Structural Investigation on Membrane Proteins: the case of the KdgM Porin Family, OmpF and IIBC^{suc}

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

Vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

des Universität Basel

von

Christophe Wirth

aus Frankreich

Basel, 2012

Genehmigt von der Philophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

      Prof. Dr. Tilman Schirmer

      Prof. Dr. Bernhard Erni

Basel, den 02.03.2010

                                 Prof. Dr. Eberhard Parlow

                                          Dekan

Declaration

I declare that I wrote this thesis, **Structural Investigation on Membrane Proteins: the case of the KdgM Porin Family, OmpF and IIBC**^{suc}, with the help indicated and only handed it in to the faculty of science of the University of Basel and to no other faculty and no other university.

II

Abstract

Gram-negative bacteria are surrounded by two distinct membranes delimiting the periplasm. The inner-membrane (or cytoplasmic membrane) is composed of a phospholipid bilayer whereas the outer-membrane is asymmetric and composed of lipopolysaccharides in the outer leaflet and phospholipids in the inner leaflet. Both membranes are spanned by numerous membrane proteins. These are various and their genes are representing about 30% of the genome in living organisms. Although the first membrane protein structures were solved in the 80s using X-ray crystallography, structural data is still very limited. This dissertation groups our contributions to this field in the last few years and is divided in three parts.

A first chapter will describe our work on the structural characterization of proteins from the KdgM family. This family of small porins has been first identified in *Dickeya dadantii* a plant pathogen. It has been shown that members of that specific porin family are able to transport acidic (oligo)saccharides but the structural and biochemical data remain scarce. We undertook to study this family using X-ray crystallography working in parallel on several homologues. The chapter is divided in three parts: (I) the purification of the *Escherichia coli* NanC porin, a member of the KdgM family involved in the uptake of sialic acid, its crystallization and structure determination at 1.8 Å resolution; (II) the discussions concerning the biological implications of the NanC structure that was published in a research article as well as additional discussions, and (III) the still ongoing work on the purification and crystallization of KdgM using microseeding matrix screening and surface entropy reduction.

In a second chapter, we will discuss a new crystal form of OmpF, the general porin of *E. coli*. OmpF is probably the best studied outer-membrane general porin as it is very abundant and rather easy to purify. It has also been used as a model to study translocation of ions through porins. Since it first structural determination, OmpF structures have been determined in several space groups. Here we report a new crystal form diffracting to 2.7 Å resolution and we discuss the presence of 32-symmetry related, honeycomb like layers reoccurring in almost all OmpF crystal forms and formed by identical contacts.

The last chapter will discuss the purification and crystallization attempts of the inner-membrane sucrose specific permease IIBC^{suc} from *Salmonella typhimurium* and of its complex with a Fab antibody fragment. This protein belongs to the phosphoenolpyruvate: sugar phosphotransferase system (PTS) which is ubiquitous in eubacteria, but is not found in other organisms. The PTS belongs to the group translocation super-family of transporters and is composed of proteins or domains allowing the specific uptake of a sugar and its concomitant phosphorylation. Apart from topology studies, no structural data is available on the sugar translocation proteins of the PTS whereas all other members have been characterized. In order to increase the crystallization ability of the IIBC^{suc} permease, we used monoclonal antibodies Fab fragments to enlarge the hydrophilic surface of the permease and to rigidify its structure. Fab fragment production, purification and binding will be discussed as well as crystallization trials of IIBC^{suc}-Fab complexes.

