in the 0.5 to 1

range [55-57], i.e. not so far from our experimental conditions. The KpOmpA TM domain has been demonstrated to exhibit ns- to  $\mu$ s-timescale molecular motions when embedded in detergent micelles [9]. Heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  NOE relaxation of the backbone  $^{15}\text{N}$  spins was used to probe the picosecond to nanosecond dynamic fluctuations of individual N-H bonds, chemical exchange line-broadening of the backbone amide resonances were sensitive to microsecond to millisecond backbone motions, and  $^2\text{H}_2\text{O}/\text{H}_2\text{O}$  solvent exchange experiments were used to probe slower dynamic processes (seconds to days). A mobility gradient was observed, from the rigid  $\beta$ -barrel embedded in the hydrophobic core of the micelles to the disordered and highly dynamic loops. At the boundary of these two regions (i.e., the interface region), residues exhibited intermediate motions with characteristic times of micro- to milliseconds, as reflected by the extended conformational exchange-induced line broadening. In most of the  $\beta$ -barrel-to-loop interface, and despite the large heteronuclear NOE values that reflect an ordered structure, residues belonging to the  $\beta$ -barrel (**Fig. 6**, squares in orange) exhibit conformational exchange, characterized by line-broadening at different temperatures and static magnetic field strengths. This conformational exchange is propagated toward the loops, in which regions the induced line-broadening increased to the extent that it hampered the assignment of resonances (**Fig. 6**, orange circles filled with gray). Moreover, restricted mobility was measured from heteronuclear NOE values (average value of  $0.6 \pm 0.1$ ) for residues with random coil chemical shifts from S28 to N38 in L1, from D122 to Y127 in L3 and for the entire L2 and L4 loops (Fig. S3 in [9]).

In lipid bilayers, KpOmpA dynamics had to be studied by ssNMR approaches, such as  $T_2$ -filter, dipolar- and scalar-based polarization transfer, to identify mobile and static moieties [54, 58]. Several approaches to assess internal dynamics in a site-specific manner have been reported, such as  $^{15}\text{N}$  spin-lattice relaxation rates and averaged dipolar and/or CSA interactions [49, 50, 59]. However, due to the relatively large molecular weight of KpOmpA, a residue type specific approach was used, with observations focused on the Thr and Ser resonances, which are well-distributed within the 3D fold, as shown in **Fig. 3**. All of the evidence described in the results section indicates that KpOmpA has an identical structure and dynamics in micelles and in lipid bilayers. The 2D PDSD experiments demonstrate that the carbon chemical shifts are highly similar in both environments, as shown in **Fig. S1**. Ser and Thr are well-distributed throughout the  $\beta$ -sheet and random coil environments. The resonances from residues in the  $\beta$ -barrel possess all of the characteristics of rigid segments; for example, shorter  $T_{CP}$  build-up rates, the absence of signal in INEPT-based excitation and longer  $T_1$  relaxation times. Loops L2 and L4, and the N-terminal parts of loops L1 and L3 are

subjected to millisecond to microsecond motions (as revealed by longer  $T_{CP}$  and shorter  $T_{1\rho}$  times for Thr signals), whereas the C-terminal part of loops L1 and L3 are subjected to motions of higher frequency and/or higher amplitude (as revealed by the Thr signals observed in an INEPT excitation scheme, and by the longer  $T_{1\rho}$  of the Ser 129, 133, 135 resonances in L3).

Taken together, the differences observed in the behavior of the Ser and Thr residues suggest that the dynamics with an intermediate time scale that were observed for KpOmpA in micelles [9] persist at the membrane interface in presence of lipid bilayers, where there are specific constraints such as lipid lateral pressure and membrane strain curvature. This conservation of dynamics in different conditions could be related to the amino acid composition of the protein at the interface. Indeed, it has been demonstrated that the mutation of Trp residues to Phe in OmpA from *E. coli* reduces the conformational exchange phenomena [19] and allows the assignment of more residues at the interface. However, no conformational exchange processes probed by relaxation dispersion experiments in the millisecond to microsecond range were revealed for any of the observable residues of *E. coli* OmpA, suggesting the absence of global backbone conformational changes [60]. Such global changes have been reported for PagP in micelles [61].

The C-terminal halves of L1 and L3 in KpOmpA exhibit typical loop behavior, with random coil chemical shifts, few unassigned residues and fast isotropic dynamics (**Fig. 6**). Approximately half of the residues in L2 and L4 remain unassigned, due to millisecond to microsecond conformational exchange processes. These exchange phenomena seem to propagate up to the center of the loops that presents a restricted mobility. Interestingly, the alignment of KpOmpA with its *E. coli* homologue shows a high level of similarity between their respective loops L2 and L4, whereas the other two loops (L1 and L3) are more diverse (**Fig. S2**). It is unclear whether the conservation of L2 and L4 is the result of evolutionary pressure, similar to that exerted on the conserved  $\beta$ -barrels of these molecules. The structural importance of the extracellular loops is still discussed. Datta et al. [62] suggested that inter-loop interactions are critical for maintaining the 3D structure of OmpA from *E. coli*, whereas Koebnik claimed that the structural integrity of the protein is entirely attributed to the  $\beta$ -barrel and turns region, because loop-truncated mutants did not exhibit thermal instability [40]. Surprisingly, in the same study by Koebnik et al, these highly disordered loops were not digested by subtilisin, provided that the full complement of loops was present, although the truncated mutants were more susceptible to digestion. We also observed (data not shown) for KpOmpA that Lysine-C endoprotease is not able to cleave the loops, despite the fact that

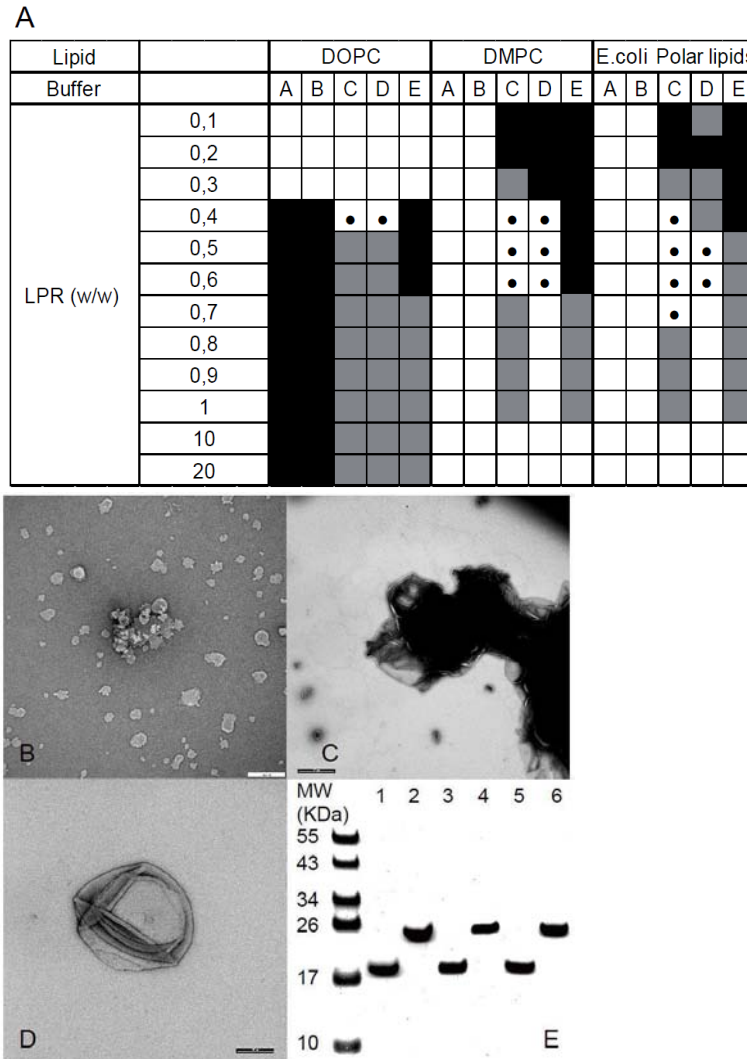
there are two Lys residues in L2 and one in the L3 N-terminal half. Hence, our NMR data correlates well with the results of biochemical experiments. It was previously reported that *E. coli* OmpA and KpOmpA are necessary for invading brain endothelial cells [63] and bronchial epithelial cells [64], respectively, and that both proteins participate in bacterial survival in macrophages. A detailed study with a variety of mutants established the importance of the different loops of *E. coli* OmpA during these immunological processes [65, 66]. L2 contributes to bacterial survival in dendritic cells and polymorphonuclear lymphocytes, as well as influencing the immune response by binding, together with the more conserved L4, to the complement pathway regulator C4bp. Interestingly, the loop mutants that exhibited reduced invasiveness were those with altered amino acids in the interfacial, less dynamic regions of L2 and L4, as well as the less flexible segment of L1 emanating from the  $\beta$ -barrel immediately after the first  $\beta$ -sheet (**Fig. 6**).

Taken together, these data suggest that the reduced mobility in certain loop areas correlates with their level of evolutionary conservation (highest in L4) and their propensity for involvement in bacteria invasion. This could be explained either by the contribution of these regions to the stability of the underlying, and similarly well-conserved  $\beta$ -barrel and/or by their function as “antigen presenting” segments, responsible for proper positioning of the host receptor-binding motifs found in the highly disordered extremities of the loops. The residues found “on top” of the loops are both the most variable and the most mobile. More precisely, our observations in solution and ssNMR indicate that two regions of loop L1 (from G39 to T47) and of loop L3 (from A128 to E136) are the most mobile segments of the outer surface of KpOmpA.

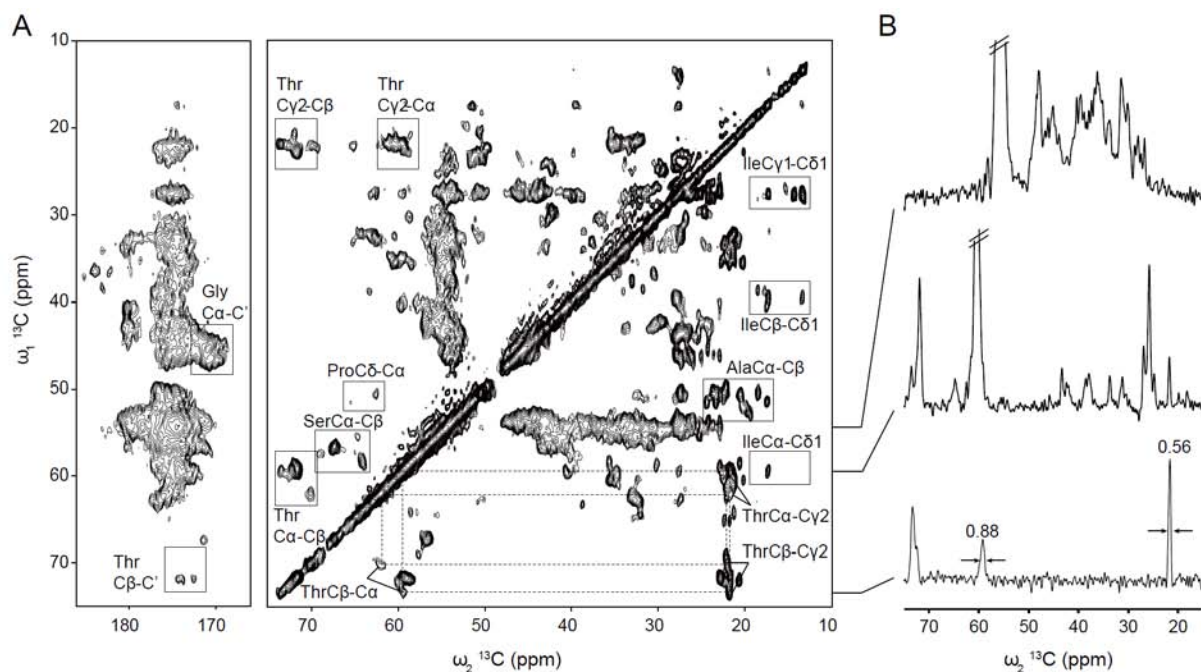
### **Acknowledgement:**

These studies and the IPBS NMR equipment were financed by the French Research Ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées and European structural funds. The NMR data were acquired on the spectrometers of the PICT – IBISA national facility. The research leading to these results has received funding from the European Community's Seventh Framework Program FP7/2007-2013 under grant agreement n° 211800, and from the European Drug Initiative on Channels and Transporters grant HEALTH-F4-2007-201924. Deborah Gater is acknowledged for the final reading of the manuscript.

*Figures and figure texts are on the next 6 pages. SI is available after the figures.*

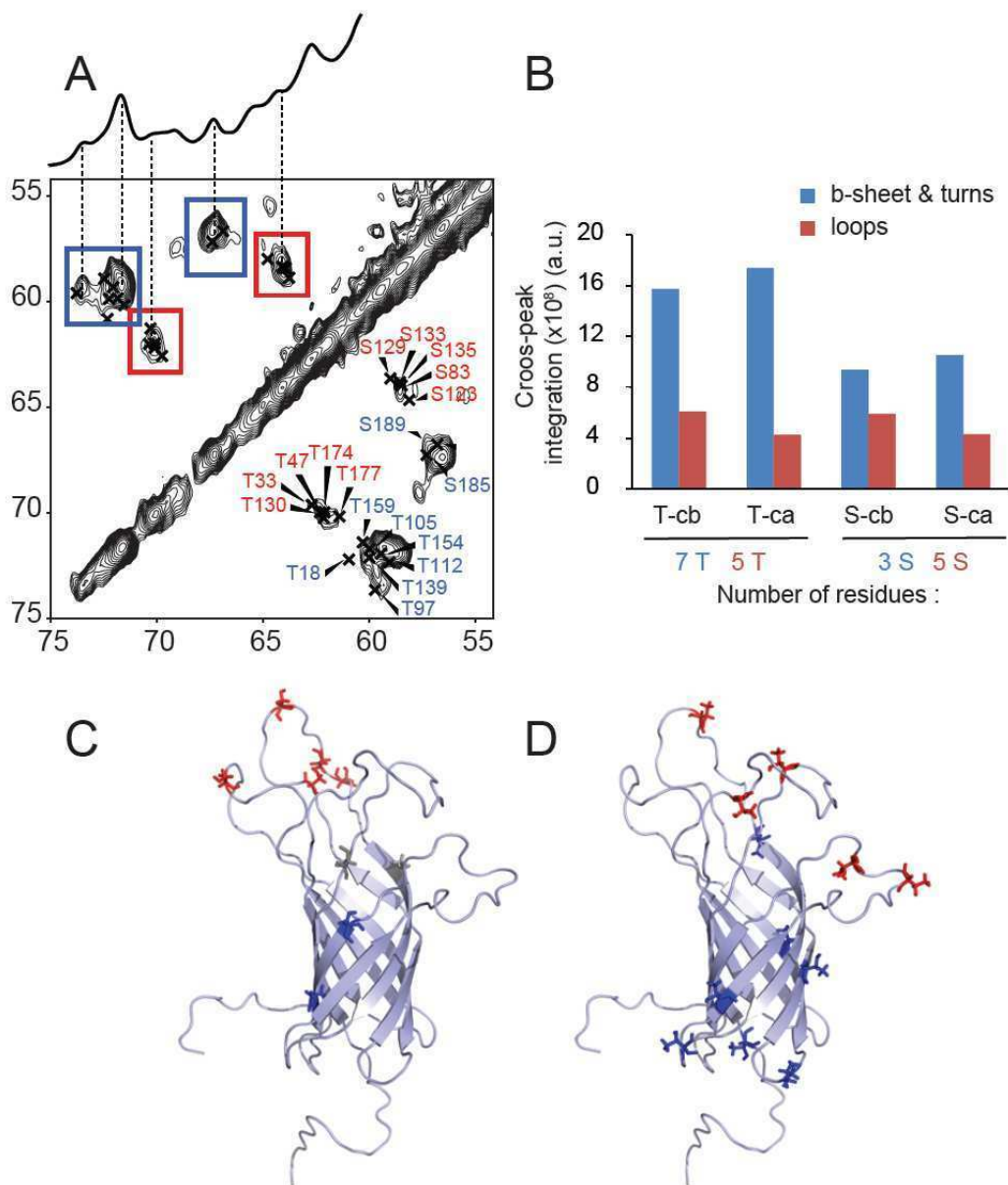


**Fig. 1: Screening for the reconstitution of KpOmpA TM domain in lipid vesicles.** **(A)** Summary of experimental conditions screened for the reconstitution of KpOmpA in lipid vesicles. Positive hits are identified with a white box and a black dot. Black boxes represent rejected conditions in which the samples were heavily aggregated. Gray boxes represent intermediate conditions: sample exhibiting moderate aggregation and/or presenting a threat to the overall homogeneity (too low LPR) or to the efficiency of centrifugation (too high LPR). Reconstitution buffers are labeled from A to E: *Buffer A*, 20 mM Hepes (pH 7.0), 140 mM NaCl; *Buffer B*, 20 mM Hepes (pH 7.0), 300 mM NaCl; *Buffer C*, 20 mM Tris-HCl (pH 8.5), 100 mM NaCl; *Buffer D*, 20 mM Tris-HCl (pH 8.5), 140 mM NaCl; *Buffer E*, 20 mM Tris-HCl (pH 8.5), 300 mM NaCl. **(B-D)** Electron micrographs of negatively-stained KpOmpA-containing vesicles with different conditions (the scale bar is 200 nm); **(B)** The lower pH reduces the sample quality, resulting in small, aggregated vesicles when the protein is reconstituted at LPR of 0.4 (w/w) in DOPC (20 mM HEPES at pH 7.0, 140 mM NaCl). **(C)** Sample aggregation at LPR of 0.2 (w/w) in DMPC (20 mM Tris at pH 8.5, 100 mM NaCl). **(D)** Vesicle with reconstituted KpOmpA TM domain at LPR of 0.5 (w/w) in *E. coli* polar lipids extract (PLE, 20 mM Tris at pH 8.5, 140 mM NaCl). **(E)** Coomassie stained SDS PAGE gel of KpOmpA TM domain in different environments. The KpOmpA TM domain purified in 0.1% Zwittergent 3-14 (lanes 1 and 2), in 2% OG after detergent exchange (lanes 3 and 4) and in *E. coli* PLE after reconstitution for ssNMR experiments (lanes 5 and 6). The aliquots on lanes 1, 3 and 5 represent the native fold of the  $\beta$ -barrel, whereas those on lanes 2, 4 and 6 are heat-denatured (100°C, 5 min) and the protein bands appear at their expected positions (~23.4 kDa) from the marker ladder on the left lane.

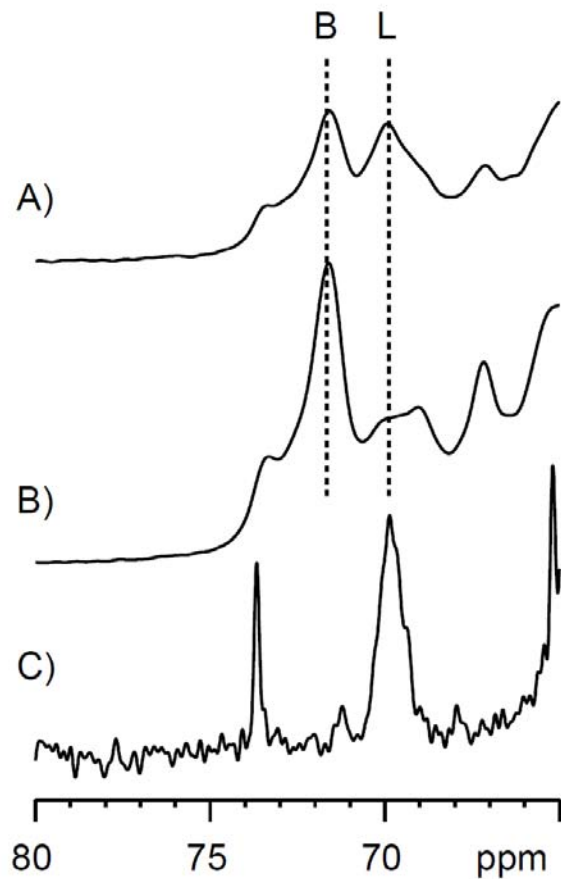


**Fig. 2: (A)** Carbonyl and aliphatic regions of the 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  PDSD correlation spectrum (40 ms mixing time) recorded on reconstituted ( $\text{U-}^{13}\text{C}$ ,  $^{15}\text{N}$ )-labeled KpOmpA TM domain in *E. coli* polar lipids extract (PLE) vesicles at a sample temperature of 20°C with MAS 12 kHz and a proton Larmor frequency of 700 MHz. Amino acid-specific assignments are indicated. Dashed lines indicate the characteristic intraresidue CC connectivities of Thr within  $\beta$ -sheet and loop protein regions, respectively. **(B)** Series of 1D slices extracted from the 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  PDSD correlation spectrum at different  $\omega_1$   $^{13}\text{C}$  frequencies. The carbon line-widths of the Thr C $^\alpha$  and CH $_3$  resonances are indicated in ppm. The NMR samples typically contained 22 mg of protein, 11 mg of lipids and 17 mg of water (lipid/protein molar ratio = 15/1).

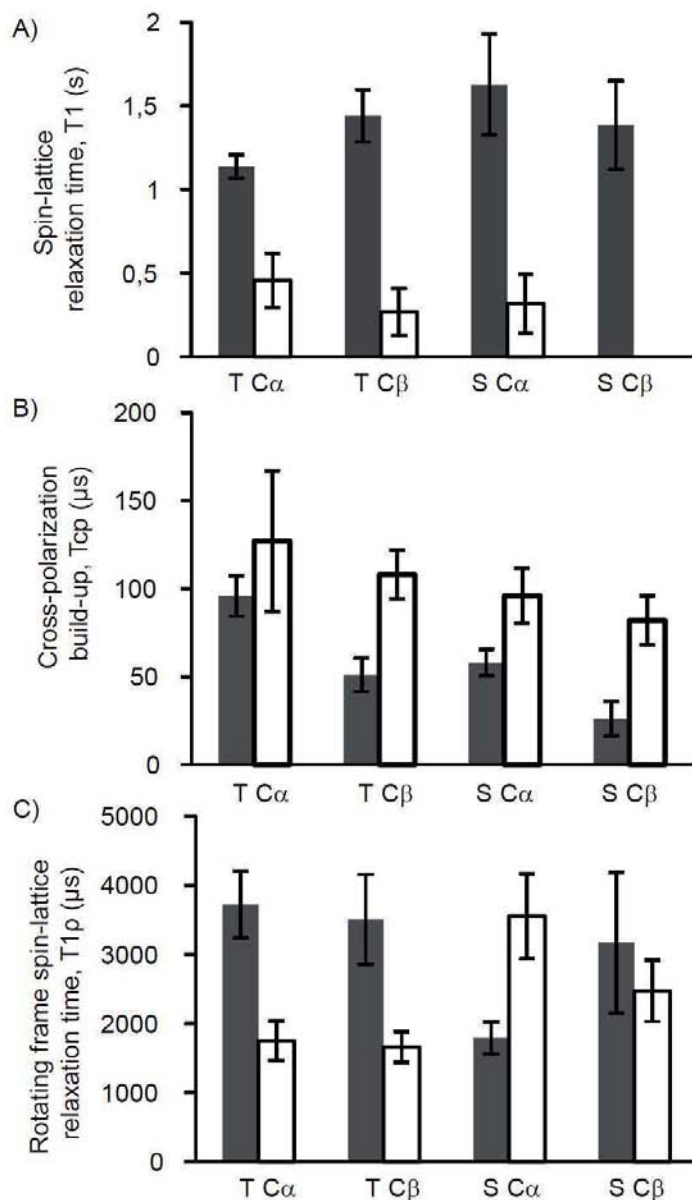




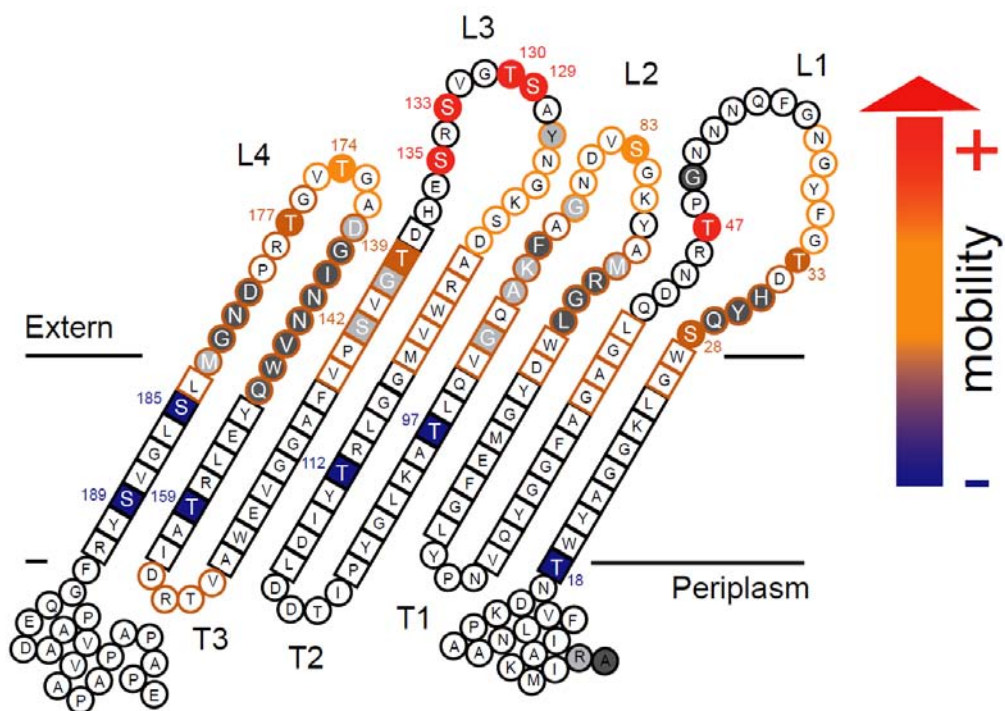
**Fig. 3:** (A) Selected region of the 2D  $^{13}\text{C},^{13}\text{C}$  PDS correlation spectrum showing backbone  $\text{C}^\beta\text{-C}^\alpha$  correlations of the Thr and Ser residues of the KpOmpA TM domain. Solution NMR assignments are indicated when available. Open rectangles represent spectral regions used for signal integration, expected correlations were obtained from available solution NMR chemical shifts of KpOmpA TM domain (BMRB 15651 [9]) after correction due to  $^2\text{H}$  isotope effect. (B) Integrations of spectral regions corresponding to Thr and Ser residues within random-coil (red bars) and  $\beta$ -sheet-like (blue bars) protein segments plotted onto the number of residues. (C) Distribution of Ser residues within KpOmpA TM domain (PDB ID: 2K0L). (D) Distribution of Thr residues within KpOmpA TM domain. For (C) and (D) the color code is: residues within  $\beta$ -sheet TM and turns segments are in blue, those within the extracellular loops are in red. Two unassigned residues are labeled in dark gray.



**Fig. 4:** 1D proton decoupled, <sup>13</sup>C-detected MAS NMR spectra of reconstituted (U-<sup>13</sup>C,<sup>15</sup>N)-labeled KpOmpA in *E. coli* PLE vesicles recorded at a MAS frequency of 12 kHz and a sample temperature of 20°C. **(A)** Direct excitation recorded with a 90° pulse on <sup>13</sup>C and accumulated 2K scans; **(B)** Cross-polarization (CP) transfer MAS recorded with a CP contact time of 400 μs, with 2K scans; **(C)** Refocused-INEPT excitation recorded with 12K scans. NMR chemical shifts of Thr C<sup>β</sup> resonances characteristic of β-sheet TM and random coil protein regions (loops) are labeled B and L, respectively.



**Fig. 5:** (A)  $^{13}\text{C}$  spin-lattice relaxation time ( $T_1$ ) of  $\text{C}^\alpha$  and  $\text{C}^\beta$  of serine (S) and threonine (T) residues. (B) and (C) correspond respectively to the cross-polarization build-up ( $T_{CP}$ ) and proton spin-lattice relaxation time in the rotating frame ( $T_{1\rho}$ ) determined from the cross-polarization dynamics. Gray bars correspond to carbons within  $\beta$ -sheet TM and turns segments; white bars correspond to carbons within the extracellular loops.