Abbreviations

AMP / ATP: adenosine mono-phosphate / adenosine tri-phosphate

C_8E_4: tetraethylene glycol monooctyl ether

C_8E_5: pentaethylene glycol monooctyl ether

C_8HESO: 2-hydroxyethyloctylsulfoxide

CRP: cyclic AMP receptor protein

Da / kDa: Dalton / kiloDalton

DNA: desoxynucleic acid

DM: decylmaltoside

ELISA: Enzyme Linked ImmunoSorbent Assay

Fos12: foscholine 12

FT: flow through

g: gravitation unit

GA2 / GA3 / GA4 / GAn: di- / tri- / tetra- / oligogalactoronate

GlcNAc-6P: N-acetyl glucosamine-6-phosphate

h: hour

His-tag: hexahistidine tag

IgG: immunoglobulin G

K: Kelvin

kb: kilo base

K_D: dissociation constant

LB: Luria-Broth medium

LDAO: N-dodecyl-N,N-dimethylamine-N-oxide

LPS: Lipopolysaccharides

LS: N-lauroyl sarcosine

mAb: monoclonal antibody

mg: milligram

Min: minute

MIRAS: multiple isomorphous replacement and anomalous scattering

ml / µl: milliliter / microliter

mM / M: millimolar / molar

µm: micrometer

MW: molecular weight

Neu5Ac: N-acetyl neuraminic acid

Neu5,9Ac$_2$: 9-O-acetyl N-acetylneuraminic acid

NG: nonylglucoside

NMR: nucleic magnetic resonance

OD: optical density

OG: octylglucoside

OM: octylmaltoside

OMP: outer-membrane channel forming proteins

OPOE: octylpolyoxyethylene

OTG: octylthioglucoside

PAGE: polyacrylamide gel electrophoresis

PEG: polyethylene glycol

PEP: phosphoenolpyruvate

PDB: Protein Data Bank

pS: pico Siemens

psi: pound per square inch

PTS: phosphoenolpyruvate-dependent carbohydrate transport system

r.m.s.d.: root mean square deviation

rpm: revolutions per minute

SDS: sodium dodecylsulfate

SDS-PAGE: sodium dodecylsulfate polyacrylamide gel electrophoresis

UV: ultraviolet

Acknowledgements

My grateful thanks go to Prof. Tilman Schirmer for giving me the opportunity to work in his research group and on these interesting projects. I was very happy to be guided by him throughout my PhD study and that he let me enough liberty in research to allow me to acquire decisional skills. His help and experience was especially critical to solve crystallographic problems encountered during this study.

This work was done in collaboration with several people in house or from abroad. I'm very grateful to Dr. Guy Condemine (Villeurbanne) for our fruitful collaboration on small porins. Many thanks to Prof. Simon Bernèche and Céline Boiteux (Basel) for the MD simulations and to Prof. Bob Eisenberg and Janhavi Giri (Chicago) for the lipidic bilayer studies.

I am also very grateful to Dr. Caroline Peneff-Verheyden for her direct supervision during these four years of PhD. First having different projects, we rapidly started working together, scientifically complementing each other, and forming the "membrane protein team". Thanks a lot.

I would like to thank the whole research group, past and present, and especially Dr. Paul Wassmann. We almost started and finished together this very important period in both our lives. Sharing the same office, we were kind of PhD brothers, dealing with similar problems and searching solutions together. Thanks also to Franziska Zaehringer for her constant availability for helping others. Many thanks also go to both my neighbors during this work, the two Arnaud (Dr. Arnaud Baslé and Arnaud Goepfert). The many interesting discussions we had together helped me a lot. I would also like to thank Dr. Zora Markovic-Housley, Dr. Claudia Massa, Dr. Dinesh Palanivelu, Dr. Sivaraman Padavattan, Nisha Vanayak and Tillmann Heinisch for helpful discussions and the nice lab atmosphere as well as Dietrich Samoray for the technical assistance. Finally, during my PhD work we had several master

students in the lab: Marcel Meury, Roman Lehner, Ludwig Zumthor Mathias Barone and Aline Borer. I enjoyed a lot working with them.

Additionally, I would also like to thank Dr. David Cobessi, now scientist at the IBS (Grenoble). He was my mentor during my Master training in Strasbourg and taught me membrane proteins purification and crystallography. Probably more important, he gave me the desire to continue in that direction and to carry on working on membrane proteins.

Of course, I would like to show all my gratitude to my parents, sister and brother for their assistance and constant encouragements during my studies and PhD.

Finally, my very special thanks are going to Carole, my wife, who supported me during this period and was beside me during all good and less good moments. Thank you so much…

VIII

Contents

III. Expression, purification and crystallization assays of sucrose permease IIBC and its complex with a Fab fragment 125

Preamble: Membrane proteins

The cytoplasm of a cell is separated from the extracellular environment by a (or several) lipidic membrane(s). The proteins located in the membrane are playing an essential role in cell life, as they assume critical functions such as the uptake of the nutrients necessary for its survival or the sensing of all types of signals coming from the environment. Because of these critical functions, membrane proteins are also important pharmaceutical drug targets.

Genes coding for membrane proteins are representing about 30% of the genome contents. Today, more than 60000 structures of proteins are deposited in the Protein Data Bank (PDB), therefore, there should be about 20000 membrane protein structures in this depository. However, the reality is very different. Compared to the vast majority of soluble proteins, membrane proteins are more difficult to express, to solubilise and to crystallize. Furthermore, there are often additional problems such as low protein solubility or stability. Therefore, on membrane proteins structures are representing less than 1% of the total structures known. The difficulty to obtain their 3-dimensional structure and therefore to obtain fundamental knowledge as well as their therapeutic potential, makes structural studies on membrane proteins challenging but clinically relevant.