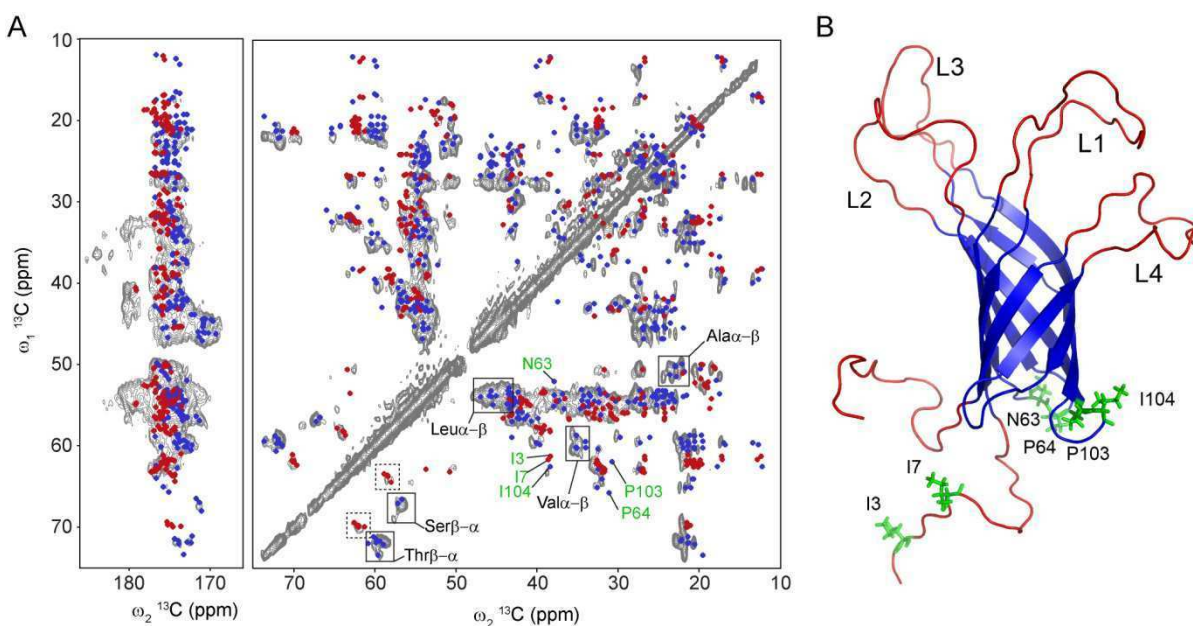
**Fig. 6:** Dynamics of KpOmpA TM domain in lipid bilayers. Topological representation of KpOmpA TM domain (PDB ID: 2K0L) illustrating residue-specific mobility within distinct protein segments. Residues within  $\beta$ -sheet and random coil regions are represented by squares and circles, respectively. Amino acids are given in single-letter notation. In both DHPC micelles and *E. coli* polar lipids extract bilayers, residues that experience fast isotropic motions are colored in red and rigid protein segments are colored in blue. Residues that exhibit restricted mobility in DHPC micelles are highlighted; they correspond to residues in conformational exchange (dark orange) and characterized by  $^1\text{H}$ - $^{15}\text{N}$  NOEs at  $0.6 \pm 0.1$  (light orange). Unassigned residues are colored in gray.

*SI is available on the next 2 pages.*

## Supplementary Information

Dynamics of *Klebsiella pneumoniae* OmpA transmembrane domain: The four extracellular loops display restricted motion behavior, in micelles and in lipid bilayers.

Iordan Iordanov, Marie Renault, Valérie Réat, Patrick D. Bosshart, Andreas Engel, Olivier Saurel, Alain Milon



**Fig. S1.** (A) Two-dimensional  $^{13}\text{C}$ - $^{13}\text{C}$  PDSD correlation spectrum of the membrane-embedded KpOmpA TM domain reconstituted in *E. coli* PLE vesicles (PLR=0.5, w/w) recorded with a PDSD mixing time of 40 ms at a MAS frequency of 12 kHz and at 20°C (sample temperature). Expected intraresidue C-C correlations predicted from solution NMR chemical shifts of KpOmpA TM domain (BMRB accession code: 15651) after correction for isotope effect (squares). (B) Cartoon representation of the solution NMR structure of KpOmpA TM domain (PDB accession code: 2K0L). Color-coding in (A) and (B): red, extracellular loops (L1 to L4), N- and C-term extremities; blue, transmembrane  $\beta$ -sheets ( $\beta$ 1 to  $\beta$ 8) and periplasmic turns (T1 to T3). Residues with backbone  $\text{C}^\alpha/\text{C}^\beta$  correlations that could not be observed at similar peak position than those predicted by solution NMR chemical shifts are labeled in green.



## References

- [1] E. Wallin, G. von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, *Protein Sci.*, 7 (1998) 1029-1038.
- [2] G.C. Terstappen, A. Reggiani, In silico research in drug discovery, *Trends Pharmacol. Sci.*, 22 (2001) 23-26.
- [3] P.J. Judge, A. Watts, Recent contributions from solid-state NMR to the understanding of membrane protein structure and function, *Curr. Opin. Chem. Biol.*, 15 (2011) 690-695.
- [4] J.H. Chill, F. Naider, A solution NMR view of protein dynamics in the biological membrane, *Curr. Opin. Struct. Biol.*, 21 (2011) 627-633.
- [5] L.K. Tamm, B.Y. Liang, NMR of membrane proteins in solution, *Prog. Nucl. Magn. Reson. Spectrosc.*, 48 (2006) 201-210.
- [6] H.J. Kim, S.C. Howell, W.D. Van Horn, Y.H. Jeon, C.R. Sanders, Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins, *Prog. Nucl. Magn. Reson. Spectrosc.*, 55 (2009) 335-360.
- [7] A. McDermott, Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR, *Annu. Rev. Biophys.*, 38 (2009) 385-403.
- [8] A. Naito, Structure elucidation of membrane-associated peptides and proteins in oriented bilayers by solid-state NMR spectroscopy, *Solid State Nucl. Magn. Reson.*, 36 (2009) 67-76.
- [9] M. Renault, O. Saurel, J. Czaplicki, P. Demange, V. Gervais, F. Lohr, V. Reat, M. Piotto, A. Milon, Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications, *J. Mol. Biol.*, 385 (2009) 117-130.
- [10] T.N. Nguyen, P. Samuelson, F. Sterky, C. Merle-Poitte, A. Robert, T. Baussant, J.F. Haeuw, M. Uhlen, H. Binz, S. Stahl, Chromosomal sequencing using a PCR-based biotin-capture method allowed isolation of the complete gene for the outer membrane protein A of *Klebsiella pneumoniae*, *Gene*, 210 (1998) 93-101.
- [11] P. Jeannin, B. Bottazzi, M. Sironi, A. Doni, M. Rusnati, M. Presta, V. Maina, G. Magistrelli, J.F. Haeuw, G. Hoeffel, N. Thieblemont, N. Corvaia, C. Garlanda, Y. Delneste, A. Mantovani, Complexity and complementarity of outer membrane protein a recognition by cellular and humoral innate immunity receptors, *Immunity*, 22 (2005) 551-560.
- [12] C. March, D. Moranta, V. Regueiro, E. Llobet, A. Tomas, J. Garmendia, J.A. Bengoechea, *Klebsiella pneumoniae* outer membrane protein A is required to prevent the activation of airway epithelial cells, *J. Biol. Chem.*, 286 (2011) 9956-9967.
- [13] S.G. Smith, V. Mahon, M.A. Lambert, R.P. Fagan, A molecular Swiss army knife: OmpA structure, function and expression, *FEMS Microbiol. Lett.*, 273 (2007) 1-11.
- [14] E. Sugawara, H. Nikaido, Pore-forming activity of OmpA protein of *Escherichia coli*, *J. Biol. Chem.*, 267 (1992) 2507-2511.
- [15] A. Negoda, E. Negoda, R.N. Reusch, Resolving the native conformation of *Escherichia coli* OmpA, *FEBS J.*, 277 (2010) 4427-4437.
- [16] S.G.J. Smith, V. Mahon, M.A. Lambert, R.P. Fagan, A molecular Swiss army knife: OmpA structure, function and expression, *FEMS Microbiol. Lett.*, 273 (2007) 1-11.
- [17] A. Pautsch, G.E. Schulz, Structure of the outer membrane protein A transmembrane domain, *Nat. Struct. Biol.*, 5 (1998) 1013-1017.
- [18] A. Pautsch, G.E. Schulz, High-resolution structure of the OmpA membrane domain, *J. Mol. Biol.*, 298 (2000) 273-282.
- [19] A. Arora, F. Abildgaard, J.H. Bushweller, L.K. Tamm, Structure of outer membrane protein A transmembrane domain by NMR spectroscopy, *Nat. Struct. Biol.*, 8 (2001) 334-338.
- [20] K. Cox, P.J. Bond, A. Grottesi, M. Baaden, M.S. Sansom, Outer membrane proteins: comparing X-ray and NMR structures by MD simulations in lipid bilayers, *Eur. Biophys. J.*, 37 (2008) 131-141.

- [21] R.S. Cantor, Lipid composition and the lateral pressure profile in bilayers, *Biophys. J.*, 76 (1999) 2625-2639.
- [22] J.T. Groves, The physical chemistry of membrane curvature, *Nat. Chem. Biol.*, 5 (2009) 783-784.
- [23] P. Garidel, C. Johann, A. Blume, Thermodynamics of lipid organization and domain formation in phospholipid bilayers, *J. Liposome Res.*, 10 (2000) 131-158.
- [24] D. Fourel, A. Bernadac, J.M. Pages, Involvement of Exposed Polypeptide Loops in Trimeric Stability and Membrane Insertion of Escherichia-Coli Ompf Porin, *Eur. J. Biochem.*, 222 (1994) 625-630.
- [25] A. Basle, G. Rummel, P. Storici, J.P. Rosenbusch, T. Schirmer, Crystal structure of osmoporin OmpC from E-coli at 2.0 angstrom, *J. Mol. Biol.*, 362 (2006) 933-942.
- [26] K. Nakamura, S. Mizushima, Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12, *J. Biochem.*, 80 (1976) 1411-1422.
- [27] J.L. Rigaud, M.T. Paternostre, A. Bluzat, Mechanisms of membrane protein insertion into liposomes during reconstitution procedures involving the use of detergents. 2. Incorporation of the light-driven proton pump bacteriorhodopsin, *Biochemistry*, 27 (1988) 2677-2688.
- [28] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent, *J. Biol. Chem.*, 193 (1951) 265-275.
- [29] J. Murphy, J.P. Riley, A Single-Solution Method for the Determination of Soluble Phosphate in Sea Water, *J. Mar. Biol. Ass. U.K.*, 37 (1958) 9-14.
- [30] A.E. Bennett, C.M. Rienstra, M. Auger, K.V. Lakshmi, R.G. Griffin, Heteronuclear Decoupling in Rotating Solids, *J. Chem. Phys.*, 103 (1995) 6951-6958.
- [31] N. Bloembergen, On the interaction of nuclear spins in crystalline lattice, *Physica*, 15 (1949) 386-426.
- [32] D.M.J. Duer, Solid-state NMR spectroscopy principles and applications, in, Blackwell science, 2002.
- [33] B.M. Fung, A.K. Khitrin, K. Ermolaev, An improved broadband decoupling sequence for liquid crystals and solids, *J. Magn. Reson.*, 142 (2000) 97-101.
- [34] J. Czaplicki, G. Cornelissen, F. Halberg, GOSA, a simulated annealing-based program for global optimization of nonlinear problems, also reveals transyears, *J. Appl. Biomed.*, 4 (2006) 87-94.
- [35] T. Surrey, F. Jahng, Refolding and Oriented Insertion of a Membrane-Protein into a Lipid Bilayer, *Proc. Natl. Acad. Sci. U. S. A.*, 89 (1992) 7457-7461.
- [36] N.K. Burgess, T.P. Dao, A.M. Stanley, K.G. Fleming, beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro, *J. Biol. Chem.*, 283 (2008) 26748-26758.
- [37] G.J. Patel, S. Behrens-Kneip, O. Holst, J.H. Kleinschmidt, The Periplasmic Chaperone Skp Facilitates Targeting, Insertion, and Folding of OmpA into Lipid Membranes with a Negative Membrane Surface Potential, *Biochemistry*, 48 (2009) 10235-10245.
- [38] Z.V. Leonenko, E. Finot, H. Ma, T.E. Dahms, D.T. Cramb, Investigation of temperature-induced phase transitions in DOPC and DPPC phospholipid bilayers using temperature-controlled scanning force microscopy, *Biophys. J.*, 86 (2004) 3783-3793.
- [39] M. Hiller, L. Krabben, K.R. Vinothkumar, F. Castellani, B.-J. van Rossum, W. Kuhlbrandt, H. Oschkinat, Solid-state magic-angle spinning NMR of outer-membrane protein G from Escherichia coli, *ChemBioChem*, 6 (2005) 1679-1684.
- [40] R. Koebnik, Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from Escherichia coli, *J. Bacteriol.*, 181 (1999) 3688-3694.
- [41] Y. Wang, O. Jardetzky, Probability-based protein secondary structure identification using combined NMR chemical-shift data, *Protein Sci.*, 11 (2002) 852-861.
- [42] Y. Li, D.A. Berthold, R.B. Gennis, C.M. Rienstra, Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy, *Protein Sci.*, 17 (2008) 199-204.

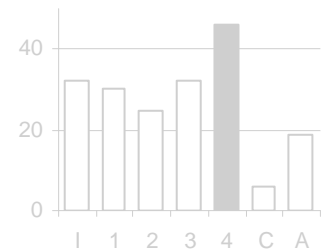


- [43] Y. Han, J. Ahn, J. Concel, I.-J.L. Byeon, A.M. Gronenborn, J. Yang, T. Polenova, Solid-State NMR Studies of HIV-1 Capsid Protein Assemblies, *J. Am. Chem. Soc.*, 132 (2010) 1976-1987.
- [44] A. McDermott, T. Polenova, A. Bockmann, K.W. Zilm, E.K. Paulsen, R.W. Martin, G.T. Montelione, Partial NMR assignments for uniformly (C-13, N-15)-enriched BPTI in the solid state, *J. Biomol. NMR*, 16 (2000) 209-219.
- [45] M. Etzkorn, S. Martell, O.C. Andronesi, K. Seidel, M. Engelhard, M. Baldus, Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy, *Angew. Chem.*, 46 (2007) 459-462.
- [46] R.A. Venters, B.T. Farmer, C.A. Fierke, L.D. Spicer, Characterizing the use of perdeuteration in NMR studies of large proteins C-13, N-15 and H-1 assignments of human carbonic anhydrase II, *J. Mol. Biol.*, 264 (1996) 1101-1116.
- [47] P.J. Spooner, R.H. Friesen, J. Knol, B. Poolman, A. Watts, Rotational mobility and orientational stability of a transport protein in lipid membranes, *Biophys. J.*, 79 (2000) 756-766.
- [48] A.G. Palmer, J. Williams, A. McDermott, Nuclear magnetic resonance studies of biopolymer dynamics, *J. Phys. Chem.*, 100 (1996) 13293-13310.
- [49] N. Giraud, M. Blackledge, M. Goldman, A. Bockmann, A. Lesage, F. Penin, L. Emsley, Quantitative analysis of backbone dynamics in a crystalline protein from nitrogen-15 spin-lattice relaxation, *J. Am. Chem. Soc.*, 127 (2005) 18190-18201.
- [50] J. Yang, M.L. Tasayco, T. Polenova, Dynamics of Reassembled Thioredoxin Studied by Magic Angle Spinning NMR: Snapshots from Different Time Scales, *J. Am. Chem. Soc.*, 131 (2009) 13690-13702.
- [51] V. Chevelkov, U. Fink, B. Reif, Quantitative analysis of backbone motion in proteins using MAS solid-state NMR spectroscopy, *J. Biomol. NMR*, 45 (2009) 197-206.
- [52] G. Reuther, K.T. Tan, A. Vogel, C. Nowak, K. Arnold, J. Kuhlmann, H. Waldmann, D. Huster, The lipidated membrane anchor of full length N-Ras protein shows an extensive dynamics as revealed by solid-state NMR spectroscopy, *J. Am. Chem. Soc.*, 128 (2006) 13840-13846.
- [53] J.R. Lewandowski, J. Sein, H.J. Sass, S. Grzesiek, M. Blackledge, L. Emsley, Measurement of site-specific <sup>13</sup>C spin-lattice relaxation in a crystalline protein, *J. Am. Chem. Soc.*, 132 (2010) 8252-8254.
- [54] J. Yang, L. Aslimovska, C. Glaubitz, Molecular Dynamics of Proteorhodopsin in Lipid Bilayers by Solid-State NMR, *J. Am. Chem. Soc.*, 133 (2011) 4874-4881.
- [55] H. Nikaido, M. Vaara, Molecular-Basis of Bacterial Outer-Membrane Permeability, *Microbiol. Rev.*, 49 (1985) 1-32.
- [56] H. Keweloh, G. Weyrauch, H.J. Rehm, Phenol-Induced Membrane-Changes in Free and Immobilized Escherichia-Coli, *Appl. Microbiol. Biotechnol.*, 33 (1990) 66-71.
- [57] H. Nikaido, Molecular basis of bacterial outer membrane permeability revisited, *Microbiol. Mol. Biol. Rev.*, 67 (2003) 593-+.
- [58] H. Heise, W. Hoyer, S. Becker, O.C. Andronesi, D. Riedel, M. Baldus, Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR, *Proc. Natl. Acad. Sci. U. S. A.*, 102 (2005) 15871-15876.
- [59] O.C. Andronesi, S. Becker, K. Seidel, H. Heise, H.S. Young, M. Baldus, Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy, *J. Am. Chem. Soc.*, 127 (2005) 12965-12974.
- [60] B. Liang, A. Arora, L.K. Tamm, Fast-time scale dynamics of outer membrane protein A by extended model-free analysis of NMR relaxation data, *Biochim. Biophys. Acta - Biomembranes*, 1798 (2010) 68-76.
- [61] P.M. Hwang, R.E. Bishop, L.E. Kay, The integral membrane enzyme PagP alternates between two dynamically distinct states, *Proc. Natl. Acad. Sci. U. S. A.*, 101 (2004) 9618-9623.
- [62] D. Datta, N. Vaidehi, W.B. Floriano, K.S. Kim, N.V. Prasadarao, W.A. Goddard, 3rd, Interaction of E. coli outer-membrane protein A with sugars on the receptors of the brain microvascular endothelial cells, *Proteins*, 50 (2003) 213-221.

- [63] N.V. Prasadaraο, Identification of Escherichia coli outer membrane protein A receptor on human brain microvascular endothelial cells, *Infect. Immun.*, 70 (2002) 4556-4563.
- [64] M. Pichavant, Y. Delneste, P. Jeannin, C. Fourneau, A. Bricet, A.B. Tonnel, P. Gosset, Outer membrane protein a from Klebsiella pneumoniae activates bronchial epithelial cells: Implication in neutrophil recruitment, *J. Immunol.*, 171 (2003) 6697-6705.
- [65] T.A. Pascal, R. Abrol, R. Mittal, Y. Wang, N.V. Prasadaraο, W.A. Goddard, Experimental Validation of the Predicted Binding Site of Escherichia coli K1 Outer Membrane Protein A to Human Brain Microvascular Endothelial Cells: identifications of critical mutations that prevent E. Coli meningitis, *J. Biol. Chem.*, 285 (2010) 37753-37761.
- [66] N.V. Prasadaraο, R. Mittal, S. Krishnan, I. Gonzalez-Gomez, Deciphering the Roles of Outer Membrane Protein A Extracellular Loops in the Pathogenesis of Escherichia coli K1 Meningitis, *J. Biol. Chem.*, 286 (2011) 2183-2193.

# Chapter 4

**Enzyme digestion experiments on the KpOmpA membrane domain,  
periplasmic domain and full-length proteins**



## Summary

This chapter describes enzyme-based digestion/proteolysis experiments of unfolded and folded KpOmpA in different environments (i.e. detergent micelles and lipid bilayers) and buffer solutions. The susceptibility or the resistance of the KpOmpA extracellular loops to protease cleavage is discussed in relation to the dynamic behavior of this protein region. It is shown that, despite their overall higher mobility levels (in comparison to the barrel core) found in Chapter 3 for both the micellar and bilayer environments, the loops of the protein are resistant to a large-scale cleavage by the Lysine-C endoprotease which is specific for Lysines. Trypsin however may cleave an Arginine residue which is located in the most mobile part of loop L3.

## 4.1. Introduction

As shown in Chapter 1, the proper  $\beta$ -barrel fold of N- and F-KpOmpA is preserved when these protein constructs are reconstituted in PLE liposomes by dialysis-driven detergent removal. This was further confirmed for N-KpOmpA in Chapter 3 and, in addition, the mobility of the extracellular loops of N-KpOmpA was evaluated and compared to that in micellar environment [1]. It was thus revealed that the increased (compared to the barrel core of the molecule) dynamics of the extracellular region is an intrinsic property of the molecule and not just a function of its environment. However, it was also determined that these loops do not exhibit a homogenous dynamics (i.e. random coil behavior for the entire loops region), but they follow a gradient of dynamics emanating from the center of the barrel and in a direction of the outmost segments of the molecule (parallel to the bilayer normal). It is thus interesting to ask how stable, or unstructured, are those segments and do they play a role in the overall stability of the molecule, apart from possessing inter-molecular recognition properties. One possible approach in this direction is observing the gradual protein digestion (or fragmentation) by limited proteolysis ([2, 3]). The appearing protein fragments can then be tracked by mass spectrometry and gel electrophoresis, thus pointing to the enzyme-accessible points on the polypeptide chain. Such data can establish the 'protection' value of the environment at the level of the transmembrane protein section (such as detergent micelle or lipid bilayer), it can explore the propensity of the protein loops (outside of this environment) to get cleaved and, finally, it can explore the susceptibility of the soluble protein domain (C-KpOmpA) to the same enzyme digestion. Masae Sugawara, during her PhD in the lab ([4]), had realized preliminary experiments with the Lys-C endoprotease and she had shown that specific fragments may indeed be observed and identified by mass spectrometry.

## 4.2. Material and methods

### 4.2.1. Commercially available proteases used in this chapter

(A) The highly purified commercial product 'Endoproteinase Lys-C sequencing grade' (Roche) is a 269 a.a. (~28 kDa) enzyme derived from *Lysobacter enzymogenes* and provided in 5- $\mu$ g lyophilized stocks (stored at 4°C). Following the manufacturer's instructions, each such powder stock was resuspended in 50  $\mu$ l Milli-Q water, which results in the following buffer: 50 mM HEPES (pH 8), 10 mM EDTA, 5 mg/ml raffinose and 0.1 mg/ml Lys-C enzyme.

This 50- $\mu$ l stock solution was divided in 5x10- $\mu$ l aliquots (each containing 1  $\mu$ g of the enzyme) and frozen at  $-20^{\circ}\text{C}$  until used. Only fresh preparations or first-time defrosted enzyme aliquots were used for each experiment. Every digestion reaction (with a single exception, mentioned below) was performed with N-, F- or C-KpOmpA buffer solutions of at least an order of magnitude larger volume compared to the used Lys-C aliquot, in order to ensure reaction conditions determined by the KpOmpA buffer and not by the enzyme buffer. Every digestion reaction was performed at a temperature not higher than  $37^{\circ}\text{C}$ , in order to avoid autolysis (Lys-C itself has five Lysine residues).

**(B)** The Sequencing Grade Modified Trypsin (Promega) is a product with high activity, high specificity to C-terminal cleavages after both Lysine and Arginine residues and resistance to autolysis. The lyophilized powder of 20  $\mu$ g Trypsin was dissolved in 200  $\mu$ l of commercially supplied buffer (50 mM acetic acid), distributed in 10- $\mu$ l aliquots (1  $\mu$ g enzyme) and frozen at  $-20^{\circ}\text{C}$  until used. Only fresh preparations or first-time defrosted enzyme aliquots were used for each experiment.

#### 4.2.2. Lys-C digestion of urea-denatured F-KpOmpA

100  $\mu$ l of F-KpOmpA stock solution (14.7 mg/ml in 0.1% Zwittergent 3-14 buffer, pH 8.5) were precipitated by adding 5.3  $\mu$ l of 5M NaCl and 2 ml cold ethanol, after which the mixture was left for 4 hours at  $-20^{\circ}\text{C}$ . Upon removal of the ethanol supernatant (5000 g, 10 min,  $4^{\circ}\text{C}$ ), three steps of 2-ml Milli-Q water washings and centrifugations (20 000 g, 10 min,  $4^{\circ}\text{C}$ ) removed the rest of the ethanol and separated the precipitated protein pellet, which was then solubilized in  $\sim$ 35  $\mu$ l of 7M urea. The protein concentration was estimated to be 8.8 mg/ml. From this protein/urea solution, 22.7  $\mu$ l ( $\sim$ 200  $\mu$ g F-KpOmpA) were brought up to 100  $\mu$ l (2 mg/ml F-KpOmpA) by dilution with 77.3  $\mu$ l of 150 mM NaCl, which also diluted the urea down to 1.6M. An aliquot of 25  $\mu$ l ( $\sim$ 50  $\mu$ g F-KpOmpA) was further diluted with 25  $\mu$ l 150 mM NaCl and 10  $\mu$ l of Lys-C stock solution (1  $\mu$ g enzyme) were added, thus giving a 60- $\mu$ l sample containing  $\sim$ 0.7M urea, 150 mM NaCl as a predominant salt component, 50  $\mu$ g F-KpOmpA and 1  $\mu$ g Lys-C (protein-to-enzyme ratio of 50:1 w/w, equivalent to  $\sim$ 30:1 mol/mol). The stock buffer of Lys-C in this experiment, by exception, was diluted only 6 times, rather than the usual (at least) an order of magnitude dilution, described above. The mixture was left overnight at  $37^{\circ}\text{C}$ .

#### 4.2.3. Lys-C digestion of folded N- and F-KpOmpA

For the micellar samples, 50-500 µg (indicated in text) of N- or F-KpOmpA (in 0.1% Zwittergent 3-14 buffer, brought to the desired protein concentration of 1-3 mg/ml) were directly mixed with 10 µl Lys-C stock solution (1 µg enzyme) and incubated overnight at 37°C. The protein-to-enzyme ratio, varying from 50:1 to 500:1 (w/w), is indicated in text.

For the liposomes samples, 50-500 µg (indicated in text) of N- or F-KpOmpA reconstituted in PLE liposomes and found in 20 mM Tris (pH 8.5) were mixed with 10 µl Lys-C stock solution (1 µg enzyme) and incubated overnight at 37°C. The lipid-to-protein ratio (LPR, varying from 0.5 to 5 w/w) and the protein-to-enzyme ratio (varying from 50:1 to 500:1 w/w) are indicated in text.

#### 4.2.4. Buffer exchange and digestion of F-KpOmpA in micelles

Purified F-KpOmpA in 0.1% Zwittergent 3-14 buffer was diluted (with the same buffer) down to a protein concentration of 1 mg/ml, distributed in 500-µl aliquots (500 µg protein in each) and dialyzed (12-14 kDa cutoff, SpectraPor) overnight at 4°C against 0.5L of 1 from 9 different buffers, all of which contained 0.1% Zwittergent 3-14. Essentially, each of these solutions differs in its pH-value and in the concentration of NaCl, and represents a 10 mM Na-phosphate buffer brought at the desired pH by mixing and diluting different proportions of 0.2M NaH<sub>2</sub>PO<sub>4</sub> and 0.2M Na<sub>2</sub>HPO<sub>4</sub> stocks. The monitored pH did not differ with more than 0.1 units for every buffer. The nine buffers, encoded alphabetically, are as follows (the buffers are also described in text):

<u>Buffer</u>	<u>NaCl</u>	<u>pH</u>
A	0 mM	6
B	0 mM	7
C	0 mM	8
K	150 mM	6
L	150 mM	7
M	150 mM	8
X	300 mM	6
Y	300 mM	7
Z	300 mM	8

The nine 500-µl samples were then collected, centrifuged (20 000 g, 10 min, 4°C) to remove eventual precipitates and the protein concentrations were measured spectrophotometrically. The protein concentrations differed within ~10% among the nine samples. Calculated small amounts of the nine buffers (from A to Z) were then added to

some of the corresponding samples in order to equilibrate all concentrations. At the end, 100 µg F-KpOmpA were contained in 106 µl of buffer for each of the nine samples. 15-µl aliquots were then collected from each sample before subjecting it to enzyme digestion. The rest ~90 µl of each sample (~85 µg F-KpOmpA) were mixed with 2 µl of enzyme stock solution (0.2 µg Lys-C), giving protein-to-enzyme ratio of about 430:1 (w/w), equivalent to ~300 F-KpOmpA molecules per enzyme molecule. These nine reaction mixtures were incubated in conditions of 'complete digestion' (~16 hours at 37°C). At the end, two 15-µl aliquots were taken from each sample and one of them was heat-denatured (5 min, 100°C). In this way, a total of 3 aliquots from each of the nine buffers/reaction conditions were collected.

#### 4.2.5. Buffer exchange and digestion of F-KpOmpA in liposomes

Purified F-KpOmpA in 0.1% Zwittergent 3-14 buffer was used for dialysis-driven protein reconstitution as described in Chapter 1, with several modifications: (1) the detergent exchange step (from 0.1% Zwittergent 3-14 to 2% OG) was performed with 10 mM Na-phosphate buffer at pH 8 (rather than 25 mM Tris at pH 8.5) with 100 mM NaCl. The same buffer, supplemented with 400 mM imidazole, was used for the elution of F-KpOmpA in 2% OG micelles. (2) The LPR of the ternary mixture was set to 2 (w/w), equivalent to ~100 lipid molecules per protein molecule. (3) The dialysis for detergent removal was performed in 3 times of 1L 10 mM Na-phosphate buffer (pH 8) containing 100 mM NaCl, with reduction of the NaCl concentration down to zero in two 50-mM steps. Finally, the dialyzed liposomes were collected (3000 g, 5 min, 15°C) and resuspended in 0.5 ml 10 mM Na-phosphate buffer (pH 8). The protein concentration (~4.4 mg/ml) was checked photometrically in an aliquot from the sample upon vesicles destruction with 10 times volume excess of 1% Zwittergent 3-14 buffer. The final proteoliposomes stock thus contained ~2.2 mg reconstituted F-KpOmpA at LPR of 2 (w/w), in 10 mM Na-phosphate buffer (pH 8).

The proteoliposomes stock was then divided into nine 25-µl aliquots (each containing 100-110 µg F-KpOmpA) and diluted with 475 µl of one of the nine buffers, thus shifting the final solution composition 20 times in favor of the particular buffer (A-Z). The same 9 buffers (as used with the micellar samples above) were prepared, without the addition of Zwittergent 3-14. The pH of the solutions with higher salt content (buffers X, Y and Z) was found to deviate with around 0.3 units from the expected values. Those buffers had their pH-values adjusted with small amounts of the 0.2M Na-phosphate stock solutions, until all of the nine buffers reached their expected pH-values within 0.1 units of difference.



For the enzyme digestion, 2  $\mu$ l of enzyme stock solution (0.2  $\mu$ g Lys-C) were added to each of the nine 500- $\mu$ l samples (giving protein-to-enzyme ratio of  $\sim$ 500:1 w/w) and incubated overnight at 37°C with gentle shaking (to avoid gravitational pelleting of the sample material). The nine digestion reactions were then centrifuged (10 000 g, 10 min, 4°C) and their supernatants discarded to remove the enzyme. Each proteoliposomes pellet was resuspended in fresh  $\sim$ 100  $\mu$ l of its corresponding buffer.

#### 4.2.6. Digestion kinetics of N- and F-KpOmpA in different environments

N- or F-KpOmpA was subjected to treatment with Lys-C as described above. The protein-to-enzyme ratio and the temperature of the reaction are indicated in text for each particular case. At certain time-points, aliquots from the reaction mixture were collected and either immediately added to equal volume of SDS/loading dye and boiled for 5 min (for SDS PAGE), or added to equal volume of 0.1% TFA / 50% AcN (1:1 v/v) and frozen at  $-20^{\circ}\text{C}$  (for MALDI-TOF). The same procedure was applied for N- and F-KpOmpA either in micelles or in bilayers.

#### 4.2.7. Digestion kinetics of C-KpOmpA

The C-KpOmpA protein at concentration of 1.5 mg/ml is found in 10 mM Na-phosphate buffer (pH 7.5) and 50 mM NaCl. 320  $\mu$ l of this solution ( $\sim$ 500  $\mu$ g C-KpOmpA) were mixed with 10  $\mu$ l of Lys-C stock (1  $\mu$ g enzyme), giving protein-to-enzyme ratio of 500:1 (w/w). The reaction was held at 16°C. At certain time-points, aliquots from the reaction mixture were collected and either immediately added to equal volume of SDS/loading dye and boiled for 5 min (for SDS PAGE), or added to equal volume of 0.1% TFA / 50% AcN (1:1 v/v) and frozen at  $-20^{\circ}\text{C}$  (for MALDI-TOF).

#### 4.2.8. Trypsin digestion of N-, F- and C-KpOmpA

Stock solutions of N-KpOmpA (1.7 mg/ml) and F-KpOmpA (1.8 mg/ml) in 0.1% Zwittergent 3-14 (25 mM Tris (pH 8.5), 150 mM NaCl) were diluted in the same buffer down to 1.5 mg/ml. A stock solution of C-KpOmpA (6.1 mg/ml) was diluted in 10 mM Na-phosphate buffer (pH 7.5) with 50 mM NaCl down to 1.5 mg/ml as well. 120  $\mu$ g of each protein construct were then mixed with 6  $\mu$ l Trypsin stock (0.6  $\mu$ g enzyme), giving protein-to-enzyme ratio of 200:1 (w/w). The three mixtures were left for 2.5 hours at 37°C and then aliquots were collected for SDS PAGE and MALDI-TOF, similarly to the previous preparations.

#### 4.2.9. MALDI-TOF samples preparation and measurement

Aliquots of the different digestion reactions were mixed with equal volume of a two-component solution composed of 0.1% trifluoroacetic acid (TFA) and 50% acetonitrile (AcN) in 1:1 (v/v) ratio, as described above. If the sample material was considered too diluted, this mixture was dried under vacuum (miVac Duo Concentrator, GeneVac) and resuspended in 10  $\mu$ l of the same TFA/AcN solution. 1  $\mu$ l of the sample (1-2  $\mu$ g KpOmpA) was then deposited on a MALDI-plate and mixed with 1  $\mu$ l of matrix solution (either dihydroxybenzoic acid (DHB),  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA, Fluka Analytical) or 3,5-dimethoxy-4-hydroxycinnamic acid (SPA), as indicated in text). Upon drying at room temperature, the plate was loaded into the vacuum chamber of the spectrometer.

All mass spectrometry measurements were performed on a Voyager DE-STR (Applied Biosystems) MALDI-TOF spectrometer, using Voyager Instrument Control Panel software for data acquisition and Data Explorer software for data processing and analysis. The spectra were acquired either in linear (1-50 kDa range, 25 kV accelerating voltage, 300 ns delay time) or in reflectron (0.8-1 kDa range, 20 kV accelerating voltage, 240 ns delay time) mode, with positive ion detection and 1000 shots per spectrum in each case. Spectral calibration was achieved by using commercial compounds of known sizes, such as peptides mixtures (for reflectron mode acquisitions) or BSA (for linear mode).

### 4.3. Basic concept

The endoprotease Lysine-C (hereby abbreviated as Lys-C) is a sequence-specific proteolytic enzyme that cleaves unstructured polypeptide chains C-terminally after a Lysine (K) residue. The enzyme specificity of cleavage after only K-residues is quite high (~90% [5]) which is important if one is to analyze appearing protein fragments from a limited database of possibilities (explained below). The ability of Lys-C to retain its activity in the harsh environment of various denaturing agents (109% activity in 0.1% (w/v) SDS, 90% activity in 1M urea, 86% activity in 4M urea) makes possible the digestion experiments at such conditions.

The N-KpOmpA protein construct possesses a total of 7 Lysine residues (found in the barrel and loops regions) and F-KpOmpA has 18 such K-sites (found in one or the other domain of the molecule), as shown below.

The K-sites present on the chain of N-KpOmpA are thus

ARIMKAI FVLNAAPKDN TWYAGGKLGW SQYHDTGFYGNFQNNNGPTRNDQ LGAG  
 AFGGYQV NPY LGFEMGYDW LGRMAYKGSVDNGAFKA QGVQLTAKLGYP ITDD LDI  
 YTRLGGMVWRA DSKGNYASTGVSRSSEH DTGVSPVFAGGVEWA VTRD IATRLEY QW  
 VNNIGDAGTVGTRPDNGM LSLGVSYRF GQEDAAPVVAPAPAPAPEHHHHH

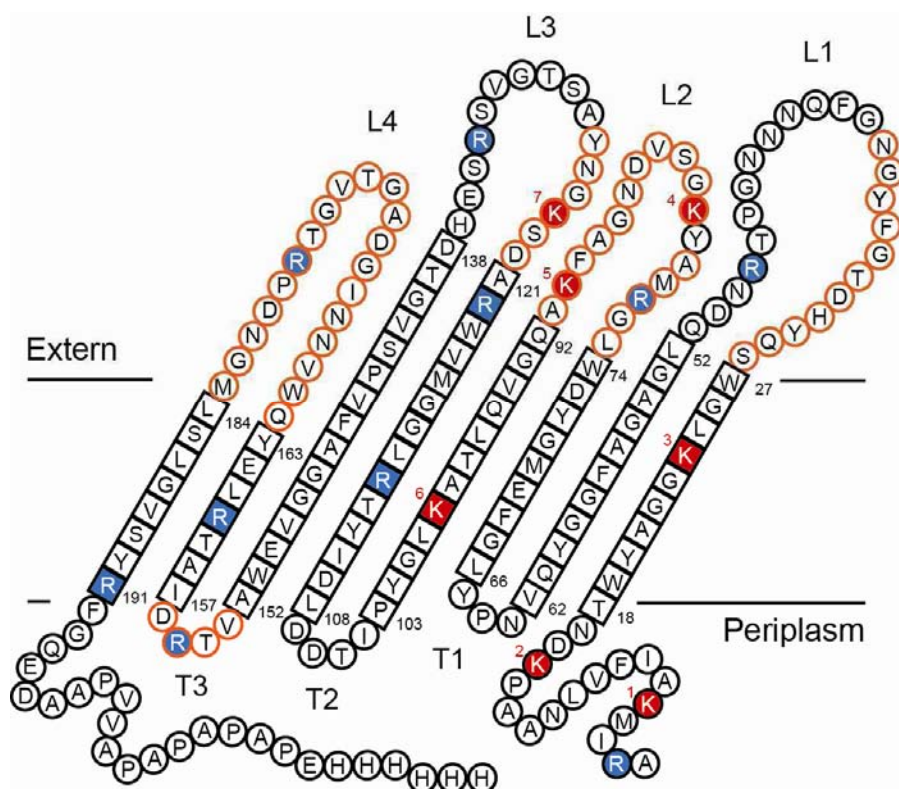
and the K-sites on the chain of F-KpOmpA are:

MKATVAQAAPKDN TWYAGGKLGW SQYHDTGFYGNFQNNNGPTRNDQ LGAGAFGG  
 YQV NPY LGFEMGYDW LGRMAYKGSVDNGAFKA QGVQLTAKLGYP ITDD LDIYTRL  
 GGMVWRA DSKGNYASTGVSRSSEH DTGVSPVFAGGVEWA VTRD IATRLEY QWVNNI  
 GDAGTVGTRPDNGM LSLGVSYRF GQEDAAPVVAPAPAPAPEVAT KHFTL KSDVLF  
 NFNKATL KPEGQQALDQLYTQLSNMDP KDGSAVVLGYTDRI GSEAYNQQLSE KRA  
 QSVVDYLVAK KGIPAG KISARGMGESTPVTGNTCDNV KKARAALIDCLAPDRRVEIE  
 VKGY KEVVTQPAALELVPRGSVEHHHHH

In each case above, the Lysine residues (K) are indicated with bold and underlined letters, while the eight  $\beta$ -sheets (according to the solution state NMR structure [1]) are in black boxes and the four extracellular loops are in red characters. Since a structure of the C-terminal domain is not available, the positions of the Lysine residues as a function of the secondary structure elements in that domain are unknown. **Fig. 4.1** shows the positions of the K-sites on the membrane topology map of N-KpOmpA, as determined from the structure elucidation by NMR. For N-KpOmpA (as well as for the barrel region of F-kpOmpA) there are two Lysine residues found in the first and in the fourth  $\beta$ -sheets, respectively. These two K-sites are termed 'protected', because they belong to the hydrophobic section of the molecule and are normally found in micellar or bilayer surrounding, which leaves them inaccessible for the enzyme. All other K-sites are considered potential targets for Lys-C, since they belong either to the surface-exposed areas of the barrel, or to the C-terminal domain with unknown topology. Theoretically, if N-KpOmpA is digested by the enzyme to the maximal possible level (i.e. every one of the seven K-sites is cleaved), this will separate the protein chain into 8 fragments (named from N1 to N8) with different masses, as sequentially listed below (for clarity, the initial residues of the longest fragments are omitted):

Fragment ( <b>N-KpOmpA</b> )	amino acids	M.W. [Da]
N1. ARIM <u>K</u>	5	617.80
N2. AIFVLNAAP <u>K</u>	10	1043.26
N3. DNTWYAGG <u>K</u>	9	1011.06
N4. ...VNPYLGFEMGYDWLGRMAY <u>K</u>	57	6402.94
N5. GSVDNGAF <u>K</u>	9	893.95
N6. AQGVQLT <u>K</u>	9	915.05
N7. LGYPITDDLDIYTRLGGMVWRAD <u>S</u> <u>K</u>	25	2856.24
N8. ...FGQEDAAPVVAPAPAPAPEHHHHHH	92	9755.58

TOTAL SUMMED M.W. = 23496 - 7x H<sub>2</sub>O = 23370 Da (intact chain).



**Fig. 4.1.** Topology map of N-kpOmpA in DHPC micelles according to the solution state NMR data [1]. The approximate limits of the outer membrane, the four extracellular loops (L1-4) and the three periplasmic turns (T1-3) are indicated. Residues in squares are found in the  $\beta$ -barrel and those in circles - in the more flexible parts of the molecule. The first and the last amino acid in each  $\beta$ -sheet are numerated (black numbers). The Lysine residues are indicated with red background and numerated (red numbers) according to the putative fragment which ends with the respective K-site. Together with the Lysines, the Arginines (blue background) represent targets for Trypsin. Following a simplified adaptation of **Fig. 6** from Chapter 3, residues in the loops/turns areas that exhibit restricted mobility monitored by conformational exchange and those characterized by intermediate  $^1\text{H}$ - $^{15}\text{N}$  NOEs of  $0.6 \pm 0.1$  are encircled in orange.

Respectively, complete digestion of F-KpOmpA will produce a total of 19 such fragments (named from F1 to F19):

Fragment ( <b>F-KpOmpA</b> )	amino acids	M.W. [Da]
F1. <u>MK</u>	2	277.38
F2. ATVAQAAP <u>K</u>	9	855.99
F3. DNTWYAGG <u>K</u>	9	1011.06
F4. . . . VNPYLGFEMGYDWLGRMAY <u>K</u>	57	6402.94
F5. GSVDNGAF <u>K</u>	9	893.95
F6. AQGVQLT <u>A</u> <u>K</u>	9	915.05
F7. LGYPITDDLDIYTRLGGMVWRAD <u>S</u> <u>K</u>	25	2856.24
F8. . . . FGQEDAAPVVAPAPAPAPEVAT <u>K</u>	90	9332.23
F9. HFTL <u>K</u>	5	644.78
F10. SDVLFNFN <u>K</u>	9	1083.21
F11. ATL <u>K</u>	4	431.53
F12. PEGQQALDQLYTQLSNMDP <u>K</u>	20	2276.50
F13. DGSAVVLGYTDRIGSEAYNQQLSE <u>K</u>	25	2700.89
F14. RAQSVVDYLVA <u>K</u>	12	1348.56
F15. GIPAG <u>K</u>	6	541.65
F16. ISARGMGESTPVTGNTCDNV <u>K</u>	21	2137.36
F17. ARAALIDCLAPDRRVEIEV <u>K</u>	20	2238.63
F18. GY <u>K</u>	3	366.42
F19. EVVTQPAALELVPRGSVEHHHHHH	24	2716.99
TOTAL SUMMED M.W. = 39031 - 18x H <sub>2</sub> O = 38707 Da (intact chain).		

As in the case of N-KpOmpA (23 370 Da) shown previously, fragments 1-8 of F-KpOmpA also compose the  $\beta$ -barrel (in this case the barrel is 22 419 Da) with some differences found in N1/F1 and N2/F2 (different initial sequence in each construct), and in N8/F8 (different composition of the flexible hinge region, which in the case of F-KpOmpA does not end with a His-tag but continues towards the C-terminal domain). Fragments 3-7, however, are identical for the two constructs. Fragments 9-19 in F-KpOmpA compose its C-terminal domain (16 306 Da). For the sake of clarity and due to the fact that the two protein constructs would have different numbering in essentially the same amino acids (i.e. W19 in N-KpOmpA is the same as W15 in F-KpOmpA), we will now numerate only the Lysine residues in the order of their appearance on the polypeptide chain. For instance, the end of

the first fragment (either N1 or F1) will be marked as K1, while the end of the second fragment (N2 or F2) will be K2, etc. This nomenclature was used on **Fig. 4.1** as well.

Note that the masses of several fragments (particularly in the more complex case of F-KpOmpA) are quite close to each other, which can obstruct the data analysis at a later point. However, several other fragments are of larger sizes and are easily distinguished from the rest. Such fragments are N4/F4 which is bordered by one 'protected' Lysine (K3) and N8/F8 which is the sole individual fragment with such notably large size. These can serve as 'reference points' during the mass spectroscopy measurements.

Nevertheless, an incomplete enzyme digestion (for example, when certain K-sites are 'protected') will produce not only individual fragments, but also a variety of sequential combinations of these fragments. Therefore one needs to consider all possible such combinations, which can later be found (or not) on the spectra. Complete sets of these theoretical combinations (as well as the sizes of the individual fragments) are shown for N- and F-kpOmpA on **Table 4.1** and **Table 4.2**, respectively.

**Table 4.1. N-KpOmpA** - all possible fragment combinations after Lys-C digestion.

Start N1		Start N2		Start N3		Start N4		Start N5		Start N6		Start N7		Start N8	
Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.
N1-2	1643														
N1-3	2636	N2-3	2036												
N1-4	9021	N2-4	8421	N3-4	7396										
N1-5	9897	N2-5	9297	N3-5	8272	N4-5	7279								
N1-6	10794	N2-6	10265	N3-6	9169	N4-6	8176	N5-6	1791						
N1-7	13632	N2-7	13032	N3-7	12007	N4-7	11014	N5-7	4629	N6-7	3753				
N1-8	23370	N2-8	22770	N3-8	21745	N4-8	20752	N5-8	14367	N6-8	13491	N7-8	12594		
<b>N1 =</b>	<b>618</b>	<b>N2 =</b>	<b>1043</b>	<b>N3 =</b>	<b>1011</b>	<b>N4 =</b>	<b>6403</b>	<b>N5 =</b>	<b>894</b>	<b>N6 =</b>	<b>915</b>	<b>N7 =</b>	<b>2856</b>	<b>N8 =</b>	<b>9756</b>

'Frag' denotes the combination of fragments (ex. N1-2, N1-3, etc.) from the chain, starting from (and including) a given fragment (ex. N1).

'Start' denotes the number of the starting fragment (ex. N1, N2, etc.), after which the weights of all combinations are estimated.

'M.W.' denotes the molecular weight (in Daltons) of each combination of fragments.

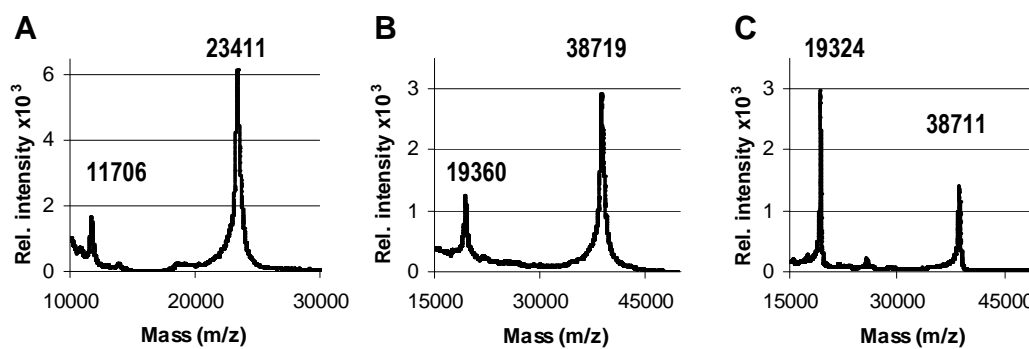
The sizes of the individual fragments (shaded in grey) are shown at the bottom.

**Table 4.2. F-KpOmpA** - all possible fragment combinations after Lys-C digestion (the indications are as in Table 4.1).

Start	F1	Start	F2	Start	F3	Start	F4	Start	F5	Start	F6	Start	F7	Start	F8	Start	F9
Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.
F1-2	1115																
F1-3	2108	F2-3	1849														
F1-4	8493	F2-4	8234	F3-4	7396												
F1-5	9369	F2-5	9110	F3-5	8272	F4-5	7279										
F1-6	10266	F2-6	10007	F3-6	9169	F4-6	8176	F5-6	1791								
F1-7	13105	F2-7	12845	F3-7	12007	F4-7	11014	F5-7	4629	F6-7	3753						
F1-8	22419	F2-8	22159	F3-8	21321	F4-8	20328	F5-8	13943	F6-8	13068	F7-8	12170				
F1-9	23046	F2-9	22786	F3-9	21948	F4-9	20955	F5-9	14570	F6-9	13694	F7-9	12797	F8-9	9959		
F1-10	24111	F2-10	23851	F3-10	23013	F4-10	22020	F5-10	15635	F6-10	14760	F7-10	13862	F8-10	11024	F9-10	1710
F1-11	24524	F2-11	24265	F3-11	23427	F4-11	22434	F5-11	16049	F6-11	15173	F7-11	14276	F8-11	11438	F9-11	2124
F1-12	26783	F2-12	26523	F3-12	25685	F4-12	24692	F5-12	18307	F6-12	17432	F7-12	16534	F8-12	13696	F9-12	4382
F1-13	29466	F2-13	29206	F3-13	28368	F4-13	27375	F5-13	20990	F6-13	20114	F7-13	19217	F8-13	16379	F9-13	7065
F1-14	30796	F2-14	30537	F3-14	29699	F4-14	28706	F5-14	22321	F6-14	21445	F7-14	20548	F8-14	17710	F9-14	8395
F1-15	31320	F2-15	31061	F3-15	30223	F4-15	29230	F5-15	22845	F6-15	21969	F7-15	21072	F8-15	18233	F9-15	8919
F1-16	33439	F2-16	33180	F3-16	32342	F4-16	31349	F5-16	24964	F6-16	24088	F7-16	23191	F8-16	20353	F9-16	11038
F1-17	35660	F2-17	35401	F3-17	34563	F4-17	33570	F5-17	27185	F6-17	26309	F7-17	25412	F8-17	22573	F9-17	13259
F1-18	36008	F2-18	35749	F3-18	34911	F4-18	33918	F5-18	27533	F6-18	26657	F7-18	25760	F8-18	22922	F9-18	13608
F1-19	38707	F2-19	38448	F3-19	37610	F4-19	36617	F5-19	30232	F6-19	29356	F7-19	28459	F8-19	25621	F9-19	16307
<b>F1 =</b>	<b>277</b>	<b>F2 =</b>	<b>856</b>	<b>F3 =</b>	<b>1011</b>	<b>F4 =</b>	<b>6403</b>	<b>F5 =</b>	<b>894</b>	<b>F6 =</b>	<b>915</b>	<b>F7 =</b>	<b>2856</b>	<b>F8 =</b>	<b>9332</b>	<b>F9 =</b>	<b>645</b>
Start	F10	Start	F11	Start	F12	Start	F13	Start	F14	Start	F15	Start	F16	Start	F17	Start	F18
Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.
F10-11	1497																
F10-12	3755	F11-12	2690														
F10-13	6438	F11-13	5373	F12-13	4959												
F10-14	7769	F11-14	6703	F12-14	6290	F13-14	4031										
F10-15	8292	F11-15	7227	F12-15	6814	F13-15	4555	F14-15	1872								
F10-16	10412	F11-16	9346	F12-16	8933	F13-16	6674	F14-16	3992	F15-16	2661						
F10-17	12632	F11-17	11567	F12-17	11154	F13-17	8895	F14-17	6212	F15-17	4882	F16-17	4358				
F10-18	12981	F11-18	11916	F12-18	11502	F13-18	9244	F14-18	6561	F15-18	5230	F16-18	4706	F17-18	2587		
F10-19	15680	F11-19	14615	F12-19	14201	F13-19	11943	F14-19	9260	F15-19	7929	F16-19	7405	F17-19	5286	F18-19	3065
<b>F10 =</b>	<b>1083</b>	<b>F11 =</b>	<b>432</b>	<b>F12 =</b>	<b>2277</b>	<b>F13 =</b>	<b>2700</b>	<b>F14 =</b>	<b>1349</b>	<b>F15 =</b>	<b>542</b>	<b>F16 =</b>	<b>2137</b>	<b>F17 =</b>	<b>2239</b>	<b>F18 =</b>	<b>366</b>
																<b>F19 =</b>	<b>2717</b>

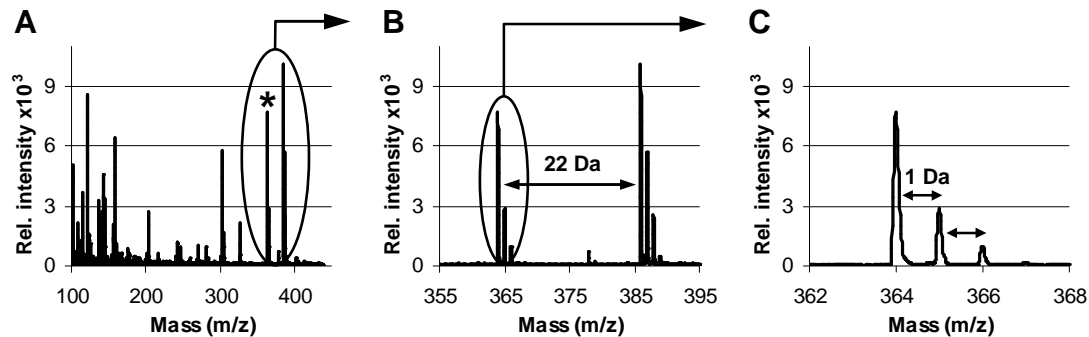


As a starting point (and before initiating the enzyme digestion experiments), the visualization of the two protein constructs in different environments was tested by MALDI-TOF. **Fig. 4.2** shows spectra of N- and F-KpOmpA in Zwittergent 3-14 micelles or reconstituted in *E. coli* Polar Lipids Extract (PLE) bilayers.



**Fig. 4.2.** MALDI-TOF spectra (acquired in linear mode) of different KpOmpA preparations. **(A)** N-KpOmpA spectrum (matrix: SPA) in 0.1% Zwittergent 3-14 micelles exhibiting the protein peak at the expected molecular weight and a 'double ion' peak with twice smaller mass-to-charge ratio. **(B)** F-KpOmpA spectrum (matrix: SPA) in 0.1% Zwittergent 3-14 showing the protein at 38 719 Da and a 'double ion' peak with twice smaller size. **(C)** F-KpOmpA spectrum (matrix: CHCA) of the protein reconstituted in PLE liposomes, exhibiting similar values as in (B).

The measured values are slightly different than the theoretically calculated ones due to the lower accuracy of the MALDI-TOF linear measurement mode. Significantly higher resolution can be obtained by using the reflectron mode of the spectrometer, as exemplified on **Fig. 4.3**. In this case, only the detergent buffer of 1% Zwittergent 3-14 (25 mM Tris (pH 8.5), 150 mM NaCl) was measured in the range of small molecular weights. It is noticeable the large amount of signals (**Fig. 4.3A**) emanating from the simple mixture of the detergent buffer and the matrix (in this case - dihydroxybenzoic acid, DHB). Some of this variance is produced by the very same chemical compounds, detected as their sodium salts and/or as a function of their isotopic distribution. For instance, the peak of the Zwittergent 3-14 molecules (M.W. 363.6) is identified on **Fig. 4.3A** and a zoom-in on that area (on **Fig. 4.3B**) shows another signal from the detergent, but shifted with the mass of one sodium atom (~22 Da). Further zoom-in into the main detergent peak (**Fig. 4.3C**) reveals the 1-Da differences among the isotopic distributions of the atoms which compose the Zwittergent 3-14 molecule.



**Fig. 4.3.** Reflectron mode spectrum (matrix: DHB) of 1% Zwittergent 3-14 buffer (25 mM Tris (pH 8.5), 150 mM NaCl). (A), (B) and (C) are three subsequent zooms on the indicated encircled areas. The main detergent peak (364 Da, shown on (C)) is denoted with an asterisk on (A).

It is thus clear that adding protein fragments to such a sample and acquiring under reflectron mode will produce heavily overcrowded spectra, turning its great resolution into a double-edged knife. On the other hand, the sizes of the observed protein fragments and combinations of fragments will often expand beyond the instrumental range of the reflectron mode (up to ~10 kDa). Therefore the linear mode was used for the most of the MALDI-TOF measurements throughout this chapter, with some complementary acquisitions in reflectron mode when that was necessary and appropriate. One possibility to reduce the overcrowding of the spectra is to subject the digested protein sample to a chromatography in a reversed phase column, in order to remove the detergent/lipids prior to the measurement. However, as it is shown further down this chapter, very often the protein was subjected to incomplete digestion (particularly during the digestion kinetics experiments) which left large population of full-size molecules (or protein domains) that appeared to be difficult to elute from such columns by the acetonitrile (AcN) solution. Considering the non-quantitative nature of the MALDI-TOF, one such additional scrambling with the total amount of eluted material (and the proportions of the different protein fragments) interfered heavily with the observable signals, sometimes to the extent of complete disappearance of certain fragments from the spectra (data not shown). This type of sample treatment was therefore avoided and all samples were prepared as described in Material and methods.