During this work, we focussed on Gram-negative bacteria that are surrounded by two membranes delimiting a cellular compartment called periplasm. The outer-membrane is an asymmetric selective barrier composed of a layer of phospholipids on the periplasmic side and a layer of lipopolysaccharide turned to the extracellular environment. The inner-membrane (or cytoplasmic membrane) is composed of a phospholipid bilayer (Figure 1). Both membranes are spanned by many proteins allowing signalling and passage of nutrients or waste molecules. The outer-membrane is a semi-permeable barrier. Small molecules with a typical size of less than 600 Da can diffuse through β-barrel proteins called porins that can be specific or not for one particular molecule or class of molecules. Bigger nutrients (such as vitamin B12) can also cross the outer-membrane *via* bigger protein channels but are not diffusing freely as this process is energy dependent. The inner-membrane is much more impermeable as no wide

open water filled channels exist. To allow translocation, flexibility is required and likely obtained thank to their α-helical folds. In the inner-membrane, proteins are involved in other cellular processes such as signal transduction or energy formation.
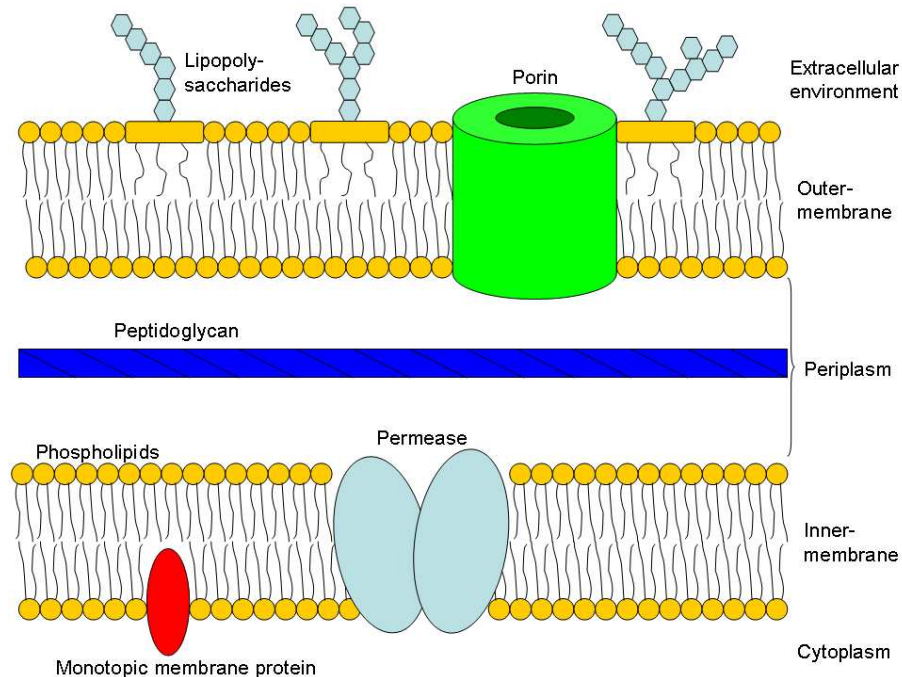
Figure 1: Schematic representation of a Gram-negative bacterium membrane organisation

In this thesis, results concerning inner- and outer-membrane proteins will be presented in three main parts: (I) the KdgM family of specific porins including the crystal structure of NanC, a member of that family responsible for the uptake of sialic acid in *Escherichia coli*, (II) the crystal structure of OmpF, a general porin of *E. coli* in a new space group and, finally, (III) the expression, purification and crystallization trials of enzyme IIBC^{suc}, a sugar specific permease from *Salmonella typhimurium*, with and without the help of antibody Fab fragments.

Chapter I

Structural characterisation of the KdgM outer-membrane protein family

Introduction

Outer-membrane pore forming proteins: functional classification

Gram-negative bacteria are surrounded by two membranes delimiting the periplasm. The outer-membrane is an asymmetric bilayer and is mostly composed of phospholipids on the periplasmic leaflet and of lipipolysaccharides (LPS) on the outer leaflet. This efficient protection barrier is spanned by numerous outer-membrane proteins. Amongst the variety of functions, pore forming proteins called porins, are allowing uptake of nutrients and removal of waste or toxic molecules (Nikaido & Vaara, 1985). In Gram-negative bacteria, proteins involved in nutrient translocation can be classified in three distinct categories (Delcour, 2003; Nikaido, 2003).

The so-called general (or non-specific) porins, constitute the majority of expressed proteins present in the outer-membrane. Proteins from this well described category are responsible for the translocation of small (less than 600 Da of molecular weight) solutes through the outer-membrane with very little selectivity (Nikaido, 2003). The flux of molecules through a non-specific pore forming protein is driven by the concentration gradient between the extracellular environment and the periplasm and is directly proportional to the concentration difference between these two compartments. The well studied OmpF from *Escherichia coli* belongs to this subcategory (see Chapter II).

The members of the second category, called specific porins or outer-membrane substrate specific channels, differ from the general porin by their ability to bind a given solute *via* a saturable binding site (Schirmer, 1998). The presence of this binding site confers to channels a higher efficiency in translocating molecules at a low extracellular concentration compared to general porins. As specific porins have a binding site with a given affinity, an analogy with enzymes has been proposed and therefore the translocated solute is also sometimes called the substrate of the specific porin (Schirmer, 1998). The K_D values range from millimolar (for the maltose specific porin from *E. coli* (Benz et al, 1987)) to sub-micromolar (for the phosphate specific porin OprP from *Pseudomonas aeruginosa* (Moraes et al, 2007)).

Finally, the third category, called outer-membrane receptors, comprises proteins able to translocate larger substrates (600 to 1500 Da) such as siderophores or vitamine B12 (Ferguson & Deisenhofer, 2002). In contrast with other porins, the members of this outer-membrane protein family are not forming continuous open channels as they are occluded by a so-called

"plug" domain. In addition, these proteins require energy provided by the inner-membrane TonB complex in order to transport their substrate (Braun & Endriss, 2007).

Outer-membrane pore forming proteins: structural classification

During the last 20 years, 3-dimensional structures from outer-membrane pore forming proteins (OMPs) belonging to all the functional categories have been obtained. OMPs can therefore also be classified according to their structures and oligomeric states (Schulz, 2002).

A first structural subfamily comprises OMPs belonging to the non-specific and specific porins categories. Proteins belonging to this family all form water-filled hollow β-barrels, composed of 16 to 18 strands and are assembled in very stable trimers (Fig. I.1). Another common feature of this family is the presence of a loop folding back into the pore about half way between the periplasm and the extracellular vestibule. This loop is also called constriction loop, as it significantly diminishes the pore radius. This subfamily comprises porins from *Rhodobacter capsulatus* (Weiss et al, 1991), *Rhodobacter blasticus* (Kreusch & Schulz, 1994), *Paracoccus denitrificans* (Hirsch et al, 1997), the *E. coli* general porins OmpF, PhoE and OmpC (Basle et al, 2006; Cowan et al, 1992) as well as the general porins OmpK36 of *Klebsiella blasticus* (Dutzler et al, 1999) and Omp32 of *Comamonas acidovorans* (Zeth et al, 2000). Other structures of specific outer-membrane channels such as the maltose specific porin LamB from *E. coli* (Schirmer et al, 1995) or the sucrose specific porin ScrY from *Salmonella typhimurium* (Forst et al, 1998) were also determined and belong to that structural subfamily. More recently, three additional specific porins of *P. aeruginosa*: OprP (Moraes et al, 2007), OprD (Biswas et al, 2007), OpdK (Biswas et al, 2008) were also structurally characterized.

In the last years, several other outer-membrane channel proteins have been structurally characterized that clearly belong to a new structural subfamily. This second structural group is defined by smaller proteins forming monomeric 12 to 14-stranded β-barrels (Fig. I.1). In this case, no loop is responsible for the channel constriction. This subfamily regroups the putative oligosaccharide specific OmpG (Subbarao & van den Berg, 2006; Yildiz et al, 2006), the nucleoside specific Tsx from *E. coli* (Ye & van den Berg, 2004) and three specific porins