The non-quantitative nature of the MALDI-TOF arises most prominently from the fact that the different proteins or protein fragments have different excitation maxima as a function of the laser power. For this reason, an attempt was made for each measurement to visualize as many components of the sample as possible, by slowly modifying the laser power

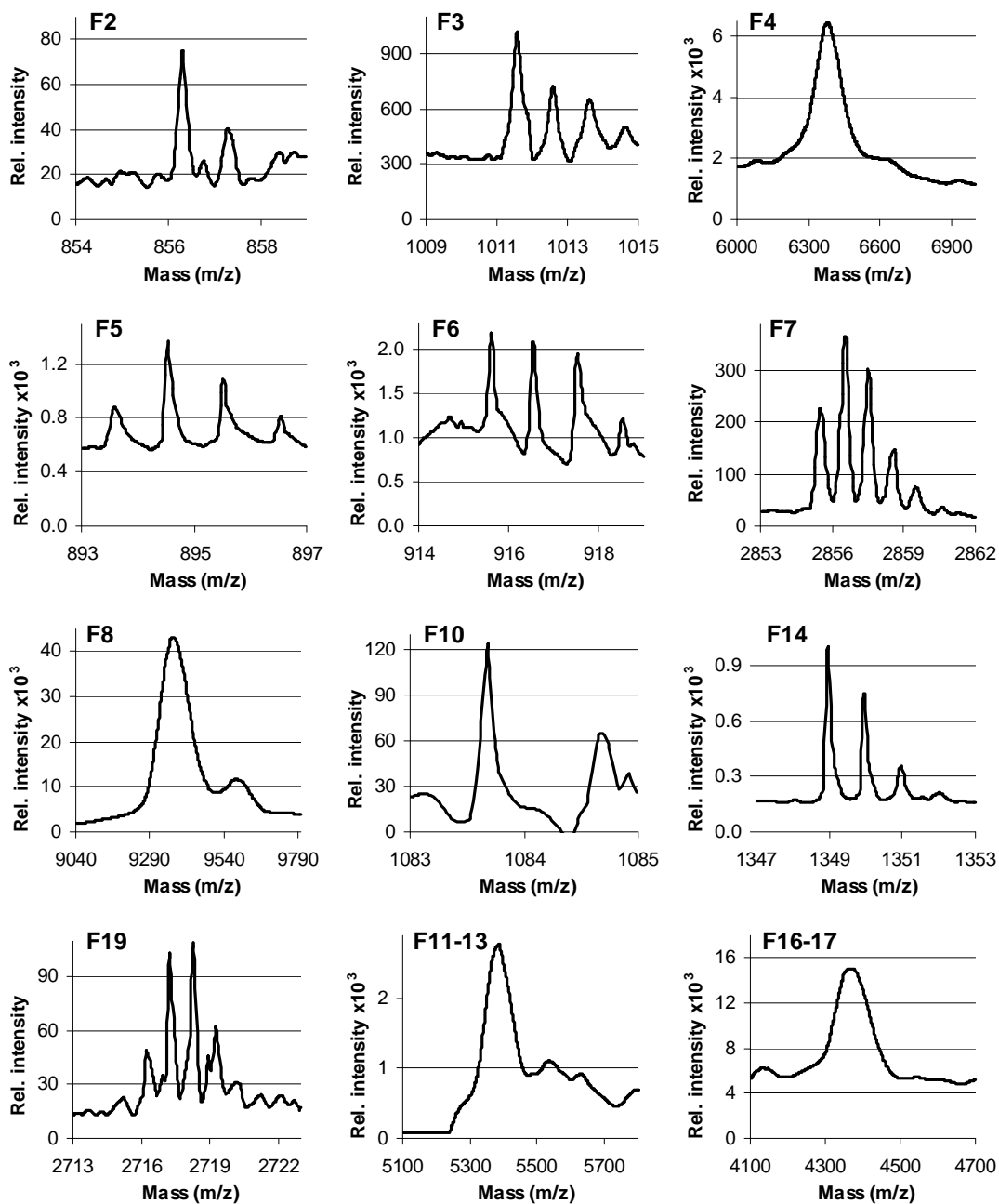
throughout the 1000 shots of acquisition. Using such a 'sweep mode' did not, of course, provide quantitative value to the measurement, but only aimed to qualitatively detect the maximal number of protein fragments contained in the sample. In addition to the laser power, the type of utilized matrix can also influence the measurement by preferentially visualizing certain fragments at the expense of others. An effort was made for each sample to be measured in at least two matrix environments (usually SPA and CHCA, in the majority of the cases), with subsequent presentation of the stronger signals for any given protein fragment or combination of fragments.

Since the Lys-C enzyme is highly active (particularly at reaction temperature of 37°C), the fragmentation of the protein chain usually happened within minutes or 1 hour. In general, longer reaction times often lead to complete degradation of the digested protein down to its individual fragments, thus masking the sequential steps of cleavage and preventing the observation of the protein digestion as a function of time, i.e. the kinetics of the digestion. This type of experiments is discussed separately in this chapter. The next section analyzes the KpOmpA digestion after long times at high temperature, allowing maximal fragmentation of the polypeptide chain. These can be described as conditions of 'complete digestion' of the protein. Since Lys-C is an enzyme with high specificity for Lysine residues, the smallest protein pieces that can be produced in this way are the individual fragments described above. These fragments, regardless of their variation in size, are the 'building blocks' of the protein chain in the context of the digestion experiments. Larger protein segments (containing more than one fragment) found at the end of the 'complete digestion' process can be considered as 'infinitely stable' (if present at all). With other words, the inability of the enzyme to cleave certain K-sites for one reason or another can be monitored at such reaction conditions. This is useful in the case of folded membrane proteins, whose hydrophobic core is embedded within the detergent micelle or lipid bilayer - environments that physically obstruct the access of the enzyme to some sections of the protein, while still allowing access to others (**Fig. 4.1**).

## 4.4. Complete enzyme digestion of KpOmpA with Lys-C

### 4.4.1. Digestion of urea-denatured F-KpOmpA

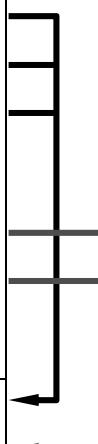
Before analyzing the enzyme cleavage on KpOmpA in different hydrophobic environments, the protein was subjected to 'complete digestion' in its denatured form in order to probe the visualization of all individual fragments. F-KpOmpA was urea-denatured and subjected to digestion as described in Material and methods. As seen on **Fig. 4.4**, the majority of the individual fragments could be tracked down and nearly complete reconstruction of the protein chain could be established (**Table 4.3**). Interestingly, despite of the long reaction time and high temperature, the enzyme did not manage to cleave certain K-sites even on a completely denatured protein chain. Fragments F11, 12 and 13 were not observed alone but in combination, and the same was sometimes observed for fragments F16 and 17. A possible explanation for this may be found in a work investigating statistically large databases of produced protein fragments after digestion with the Lys-C enzyme [5]. Of note, the authors point out that certain amino acids combination on the chain, found in close proximity to the K-site, may affect the cleavage capacity of the enzyme. For instance, protein fragments (ending with a given K-site) are very rarely observed when there is a Proline or Glutamate residue immediately after the targeted Lysine (position +1). Similar is the outcome when there is a Glutamate or Aspartate residue at position +2 after the Lysine.



**Fig. 4.4.** Complete digestion of unfolded F-KpOmpA in ~0.7M urea and tracking of the individual fragments. Each fragment (FX) or combination of fragments is indicated on the corresponding truncated spectrum. Compare with **Table 4.3** on the next page.

**Table 4.3.** Details related to the tracking of the individual F-KpOmpA fragments presented on **Fig. 4.4**. The individual fragments are listed in the first column, while the second column denotes their calculated (i.e. expected) molecular weights ( $M.W._{theor}$ ). The third column presents the measured mass-to-charge ratios ( $M.W._{meas}$ ), if available. ND stands for 'not detected'. Note that some fragments (in bold letters) are often found in combinations (as shown at the bottom of the table), regardless of the denatured state of the protein during the digestion. The column 'Mode' denotes if the fragment was detected during acquisition in linear (L) or in reflectron (R) mode.

Fragment	$M.W._{theor}$	$M.W._{meas}$	Mode
F1	277	ND	----
F2	856	857	R
F3	1011	1012	R
F4	6403	6383	L
F5	894	895	R
F6	915	916	R
F7	2856	2856	R
F8	9332	9374	L
F9	645	ND	----
F10	1083	1083	R
<b>F11</b>	432	ND	----
<b>F12</b>	2277	ND alone	----
<b>F13</b>	2700	ND alone	----
F14	1349	1349	R
F15	542	ND	----
<b>F16</b>	2137	ND alone	----
<b>F17</b>	2239	ND alone	----
F18	366	ND	----
F19	2717	2717	R
F11-13	5373	5386	L
F16-17	4358	4372	L



A close inspection on the F-KpOmpA chain in the area surrounding K11, K12 and K16 (which are the borders between fragments F11/12, F12/13 and F16/17, respectively) reveals the presence of such unfavorable (for enzyme cleavage) combinations, as shown below:

Lysine	Border	Surroundings
K11	F11/12	ATL <b>K</b> PE <b>G</b> Q
K12	F12/13	...SNM <b>D</b> P <b>K</b> D <b>G</b> SA
K16	F16/17	...NTC <b>D</b> N <b>V</b> <b>K</b> ARAA

The fragments-bordering K-sites are denoted with bold and underlined letters, while amino acids that obstruct the enzyme cleavage when present in a particular position are denoted in black boxes. The presence of Proline immediately after K11 is particularly unfavorable for a Lys-C cleavage, even if this segment is physically exposed to the enzyme. The case of K12 is not so restrictive [5], but still fragment F13 is rarely observed (if at all). Surprisingly, the surroundings of K16 should not interfere heavily with a digestion at that point. This may represent a new ‘forbidden for cleavage’ site for the Lys-C enzyme in the presence of urea, or the explanation lies elsewhere. Of note, digestion of the folded C-KpOmpA construct with Lys-C (described further down this chapter) managed to cleave efficiently this K-site.

#### 4.4.2. Digestion of N- and F-KpOmpA in detergent micelles

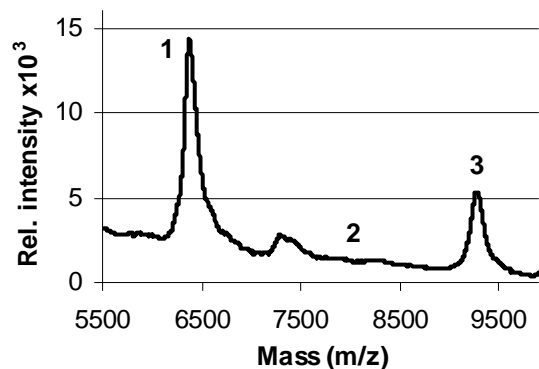
With the signals of the individual fragments established, the digested experiments can be performed on folded protein in order to compare the cleavage patterns of folded and unfolded polypeptides in conditions of ‘complete digestion’. As seen on **Fig. 4.1**, the hydrophobic environment (micelle or bilayer) surrounding the barrel core of the protein should protect some of the Lysine residues by denying access to the enzyme. Such residues are K3 and K6 (bordering fragments 3/4 and 6/7, respectively), while all others K-sites are considered potentially cleavable targets. As observed during the denatured protein digestion, K3 and K4 are also cleavable points, provided that Lys-C has access to unfolded chain in these areas. Therefore, in this nomenclature, K-sites like K11 are termed ‘forbidden’ (Lys-C does not cleave them even if it has access to them), while K-sites like K3 are termed ‘protected’ (cleavable by Lys-C, but only if reached). In particular, the potentially cleavable K-sites found on the extracellular loops of KpOmpA (K4, K5 and K7, **Fig. 4.1**) can be thus analyzed in terms of enzyme accessibility and the related to that mobility of the loops area.

**Fig. 4.5** compares linear mode spectra of the two digested KpOmpA constructs in their folded state with the previously discussed digestion of denatured F-KpOmpA (in 0.7M urea). The signal of Peak 1 on each spectrum corresponds to fragment No. 4 which is bordered by one ‘protected’ Lysine residue (K3) found in the first  $\beta$ -sheet of the barrel. This fragment is with identical size in the two protein constructs (F4 = N4 = 6403 Da). The presence of F4 alone on **Fig. 4.5A** (unfolded F-KpOmpA) is indicative for missing protection (i.e. exposure) of the protein chain to the Lys-C, as expected. The properly folded  $\beta$ -barrel will keep K3 hidden in the micelle, but it will expose the previous K2 (N-terminus) and the following K4 (Loop 2) to the enzyme, thus producing the combination of fragments F3-4. In fact, K2 itself falls to a

certain extent into the 'forbidden for cleavage' category, because this Lysine is bordered by Proline and Aspartate residues - a situation that does not favor cleavage by Lys-C [5]. Since F2 was detected in the urea-denatured sample, perhaps not only the amino acids composition of the chain but also the presence of the detergent micelle in close proximity to K2 further enhances the miss-cleavage of that K-site. Hence the actually observed 'protected' combination of fragments in the micellar sample is F2-4, which is detected only in the case of folded protein (Peak 2 on **Fig. 4.5B**, missing on **A**). It should be noted that F4 alone is also detected in the micellar sample. It is likely that small fraction of the protein population loses the fold of the  $\beta$ -barrel and becomes susceptible to digestion, especially after ~16 hours at 37°C during the reaction incubation. The amount of this fraction can not be estimated with these experiments since quantitative values can not be attributed to the MALDI-TOF. It is thus possible that the method preferentially visualizes this subpopulation. For the moment it can only be mentioned that the combination F2-4 is found exclusively in the micellar sample and not in the urea-denatured one, which provides a way to distinguish between these two states. The same conclusions apply for the digestion of the transmembrane domain alone, N-KpOmpA, shown on **Fig. 4.5C**. The third signal described on **Fig. 4.5** (Peak 3) belongs to fragment No. 8 (F8 or N8). F8/N8 is used as a good reference point since it is the largest individual fragment (>9000 Da) and it is bordered by two potentially cleavable Lysine residues (in F-KpOmpA) or only one such residue (in N-KpOmpA), i.e. this fragment can always be observed, regardless of the sample particularities. Of note, the sizes of F8 and N8 are slightly different, as are those of F2 and N2. This is reflected as a visible difference between the peak positions of F2-4 and N2-4, and between these of F8 and N8 on **Fig. 4.5 B** and **C**, respectively (compare the values presented to the right of each spectrum). Such a small (but detectable) 'shift' of these peaks further confirms that they were correctly assigned and points out their suitability to serve as reference signals.



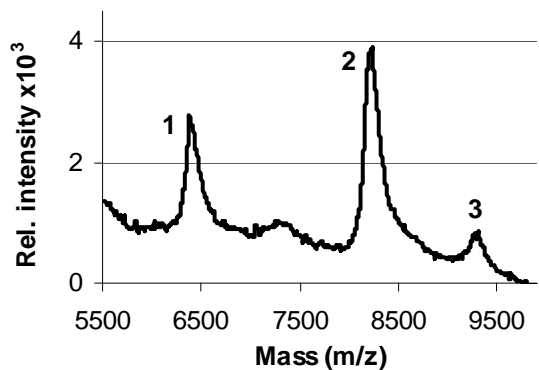
**A) F-KpOmpA / 0.7M urea**



Unfolded protein. Matrix: SPA.

Peak	M.W. <sub>.meas</sub>	M.W. <sub>.theor</sub>
1. F4	6390	6403
2. F2-4	-----	8234
3. F8	9281	9332

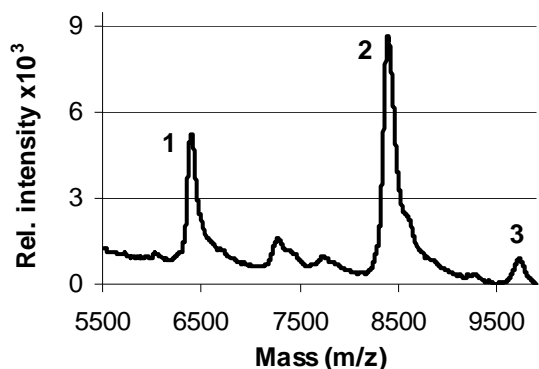
**B) F-KpOmpA / Zwittergent 3-14**



Folded protein. Matrix: CHCA.

Peak	M.W. <sub>.meas</sub>	M.W. <sub>.theor</sub>
1. F4	6408	6403
2. F2-4	8236	8234
3. F8	9311	9332

**C) N-KpOmpA / Zwittergent 3-14**



Folded protein. Matrix: SPA.

Peak	M.W. <sub>.meas</sub>	M.W. <sub>.theor</sub>
1. N4	6405	6403
2. N2-4	8406	8421
3. N8	9723	9756

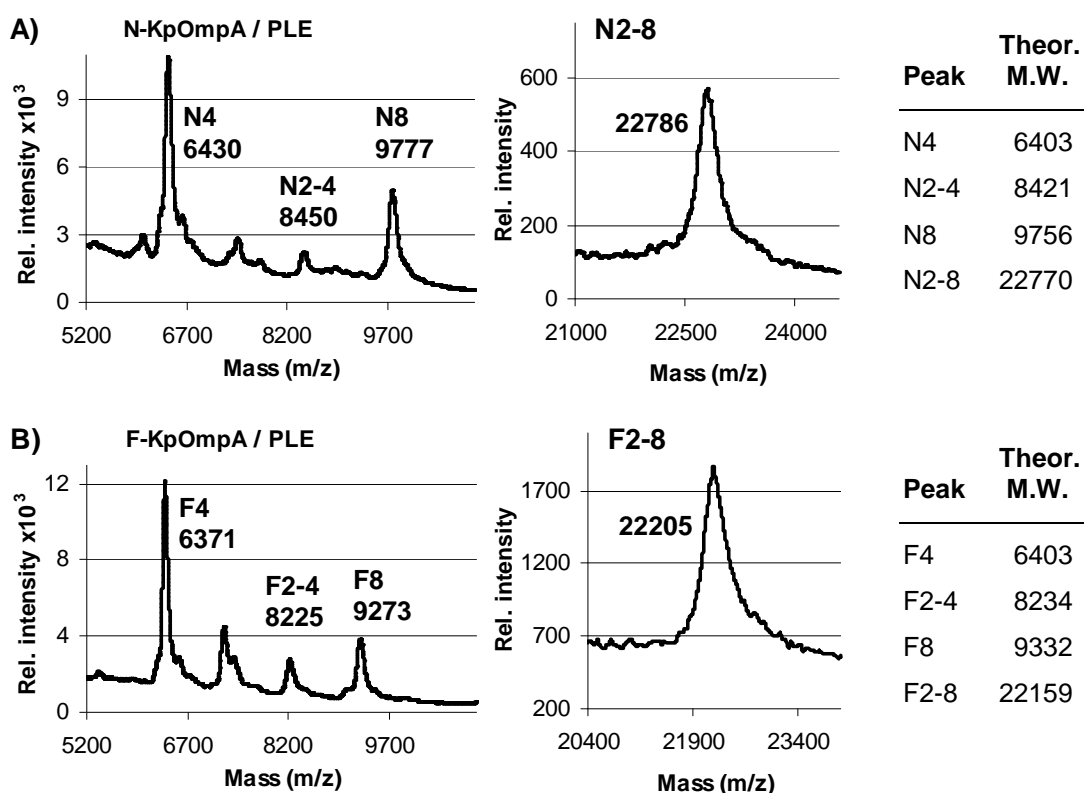
**Fig. 4.5.** Comparison of the digestion patterns of folded and unfolded KpOmpA on three spectra acquired in linear mode. In each case, the protein-to-enzyme ratio is 50:1 (w/w). **(A)** Denatured F-KpOmpA digested in 0.7M urea. **(B)** Folded F-KpOmpA digested in 0.1% Zwittergent 3-14. **(C)** Folded N-KpOmpA digested in 0.1% Zwittergent 3-14.

#### 4.4.3. Digestion of N- and F-KpOmpA in liposomes

Up to this point, the micellar KpOmpA preparations were analyzed and the 'protective effect' of the detergent micelle was visualized. The presence of both F4 (alone or in the F2-4 combination) and F8 pointed out that K4 and K7 (found in the loops area of the  $\beta$ -barrel) are subjected (at least to a certain extent) to enzyme cleavage, as expected. However, K5 (on Loop 2) was not found to be digested, which often produced the combination F5-7 (data not shown). In this combination, K6 itself is 'protected' since it belongs to the fourth  $\beta$ -sheet (**Fig. 4.1**). In micelles, the very close proximity of K5 to the structured barrel core is presumably the reason for the miss-cleavage at that point. This brings the question whether the cleavage pattern (of the protein in general and of the loops in particular) will remain the same when the protein is digested after reconstitution in liposomes - a different environment that can provide different constraints on the extracellular loops. For this reason, dialysis-driven reconstitution of N- and F-KpOmpA in Polar Lipids Extract (PLE) was performed as described in Chapter 1. The digestion experiments of these samples are described in Material and methods, with the lipid-to-protein ratios (LPR) indicated in text for the individual samples.

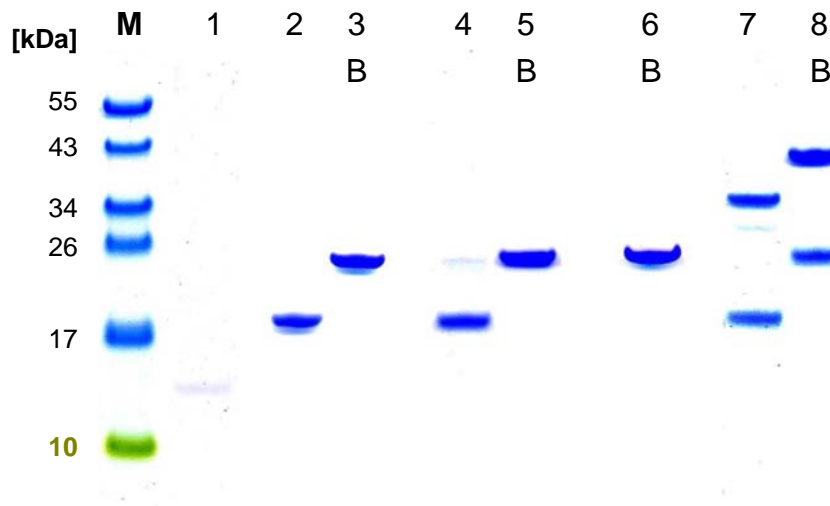
**Fig. 4.6** shows the resulting linear mode spectra for N- and F-KpOmpA digested at protein-to-enzyme ratio of 50:1 (w/w). Cleavages on Lysine sites K4 and K7 were again detected, suggesting that the extracellular loops of the proteins are still (at least partially) accessible to the enzyme. Surprisingly, in both cases of reconstituted proteins the signals from fragment No. 4 (lack of protection on K3) were clearly distinguished, while these indicating combinations of fragments 2-4 were greatly reduced. With other words, the cleavage patterns of the reconstituted proteins resembled more closely those from the digestion of the denatured protein (in 0.7M urea), rather than the expected similarity with the micellar samples. Weaker signals attributed to the nearly complete transmembrane domains (N2-8 and F2-8) were also detected at higher molecular weights. Interestingly, similar results were obtained previously with another KpOmpA construct ([4]). The outcome of the proteoliposomes digestion presented above was thus not new, and yet it was unexpected since the protein reconstitution was considered successful, as described in Chapter 1. Nevertheless, the same outcome was systematically observed for both N- and F-KpOmpA and at different LPRs as well (data not shown). The possible explanation for this is twofold: (i) The reconstitution of the protein is perhaps not 100% efficient which traps certain (albeit rather small) protein population outside the membranes, leaving these molecules vulnerable to enzyme digestion. (ii) The MALDI-TOF is not a quantitative method and may, under certain conditions, preferentially visualize a population in the sample which is not the main one.

Since the experimental interest up to this point was mostly focused in observing the protein fragments in the range of 1 to 10 kDa, the laser power was manually adjusted throughout each measurement for maximal appearance of those signals, eventually leaving the intact barrel (which is with larger molecular weight) nearly invisible (**Fig. 4.6 A and B**, the spectra on the right). The abovementioned considerations thus raise the question for the true state of the majority of the protein population and what is the large-scale response of this population (in general) and the extracellular loops (in particular) to the enzyme treatment. To answer these questions one needs to couple the non-quantitative (but sensitive) mass spectrometry with another, bulk-sample representative method.



**Fig. 4.6.** Complete digestion of N- and F-KpOmpA reconstituted in PLE lipid bilayers at LPR of 0.5 (w/w). For the two cases, the protein-to-enzyme ratio is 50:1 (w/w). **(A)** Digestion of N-KpOmpA exhibiting the main reference peaks as indicated (left spectrum; matrix: SPA) and the nearly entire barrel section of the molecule (right spectrum; matrix: CHCA). **(B)** Digestion of F-KpOmpA exhibiting the main reference peaks as indicated (left spectrum; matrix: SPA) and the nearly entire barrel section of the molecule (right spectrum; matrix: CHCA). The measured weights (m/z) are indicated next to each peak and the theoretical ones - on the right side of the figure.

**Fig. 4.7** shows SDS PAGE image of aliquots of 3-5  $\mu\text{g}$  protein from each of the digestion experiments described previously. Unlike the MALDI-TOF aliquots that were mixed with TFA/AcN solution, these aliquots (from the same reaction mixtures) were added to the SDS-containing loading dye as a preparative step before the gel electrophoresis. As expected, the urea-denatured F-KpOmpA (lane 1) was completely digested by Lys-C during the reaction time. In contrast, the micelle-embedded protein (either N- or F-KpOmpA) did not disappear entirely even after an overnight incubation with the enzyme (lanes 2-5). For both N- and F-KpOmpA, their  $\beta$ -barrel domains are found at the end of the respective reactions. In the case of N-KpOmpA, the entire molecule virtually consists of the  $\beta$ -barrel, while F-KpOmpA initially contains in addition the C-terminal domain.

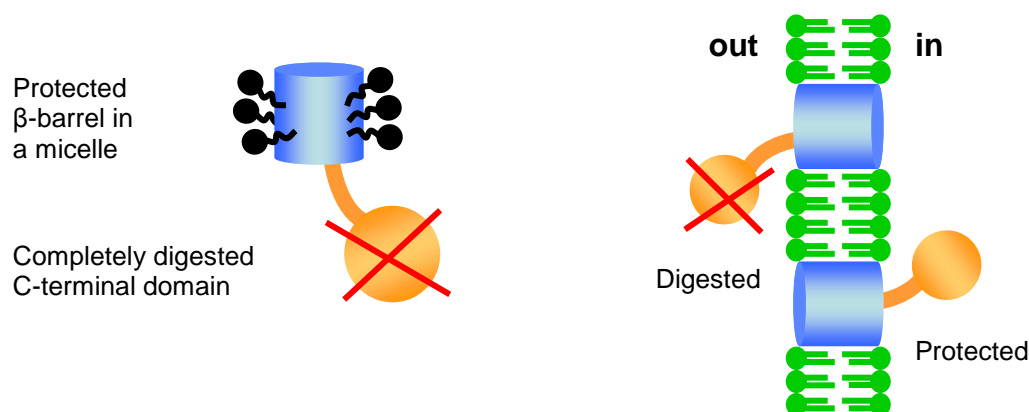


**Fig. 4.7.** SDS PAGE gel of aliquots from the digestion reactions described in text: lane 1 - urea-denatured F-KpOmpA; lanes 2 and 3 - N-KpOmpA in Zwittergent 3-14 micelles; lanes 4 and 5 - F-KpOmpA in Zwittergent 3-14 micelles; lane 6 - N-KpOmpA reconstituted in PLE (LPR 0.5 w/w); lanes 7 and 8 - F-KpOmpA reconstituted in PLE (LPR 0.5 w/w). M - marker ladder. The letter 'B' (on lanes 3, 5, 6 and 8) denotes that the aliquot was heated for 5 min at 100°C. All digestion reactions were performed at 37°C for ~16 hours, with protein-to-enzyme ratio of 50:1 (w/w).

It is therefore shown that, in the case of each protein construct, the micelle-embedded barrel core remains protected from enzyme digestion, while the soluble C-terminal domain of F-KpOmpA is successfully targeted by Lys-C and is completely absent at the end of the reaction. The bands on lanes 2 and 4 clearly belong to the  $\beta$ -barrels of the respective protein constructs, which is proved by the shift in their apparent molecular weights after boiling of the aliquots (lanes 3 and 5). Such shift of the band migration, also known as 'heat modifiability', is

a well known feature of the  $\beta$ -barrel proteins [6]. It should be noted at this point that eventual cleavage in the extracellular loops area of the barrel can remain 'invisible' in the SDS PAGE gel, because the folded barrel core structure will still migrate as an intact molecule (lanes 2 and 4). However, if the aliquot is heat-denatured, any such chemical interruption of the polypeptide chain should be visualized as disappearance of the full-length protein band and its substitution by at least two other bands of smaller sizes. Since this is not observed (lanes 3 and 5), the remaining conclusion is that the protein loops avoid enzyme cleavage on a large scale, although some minor protein population is subjected to digestion, occurring predominantly on Lysines K1, K4 and K7 (**Fig. 4.5**).

Similar is the outcome from the digestion reactions with reconstituted N- and F-KpOmpA (**Fig. 4.7**, lanes 6-8). Again, the  $\beta$ -barrels of both protein constructs remain largely intact and again, cleavage on the loops is not observed. A notable difference is the presence of two bands in the gel aliquots of the reconstituted F-kpOmpA (lanes 7 and 8). The smaller fragment corresponds to the barrel core of the molecule and the larger one - to the entire polypeptide chain (the presence of the  $\beta$ -barrel in both bands is confirmed by their migrational shift after boiling of the aliquot - lane 8). The reason for this observation is found in the conditions for protein reconstitution. Since the procedure relies on dialysis-driven detergent removal from the mixed protein / detergent / lipid micelles and the formation of lipid bilayers entrapping the protein, it is expected that the positioning of the protein molecules in the newly formed membranes is random. The absence of preferred oriented insertion leaves about one



**Fig. 4.8.** Schematic representation of the enzyme accessibility to the soluble C-terminal domain (orange color) of F-KpOmpA. In detergent micelles (left side), that domain will always be vulnerable to enzyme digestion, but when the protein is reconstituted in liposomes (right side) around one half of the C-terminal domains will remain protected in the vesicle lumen.

half of the protein population with the loops facing the outside of the vesicles, while the other half points those inwards. Respectively, about one half of the C-terminal domains will remain protected in the liposome's lumen, while the other half will be exposed to the external environment and therefore to the Lys-C enzyme (**Fig. 4.8**). Of note, the extracellular loops avoid enzyme cleavage when the proteins are reconstituted in PLE (even when the loops point towards the vesicle exterior), similarly to the micellar samples. Another aspect of similarities between these two kinds of samples is the complete digestion of the C-terminal domain, if it is accessible to the enzyme.

It was thus shown that the majority of the protein populations, for both N- and F-KpOmpA and for both the micelle and bilayer environments remain largely intact after treatment with Lys-C, with well preserved loops regions and vulnerable C-terminal domains. The complete disappearance of the soluble C-terminal domain (when exposed to the Lys-C) is somewhat expected in the conditions of 'complete digestion' (~16 hours at 37°C). On the other hand, the ability of the micelle- or lipids-protected barrel core of KpOmpA to resist cleavage even at such longer times is perhaps not surprising as well, albeit formidable in terms of protein/micelle stability. However, the lack of a large scale cleavage in the loops region of the  $\beta$ -barrel, even if the mass spectrometry clearly suggests a small scale digestion of that area, is rather intriguing. It appears that Lysine sites K4, K5 and K7 are not prone to fragmentation, presumably due to inaccessibility of the enzyme.