involved in the uptake of hydrophobic compounds: FadL from *E. coli* (van den Berg et al, 2004), TodX from *Pseudomonas putida* and TbuX from *Ralstonia pickettii* (Hearn et al, 2008).
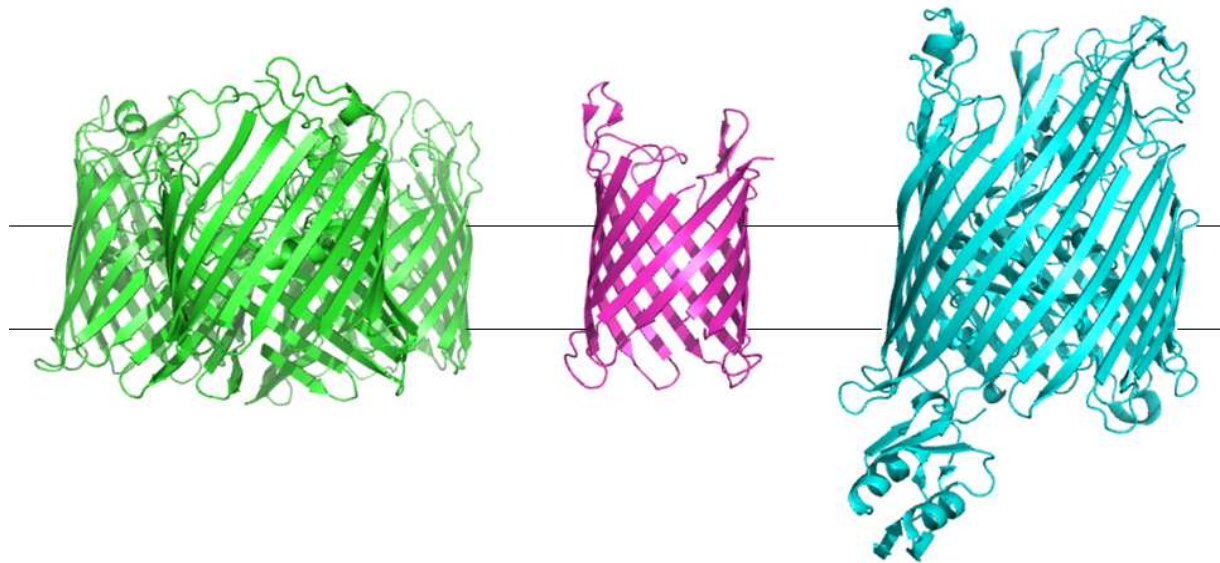
Figure I.1: Outer-membrane β-barrel channel proteins

Cartoon representation of the three structural subclasses of outer-membrane channel proteins: OmpF, the trimeric, 16-stranded general porin from *E. coli* (green) (Cowan et al, 1992), the small monomeric Tsx protein from *E. coli* (magenta) (Ye & van den Berg, 2004), and FpvA a 22-stranded TonB dependent transducer from *P. aeruginosa* (cyan) (Wirth et al, 2007). In this representation, the extra-cellular side is at the top and the periplasm is at the bottom of the figure.

Finally, a third structural group can be defined including the TonB dependent receptors and transducers such as FhuA from *E. coli* (Ferguson et al, 1998; Locher et al, 1998) or FpvA from *P. aeruginosa* (Cobessi et al, 2005). These proteins form huge, 22-stranded β-barrel occluded by a so-called "plug" domain (Fig. I.1). Some proteins, then called "transducers" belonging to this structural group are also acting as signal transducer for sensing of extracellular molecules. This subgroup is carrying an additional domain allowing the transduction of the signal to inner-membrane located proteins (Brillet et al, 2007; Wirth et al, 2007).

The KdgM family

Recently, a new outer-membrane protein family, called the KdgM family, has been identified (Blot et al, 2002). Members of this family have the shortest sequence amongst the described outer-membrane channels with an average mature protein length of about 220 amino-acids. The first member of the family has been identified in the plant pathogen *Dickeya dadantii* (formerly called *Erwinia chrysanthemi*) but the protein family (Pfam accession number: PF06178) comprises about 300 other members today. These proteins are regrouped in a family based on the assumption that structural similarity can be expected due to conserved profiles in their sequences. KdgM homologues are present in 56, mostly Enterobacteriaceae, species including plant and human pathogens such as uropathogenic or food poisoning *E. coli*, *Salmonella typhimurium*, *Yersinia pestis*, *Yersinia enterocolitica* and *Vibrio*, *Shigella* and *Pseudomonas* species. In most cases two and up to four homologous sequences are present in the genome indicating a functional advantage for these bacteria to maintain several paralogues.

D. dadantii pathogenicity

D. dadantii is a plant pathogen responsible for the so-called soft rot disease (Hugouvieux-Cotte-Pattat et al, 1996) causing huge economical lost in agricultural plants. The plant cell wall is, amongst others, composed of pectin, a polymer of α-1-4-linked galacturonate residues, some of which can be either methyl- or acetyl-esterified. *D. dadantii* is able to colonize wounded plants. During the course of infection, *D. dadantii* is massively secreting a pool of plant cell wall degrading enzymes, mostly pectinases, *via* the Out system (Condemine et al, 1992; Pugsley, 1993), a type II secretion pathway (Driessen & Nouwen, 2008). More than 25 enzymes are involved in the pectin degradation and catabolism (see Fig. I.2). Several of these enzymes have redundant functions but are synthesized under various environmental contexts or have different locations (secreted, membrane attached, periplasmic or cytoplasmic) (Condemine et al, 1992; Pugsley, 1993). The expression of these numerous enzymes with slight functional and/or regulation differences accounts for the ability of *D. dadantii* to completely degrade the plant cell wall pectin polymer into small oligogalactoronates molecules, leading to the soft rot disease. Oligogalacturonate (GAn), the pectin degradation product, is composed of heterogeneous oligomers of galacturonides ranging mostly from dimers (GA2) to tetramers (GA4) but also some longer chains.
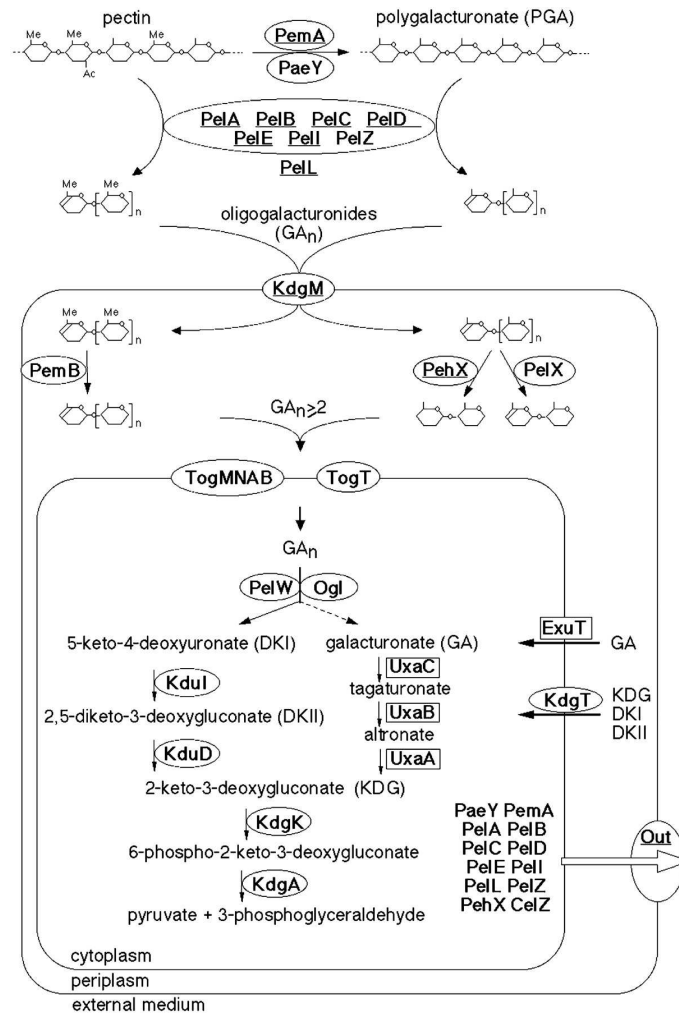
Figure I.2: Schematic representation of the enzymes involved in pectin degradation of pectin and catabolism of galacturonic acid in *D. dadantii* (Blot et al, 2002).

GAn can be used as nutrient by *D. dadantii* which possesses all enzymes required for its uptake and catabolism. Amongst these enzymes, PelX is a periplasmic exo-pectate lysase able to cleave GAn (n≥4) into shorter galacturonates (Shevchik et al, 1999b). The GAn degradation products can then enter the cytoplasm *via* TogMNAB and TogT (Hugouvieux-Cotte-Pattat & Reverchon, 2001), two inner-membrane transporters. The first is an ATP-binding cassette transporter belonging to the carbohydrate uptake transporter-1 family whereas the second belongs to the glycoside-pentoside-hexuronide transporter symporter family. In the cytoplasm PelW pectate-lyase is able to cleave GA4 and GA3, whilst GA2 is preferentially processed by Ogl an oligogalacturonate lyase (Shevchik et al, 1999a). The final products of GAn degradation are galacturonate and 5-keto-4-desoxyuronate which are further degraded to pyruvate and 3-phosphoglyceraldehyde. These two molecules finally enter the

general cellular metabolism in order to produce energy for the bacterial cell (Shevchik et al, 1999b). All the proteins involved in pectin degradation as well as GAn entry in the cell and catabolism have been identified including proteins responsible for the translocation of GAn at the outer-membrane level.