#### 4.4.4. Influence of pH and ionic strength on the F-KpOmpA digestion

The digestion experiments described so far were each performed in a single buffer condition - either the 0.1% Zwittergent 3-14 solution (25 mM Tris (pH 8.5), 150 mM NaCl) of the micellar protein stocks, or the 20 mM Tris (pH 8.5) of the proteoliposomes preparations. Since the mobility of the extracellular loops (and their enzyme digestion as a function of that) may be influenced by the reaction conditions, it is interesting to probe the protein vulnerability to Lys-C in different buffers, such as varying in salt concentration and acidity. Aliquots of F-KpOmpA in micelles had their buffer environment exchanged to 1 from 9 different Na-phosphate buffers, as described in Material and methods. The reason for substituting the so far used Tris is based on the more flexible pH-range obtainable with the Na-phosphate buffers (especially below pH of 7). In this way the chemical nature of the buffering compound itself is eliminated as a variable parameter, leaving that function only to the pH-value and the NaCl concentration. For the same reason, the pH was not adjusted with external compounds (such as HCl or NaOH) after the addition of NaCl and Zwittergent 3-14 powders. The

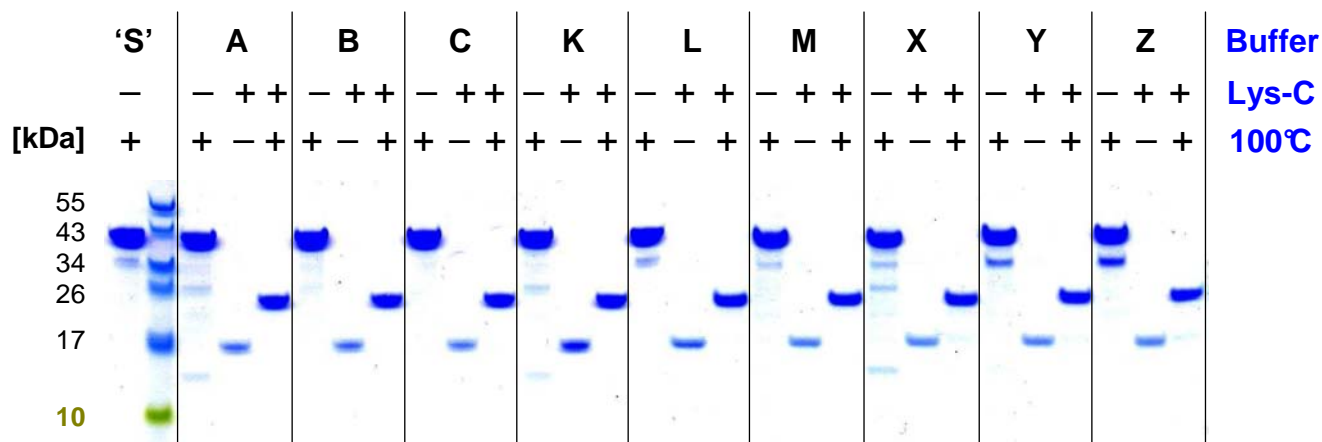
monitored pH did not differ with more than 0.1 units for every buffer. For simplicity, these 9 different buffers are encoded in ‘smaller alphabetic steps’ as a function of their pH and in ‘larger alphabetic steps’ as a function of the NaCl concentration, as schematized in **Table 4.4**. **Fig 4.9** shows the outcome of the nine digestion reactions, concluding that the final digestion products from each reaction condition are all identical with the previous observations (**Fig. 4.7**).

<b>[NaCl]</b> \ <b>pH</b>	6	7	8
0 mM	<b>A</b>	<b>B</b>	<b>C</b>
150 mM	<b>K</b>	<b>L</b>	<b>M</b>
300 mM	<b>X</b>	<b>Y</b>	<b>Z</b>

**Table 4.4.** Set of the 9 different Naphosphate buffers used for exploring the F-KpOmpA digestion with Lys-C in 0.1% Zwittergent 3-14 micelles. Each buffer is encoded with a single letter.

The variation in salt concentration and acidity (within the physiological range of pH 6-8) apparently did not provoke large scale cleavage in the loops area, as before. Again, the C-terminal domain is completely digested in all reaction conditions. The preservation of the  $\beta$ -barrel is confirmed by the ‘heat modifiable’ band present at the end of the experiment (the second and the third lane for each buffer condition). The weaker band intensities of the non-boiled enzyme-digested aliquots can be explained by: (i) smaller binding area for the Coomassie dye when the molecule preserves its barrel fold; (ii) sample concentration via water evaporation during the heating of the respective denatured aliquots. Therefore it is correct to compare band intensities only between the heat-denatured non-digested and digested aliquots (the first and the third lane for each buffer condition).

To probe the response to digestion of the protein in lipid bilayers as a function of the same 9 buffer conditions, essentially the same experiment was performed with F-KpOmpA-containing PLE liposomes (described in details in Material and methods). The LPR in this case was increased to 2 (w/w), aiming to allow easier accommodation of the protein in the newly formed bilayers, especially since this type of sample is not as requiring in terms of high protein content, as is the ssNMR sample (LPR of 0.5 w/w).



**Fig. 4.9.** Coomassie-stained SDS PAGE gel of the nine digestion reactions. The buffer conditions (from A to Z, as in **Table 4.4**) are indicated on top. The absence or presence of the enzyme (Lys-C) and of heat-denaturation (100°C) for the different aliquots are denoted with ‘-’ and ‘+’, respectively. Lane ‘S’ represents the heat-denatured F-KpOmpA starting stock solution (in 0.1% Zwittergent 3-14, 25 mM Tris (pH 8.5) and 150 mM NaCl).

Upon mixing of the prepared liposomes aliquots with one of the 9 buffers, a certain level of vesicles aggregation was macroscopically observed in some of the reaction conditions, predominantly in those at higher salt concentrations (**Table 4.5**). The higher pH, on the other hand, tended to counteract this precipitation. This effect was detected previously and discussed in Chapter 2.

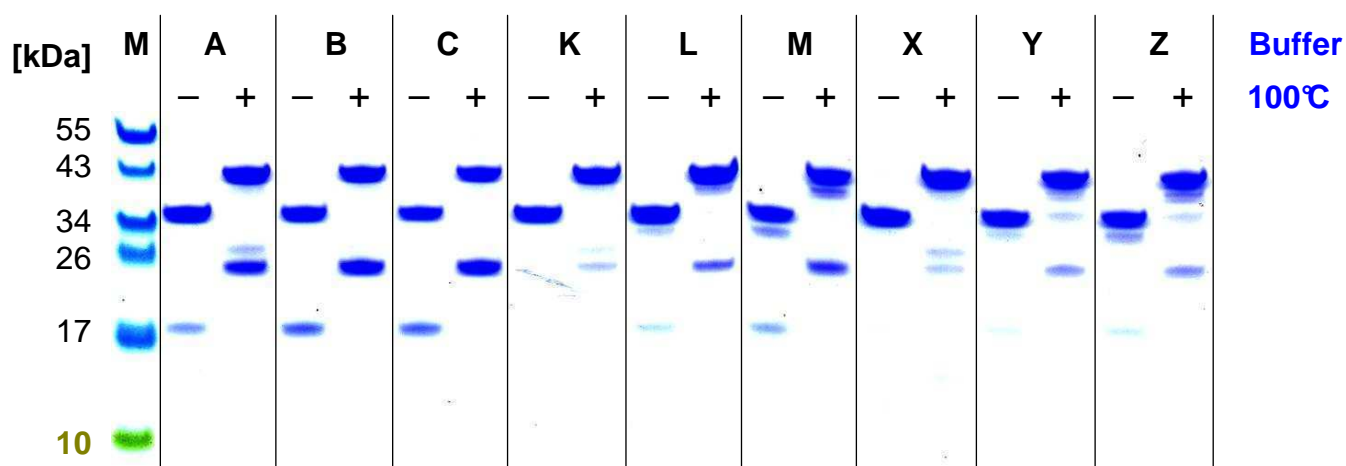
[NaCl] \ pH	6	7	8
0 mM	<b>A</b>	<b>B</b>	<b>C</b>
150 mM	<b>K</b>	<b>L</b>	<b>M</b>
300 mM	<b>X</b>	<b>Y</b>	<b>Z</b>

**Table 4.5.** Macroscopically observed proteoliposomes aggregation in the 9 buffers, formally divided into three groups: heavy precipitation (black background), moderate precipitation (grey background) and no observed precipitation (white background).

Aliquots (heat-denatured or not) from the nine digestion reactions were visualized on SDS PAGE (**Fig. 4.10**). The ‘heat-modifiable’ two-band cleavage pattern observed before (**Fig. 4.7**) appeared again, thus pointing towards similar conclusions. Namely, in all of the nine reaction conditions, the  $\beta$ -barrel of the molecule remains as an ‘infinitely stable’ unit with intact extracellular loops, while around one half of the C-terminal domains is completely degraded. The complete (or nearly complete) absence of the  $\beta$ -barrel band in all salt-



containing buffers (K-Z) is not a meaningful difference since it only reflects the aggregation propensity of these samples and the subsequent obstruction for the Lys-C to reach the exposed half of the C-terminal domains, leaving more than a half of the protein population as intact full-length chains. It is notable how this effect is counteracted by the increase in pH, as the  $\beta$ -barrel band partially reappears throughout the sequences K-to-M and X-to-Z. The three samples without NaCl (A, K and X) exhibited two additional and very weak bands (around 26 and below 17 kDa), but only when boiled. These bands represent small amounts of non-specific, thermal-induced cleavage of the denatured full-length protein somewhere in the area of its hinge region (presumably a DP-site in the beginning of the C-terminal domain). However, increasing the salt concentration and in the same time the pH (M and Z) produced one additional band (immediately below that of the full-length protein at ~39 kDa), which is present on both the heated and not heated aliquots. The appearance of this band occurs in parallel with the appearance of the  $\beta$ -barrel band in the sequences K-to-M and X-to-Z. Since the enzyme accessibility is reduced at higher salt but partially compensated at higher pH, this nearly intact F-KpOmpA chain is perhaps reflecting cleavage of the most enzyme-accessible piece of the molecule, presumably somewhere close to the C-terminus of the C-terminal domain (i.e. K17 or K18). This suggestion is further discussed below in this chapter.



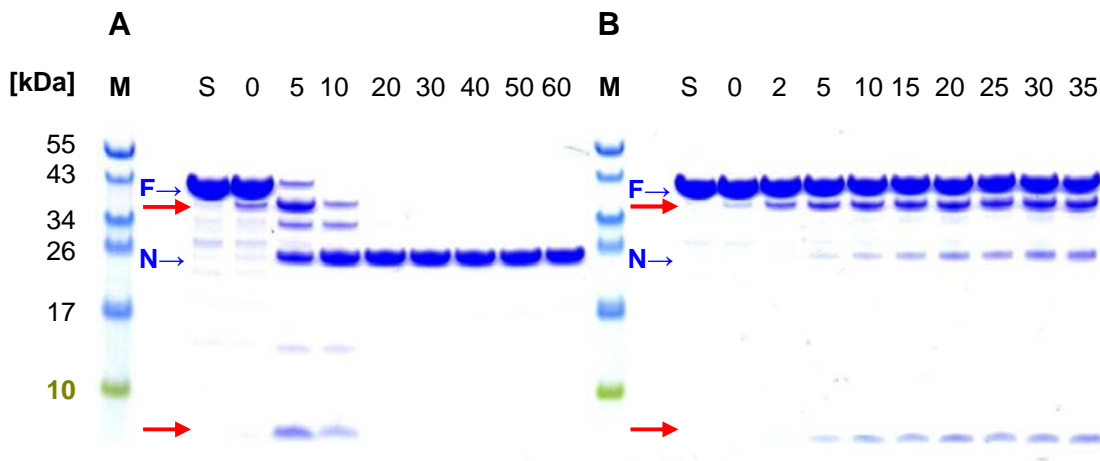
**Fig. 4.10.** SDS PAGE gel of the nine digestion reactions (F-KpOmpA in liposomes). The buffer conditions (from A to Z, as in **Table 4.4** but without detergent) are indicated on top. The absence or presence of heat-denaturation (100°C) for the different aliquots is denoted with '-' and '+', respectively. M - marker ladder.

## 4.5. Enzyme digestion kinetics in different environments

The long-time/high-temperature reaction conditions, used in all of the experiments above, ensure the maximal possible fragmentation of the polypeptide chain, thus leaving intact only the 'infinitely stable' sections of the molecule (the  $\beta$ -barrel). On the other hand, observing the gradual fragmentation of the protein as a function of time (i.e. the kinetics of digestion) may reveal which are the most sensitive, quickly degraded protein sections. An attempt for accomplishing this task is described below.

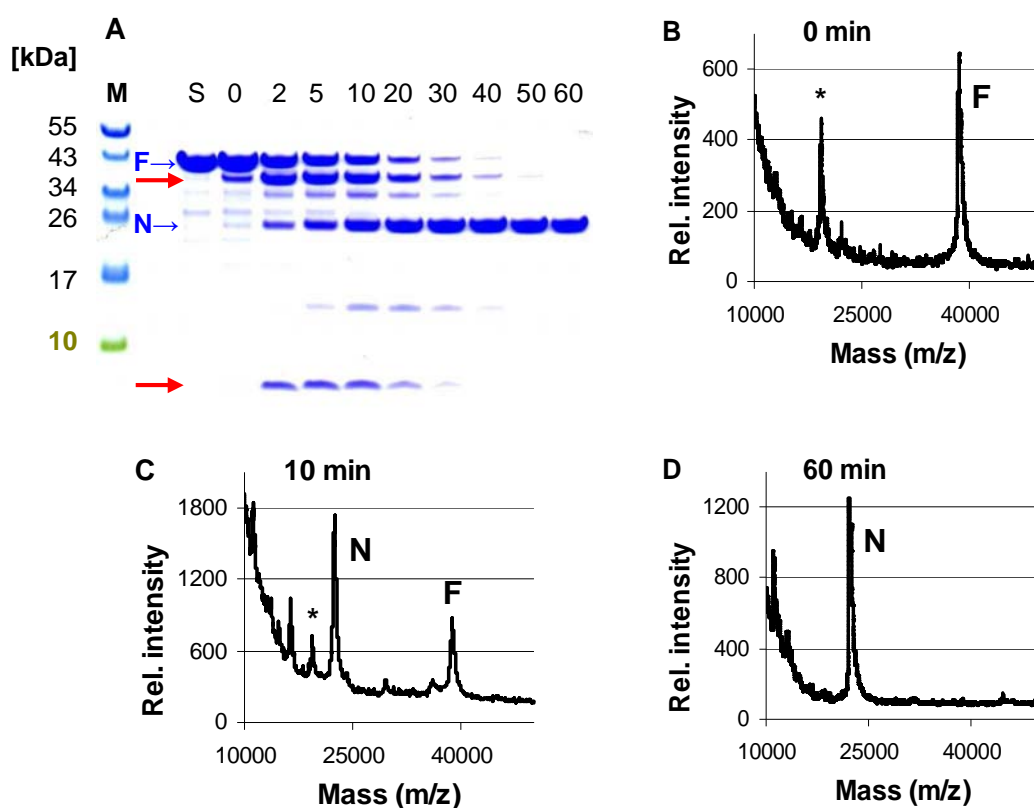
### 4.5.1. Digestion kinetics of KpOmpA in detergent micelles

In order to observe the gradual fragmentation of KpOmpA as a function of time, we optimized the temperature and enzyme/substrate ratio. Since N-KpOmpA does not exhibit cleavages in the loops area, the more suitable protein construct for these experiments is thus F-KpOmpA, which will allow observation of the reduction of the full-length polypeptide chain down to the  $\beta$ -barrel domain (i.e. the disappearance of the C-terminal domain). **Fig. 4.11** shows the digestion kinetics of F-KpOmpA within an hour of time at two different temperatures (37°C and 0°C).



**Fig. 4.11.** Digestion kinetics of F-KpOmpA in 0.1% Zwittergent 3-14 micelles at protein-to-enzyme ratio of 450:1 (w/w), at 37°C ( **A** ) and 0°C ( **B** ). The fastest appearing cleavage products are denoted with red arrows. The intact (full-length) chain is marked as 'F' and the  $\beta$ -barrel as 'N'. The time points (in minutes) are indicated above the respective gel wells. The 'zero time-point' (0 min) indicates that the aliquot was collected immediately after addition of the enzyme. 'S' denotes the protein stock solution before subjecting it to digestion. All aliquots were heat denatured (5 min, 100°C). M - marker ladder.

Clearly the high temperature (**Fig. 4.11A**) promotes high speed of the reaction with complete degradation of the C-terminal domain within 20 min of time. The red arrows indicate the fastest cleavage products - the protein fragments that are produced first upon cleavage on a certain K-site, presumably somewhere close to the C-terminus of the C-terminal domain. The very low molecular weight piece consists of more than one single fragment, since it disappears in the time course. The low temperature (**Fig. 4.11B**), on the other hand, considerably slows down the reaction and further confirms the accumulation of the first cleavage products (red arrows).



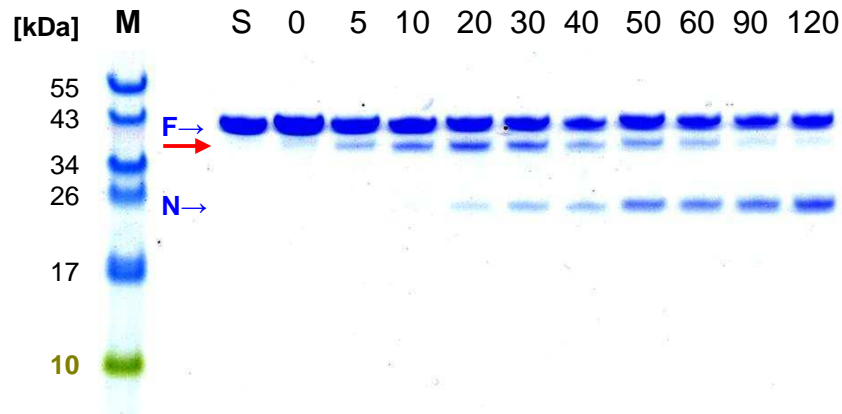
**Fig. 4.12.** Digestion kinetics of F-KpOmpA in 0.1% Zwittergent 3-14 micelles at protein-to-enzyme ratio of 450:1 (w/w) at 16°C. (**A**) SDS PAGE image of the complete transition of the protein from full-length chain to the stable  $\beta$ -barrel. The fastest appearing cleavage products are denoted with red arrows. The intact (full-length) chain is marked as 'F' and the  $\beta$ -barrel as 'N'. The time points (in minutes) are indicated above the respective gel wells. The 'zero time-point' (0 min) indicates that the aliquot was collected immediately after addition of the enzyme. 'S' denotes the protein stock solution before subjecting it to digestion. All aliquots were heat denatured (5 min, 100°C). M - marker ladder. (**B**), (**C**) and (**D**) show MALDI-TOF linear mode spectra (matrix: SPA) at time-points 0, 10 and 60 min, respectively. 'F' denotes the full-length of the protein and 'N' - the  $\beta$ -barrel alone. The asterisk (\*) indicates the 2<sup>+</sup> ion of 'F'.

Exactly the same experiment held at the intermediate temperature of 16°C showed the better resolved, full kinetic window of the F-KpOmpA digestion within an hour (**Fig. 4.12A**). The degradation of the intact full-length polypeptide ('F') down to the stable  $\beta$ -barrel ('N') is tracked by MALDI-TOF for each time-point and representative spectra for 0, 10 and 60 min are shown on **Fig. 4.12 B, C** and **D**, respectively.

The intermediate cleavage products of large sizes (between 'F' and 'N' on **Fig. 4.12A**) were not clearly identified on the mass spectra. Their weak signals, observed on **Fig. 4.12C** (10 min of digestion reaction), prevented the more accurate determination of the sizes of these fragments, although they appear to correspond to the  $\beta$ -barrel with a piece of the C-terminal domain still attached to it. It is speculated that the higher molecular weight intermediate product resembles one of the combinations F1-16 or F2-16, while the lower molecular weight product might be F1-13 or F2-13. The degradation of the entire protein thus follows a pathway starting from its C-terminus and reducing the polypeptide in several discrete steps towards its N-terminus, leaving the stable  $\beta$ -barrel intact at the end. The products identification from these first digestion steps is analyzed further below with the help of the C-KpOmpA construct.

#### 4.5.2. Digestion kinetics of KpOmpA in lipid bilayers

For comparison of the digestion pathways in detergent micelles and in lipid bilayer environment, the same experiment was conducted with F-KpOmpA reconstituted in Polar Lipids Extract (PLE) at LPR of 5 (w/w). The reconstitution did not differ from the previously described procedure. In the case of this sample, the reaction at temperatures of 16°C and even 30°C appeared too slow (data not shown) and was therefore conducted at 37°C. **Fig. 4.13** shows the outcome of this experiment. As before, the two-band pattern at the end of the reaction is observed again, since half of the C-terminal domains are protected in the vesicle lumen. The first cleavage point (red arrows) resembles the one from the micellar sample and in this context refers to the exposed C-terminus of half of the soluble domains. Conducting the same experiment with micellar (at 16°C) and reconstituted (at 37°C) N-KpOmpA did not produce any fragmentation on the SDS PAGE image (data not shown), further confirming the C-terminal origin of the first cleavages. As before, the extracellular loops remained inaccessible to the enzyme due to the presence of K4, K5 and K7 in not highly mobile loop areas (**Fig. 4.1**). Note that the lower molecular weight product found in the micellar sample above (speculated to be F1-13 or F2-13) is not seen in this case.



**Fig. 4.13.** Digestion kinetics of F-KpOmpA in PLE liposomes at protein-to-enzyme ratio of 450:1 (w/w) at 37°C. The fastest appearing cleavage product is denoted with a red arrow (the lower molecular weight fragment is not seen). The intact (full-length) chain is marked as 'F' and the  $\beta$ -barrel as 'N'. The time points (in minutes) are indicated above the respective gel wells. The 'zero time-point' (0 min) indicates that the aliquot was collected immediately after addition of the enzyme. 'S' denotes the protein stock solution before subjecting it to digestion. All aliquots were heat denatured (5 min, 100°C). M - marker ladder.

#### 4.5.3. Digestion kinetics of C-KpOmpA

The first cleavage products originating from the C-terminal domain of KpOmpA were defined by performing Lys-C digestion kinetics experiment on the C-KpOmpA construct (~17.2 kDa, expressed and purified as described in Chapter 1). All theoretical fragments (C1-12) and fragments combinations that can originate from the K-sites of that construct are presented on **Table 4.6**, similarly to the previous tables for N- (**Table 4.1**) and F-KpOmpA (**Table 4.2**).

**Fig. 4.14A** shows the SDS PAGE image from this experiment, exhibiting the first cleavage products of the digestion (red arrows). MALDI-TOF analysis of the two adjacent time-points (2 and 60 min, **Fig. 4.14 B-E**) revealed that the first cleavage occurs at the C-terminus of C-KpOmpA, releasing the fragment combination C10-12 (3383 Da). In the context of F-KpOmpA, that would correspond to F17-19. In the time course of the experiment, C10-12 is gradually degraded down to C10-11 and individual C10 and this is coupled with the fragmentation of the rest of the protein as well, generating such fragments as C1, C2-3, C7 and others (data not shown). It was shown previously that this domain is very vulnerable to digestion by Lys-C, presumably due to the absence of detergent or bilayer protective environment. Here we confirm that the fastest cleavages occur at the C-terminus of the molecule (either F- or C-KpOmpA), starting with the last three C-terminal fragments separated (individually or in combination) from the rest of the polypeptide.

**Table 4.6. C-KpOmpA** - all possible fragment combinations after Lys-C digestion.

Start 1		Start 2		Start 3		Start 4		Start 5		Start 6		Start 7		Start 8	
Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.
C1-2	3398														
C1-3	4463	C2-3	1710												
C1-4	4876	C2-4	2123	C3-4	1497										
C1-5	7135	C2-5	4382	C3-5	3755	C4-5	2690								
C1-6	9818	C2-6	7065	C3-6	6438	C4-6	5373	C5-6	4959						
C1-7	11148	C2-7	8395	C3-7	7769	C4-7	6703	C5-7	6290	C6-7	4031				
C1-8	11672	C2-8	8919	C3-8	8292	C4-8	7227	C5-8	6814	C6-8	4555	C7-8	1872		
C1-9	13804	C2-9	11051	C3-9	10425	C4-9	9359	C5-9	8946	C6-9	6687	C7-9	4005	C8-9	2674
C1-10	16025	C2-10	13272	C3-10	12645	C4-10	11580	C5-10	11167	C6-10	8908	C7-10	6225	C8-10	4895
C1-11	16373	C2-11	13620	C3-11	12994	C4-11	11928	C5-11	11515	C6-11	9256	C7-11	6574	C8-11	5243
C1-12	17169	C2-12	14416	C3-12	13790	C4-12	12724	C5-12	12311	C6-12	10052	C7-12	7369	C8-12	6039
<b>C1 =</b>	<b>2771</b>	<b>C2 =</b>	<b>645</b>	<b>C3 =</b>	<b>1083</b>	<b>C4 =</b>	<b>432</b>	<b>C5 =</b>	<b>2276</b>	<b>C6 =</b>	<b>2701</b>	<b>C7 =</b>	<b>1349</b>	<b>C8 =</b>	<b>542</b>

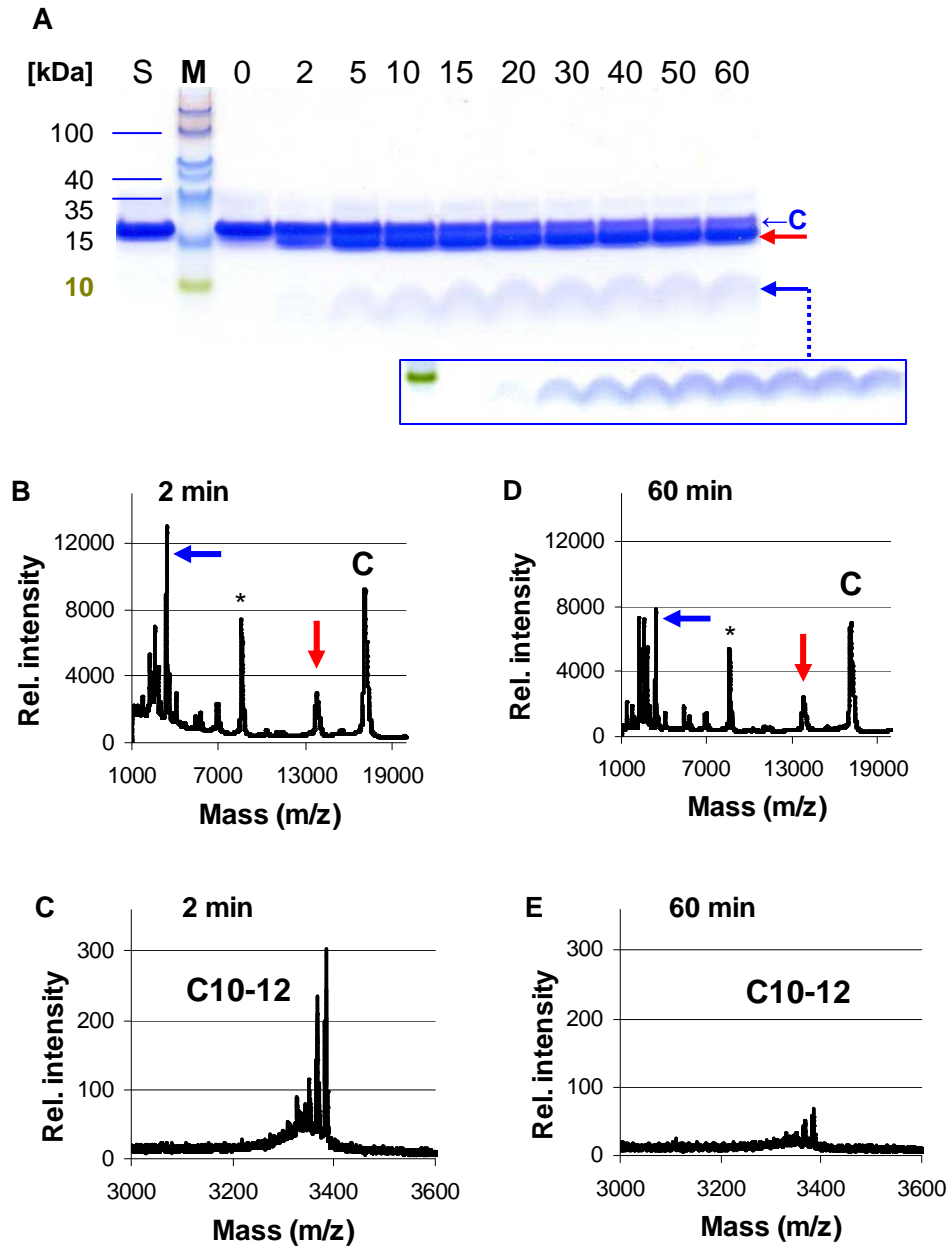
Start 9		Start 10		Start 11		Start 12	
Frag	M.W.	Frag	M.W.	Frag	M.W.	Frag	M.W.
C9-10	4371						
C9-11	4719	C10-11	2587				
C9-12	5515	C10-12	3383	C11-12	1162		
<b>C9 =</b>	<b>2150</b>	<b>C10 =</b>	<b>2239</b>	<b>C11 =</b>	<b>366</b>	<b>C12 =</b>	<b>814</b>

'Frag' denotes the combination of fragments (ex. C1-2, C1-3, etc.) from the chain, starting from (and including) a given fragment (ex. C1).

'Start' denotes the number of the starting fragment (ex. C1, C2, etc.), after which the weights of all combinations are estimated.

'M.W.' denotes the molecular weight (in Daltons) of each combination of fragments.

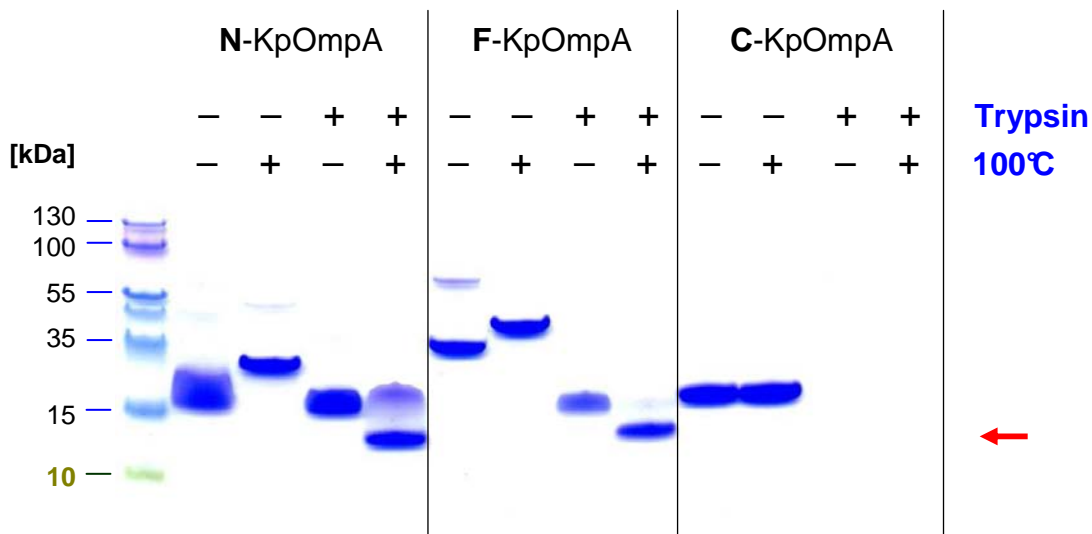
The sizes of the individual fragments (shaded in grey) are shown at the bottom.



**Fig. 4.14.** Digestion kinetics of C-KpOmpA with Lys-C at protein-to-enzyme ratio of 500:1 (w/w) at 16°C. (A) SDS PAGE image of the digestion reaction. The fastest appearing cleavage products are denoted with red and blue arrows (the inset shows contrasted lower molecular weight product for visualization purposes). The intact (full-length) chain is marked as 'C'. The time points (in minutes) are indicated above the respective gel wells. The 'zero time-point' (0 min) indicates that the aliquot was collected immediately after addition of the enzyme. 'S' denotes the protein stock solution before subjecting it to digestion. All aliquots were heat denatured (5 min, 100°C). M - marker ladder. (B-E) MALDI-TOF spectra (matrix: CHCA) at two time-points of the reaction, showing identified protein fragments: (B and D) Linear mode spectra at 2 and 60 min of digestion, respectively. 'C' denotes the intact C-KpOmpA, while the asterisk indicates its 2<sup>+</sup> ion. The vertical red arrow shows the large section of the protein that accumulates over time (compare with A) and it is identified as fragment C1-9 (~13.8 kDa). Its corresponding 'partner', fragment C10-12, is shown by the horizontal blue arrow; (C and E) Reflectron mode spectra of B and D, respectively, showing fragment C10-12 at its expected mass (3383 Da).

## 4.6. Trypsin digestion of KpOmpA

The previously described inability of Lys-C to cleave the extracellular loops of N- and F-KpOmpA regardless of their folding environment, buffer composition and lipid-to-protein ratio most likely emanates from the specific positions of the monitored 'probes', i.e. the Lysine residues on the loops. All the three K4, K5 and K7 are found in loop areas with decreased mobility and therefore restricted accessibility to the enzyme. The Arginine residues on the loops, on the other hand, are present both in the very flexible segments of L1 and L3, as well as in the more rigid areas of L2 and L4 (**Fig. 4.1**). These positions, termed R-sites, appear to be excellent additional targets for probing the loops accessibility (and therefore dynamics) to enzyme cleavage in different environments.



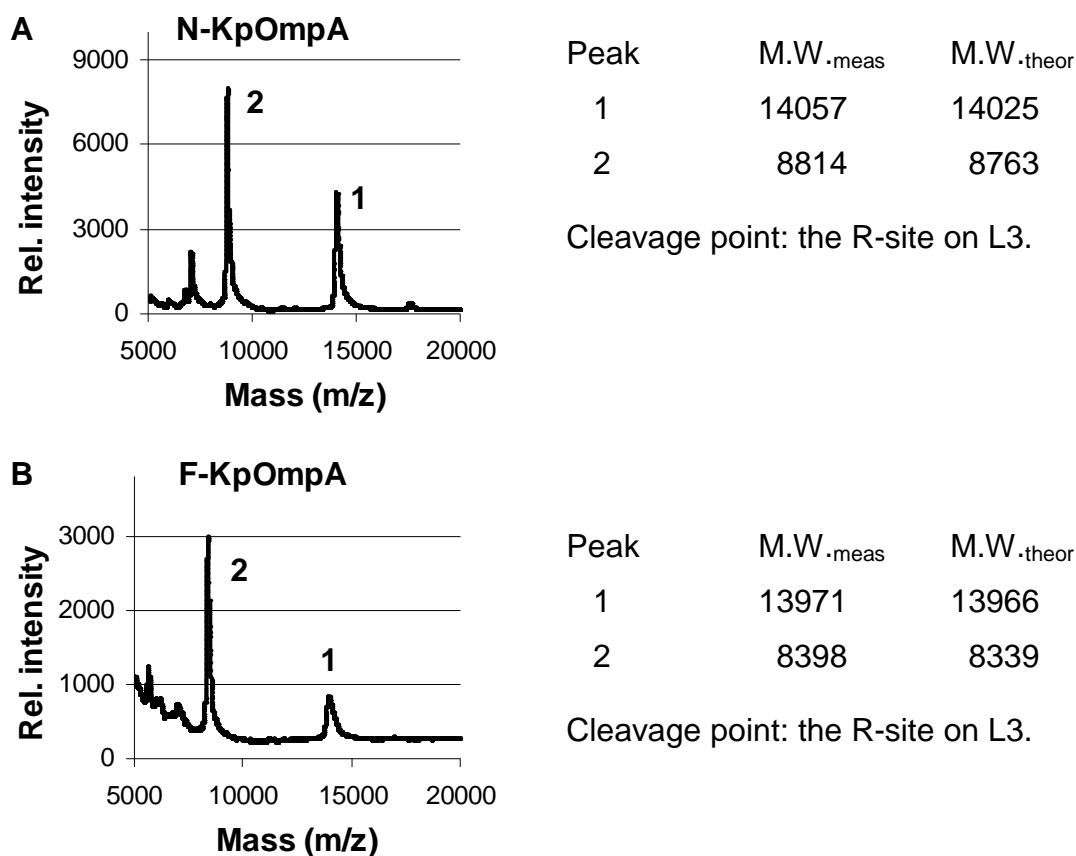
**Fig. 4.15.** Trypsin digestion (2.5 hours at 37°C) of the three KpOmpA protein constructs. The presence or absence of Trypsin and of heat denaturation (100°C) of the aliquots are indicated respectively with '+' and '-' above each well. The red arrow indicates a fragment which is the result of cleavage in the loops area of the  $\beta$ -barrel.

The three protein constructs, N-, F- and C-KpOmpA were Trypsin-treated in 'complete digestion' conditions, as described in Material and methods. **Fig. 4.15** shows the gel image from the three digestion experiments, as well as non-enzyme treated stock solutions of the proteins. Note that the  $\beta$ -barrels of N- and F-KpOmpA are structurally preserved and after the



enzyme digestion they still migrate to their folded positions. However, upon boiling of the respective aliquots, the barrels of N- and F-KpOmpA exhibit interruptions on their chains. The larger fragment from this interruption is observed between 10 and 17 kDa. Although the ‘partnering’ (smaller) fragment is not observed, this is clear indication that at least one Arginine in the loops area of the molecule has been cleaved (assuming that the K-sites are just as inaccessible as before). From its hand, the C-KpOmpA behaves similarly to the C-terminal domain of F-KpOmpA - in conditions of ‘complete digestion’ it disappears at the end of the reaction, leaving no cleavage products of large sizes.

The MALDI-TOF aliquots from the three digestion reactions were used to determine the exact cleavage point (i.e. the R-site) in the extracellular loops of the  $\beta$ -barrel. **Fig. 4.16** shows the linear mode spectra of N- and F-KpOmpA after digestion with Trypsin.



**Fig. 4.16.** MALDI-TOF linear spectra (matrix: CHCA) of N- and F-KpOmpA after Trypsin digestion. (A) Spectrum of digested N-KpOmpA exhibiting two main peaks at 14057 and 8814 Da. (B) Spectrum of digested F-KpOmpA exhibiting two main peaks at 13971 and 8398 Da. The table next to each spectrum shows the theoretical masses ( $M.W._{theor}$ ) of the observed ( $M.W._{meas}$ ) fragments.

In the case of N-KpOmpA (**Fig. 4.16A**), the first peak (1) fits very well to a cleavage product starting after the first Lysine residue on the chain and ending with R134 - the R-site on L3 (14025 Da), while the second peak (2) represents the rest of the chain (8763 Da). It should be noted that the higher molecular weight peak has a value that is quite close to another putative combination, generated after cleavages on the R-sites of both L2 and L4 (14009 Da). This possibility is overruled due to three reasons: (i) The resulting 'partnering' fragments after such cleavage are not observed; (ii) The signal at ~8.8/8.4 kDa can not be explained in this case; (iii) The R-site on L4 is immediately followed by a Proline residue which, as in the case of Lys-C, prevents cleavage even if that ('forbidden') site is exposed to the enzyme.

In the case of F-KpOmpA (**Fig. 4.16B**), the first peak (1) appears to consist of the same piece of the polypeptide (up to the R-site on L3), when taking into account the slight differences in the amino acids compositions at the N-termini of the two protein constructs. The expected value in this case would be 13966 Da. The lower molecular weight peak (2) has a different value compared to N-KpOmpA due to the same reasons - it expands from the same R-site, through the last three  $\beta$ -sheets of the barrel and the Proline/Alanine-rich hinge region that follows them, and terminates to the nearest K-site (**PAPAPAPEVATK**). The expected mass of that fragment is 8339 Da. For both N- and F-KpOmpA, it was shown previously that these K-sites (K1 and K8) are accessible for Lys-C and it is assumed that they are for the Trypsin as well. Ultimately, K8 in F-KpOmpA was constantly cleaved by Lys-C in order to produce the stable barrel (~22.2-22.4 kDa) by separating it from the vulnerable C-terminal domain.

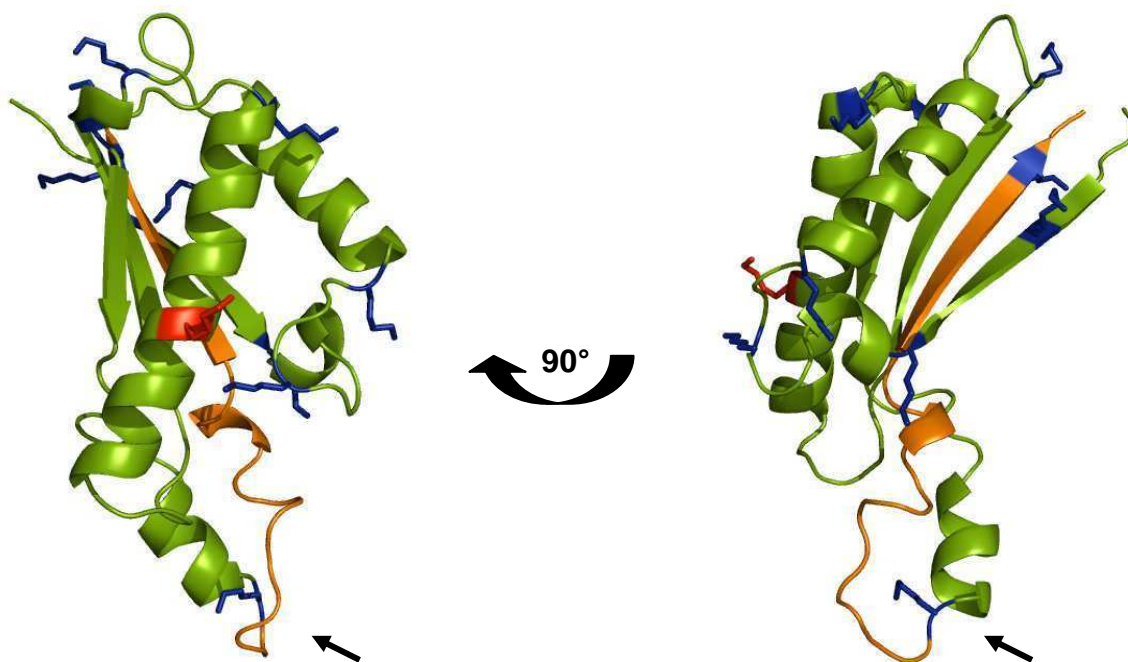
## 4.7. Conclusion

The proteolysis experiments described above successfully combined the precision of the mass spectrometry with the quantitative SDS PAGE in order to identify several protein areas with different propensity for fragmentation, which was correlated with their mobility.

Interestingly, the previously found native modification with oligo-hydroxybutyrate on the 8<sup>th</sup>  $\beta$ -sheet of *E. coli* OmpA ([7]) was never observed in our experiments. The accumulation of the overexpressed polypeptide in cytoplasmic inclusion bodies still allows the modification, as shown in [7]. Such a covalent modification on KpOmpA would not escape detection even with acquisition in the MALDI-TOF linear mode, since its size is in the range of several hundreds

of Daltons (depending on the number of monomeric units attached) which would alter significantly the size of N8/F8. Presumably, the absence of such modification could be explained with the fact that KpOmpA is not a native protein for *E. coli*.

With the help of the C-KpOmpA construct, the fastest appearing cleavage product of the molecule was identified (fragments C10-12 in the case of C-KpOmpA, corresponding to F17-19 in F-KpOmpA). The location of the cleaved K-site is not situated between the two domains of F-KpOmpA, as one might initially expect, but was shown to belong in a putative (i.e. modeled) unstructured loop in the periplasmic domain, outside of its conserved  $\beta\alpha\beta\alpha\beta$ -core motif. This core motif is seen in many other OmpA C-like domains (such as the RmpM protein, [8], PDB: 1R1M). The major variation among these proteins is found in the presence or absence of different additional amino acids inserted between the last two  $\beta$ -sheets, emanating out of the core motif and changing the global shape of the protein. Swiss-modeled (<http://swissmodel.expasy.org/>) C-KpOmpA based on the sequence of RmpM is shown in **Fig. 4.17**.



**Fig. 4.17.** Swiss-modeled (<http://swissmodel.expasy.org/>) C-KpOmpA based on the sequence of RmpM ([8], PDB: 1R1M), showing all K-sites in blue color and fragment C10-12 in orange color. The cleavage point for fragment C10-12 outside the core motif is indicated with a black arrow. The suggested K-site cleaved on a large scale only in the micellar samples (**Fig. 4.12A**) is colored red.

The fastest cleavage point from the enzyme digestion experiments is found in the extension region between the last two  $\beta$ -sheets, which is predicted to be highly unstructured and exposed. This may account for the observed fast fragmentation in that area. The speculated cleavage product F1-13 (or F2-13) from the digestion of F-KpOmpA was seen only in the micellar environment (**Fig. 4.12A**). On the modeled structure, the respective K-site (colored red on **Fig. 4.17**) for this fragmentation is found in the middle of the longest  $\alpha$ -helix. Although the side chain of this residue points outwards from the core of the structure, such a K-site (which belongs to a secondary structure element) is not expected to be readily cleaved. Presumably the detergent introduces a certain level of destabilization on the C-terminal soluble domain of the protein, which leads to the observed larger amount of intermediate products and faster digestion (compare the time points on **Fig. 4.12A** and **Fig. 4.13** and note the difference in the temperature: 16 and 37°C in the first and second cases, respectively).

Lys-C systematically failed to produce a large scale cleavage of the three K-sites on the extracellular loops of N- and F-KpOmpA (K4, K5 and K7). To a certain extent this is expected for the abovementioned K5, which is positioned too close to (i.e. one residue away from) the fourth  $\beta$ -sheet. Furthermore, the crystal structure of *E. coli* OmpA [9] positions this Lysine inside the fourth  $\beta$ -sheet, although the solution state NMR structure does not [10]. Found in this interface region of the molecule (between the barrel core and the most mobile parts of the loops), K5 is perhaps inaccessible to the enzyme. However, K4 is found nearly 'on the top' of L2 (**Fig. 4.1**) and one would expect it to be cleaved, although lower mobility levels were attributed to this loop as a whole (Chapter 3). The K7 site is also found in one of the not highly dynamic regions of L3 and its proximity to the fifth  $\beta$ -sheet is again raising suspicions in regard to its enzyme accessibility. In this way, the NMR-derived information for the loops dynamics is coupled with the MALDI-TOF data and the electrophoresis images of the digested protein aliquots: the reduced mobility levels in particular loop areas were highly resolved (i.e. specifically appointed) by solution state NMR [1] and then confirmed in the solid state (Chapter 2), while the MALDI-TOF relates these mobility levels with the protein susceptibility to enzyme digestion and the SDS PAGE provides an overview of the bulk sample. In conclusion, the mass spectrometry and gel electrophoresis failed to detect a massive, Lys-C-induced cleavage in the loops area of the protein regardless of the used construct and folding environment. From point of view of these two methods (and when the Lys-C was used for sample treatment), the protein loops behaved as one monolithic object together with the  $\beta$ -barrel. The proposed explanation lies in the specific positions of the three K-sites in the loops, which are all found in areas with more or less restricted mobility. The

Arginine residue present in the highly mobile area of L3 (**Fig. 4.1**), on the other hand, was readily cleaved by Trypsin, further relating the enzyme access with the loops mobility.

By performing the Lys-C-based proteolysis in several buffer conditions it was shown that, within the limits of the tested ranges of salt concentration and acidity, the extracellular loops of KpOmpA are still highly inaccessible to enzyme digestion. It is noted that these experiments do not cover broader ranges of the two parameters, as well as other factors that may or may not influence the cleavage pattern. Such factors might include further lowering of the pH (as in the environment of the digestion tract), or observing the influence of moderately high concentrations of chemical compounds like, for instance, urea (as in the urinary tract). Altered cleavage patterns of the K-sites in the loops (associated with altered dynamics in that region of the protein) can then be indirectly associated with OmpA-based invasion of *Escherichia* and *Klebsiella* species in those systems of the host organism and further probed by other means. Such expansion in the field of the KpOmpA enzyme digestion experiments may belong to a future work. The present work aimed in exploring the F-KpOmpA cleavage pattern in more native environments (0-300 mM NaCl and pH of 6-8) and concluded that, in these conditions, the extracellular Lysine residues (K4, K5 and K7, found on L2 and L3) are again not susceptible to a large scale digestion by Lys-C due to insufficient mobility and/or spatial inaccessibility of the enzyme (**Fig. 4.1**).

The experiments in this chapter thus provide a direct biochemical evidence of what had been derived from the NMR data. Let us recall here what these evidences were (see Chapter 3 and **Fig. 4.1**): from heteronuclear NOEs on the N-KpOmpA sample in DHPC micelles we observed that only the C-terminal halves of loops L1 and L3 gave weak NOEs (below 0.5) characteristic of highly mobile loops, while L2, L4 and the N-terminal halves of L1 and L3 gave intermediate NOEs. Solid state NMR on liposome samples with reconstituted N-KpOmpA confirmed the existence of several motional regimes within the loops and highlighted the high mobility of the three Serine residues present in the C-terminal half of L3. Since L2 and L4 are exhibiting limited mobility, their K- and R-sites are not accessible for Lys-C and Trypsin (although it is noted that the R-site on L4 is 'forbidden' for cleavage). Similarly, the K-site on L3 (K7) is found in the more rigid half of that loop and therefore avoids digestion by both enzymes. By contrast, the R-site on L3 was readily cleaved by Trypsin, separating the  $\beta$ -barrel into two pieces. This loop segment appears to be one of the most accessible areas in the extracellular portion of the protein, thus provoking interest in the field of drug development and targeting of bacterial surface proteins. The R-site on L1 was not cleaved by Trypsin. This section of L1, albeit very mobile, contains its R-site fairly close to the second  $\beta$ -

sheet. In addition, this section of L1 is facing the more rigid L2 and therefore might avoid exposure to Trypsin, while the R-site on L3 remains more accessible.

As discussed in Chapter 3, the more rigid protein loops (L2 and L4) appear to be evolutionary better conserved, while the mobile L1 and (the cleavable) L3 are not. Recently, a novel allele of OmpA (termed *ompA2*) was identified in phage-resistant *E. coli* strains [11]. An alignment only on L3 of OmpA from different *E. coli* isolates and *S. flexneri* is presented below, following Fig. 1 in [11], with the L3 of KpOmpA added on the last line:

Organism/strain	OmpA loop L3 region
<i>E. coli</i> OmpA (GenBank no. ECOMP)	D T K S N V Y G - - - - - K N H
<i>E. coli</i> OmpA, isolate H474 (human)	D T K S N V Y G - - - - - K N H
<i>E. coli</i> OmpA, isolate TA024 (Tasmanian devil)	D T K S N V Y G - - - - - K N H
<i>E. coli</i> OmpA, isolate B194 (bird)	D T K A N V P G - <b>G A S</b> F K D H
<i>S. flexneri</i> OmpA (GenBank no. AY305875)	D T K A N V P G - <b>G A S</b> F K D H
<i>K. pneumoniae</i> OmpA (KpOmpA)	D S K G N Y A S T G V S R S E H

The conserved residues above are presented in black boxes. The conserved K-site on L3 of KpOmpA (K7) that avoids digestion by both Lys-C and Trypsin is shaded in grey, while the cleavable (by Trypsin) R-site on L3 is on black background. The first three chains belong to isolates with the 'classic' *ompA1* allele. The last three chains clearly exhibit longer central segment in their L3 loops. These belong to OmpA from the phage-resistant *E. coli* B194 (with *ompA2* allele), to OmpA from the pathogenic *S. flexneri*, and to our OmpA from *K. pneumoniae*. The GAS-motif (bold letters) was proposed as one of the (among potentially many) reasons for phage resistance and it seems to be quite similar to the aligned KpOmpA sequence in that area (GVS). Considering that such a function would involve exposure of that segment to the extracellular environment, this further explains the Trypsin cleavage pattern observed in this work. Interestingly, the phage-receptor area on L3 of *E. coli* OmpA was identified previously ([12, 13]) and was found to be in very close proximity, N-terminally to the GAS/GVS motif (-SNVY-). Furthermore, loops L1, L2 and L3 were found to be necessary for recognition by the OmpA-specific phage K3, while all protein loops were needed for the process of F-conjugation to occur [14]. These data establish the functional value of the OmpA extracellular section in general, or dissected to the level of specific loops regions. The present work attempted to correlate the enzyme accessibility to the dynamics of certain loop areas of KpOmpA, showing that the less conserved, very mobile C-terminal half of L3 is exposed and

readily capable of interacting with other molecules. The fact that the R-site on L3 is the sole cleavage point in the extracellular section of KpOmpA (out of 7 potential targets for Trypsin in that area, **Fig. 4.1**) is intriguing. This represents an important feature of the otherwise unstructured protein loops, which apparently tend to protect each other from digestion, i.e. behave as if 'partially structured'. Since the loops of OmpA were shown to be not essential for the  $\beta$ -barrel formation in the hydrophobic protein core ([14]), the partial organization of the loops section presumably does not contribute strongly to the overall stability of the molecule, but serves a functional role in the variety of intermolecular recognitions with the participation of OmpA.

## References

1. Renault, M., et al., *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, 2009. 385(1): p. 117-30.
2. Fontana, A., et al., *Correlation between Sites of Limited Proteolysis and Segmental Mobility in Thermolysin*. Biochemistry, 1986. 25(8): p. 1847-1851.
3. Fontana, A., et al., *Probing protein structure by limited proteolysis*. Acta Biochimica Polonica, 2004. 51(2): p. 299-321.
4. Sugawara, M., *Etudes structurales relatives au developpement de nouveaux vaccins, in NMR and proteins-membranes interactions, IPBS-CNRS*. 2003, Paul Sabatier University: Toulouse. p. 174.
5. Raijmakers, R., et al., *Cleavage specificities of the brother and sister proteases Lys-C and Lys-N*. Chem Commun (Camb), 2010. 46(46): p. 8827-9.
6. Nakamura, K. and S. Mizushima, *Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12*. J Biochem, 1976. 80(6): p. 1411-22.
7. Xian, M., et al., *Sorting signal of Escherichia coli OmpA is modified by oligo-(R)-3-hydroxybutyrate*. Biochim Biophys Acta, 2007. 1768(11): p. 2660-6.
8. Grizot, S. and S.K. Buchanan, *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. Molecular Microbiology, 2004. 51(4): p. 1027-1037.
9. Pautsch, A. and G.E. Schulz, *Structure of the outer membrane protein A transmembrane domain*. Nature Structural Biology, 1998. 5(11): p. 1013-1017.
10. Arora, A., et al., *Structure of outer membrane protein A transmembrane domain by NMR spectroscopy*. Nature Structural Biology, 2001. 8(4): p. 334-338.
11. Power, M.L., et al., *A naturally occurring novel allele of Escherichia coli outer membrane protein A reduces sensitivity to bacteriophage*. Appl Environ Microbiol, 2006. 72(12): p. 7930-2.
12. Morona, R., M. Klose, and U. Henning, *Escherichia-Coli K-12 Outer-Membrane Protein (Ompa) as a Bacteriophage Receptor - Analysis of Mutant-Genes Expressing Altered Proteins*. Journal of Bacteriology, 1984. 159(2): p. 570-578.
13. Morona, R., C. Kramer, and U. Henning, *Bacteriophage Receptor Area of Outer-Membrane Protein Ompa of Escherichia-Coli K-12*. Journal of Bacteriology, 1985. 164(2): p. 539-543.
14. Koebnik, R., K.P. Locher, and P. Van Gelder, *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*. Molecular Microbiology, 2000. 37(2): p. 239-253.



## Conclusions and perspectives

The results presented in this work point towards several aspects of the structure and dynamic behavior of KpOmpA, which are related to its functions.

As rigid, stable bacterial  $\beta$ -barrel-containing proteins from the outer membrane, the N- (~23 kDa) and F-KpOmpA (~39 kDa) constructs are readily expressed and accumulated in inclusion bodies in *E. coli*, easily purified in substantial amounts and well refolded in detergent micelles. These features of OmpA previously made it a suitable target for a range of studies, varying from investigating its folding kinetics, its structure and its immunological roles. Apart from the detergent solubilized studies, an important point is to examine a membrane protein in a more native environment, i.e. in lipid bilayers. Although the protein folding in detergent micelles is readily achieved and its spontaneous insertion into preformed vesicles is feasible, the incorporation of large protein content in lipid bilayers is still a challenge. We have explored the dialysis-driven approach in this regard and have established appropriate conditions for reconstitution of substantial protein amounts into the newly formed vesicles. The high protein content, the large vesicle size and the absence of salt in the final sample turned it into a suitable object for solid-state NMR experiments, through which the protein folding in the bilayers was confirmed and the dynamics of several protein regions was examined.

In addition, the C-terminal periplasmic domain (~17 kDa) of KpOmpA was expressed as a separate product and its proper folding was confirmed by solution state NMR. A nice, well resolved and stable over time, 2D HSQC spectrum could be obtained. This result opens the path for a future 3D structure determination by solution state NMR. It will be possible to study this C-terminal domain alone in solution as well as in interaction with bilayers, bilayer mimetics and with the N-terminal domain. Since the full length protein is also available, it will be possible to compare the NMR spectra of the separated N- and C- terminal domains with the spectrum of the entire protein and thus visualize possible changes in the structures.

Because it represents one of the smallest  $\beta$ -barrels with only 8  $\beta$ -strands and tightly packed side chains in its lumen, KpOmpA was initially expected to behave as one rigid, monolithic entity in the membrane. However, unfolding experiments with SMFS have revealed that, like some other, larger  $\beta$ -barrel proteins, KpOmpA is also extracted from the membrane stepwise and not in a single rupture event. Furthermore, the N-terminal  $\beta$ -strand exhibited unusually strong resistance to extraction. These findings suggested that KpOmpA (if not other OmpA-proteins as well) may perform a force-relieving function *in vivo*, reducing

the mechanical stress on the bacteria in sequential manner with several intermediate steps. Supporting this hypothesis, we observed that upon a second extension of the AFM-tip the protein was able to spontaneously refold back into the membrane, after which the same (or similar) unfolding pattern was observed during the second retraction of the tip. Interestingly, the N-terminal (strongly anchored)  $\beta$ -strand appears to be the least preserved  $\beta$ -sheet among the OmpA family, as shown by multiple alignment of 1000 sequences with Phyre2 ([1], data not shown). It is still unclear if this is related with the 'uniquely strong anchoring' capacity of this  $\beta$ -sheet. Such an investigation may belong to a future work.