KdgM and KdgN

The translocation of GAn across the outer-membrane of *D. dadantii* is mediated by KdgM and KdgN, two recently discovered porins belonging to the KdgM family (Blot et al, 2002; Condemine & Ghazi, 2007). KdgM is a 25kDa outer-membrane protein that has been identified as being a specific channel for galacturonic acids. The *kdgM* gene is part of the *pelW-togMNAB-kdgM-paeX* operon, surrounded by genes involved in pectin uptake and degradation (Blot et al, 2002). The regulation of *kdgM* is complex. The expression of the whole operon is controled by KdgR, a protein regulating most of the genes involved in pectinolysis and by a cyclic AMP receptor protein (CRP) activator (Reverchon et al, 1997). An additional strong internal promoter located in front of the *kdgM* gene allows selective regulation of the two last genes in the operon by PecS a protein involved in the regulation of several virulence factors in *D. dadantii* (Rouanet et al, 2004). KdgM is strongly expressed together with the pectinases when pectin is present in the environment.

It has been shown that KdgM is essential for the transport of pectin degradation products, and in particular longer GAn (n ≥ 3). For smaller GAn molecules, the KdgM function can be taken over by the general porins of *D. dadantii*. As its function can be partially complemented by general porins, KdgM is important but not essential for *D. dadantii* during the course of plant infection. The transport ability of KdgM has been studied *in vitro* using electrophysiology experiments in lipidic bilayers resulting in the characterization of its porin activity. The pore of KdgM shows a conductance of 450 pS at a positive potential and is slightly selective for anions. To confirm, *in-vitro*, the translocation of GAn (n ≥ 3), experiments were realized with addition of trigalacturonate (GA3) to the bath of the electrophysiology experiment (Blot et al, 2002). This was resulting in typical records of a fast blocker molecule during the experiment indicating that GA3 is translocated by KdgM but that the translocation occurs so quickly that the single channel recording is not able to distinctly resolve them. At 100 mV membrane potential, a K_D of 34 mM could be estimated for GA3.

The genome of *D. dadantii* is also encoding for KdgN, a close homologue of KdgM. The reason of this redundancy is not clear yet; however, initial data comparing KdgM and KdgN suggest differences in the regulation of these two proteins depending which medium is used to grow the bacteria (Condemine & Ghazi, 2007). This is similar to what has been observed in the regulation of OmpF and OmpC in *E. coli* but it is the first time that it is observed for specific porins (Batchelor et al, 2005). KdgN presents a weak selectivity for anions and a conductance in the range of 450-500 pS. However, even if *in vivo* experiments are showing a GA3 translocation capacity, no blockage was observed in the electrophysiology experiments upon addition of GA3 up to 60 mM, in contrast with KdgM (Condemine & Ghazi, 2007).

NanC

In *E. coli*, a homologue to KdgM has been identified and called NanC. NanC is involved in *N*-acetyl neuraminic acid (Neu5Ac), the most common sialic acid uptake as it is essential for the survival of *E. coli* cells deficient in the general porins OmpF and OmpC when grown on minimum medium containing Neu5Ac as sole carbon source (Condemine et al, 2005). Sialic acids is a family of nine-carbon sugar acids mostly present in the glycoconjugates found at the surface of higher vertebrate cells (Angata & Varki, 2002). These acidic sugars are usually involved in cell-cell or cell-protein interactions as well as they can act as signalling molecules in the case of inflammation. Sialic acid is also a central molecule in host-pathogen interactions as some, mostly pathogenic, bacteria are able to utilize sialic acids in, at least, three distinct ways. Some bacteria, such as *E. coli* K1, are able to coat their cell surface with sialic acid molecules in order to mimic the host cell and to escape the immune system (Vimr et al, 2004). Some other bacteria can use sialic acid gained from the host organism as a valuable carbon and nitrogen rich nutrient (Severi et al, 2007; Vimr et al, 2004). Finally, it has been recently suggested that sialic acid plays an important role in signalling. During the course of inflammation, sialic acids are heavily released by host cells. This release has been shown to be sensed by some bacteria allowing them to escape the host immune system (Sohanpal et al, 2004; Sohanpal et al, 2007).

Similarly to what had been done with KdgM, electrophysiology experiments on NanC inserted in lipid bilayer were done in order to determine its pore properties. Between low

positive and low negative potential, the NanC channel is open whereas at high positive or negative voltages, the channel is closing. Similarly to KdgM, NanC has a conductance of 450 pS at a positive potential and is weakly selective for anions (Condemine et al, 2005). However, lipidic bilayer experiments could not confirm the Neu5Ac translocation activity as no channel fast-blocage could be observed when adding the putative substrate up to 50 mM. Unsuccessful tests had also been done using colominic acid a polymer molecule composed of very long chains of Neu5Ac.