In an attempt to correlate similarities and differences between the homologues in the OmpA family (and in the context of their functions), we investigated the dynamic behavior of KpOmpA by solid-state NMR and limited proteolysis. We showed that the mobility of the extracellular loops region is still preserved when the protein is reconstituted in lipid bilayers, similarly to the DHPC-solubilized sample ([2]), and that the C-terminal halves of L1 and L3 are the most mobile areas on the extracellular side of KpOmpA. Recent studies with EPR on EcOmpA also revealed that different environments may influence the experimental outcome, but this is not necessarily related with altered protein dynamics ([3]). Hence the gradient of mobility observed from the barrel core to the loops of the protein is an intrinsic property of the molecule and not an artifact arising from the micellar environment. In the case of KpOmpA, this was extensively confirmed by our limited proteolysis experiments. The Lysine residues found in the more rigid loop segments systematically avoided digestion by Lys-C and Trypsin, while only Trypsin managed to cleave an Arginine on one of the most flexible and exposed loops area - the C-terminal half of L3. In EcOmpA, three residues on L3 were shown to be critical for the bacterial survival inside dendritic cells and macrophages ([4]). Interestingly, L3 is considerably longer in KpOmpA and the 'extension' is found in the middle of the EcOmpA probed sequence in [4]. This opens the perspective for comparing the two cases by investigating *K. pneumoniae* invasion and survival in various immune cells. More appropriately, an OmpA-deficient *E. coli* strain transfected with KpOmpA (with its signal peptide) may be used for this purpose. Such an experiment would highlight the effect of the different loops sequences in the two proteins, found mostly on L1 and L3, and may suggest a structural basis for the different receptor recognition of EcOmpA (gp96 on neutrophils and Fc- $\gamma$  in macrophages, [4]) and KpOmpA (LOX-1 and SREC-I on macrophages and dendritic cells, [5]). The variable L1, from its hand, was shown as well to be very important for the bacterial survival. Mutations in the area of EcOmpA-L1 (which correlates with the most mobile segment of our KpOmpA-L1) reduced nearly completely the survival of *E. coli* in

macrophages. A peptide resembling that same sequence of L1 was able to block the *E. coli* invasion in brain microvascular endothelial cells - the first step in the development of meningitis ([6]). This highlights the immunological importance of the mobile loop areas in EcOmpA. In the case of KpOmpA, however, the Arginine site in the mobile C-terminal half of L1 was not cleaved by Trypsin. In a similar fashion, the two Lysine sites on L2 were also not cleaved. This was somehow expected for K5 (too close to the fourth  $\beta$ -strand), but not for K4 (which was cleaved in minor amounts, observed only with MALDI-TOF). Although we observed it as a less flexible segment in both NMR and proteolysis experiments, the N-terminal half of L2 was found to be crucial for the survival of *E. coli* in dendritic cells and polymorphonuclear leukocytes ([4]). Thus it appears that the *mobility* and the *exposure* of the different extracellular portions of KpOmpA (and perhaps EcOmpA as well) are not strictly related (i.e. overlapping) features of the protein. In this context, an interesting perspective for these studies would be further examination of the relation between dynamics and accessibility in the different loops regions, for instance by introducing single cleavable K- or R-sites in different loops positions. Of note, we point out that the miss-cleavage at the R-site on L1 may also be due to the proximity of this residue to the second  $\beta$ -sheet, which might obstruct the cleavage capability of the enzyme and not its access. Perhaps for similar reasons we did not observe cleavages on K2, found in the unstructured N-terminus of the molecule but positioned only 2 residues before the first  $\beta$ -sheet.

The functional role of the protein loops is, of course, not limited in their most flexible regions. For instance, alterations in the beginning of L1 and the end of L4 (which are adjacent in the 3D space and correspond to more rigid loop segments in KpOmpA) partially reduced the *E. coli* survival in macrophages ([4]). Those loop areas are identical in EcOmpA and KpOmpA and are presumably related to the structuring of the underlying  $\beta$ -barrel, as suggested by their limited mobility. This suggests that preserving the native structural integrity of OmpA may also play a role in the various immunological processes, apart from the exposed and/or mobile loop segments. Alternatively, the closer-to-the barrel loops segments play a role of 'antigen-presenting' support units, responsible for proper positioning of the outmost loop sections. Some of these very flexible areas are close to the barrel and therefore inaccessible, while some others are perfectly exposed (the R-site on L3). In the same way, the more rigid but exposed parts of the loops also participate in recognition processes, like in the cases of the 'top' sides of L2 and L4 (conserved in EcOmpA and KpOmpA). Therefore we suggest a relation between the evolutionary preservation and the mobility of the different loop segments of OmpA, for the moment extended only between *E. coli* and *K. pneumoniae*.

Searching among 1000 OmpA-like sequences (as mentioned above) certainly can not show such a correlation clearly, but even in that case L1 and L3 appeared, as overall, more variable than L2 and L4. Assuming that similar loops (in terms of sequence and dynamics) might serve similar purposes in different organisms (which can attack the same systems in the host), then their specific immunological features may be searched in the other, less preserved areas, which may explain the interactions with specific receptors and, eventually, allow specific targeting of a given pathogen. Such differences account not only at the species-, but at the strain-level as well, as was shown for two alleles of EcOmpA ([7]). In this case, a modification in L3 (similar to the 'extension' in the mobile half of KpOmpA-L3) is probably responsible for the phage resistance attributed to the novel *ompA2* allele, which correlated well with findings nearly 20 years before ([8]). The 'extension' in KpOmpA-L1, on the other hand, is found partially in the less mobile part of the loop. The detailed study with different EcOmpA mutants mentioned previously ([4]) did not investigate that loop section. Hence it would be interesting to delete this sequence in both EcOmpA and KpOmpA and monitor the bacterial survival (or phage resistance), as well as to check its accessibility by mutation-introduced cleavable site. Of note, the same mutants can serve the purpose for exploring the adhesive capability of the bacteria against various surfaces, i.e. the formation of biofilms.

In more practical terms (and achievable in more reasonable time), the future perspectives for investigation of KpOmpA are proposed to follow one of the possibilities:

- Aiming at the ultimate step towards membrane protein analysis in native-like environment, whole bacteria cells (expressing KpOmpA on their surface) can be investigated with 'in-cell NMR' ([9, 10]). Alternatively, isolated outer membranes can serve that purpose. A choice of a suitable OMP-deficient *E. coli* strain for the protein expression is important. The major benefit of such a system would be the orientation of the protein in the membrane, so that the extracellular loops will always face the lipopolysaccharides in the highly asymmetric outer membrane. The dynamic behavior and the enzyme accessibility of the loops in this case may vary from the *in vitro* samples described here, which may reveal more precisely the loops behavior during attack on a host cell, interaction with a phage, etc.

- The structure elucidation of the C-terminal, periplasmic domain of KpOmpA should now be feasible. Despite the abundant information available for the  $\beta$ -barrel of OmpA, little is known about this domain of the protein. In addition to its structure, a 'functional assay' for this

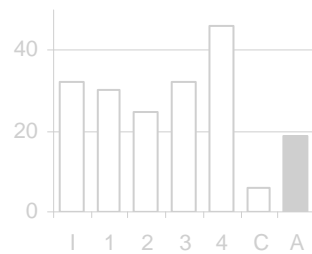
protein construct would be monitoring its interaction with peptidoglycan. The peptidoglycan can be isolated as large (polymeric) pieces or broken down to smaller units ([11]). If large (insoluble) glycan isolates are used, the bound protein can be pulled-down with them. Small (soluble) protein-bound glycan chains may be visualized by antibodies after extensive washing. The latter kind of samples can be also probed by solution state NMR for mapping the peptidoglycan-interaction site on the protein, which then can be compared with the simulated experiment with the homologous domain of RmpM from *N. meningitidis* ([12]) or with the detailed data on OmpA from *A. baumannii* ([13]) - two works with substantially different outcome in the context of protein-peptidoglycan binding. Furthermore, the proposed dimerization of the C-terminal domains of RmpM ([12]) and of MotB ([14]) can be explored in and compared with KpOmpA as well. Such a finding would be an interesting example of protein oligomerization at the level of the periplasmic domain, while the membrane domain remains monomeric.

- Another direction for future work would be to study the interaction of KpOmpA with its molecular partners, i.e. with host cell receptors or secreted products of the immune system. Several candidates for KpOmpA have been identified so far, such as the long pentraxin PTX3 or the scavenger receptors LOX-1 and SREC-I on macrophages and dendritic cells ([5]). EcOmpA, on the other hand, was shown to bind to Ecgp (a gp96-like receptor on brain microvascular endothelial cells, [15]) and (presumably) to Fc- $\gamma$  on macrophages ([16]). Interestingly, Serum Amyloid A (produced by the liver during inflammation) was shown to bind with high affinity to the OmpA proteins from several Gram-negative bacteria, including *E. coli* and *K. pneumoniae* ([17]). KpOmpA-receptor interactions like those mentioned above could be studied by SMFS, in order to see if the profile of denaturation - renaturation is affected by the interaction. It could be examined also in terms of its impact on the loops mobility by NMR or by proteolysis-MS. Thanks to the advances in this thesis, the latter approach would now be extremely efficient, particularly if combined with the reintroduction of specific Lysine mutation in order to probe specific positions in the loops.

## References

1. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. Nat Protoc, 2009. 4(3): p. 363-71.
2. Renault, M., et al., *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, 2009. 385(1): p. 117-30.
3. Jimenez, R.H.F., D.M. Freed, and D.S. Cafiso, *Lipid and Membrane Mimetic Environments Modulate Spin Label Side Chain Configuration in the Outer Membrane Protein A*. Journal of Physical Chemistry B, 2011. 115(49): p. 14822-14830.
4. Prasadarao, N.V., et al., *Deciphering the Roles of Outer Membrane Protein A Extracellular Loops in the Pathogenesis of Escherichia coli K1 Meningitis*. Journal of Biological Chemistry, 2011. 286(3): p. 2183-2193.
5. Jeannin, P., et al., *Complexity and complementarity of outer membrane protein a recognition by cellular and humoral innate immunity receptors*. Immunity, 2005. 22(5): p. 551-560.
6. Prasadarao, N.V., C.A. Wass, and K.S. Kim, *Endothelial cell GlcNAc beta 1-4GlcNAc epitopes for outer membrane protein A enhance traversal of Escherichia coli across the blood-brain barrier*. Infect Immun, 1996. 64(1): p. 154-60.
7. Power, M.L., et al., *A naturally occurring novel allele of Escherichia coli outer membrane protein A reduces sensitivity to bacteriophage*. Appl Environ Microbiol, 2006. 72(12): p. 7930-2.
8. Morona, R., C. Kramer, and U. Henning, *Bacteriophage Receptor Area of Outer-Membrane Protein Ompa of Escherichia-Coli K-12*. Journal of Bacteriology, 1985. 164(2): p. 539-543.
9. Renault, M., et al., *Cellular solid-state nuclear magnetic resonance spectroscopy*. Proc Natl Acad Sci U S A, 2011.
10. Renault, M., et al., *Solid-State NMR Spectroscopy on Cellular Preparations Enhanced by Dynamic Nuclear Polarization*. Angew Chem Int Ed Engl, 2012, 51: p. 2998-3001.
11. Christensson, B., et al., *Methodological Aspects of Staphylococcus-Aureus Peptidoglycan Serology - Comparisons between Solid-Phase Radioimmunoassay and Enzyme-Linked Immunosorbent-Assay*. Journal of Clinical Microbiology, 1984. 19(5): p. 680-686.
12. Grizot, S. and S.K. Buchanan, *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. Molecular Microbiology, 2004. 51(4): p. 1027-1037.
13. Park, J.S., et al., *Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane*. Faseb J, 2012. 26(1): p. 219-28.
14. Imada, K., et al., *Stator assembly and activation mechanism of the flagellar motor by the periplasmic region of MotB*. Molecular Microbiology, 2009. 73(4): p. 710-718.
15. Prasadarao, N.V., *Identification of Escherichia coli outer membrane protein A receptor on human brain microvascular endothelial cells*. Infection and Immunity, 2002. 70(8): p. 4556-4563.
16. Sukumaran, S.K., H. Shimada, and N.V. Prasadarao, *Entry and intracellular replication of Escherichia coli K1 in macrophages require expression of outer membrane protein A*. Infection and Immunity, 2003. 71(10): p. 5951-5961.
17. Hari-Dass, R., et al., *Serum amyloid A protein binds to outer membrane protein A of Gram-negative bacteria*. Journal of Biological Chemistry, 2005. 280(19): p. 18562-18567.

# ANNEXES



## **Summary**

The ANNEXES section represents technical information related to Chapter 1. The four sections of ANNEXES, which correspond respectively to the four sections of Chapter 1, provide step-by-step protocols for the main procedures used. By having these protocols “all in one place”, the future KpOmpA investigator is believed to be facilitated in finding and applying the different procedures, and reproducing the results from the techniques used for purification and reconstitution of the different protein constructs.



# Annex 1

## Construction and validation of the pET26bII-F-KpOmpA plasmid

### I. Cloning and purification of the *kpOmpA* gene

1. Genomic DNA of *K.p.* strain 52145 is available (~50 ng/μl). The used primers are:  
kpmpa5' GTGGCACTGCCATATGAAAGCTACCGTAGCGCAGGCCGCTCCGAAAGATAAC  
kpmpa3' GGTACCAGCTCTCGAGAGCCGCCGCTGAGTTACAACTTCT  
The NdeI (CATATG) and XhoI (CTCGAG) restriction sites are underlined.  
The concentration of each primer is 100 μM.
2. Mix 35.5 μl Milli-Q water, 2 μl genomic DNA, 2x 0.5 μl from each primer, 1 μl dNTPs (10 mM mix), 10 μl 5x Phusion HF Buffer and 1 μl Phusion Hot Start II DNA Polymerase.
3. Subject the 50-μl reaction mixture to PCR with the following scheme:
  - 30 sec of initial denaturation at 98°C.
  - 10 sec of denaturation at 98°C.
  - 60 sec of annealing/extension at 72°C.
  - 5 min of final extension at 72°C. → Leave at 12°C until collected.
4. Purify the amplified gene product with QIAquick PCR Purification kit. Briefly:
  - Add 5 volumes of Buffer PBI to 1 volume of the reaction.
  - Apply the sample to the provided QIAquick column, placed in 2-ml tube and centrifuge 1 min (all centrifugation are performed at ~18 000 g (~13 000 rpm of tabletop centrifuge)).
  - Discard the flowthrough, add 750 μl Buffer PE (add ethanol) and centrifuge for 1 min.
  - Discard the flowthrough and centrifuge again for 1 min.
  - Place the column into a clean 1.5-ml tube and elute the PCR product with 25 μl Buffer EB and centrifugation for 1 min. Add 25 μl Milli-Q water to the column and centrifuge again, adding this volume to the EB-eluate.
  - Measure the concentration (UV) and check the purity (agarose gel) of the amplicon.

### II. Purification of the empty plasmid (pET26bII)

1. Grow a stock of *E. coli* Top10 transfected with pET26bII in 50 ml LB medium (kanamycin).
2. Inoculate a small aliquot of the culture into fresh 50 ml medium and grow again.
3. Collect the bacteria (5000 g, 10 min, 4°C) and put the pellet on ice.
4. Purify the plasmid with NucleoBond Plasmid Purification kit (MN). Briefly:
  - Gently resuspend the bacteria in 8 ml Buffer S1.
  - Add 8 ml Buffer S2 to the suspension and mix gently (no vortexing).
  - Leave for 3 min at room temperature (RT).
  - Add 8 ml of cold (4°C) Buffer S3 and mix quickly, but gently.
  - Leave on ice for 5 min.
  - Equilibrate (by gravity flow) a Nucleobond AX 100 column with 2.5 ml of Buffer N2.
  - Discard the flowthrough.
  - Centrifuge the cell lysate (13 000 g, 25 min, 4°C). Discard the pellet.
  - Load the cleared cell lysate onto the column. Leave for gravity flow.

- Wash the column with 12 ml of Buffer N3. Discard the flowthrough.
- Elute the plasmid with 5 ml of Buffer N5.
- Add 3.5 ml isopropanol for plasmid DNA precipitation.
- Centrifuge (20 000 g, 30 min, 4°C) and discard the supernatant.
- Add 2 ml of 70% ethanol to the pellet and vortex shortly.
- Centrifuge (20 000 g, 10 min, RT) and discard the supernatant. Dry for 10 min at RT.
- Dissolve the plasmid DNA in few tens of µl of Milli-Q water.
- Measure the concentration (UV) and check the purity (agarose gel) of the plasmid.

### III. Restriction and ligation of pET26bII-F-KpOmpA

1. Digest the pET26bII vector by mixing:
  - 10 µl purified plasmid (here at ~280 ng/µl).
  - 10 µl Milli-Q water.
  - 2.5 µl of 10x Buffer 4 (New England Biolabs).
  - 1 µl NdeI stock and 1 µl XhoI stock solutions (New England Biolabs).
  - Leave for 30 min at 37°C.
  - Add 1 more µl from each enzyme stock and leave for another 30 min at 37°C.
  - Inactivate the reaction for 10 min at 65°C. Visualize an aliquot on agarose gel.
2. Digest the *kpOmpA* amplicon by mixing:
  - 10 µl purified PCR product (here at ~70 ng/µl)
  - 10 µl Milli-Q water.
  - 2.5 µl of 10x Buffer 4.
  - 1 µl NdeI stock and 1 µl XhoI stock solutions.
  - Leave for 30 min at 37°C.
  - Add 1 more µl from each enzyme stock and leave for another 30 min at 37°C.
  - Inactivate the enzymes for 10 min at 65°C. Visualize an aliquot on agarose gel.
3. Ligate the two digested products by mixing:
  - 10 µl of the *kpOmpA* digestion reaction.
  - 5 µl of the pET26bII digestion reaction.
  - 2 µl Milli-Q water.
  - 2 µl of 10x T4 buffer (New England Biolabs).
  - 1 µl T4 DNA ligase stock solution (New England Biolabs).
  - Repeat this step with 10 µl water instead of the *kpOmpA* gene, for 'blank sample'.
  - Leave for 30 min at 16°C. Inactivate the enzyme for 10 min at 65°C.
  - Precipitate the two ligation reactions with ethanol and dry them.

### IV. Transformation of competent *E. coli* Top10 strain

1. Cover the precipitated ligation reaction with ~50 µl competent *E. coli* Top10 suspension.
2. Transfect the cells by electroporation (1 pulse of 12.5 kV/cm = 2.5 kV in a 2 mm cuvette).
3. Dilute the cells with 900 µl LB medium and leave for 1h at 37°C.
4. Spread the cells on LB/agar petri dishes supplemented with 30 µg/ml kanamycin.
5. Repeat the above steps for the 'blank sample' as well and plant on other petri dishes.
6. Pick up single colonies from the pET26bII-F-KpOmpA dishes and from the 'blank' dishes.

7. Screen the selected colonies by PCR by using (for instance) Dream Taq Polymerase and small amounts of the *kpompa5'* primer and a commercial 3' primer for the kanamycin-resistance gene. Use buffers and PCR conditions suitable for the used polymerase.
8. Check the PCR-screening results on an agarose gel.
9. Choose from the 'positive hits' and grow that same colony in 5 ml LB (with kanamycin).

## V. Isolation of the ligated pET26bII-F-KpOmpA

1. Collect the bacteria (5000 g, 10 min, 4°C). Discard the supernatant.
2. Purify the pET26bII-F-KpOmpA vector using QIAprep Spin Miniprep kit (QIAGEN). Briefly:
  - Resuspend the bacterial pellet in 250 µl of Buffer P1 and transfer it to a centrifuge tube.
  - Add 250 µl of Buffer P2 and mix gently for up to 5 min.
  - Add 350 µl of Buffer N3 and mix quickly, but gently.
  - Centrifuge (~18 000 g, 10 min, RT). Apply the supernatant to a QIAprep spin column.
  - Centrifuge (~18 000 g, 1 min, RT) and discard the flowthrough.
  - Add 750 µl of Buffer PE to the column and centrifuge (~18 000 g, 1 min, RT).
  - Discard the flowthrough and centrifuge for 1 more minute.
  - Place the column into a clean 1.5-ml tube.
  - Elute the plasmid by applying 25 µl of Buffer EB to the column, waiting for 1 min and centrifuging for 1 min (~18 000 g, RT).
  - Add 25 µl water to the column, wait for 1 min and centrifuge again for 1 min.
  - Measure the concentration (UV) of the purified KpOmpA-containing plasmid.

## VI. Restriction map of the isolated pET26bII-F-KpOmpA

1. Visualize the integrated *kpOmpA* gene in the vector by restricting it. Mix:
  - 3 µl Milli-Q water.
  - 1 µl 10x Buffer 4.
  - 5 µl purified plasmid (here ~130 ng/µl).
  - 2x 0.5 µl from each NdeI and XhoI enzyme stocks.
  - Leave for 1h at 37°C. Visualize the restricted *kpOmpA* gene on an agarose gel.

## VII. Transfection of pET26bII-F-KpOmpA in an inducible *E. coli* strain

1. Use, for instance, competent *E. coli* BL21 (DE3). Perform this as with the *E. coli* Top10 strain above (electroporation with the purified pET26bII-F-KpOmpA plasmid, resuspension in ~20 volumes of kanamycin-free LB medium for ~1h at 37°C and visualization of positive transformants after growth on LB/agar petri dishes with 30 µg/ml kanamycin).
2. Select colonies and grow them in LB (with kanamycin).
3. Confirm overexpression of the protein after 1mM IPTG-induction in small-scale cultures.
4. Freeze non-induced cell aliquots for future work (in LB with 10% glycerol, sterile filtered).

## VIII. Media and buffers compositions

1. Luria-Bertani (LB) medium	Per 0.5L	Per 1L
Bacto Peptone	5.0 g	10 g
Bacto Yeast Extract	2.5 g	5 g
NaCl (M.W. 58.44)	5.0 g	10 g

Adjust to pH 7. Autoclave for sterilization.

/For 1.5% agar petri dishes, add 1.5 g AGAR per 100 ml LB medium./

### 2. Commercial buffers used

- *FYNNZYMES / Thermo Scientific*

5x Phusion HF Buffer (contains 7.5 mM MgCl<sub>2</sub>, *the exact composition is confidential*).

- *QIAGEN*

Buffer PBI (contains Gdn/HCl and isopropanol, *the exact composition is confidential*).

Buffer PE (add ethanol before use, *the exact composition is confidential*).

Buffer EB (10 mM Tris, pH 8.5).

Buffer P1 (50 mM Tris (pH 8), 10 mM EDTA, add 100 µg/ml RNase A before use).

Buffer P2 (0.2M NaOH, 1% SDS).

Buffer N3 (contains Gdn/HCl and acetic acid, *the exact composition is confidential*).

- *MACHEREY-NAGEL*

Buffer S1 (50 mM Tris (pH 8), 10 mM EDTA, add 100 µg/ml RNase A before use).

Buffer S2 (0.2M NaOH, 1% SDS).

Buffer S3 (2.8M K-acetate, pH 5.2).

Buffer N2 (100 mM Tris (pH 6.3 ←H<sub>3</sub>PO<sub>4</sub>), 15% ethanol, 0.9M KCl, 0.15% triton X-100).

Buffer N3 (100 mM Tris (pH 6.3 ←H<sub>3</sub>PO<sub>4</sub>), 15% ethanol, 1.15M KCl).

Buffer N5 (100 mM Tris (pH 8.5 ←H<sub>3</sub>PO<sub>4</sub>), 15% ethanol, 1M KCl).

- *NEW ENGLAND BIOLABS*

10x Buffer 4 (0.5M K-acetate, 0.2M Tris-acetate (pH 7.9), 0.1M Mg-acetate, 10 mM DTT).

10x T4 buffer (0.5M Tris (pH 7.5), 0.1M MgCl<sub>2</sub>, 0.1M DTT, 10 mM ATP).

### 3. Agarose gel electrophoresis

1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) supplemented with 1% agarose.

6x Loading dye (0.25% Bromophenol blue, 15% Ficoll-400, in water).

● ● ●

## Annex 2

### Expression, refolding and purification of N- and F-KpOmpA

*The procedure takes around 1 week, starting with a bacterial pre-culture and ending with a purified protein in 0.1% Zwittergent 3-14 buffer.*

#### I. Bacterial culture and induction of the protein overexpression

1. Thaw at room temperature a 50- $\mu$ l frozen ( $-80^{\circ}\text{C}$ ) aliquot of *E. coli* BL21 (DE3), transformed with an N- or F-KpOmpA-containing plasmid.
2. Inoculate the bacteria in 50 ml growth medium, either *LB* or *TB* (see below), supplemented with 50  $\mu\text{g/ml}$  Ampicillin (50 mg/ml stock) for N-KpOmpA or 30  $\mu\text{g/ml}$  Kanamycin (30 mg/ml stock) for F-KpOmpA.
3. Leave the pre-culture overnight ( $30^{\circ}\text{C}$ , 240 rpm).
4. Measure the optical density at 600 nm ( $\text{OD}_{600}$ ).
5. Inoculate some of the bacteria in 1L growth medium (*LB/TB* or *M9*, depending if a labelled or natural abundance protein is needed), so that the initial  $\text{OD}_{600}$  is  $\sim 0.1$ .
6. Leave the cell culture at  $37^{\circ}\text{C}$ , 190 rpm.
7. Check the  $\text{OD}_{600}$  every 20-40 min, until  $\text{OD}_{600} = 0.6-0.8$ .
8. Remove 1 ml of cell suspension and keep it in eppendorf on ice. Mark it as NON-induced Cells.\*
9. Add 1 ml IPTG stock solution (1M) to the 1L of cell culture.
10. Leave the cell culture ( $37^{\circ}\text{C}$ , 190 rpm) for 4 hours in the case of rich media, or overnight in the case of minimal medium used.
11. Remove 1 ml of cell suspension and keep it in eppendorf on ice. Mark it as Induced Cells.\*

\* Optional steps for monitoring the protein overexpression with SDS PAGE.

#### II. Protein purification

##### A) Separating the inclusion bodies (IB):

1. Centrifuge the bacterial culture (5000 g, 10 min,  $4^{\circ}\text{C}$ ). Discard the supernatant.
2. Measure the weight of the bacterial pellet. *The procedure can be interrupted at this point and the pellet left overnight at  $-20^{\circ}\text{C}$ .*
3. Resuspend the bacterial pellet in *lysis buffer* (see below), using  $\sim 10$  ml buffer per 1 g of pellet. Use tissue grinder if necessary.
4. Add lysozyme solution (20 mg/ml stock kept at  $-20^{\circ}\text{C}$ ) up to 0.5 mg/ml (40x dilution).
5. Leave for 2 hours of incubation with gentle shaking at room temperature.
6. Sonicate the lysate with a tip sonifier (5x1 min, 50% active cycle). Keep the sample on ice.
7. Centrifuge the sonicated suspension (10 000 g, 30 min,  $4^{\circ}\text{C}$ ).
8. Keep 1 ml from the supernatant in an eppendorf at  $4^{\circ}\text{C}$ .\*
9. Homogenize the pellet in *lysis buffer* supplemented with 2M urea. Use tissue grinder.
10. Perform the IB-washing (steps 7-9) two more times (3 centrifugations in total). After the final centrifugation, do not resuspend the IB-pellet. Measure the weight of the IB-pellet. *The IB-pellet can be stored for long time ( $\sim$ months) at  $-20^{\circ}\text{C}$ .*

## B) Dissolving the inclusion bodies (IB):

11. Homogenize the IB in 4-5 ml of solution per 1 g of IB-pellet. This solution is:
  - if the IB are from bacteria grown in LB or TB media, then the solution is 7M urea with 10 mM DTT (added before use from 1M stock). Leave for 4 hours of gentle shaking at room temperature. *The pellet must be dissolved slowly, do not sonicate. Avoid heating at more than 50°C. If the pellet is not dissolved in 4 hours, it can be left overnight.*
  - if the IB are from bacteria grown in M9 medium, then the solution is 6M Guanidine/HCl, 25 mM Tris (pH 8.3), 5 mM EDTA, with 10 mM DTT (added before use from 1M stock), Leave for 4-12 hours of gentle shaking at room temperature, depending on the efficiency of dissolving of the pellet. Heat (~37°C) if necessary.
12. Centrifuge the dissolved IB (10 000 g, 30 min, 4°C). Keep the supernatant for the next step. Optionally, keep the pellet at 4°C.\*

## C) Protein renaturation in 1% Zwittergent 3-14 micelles:

13. Add *renaturation buffer 1* (see below) to the dissolved IB, with a volume = 14x the volume of the supernatant from step 12. In the beginning, add the detergent buffer dropwise to the chaotropic agent/protein solution while stirring on a magnet, until the total volume is at least doubled. At the end, leave the mixture overnight on a shaker at 4°C.

## D) Buffer exchange to 0.1% Zwittergent 3-14:

14. Centrifuge the renaturation mixture (10 000 g, 30 min, 4°C) in order to remove aggregated protein. Optionally, keep any pellet at 4°C.\*
15. Dialyze (12-14 kDa cutoff) the supernatant for ~8-12 hours at 4°C against 2L of *renaturation buffer 2* (see below). Exchange to fresh 2L for another ~8-12 hours.
16. Centrifuge the renaturation mixture (10 000 g, 30 min, 4°C) in order to remove aggregated protein. Optionally, keep any pellet at 4°C.\* Proceed with the supernatant.

## E) Protein purification with discontinuous imidazole gradient:

17. Estimate photometrically the approximate protein amount in the supernatant from the previous step. Mix the supernatant with washed nickel-charged resin (such as Ni-NTA Superflow, QIAGEN) at a ratio of 1 ml resin (= 2 ml slurry) for 5 mg protein. Batch the protein-resin mixture for 3-4 hours of gentle mixing at 4°C.
18. Depending on its total amount (i.e. column volume, CV), load the resin in chromatography column of appropriate size. Initiate the washing/elution steps at ~1 ml/min speed, as follows:
  - Collect the flowthrough.
  - Wash with 6-8 CV of *renaturation buffer 2* without imidazole.
  - Wash with 6-8 CV of *renaturation buffer 2* with 20 mM imidazole.
  - Elute with *renaturation buffer 2* with 400 mM imidazole. Depending on the amount of resin, this step can be fractionated or not.
  - If the resin is about to be re-used, wash with 10 CV of *renaturation buffer 2* with 1M imidazole and then with 10 CV of imidazole-free *renaturation buffer 2*. If necessary, follow the manufacturer's prescription for thorough resin washing and recharging.

19. The protein content in each chromatography step can be determined spectro-photometrically using the dependency:

$$C_{\text{protein}} [\text{mg/ml}] = A_{280} / \epsilon,$$

where  $\epsilon$  is 2.18 L mol<sup>-1</sup>cm<sup>-1</sup> for N-KpOmpA and 1.45 L mol<sup>-1</sup>cm<sup>-1</sup> for F-KpOmpA

Dialyze again the protein-containing eluate two times against 2L *renaturation buffer 2* in order to remove the imidazole.

20. Subject all the purification steps, the fractions and the pellets to SDS PAGE. Take aliquot of the eluate and heat it (100°C, 5 min) for tracking the unfolded state of the protein.

\* Optional step for monitoring eventual protein loss with SDS PAGE (see below for gel composition and sample preparation).

### III. Media and buffers compositions

#### 1. Luria-Bertani (LB) medium

	Per 0.5L	Per 1L
Bacto Peptone	5.0 g	10 g
Bacto Yeast Extract	2.5 g	5 g
NaCl (M.W. 58.44)	5.0 g	10 g
Adjust to pH 7. Autoclave for sterilization.		
/For 1.5% agar petri dishes, add 1.5 g AGAR per 100 ml LB medium./		

#### 2. Terrific Broth (TB) medium

	Per 0.5L	Per 1L
Terrific Broth (Invitrogen)	23.5 g	47.0 g
4 ml/L glycerol	2.0 ml	4.0 ml
Autoclave for sterilization.		

#### 3. M9 (minimal) medium with <sup>13</sup>C and <sup>15</sup>N

	Per 0.5L	Per 1L
<u>Sterile solutions prepared in advance:</u>		
Milli-Q water, autoclaved	395 ml	790 ml
(5x M9 stock) (see below), autoclaved	100 ml	200 ml
<sup>13</sup> C <sub>6</sub> Glucose (powder)	1 g	2 g
<sup>15</sup> NH <sub>4</sub> Cl (powder)	1 g	2 g
MgSO <sub>4</sub> (1M stock), filtered	1 ml	2 ml
CaCl <sub>2</sub> (1M stock), filtered	0.050 ml	0.100 ml
Vitamins mix (1000x) (see below)	0.5 ml	1.0 ml
Traces (see below)	5 ml	10 ml

##### • 5x M9 stock

	Per 0.5L	Per 1L
450 mM Na <sub>2</sub> HPO <sub>4</sub> (powder)	32 g	64 g
(if as Na <sub>2</sub> HPO <sub>4</sub> x 2H <sub>2</sub> O)	40.05 g	80.1 g
(if as Na <sub>2</sub> HPO <sub>4</sub> x 7H <sub>2</sub> O)	60.3 g	120.6 g
KH <sub>2</sub> PO <sub>4</sub> (powder)	7.5 g	15 g
NaCl (powder)	1.25 g	2.5 g

<b>• Vitamins Mix</b>	<b>[C]<sub>stock</sub>, mg/ml</b>	<b>Mix (1000x)</b>	<b>[C]<sub>final</sub>, mg/ml</b>
Thiamine (vit. B1)	500	5 ml	0.23
Riboflavin (vit. B2)	100	1 ml	0.009
Nicotinamide (vit. B3)	100	1 ml	0.009
Pyridoxal (vit. B6)	10	1 ml	0.0009
Cobalamin (vit. B12)	10	0.1 ml	0.0001
Biotin (vit. B7 / H)	100	1 ml	0.009
Pantothenic acid (vit. B5)	100	1 ml	0.009
Folic acid (vit. B9)	100	1 ml	0.009

<b>• Traces</b>	<b>Per 200 ml stock</b>	<b>[C]<sub>final</sub>, mg/ml</b>
EDTA	1 g	0.05
CaCl <sub>2</sub> x 2H <sub>2</sub> O	1.2 g	0.06
CuSO <sub>4</sub> x 5H <sub>2</sub> O	0.088 g	0.0044
MnCl <sub>2</sub> x 4H <sub>2</sub> O	0.24 g	0.012
H <sub>3</sub> BO <sub>3</sub>	0.004 g	0.0002
ZnSO <sub>4</sub> x 7H <sub>2</sub> O	0.14 g	0.007
FeSO <sub>4</sub> x 7H <sub>2</sub> O	1.2 g	0.06
Ascorbic acid (vit. C)	0.04 g	0.002

<b>4. Lysis buffer (pH 8.5)</b>	<b>Per 0.5L</b>	<b>Per 1L</b>
50 mM Tris (M.W. 121.14)	3.03 g	6.06 g
200 mM NaCl (M.W. 58.44)	5.84 g	11.68 g
5 mM EDTA	0.93 g	1.86 g
0.2% Tween 20 (add immediately before use).		

<b>5. Renaturation buffer 1 (pH 8.5)</b>	<b>Per 0.5L</b>	<b>Per 1L</b>
25 mM Tris (M.W. 121.14)	1.51 g	3.03 g
150 mM NaCl (M.W. 58.44)	4.38 g	8.76 g
1% (w/v) Zwittergent 3-14 (powder)	5.00 g	10.00 g

<b>6. Renaturation buffer 2 (pH 8.5)</b>	<b>Per 1L</b>	<b>Per 5L</b>
25 mM Tris (M.W. 121.14)	3.03 g	15.14 g
150 mM NaCl (M.W. 58.44)	8.77 g	43.80 g
0.1% (w/v) Zwittergent 3-14 (powder)	1.00 g	5.00 g



## 7. SDS PAGE gel

For 1 gel

For 2 gels

(The calculations are made for 0.75 mm glasses interspace, chambers from Bio-Rad)

### • Running gel (10%)

30% Acryl/Bis	2.0 ml	3.0 ml
3M Tris (M.W. 121.14), pH 8.4	2.0 ml	3.0 ml
50% Glycerol	2.0 ml	3.0 ml
10% SDS	60.0 µl	90.0 µl
10% APS	75.0 µl	112.5 µl
TEMED	7.5 µl	11.3 µl

### • Stacking gel (7.5%)

30% Acryl/Bis	0.75 ml	0.75 ml
3M Tris (M.W. 121.14), pH 8.4	0.75 ml	0.75 ml
Water	1.50 ml	1.50 ml
10% SDS	30.0 µl	30.0 µl
10% APS	30.0 µl	30.0 µl
TEMED	3.0 µl	3.0 µl

### • Sample preparation (apply 5-10 µl per gel well)

1M DTT	5 µl
Sample	20 µl
SDS-loading dye buffer (2x)	25 µl

• • •

Summarized technical information for the two protein constructs, N- and F-KpOmpA (such as total length and amino acids composition, molecular weight, sequence, isoelectric point, etc.) is presented below:

Construct	<b>N-KpOmpA</b>
Length	216 a.a.
Molecular Weight	23369.8 Da
1 µg corresponds to	42.790 pmoles
Molar extinction coefficient	50880 M <sup>-1</sup> cm <sup>-1</sup>
A <sub>280</sub> of 1 corresponds to	0.46 mg/ml
A <sub>280</sub> corresponding to 1 mg/ml	2.177 AU
Isoelectric Point	6.24
Charge at pH 7	-2.53

Amino Acid(s)	Number count	% by weight	% by frequency
Charged (RKHYCDE)	57	32.77	26.39
Acidic (DE)	20	10.08	9.26
Basic (KR)	17	10.15	7.87
Polar (NCQSTY)	54	27.29	25.00
Hydrophobic (AILFWV)	70	31.99	32.41
A Ala	23	7.52	10.65
C Cys	0	0.00	0.00
D Asp	14	6.84	6.48
E Glu	6	3.24	2.78
F Phe	8	4.85	3.70
G Gly	31	8.54	14.35
H His	8	4.56	3.70
I Ile	6	2.89	2.78
K Lys	7	3.76	3.24
L Leu	12	5.78	5.56
M Met	5	2.74	2.31
N Asn	13	6.30	6.02
P Pro	11	4.65	5.09
Q Gln	8	4.29	3.70
R Arg	10	6.39	4.63
S Ser	9	3.47	4.17
T Thr	12	5.25	5.56
V Val	15	6.45	6.94
W Trp	6	4.50	2.78
Y Tyr	12	7.98	5.56
B Asx (Asp + Asn)	27	13.15	12.50
Z Glx (Glu + Gln)	14	7.53	6.48

```

1  ARIMKAIFVL NAAPKDNTWY AGGKLGWSQY HDTGFYGNF QNNNGPTRND
51  QLGAGAFGGY QVNPYLGFM GDWLGRMAY KGSVDNGAFK AQQVQLTAKL
101 GYPITDDLDI YTRLGGMVWR ADSKGNAST GVSRSSEHDTG VSPVFAGGVE
151 WAVTRDIATR LEYQWVNNIG DAGTVGTRPD NGMLSLGVSY RFGQEDAAPV
201 VAPAPAPAPE HHHHHH

```

Construct	<b>F-KpOmpA</b>
Length	359 a.a.
Molecular Weight	38703.9 Da
1 µg corresponds to	25.837 pmoles
Molar extinction coefficient	56140 M <sup>-1</sup> cm <sup>-1</sup>
A <sub>280</sub> of 1 corresponds to	0.69 mg/ml
A <sub>280</sub> corresponding to 1 mg/ml	1.45 AU
Isoelectric Point	6.16
Charge at pH 7	-3.44

Amino Acid(s)	Number count	% by weight	% by frequency
Charged (RKHYCDE)	100	34.11	27.86
Acidic (DE)	38	11.67	10.58
Basic (KR)	34	12.00	9.47
Polar (NCQSTY)	94	28.12	26.18
Hydrophobic (AILFWV)	117	31.24	32.59
A Ala	38	7.50	10.58
C Cys	2	0.54	0.56
D Asp	23	6.78	6.41
E Glu	15	4.89	4.18
F Phe	10	3.66	2.79
G Gly	42	6.98	11.70
H His	9	3.09	2.51
I Ile	9	2.61	2.51
K Lys	18	5.83	5.01
L Leu	24	6.97	6.69
M Met	7	2.31	1.95
N Asn	18	5.27	5.01
P Pro	18	4.59	5.01
Q Gln	17	5.50	4.74
R Arg	16	6.17	4.46
S Ser	18	4.19	5.01
T Thr	22	5.80	6.13
V Val	30	7.78	8.36
W Trp	6	2.71	1.67
Y Tyr	17	6.82	4.74
B Asx (Asp + Asn)	41	12.05	11.42
Z Glx (Glu + Gln)	32	10.39	8.91

```

1  MKATVAQAAP KDNTWYAGGK LGWSQYHDTG FYGNGFQNNN GPTRNDQLGA
51  GAFGGYQVNP YLGFEMGYDW LGRMAYKGSV DNGAFKAQGV QLTAKLGYP I
101 TDDLDIYTRL GGMVWRADSK GNYASTGVSR SEHDTGVSPV FAGGVEWAVT
151 RDIATRLEYQ WVNNIGDAGT VGTRPDNGML SLGVSYRFGQ EDAAPVVAPA
201 PAPAPEVATK HFTLKSDVLF NFNKATLKPE GQQALDQLYT QLSNMDPKDG
251 SAVVLGYTDR IGSEAYNQQ LSEKRAQSVVD YLVAKGIPAG KISARGMGES
301 TPVTGNTCDN VKARAALIDC LAPDRRVEIE VKGYKEVVTQ PAALELVPRG
351 SVEHHHHHH

```

## Annex 3

### Expression and purification of C-KpOmpA

#### I. Bacterial culture and induction of the protein overexpression

1. Thaw at room temperature a 50- $\mu$ l frozen ( $-80^{\circ}\text{C}$ ) aliquot of *E. coli* BL21 (Star or DE3), transformed with a C-KpOmpA-containing plasmid (pJexpress411-C-KpOmpA).
2. Inoculate the bacteria in 50 ml *LB* growth medium (see below), supplemented with 30  $\mu\text{g/ml}$  Kanamycin (30 mg/ml stock).
3. Leave the pre-culture overnight ( $30^{\circ}\text{C}$ , 240 rpm).
4. Measure the optical density at 600 nm ( $\text{OD}_{600}$ ).
5. Inoculate some of the bacteria in 1L *LB* medium, so that the initial  $\text{OD}_{600}$  is  $\sim 0.1$ .
6. Leave the cell culture at  $37^{\circ}\text{C}$ , 190 rpm.
7. Check the  $\text{OD}_{600}$  every 20-40 min until  $\text{OD}_{600} = 0.6-0.8$ .
8. Remove 1 ml of cell suspension and keep it in eppendorf on ice. Mark it as NON-induced Cells.\*
9. Add 1 ml IPTG stock solution (1M) to the 1L of cell culture.
10. Leave the cell culture ( $37^{\circ}\text{C}$ , 190 rpm) for 4 hours.
11. Remove 1 ml of cell suspension and keep it in eppendorf on ice. Mark it as Induced Cells.\*

\* Optional steps for monitoring the protein overexpression with SDS PAGE.

#### II. Protein purification

1. Centrifuge the bacterial culture (5000 g, 10 min,  $4^{\circ}\text{C}$ ). Discard the supernatant.
2. Measure the weight of the bacterial pellet. *The procedure can be interrupted at this point and the pellet left overnight at  $-20^{\circ}\text{C}$ .*
3. Resuspend the bacterial pellet in *lysis buffer* (see below), using  $\sim 10$  ml buffer per 1 g of pellet.
4. Leave for 2 hours of incubation with gentle shaking at room temperature.
5. Sonicate the lysate with a tip sonifier (5x1 min, 50% active cycle). Keep the sample on ice.
6. Centrifuge the sonicated suspension (10 000 g, 30 min,  $4^{\circ}\text{C}$ ) in order to remove unbroken cells and large cellular debris. Continue working with the supernatant. *The pellet can be kept for checking for protein loss due to incomplete cell lysis.\**
7. Optional: Dialyze the supernatant against 20 mM  $\text{Na}/\text{PO}_4$ -buffer (pH 8) and 100 mM NaCl, in order to remove small cytoplasmic components and the Tween 20 detergent.
8. Mix the supernatant with washed nickel-charged resin (such as Ni-NTA Superflow, QIAGEN) at a ratio of 1 ml resin (= 2 ml slurry) for 5-6 mg protein.  
Note: Since the protein concentration can not be estimated neither colorimetrically nor spectrophotometrically at this point (in addition to the problem with the very low molar extinction coefficient ( $\epsilon = 7920 \text{ M}^{-1}\cdot\text{cm}^{-1}$ ) of C-KpOmpA), it is advisable that an SDS PAGE is performed with the total cell supernatant in order to obtain (albeit subjective) feeling for the amount of C-KpOmpA in the cell lysate. A serie of dilutions of the gel-loaded aliquots may help to accomplish this task, since the detection limit of the Coomassie-stained gels is around 3-5  $\mu\text{g}$  of protein in the band. After finding (even if quite inaccurately) the expected amount of C-KpOmpA, the nickel resin can be added in slight excess, in order to ensure complete binding and minimal loss of protein.
9. Batch the protein-resin mixture for 3-4 hours of gentle mixing at  $4^{\circ}\text{C}$ .

10. Wash (*washing buffer*) and elute (*elution buffer*) the protein-bound resin as described for N- and F-KpOmpA (Annex 2), by using the appropriate buffers (see below).
11. Dialyze the purified protein two times against 2L of 20 mM Na/PO<sub>4</sub>-buffer (pH 8) and 100 mM NaCl, in order to remove the imidazole. Store at 4°C.
12. Check the purification steps for protein loss and the purity of the eluate by SDS PAGE (the gel preparation is described in Annex 2) and estimate the protein concentration (either from a UV spectrum, if the protein concentration is high enough, or by Lowry titration).

\* Optional step for monitoring eventual protein loss with SDS PAGE.

### III. Medium and buffers compositions

1. Luria-Bertani (LB) medium	Per 0.5L	Per 1L
Bacto Peptone	5.0 g	10 g
Bacto Yeast Extract	2.5 g	5 g
NaCl (M.W. 58.44)	5.0 g	10 g

Adjust to pH 7. Autoclave for sterilization.

/For 1.5% agar petri dishes, add 1.5 g AGAR per 100 ml LB medium./

2. Lysis buffer (pH 8)	Per 1L
20 mM Na/PO <sub>4</sub>	5.3 ml NaH <sub>2</sub> PO <sub>4</sub> (0.2M stock) + 94.7 ml Na <sub>2</sub> HPO <sub>4</sub> (0.2M stock)
100 mM NaCl (M.W. 58.44)	5.84 g
0.1% Tween 20 (add immediately before use).	

3. Washing buffers (pH 8)	Per 1L
20 mM Na/PO <sub>4</sub>	5.3 ml NaH <sub>2</sub> PO <sub>4</sub> (0.2M stock) + 94.7 ml Na <sub>2</sub> HPO <sub>4</sub> (0.2M stock)
100 mM NaCl (M.W. 58.44)	5.84 g
[20 mM imidazole /second washing step/ (M.W. 68.08)	1.36 g]

4. Elution buffer (pH 8)	Per 1L
20 mM Na/PO <sub>4</sub>	5.3 ml NaH <sub>2</sub> PO <sub>4</sub> (0.2M stock) + 94.7 ml Na <sub>2</sub> HPO <sub>4</sub> (0.2M stock)
100 mM NaCl (M.W. 58.44)	5.84 g
400 mM imidazole (M.W. 68.08)	27.23 g

• • •

Summarized technical information for the C-KpOmpA construct (such as total length and amino acids composition, molecular weight, sequence, isoelectric point, etc.) is presented below:

Construct	C-KpOmpA
Length	159 a.a.
Molecular Weight	17299.1 Da
1 µg corresponds to	57.807 pmoles
Molar extinction coefficient	7920 M <sup>-1</sup> cm <sup>-1</sup>
A <sub>280</sub> of 1 corresponds to	2.18 mg/ml
A <sub>280</sub> corresponding to 1 mg/ml	0.46 AU
Isoelectric Point	6.71
Charge at pH 7	-0.63

Amino Acid(s)	Number count	% by weight	% by frequency
Charged (RKHYCDE)	50	37.68	31.45
Acidic (DE)	18	12.52	11.32
Basic (KR)	17	13.17	10.69
Polar (NCQSTY)	46	29.88	28.93
Hydrophobic (AILFWV)	49	28.54	30.82
A Ala	15	6.63	9.43
C Cys	2	1.20	1.26
D Asp	9	5.95	5.66
E Glu	9	6.57	5.66
F Phe	4	3.28	2.52
G Gly	13	4.84	8.18
H His	7	5.39	4.40
I Ile	5	3.26	3.14
K Lys	11	7.98	6.92
L Leu	12	7.81	7.55
M Met	3	2.22	1.89
N Asn	8	5.25	5.03
P Pro	6	3.43	3.77
Q Gln	9	6.53	5.66
R Arg	6	5.19	3.77
S Ser	13	6.78	8.18
T Thr	8	4.73	5.03
V Val	13	7.56	8.18
W Trp	0	0.00	0.00
Y Tyr	6	5.39	3.77
B Asx (Asp + Asn)	17	11.19	10.69
Z Glx (Glu + Gln)	18	13.10	11.32

```

1  MGSSHHHHHH SSGENLYFQG SEVATKHFTL KSDVLFNFNK ATLKPEGQQA
51  LDQLYTQLSN MDPKDGSVV LGYTDRIGSE AYNQQLSEKR AQSVDYLVVA
101 KGIPAGKISA RGMGESNPVT GNTCDNVKAR AALIDCLAPD RRVEIEVKGY
151 KEVVTQPAA

```

## Annex 4

### Reconstitution of N- and F-KpOmpA in *E. coli* Polar Lipids Extract

#### I. Detergent exchange on the protein

1. Estimate the necessary amount of N- or F-KpOmpA from the respective stock solution in 0.1% Zwittergent 3-14 (kept at 4°C).
2. Add the estimated protein aliquot to washed nickel-charged resin (such as Ni-NTA Superflow, QIAGEN) at a ratio of 1 ml resin (= 2 ml slurry) for 5 mg of protein. Leave the protein-resin mixture for 3-4 hours at 4°C with gentle mixing.
3. Centrifuge the protein-bound resin (3000 g, 10 min, 4°C) and discard the supernatant (*it can be checked under UV for protein loss*).
4. Add ~50 ml of *OG buffer* (see below) to the pelleted resin and leave the mixture on a shaker at room temperature for 10-15 min.
5. Centrifuge the resin (3000 g, 10 min, 4°C) and discard the supernatant.
6. Repeat steps 4 and 5 two more times (3 centrifugations in total).
7. Resuspend the pelleted resin in small amount (~few ml) of *OG buffer* and transfer it to a chromatography column. *For small preparations, single-use columns can be utilized.*
8. Collect the flowthrough and wash the resin with ~5 column volumes (CV) of *OG buffer*.
9. Elute the protein with *OG buffer* containing 400 mM imidazole. *Essentially, this procedure represents affinity chromatography on the protein after the first chromatography with Zwittergent 3-14. Since the protein is already purified, the intermediate step of washing with 20 mM imidazole (to remove contaminants) can be omitted.*
10. Check the protein concentration in the eluate and store it at 4°C.

Note: Alternative approach is to exchange the detergent while the protein-bound resin is loaded onto the column, by extensive washing (~20 CV) with the *OG buffer*.

#### II. Lipids preparation and solubilization

1. Transfer the cold-stored (-20°C) powder *E. coli* Polar Lipids Extract (PLE, Avanti Polar Lipids) to room temperature and wait for temperature equilibration without opening the bottle/vial, in order to avoid water condensation.
2. Measure precisely (with analytical balance) the weight of an empty borosilicate glass vial.
3. Transfer certain amount (i.e. 20-50 mg) of PLE powder to the glass vial and dissolve it in chloroform at approximately 20 mg/ml.
4. Dry the lipid/chloroform mixture under a nitrogen (or argon) stream while spreading the solution on the walls and bottom of the glass vial.
5. Place the glass vial in a vacuum chamber and continue drying the lipid film for at least 3 hours (or overnight).
6. Measure precisely (with analytical balance) the weight of the glass vial and note the net weight of the lipids.
7. Dissolve the lipid film in *OG buffer* at exact concentration (i.e. 5 or 10 mg lipids/ml).

8. Warm (up to 42°C) and sonicate (5 min in water bath) the solution several times, until the lipid film is entirely dissolved in the detergent buffer. This lipid stock solution can be stored at -20°C for long terms (~months).

Note: Whether this solution looks completely transparent (mixed micelles) or opaque to a certain extent (some lipid vesicles) will depend on the detergent-to-lipid ratio and the ability of a given detergent (always above its CMC) to dissolve a given type of lipids. Observing opacity of this solution is not problematic, since the protein-detergent buffer (added before starting the reconstitution) is usually applied to a much larger volume than the lipid buffer. Having the same (or similar) detergent concentration as in the lipid-detergent stock, the protein-detergent buffer will greatly shift the detergent-to-lipid ratio in the final ternary mixture towards the detergent. This will destroy any remaining lipid vesicles and form homogenous, transparent mixture of protein molecules in detergent-lipid micelles.

### III. Protein reconstitution by dialysis

1. Mix certain amounts from the protein-detergent and lipid-detergent buffers, so that the lipid-to-protein ratio (LPR) and the total amount of protein are as desired. For example, mixing 5 ml of 2 mg/ml protein in OG buffer with 0.5 ml of 10 mg/ml lipids in OG buffer will produce a ternary mixture of 10 mg protein and 5 mg lipids, at LPR of 0.5 (w/w).
2. Leave the ternary mixture for ~15 min at room temperature with gentle shaking.
3. Transfer the ternary mixture to a dialysis tube with an appropriate cutoff (i.e. 12-14 kDa SpectraPor, as used here) and dialyze it for 12 hours against 2L of detergent-free buffer (25 mM Tris (pH 8.5) and 150 mM NaCl). If the sample volume is uncomfortably small, increase it with a few ml of OG buffer. *In the case of PLE as used lipids, the dialysis can be performed at room temperature or at 37°C. For many lipid types this needs to be done at 37°C (or even higher temperature), in order to stay above the gel-to-liquid crystalline phase transition point of the lipid.*
4. Exchange the dialysis buffer with fresh 2L for another 12 hours. Continue the dialysis buffers exchange until white, powder-like substance is formed in the tube.  
*To ensure maximal detergent removal, it is advisable to repeat the dialysis steps 3-4 times.  
In the case of ssNMR sample preparation, the NaCl needs to be removed as well. For this purpose, the exchange of dialysis buffers is done several more times (~6 hours per step is enough), with 25-mM reduction steps of NaCl, from 150 mM down to zero.*
5. Collect the proteoliposomes from the tube (~15 000 g, 10 min, room temperature) and resuspend them in small amount (~few ml, depending on the total sample size) of 20 mM Tris (pH 8.5).  
*In the case of ssNMR sample preparation, the proteoliposomes need to be further centrifuged at high speed (200 000 g, 90 min, 10°C), in order to stack them in as small as possible pellet. The weight of this pellet is compared with the total amount of starting material (mg protein + mg lipids) and is brought down (with controlled partial dehydration) to 2/3 of the NMR rotor volume (i.e 36 µl for 3.2-mm rotor and 50 µl for 4-mm rotor), thus leaving 1/3 of water/buffer content.*

Example: For preparing an ssNMR sample, 22 mg protein were reconstituted in 11 mg PLE, giving an LPR of 0.5 (w/w) and total material amount of 33 mg. For a 4-mm rotor with 50 µl volume, dehydrating the sample down to 50 mg will leave 17 mg (~30%) of buffer. After the ultracentrifugation of the vesicles, the pellet weight is found to be 110 mg. This pellet is thus dehydrated under nitrogen stream until its weight drops down to ~50 mg, after which the sample is loaded into the rotor by centrifugation (3000 g, 10 min, 10°C).



#### IV. Buffers compositions

<b>1. OG buffer (pH 8.5)</b>	<b>Per 0.2L</b>	<b>Per 0.5L</b>
25 mM Tris (M.W. 121.14)	0.61 g	1.51 g
150 mM NaCl (M.W. 58.44)	1.75 g	4.38 g
2% octylglucoside (OG, Carbosynth)	4 g	10 g
[400 mM imidazole /elution step/ (M.W. 68.08)	5.45 g	13.62 g]

<b>2. Dialysis buffer (pH 8.5)</b>	<b>Per 2L</b>	<b>Per 5L</b>
25 mM Tris (M.W. 121.14)	6.06 g	15.14 g
150 mM NaCl (M.W. 58.44)	17.53 g	43.83 g

[For NaCl removal, gradually reduce its concentration in 25-mM steps/~6-12 hours each.]

• • •







**Author**            Iordan Iordanov

**Title**                Structure and dynamics of the outer membrane protein A from *Klebsiella pneumoniae*: a joint NMR - SMFS - proteolysis and MS approach

**Supervisor**        Alain Milon

**Location and date of defense** IPBS Toulouse, 5 June 2012

**Summary**

KpOmpA is a two-domain membrane protein from *Klebsiella pneumoniae* belonging to the outer membrane protein A (OmpA) family. It is composed of a transmembrane  $\beta$ -barrel with 8  $\beta$ -strands and a C-terminal, soluble periplasmic domain. The transmembrane domain presents a significant homology with *E. coli* OmpA whose three dimensional structure has been determined by X-ray crystallography and by NMR. The *E. coli* homologue can function as an adhesin and invasin, participate in biofilm formation, act as both an immune target and evasin, and serves as a receptor for several bacteriophages. It is assumed that most of these functions involve the four protein loops that emanate from the protein to the exterior of the cell. The difference between KpOmpA and *E. coli* OmpA is mostly concentrated in these extracellular loops which are larger in the case of KpOmpA. KpOmpA was shown to activate macrophages and dendritic cells through the TLR2 dependent pathway, and these larger loops are supposed to play a specific role in the interactions with the immune system.

Thus the structure and dynamics of these loops is of prime functional significance. The currently available information in this regard, including the NMR structure determined in the IPBS NMR group in 2009, have been obtained so far with recombinant protein samples purified and refolded in detergent micelles. In the present work we first established a reconstitution protocol that allowed the incorporation of the membrane protein in the more native environment of the lipid bilayer and characterised our samples by electron microscopy.

SMFS experiments were used to probe the reconstituted KpOmpA unfolding-refolding pathways, exploring the folding mechanisms for  $\beta$ -barrel proteins and suggesting a novel role for OmpA in the bacterial membrane (in collaboration with the group of D. Müller, ETH Zürich).

The C-terminal periplasmic domain of KpOmpA was expressed and purified as a separate product and the feasibility of its structure elucidation by NMR was demonstrated by obtaining a high quality HSQC spectrum.

The dynamic behaviour of the extracellular portion of the KpOmpA membrane domain reconstituted in liposomes has been investigated by solid state MAS NMR relaxation experiments. We confirmed that the previously observed gradient of dynamic along the molecule axis is an intrinsic property of the protein. Limited proteolysis and MALDI-TOF experiments were coupled with the NMR information in order to assess more precisely the different mobility levels in the loops.

Evolutional preservation of the different loops regions is related to their observed flexibility, pointing towards immunologically important, variable, dynamic and accessible loops sections.

**Key words**        outer membrane protein A, protein expression and purification, protein reconstitution, single-molecule force spectroscopy, protein dynamics, spin-lattice relaxation, solution and solid-state NMR, electron microscopy, limited proteolysis, mass spectroscopy.

**Laboratory address**    Institut de Pharmacologie et de Biologie Structurale (IPBS), 205 route de Narbonne, 31077, Toulouse cedex, France.