The presence of NanC is not required to allow the uptake of Neu5Ac when general porins are expressed in *E. coli* and the Neu5Ac concentration is high enough. Once in the periplasm, Neu5ac can be transported by NanT, a protein belonging to the major facilitator superfamily and will be metabolised into 6-phosphorylated N-acetyl glucosamine (GlcNAc-6P) by the NanA, NanE and NanK proteins (Plumbridge & Vimr, 1999; Vimr & Troy, 1985). Further GlcNAc-6P catabolism by NagA and NagB are leading to fructose-6P that can enter the general cell metabolism. All these proteins belong to one operon whereas NanC is part of a separated 3 genes operon (*nanc-yjhT-yjhS*).

As for KdgM and KdgN, the regulation of NanC is complex and its promoter is forming one of the longest non coding zones (1.4 kb) of the *E. coli* genome (Fig I.3). The *nanc* operon is regulated by NanR, a DNA binding transcriptional regulator also involved in the regulation of sialic acid metabolism (operon *nana-nant-nane-nank*) (Kalivoda et al, 2003) and by NagC, a protein involved in the expression regulation of enzymes of the *N*-acetylglucosamine catabolism pathway, including NagA and NagB (Condemine et al, 2005). When Neu5Ac is present, it binds to NanR and thereby allows derepression of the NanC operon. The same mechanism occurs when GlcNAc-6P, the Neu5Ac degradation product, binds to NagC. NanC is co-regulated with the FimB recombinase, a protein involved in the off-to-on switching of type 1 fimbriation in *E. coli* (Sohanpal et al, 2004). However, expressions of FimB and NanC are regulated in opposite ways. While the expression of NanC is activated by the presence of Neu5Ac and GlcNAc-6P in the cell, the expression of FimB is repressed.

14
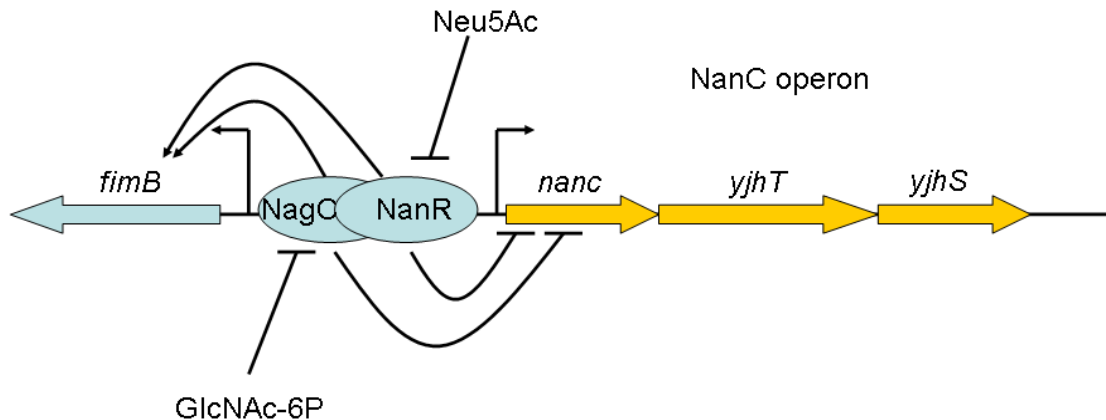
Figure I.3: NanC gene organisation and regulation.

The genes belonging to the NanC operon are colored in yellow. When NanR or NagC is bound, the expression of the NanC operon is repressed whereas fimB transcription is activated. When Neu5Ac or its degradation product GlcNAc-6P is present, the NanC operon is activated whereas the fimB expression is repressed.

The two additional genes composing the *nanc* operon are involved in sialic acid catabolism. *yjhT* encodes for a recently characterized periplasmic protein called NanM that has been shown to be a sialic acid mutarotase (Severi et al, 2008). This enzyme is able to catalyze the mutarotation of the Neu5Ac α-anomer (present in glycoconjugate) into the β-anomer that is present in solution and able to be transported by the inner-membrane sialic acid transporter NanT. This reaction is spontaneous in solution but occurs at a very slow rate. NanM allows a significant increase the reaction velocity. The protein encoded by the *yjhS* gene has been recently characterized as being a 9-O-acetyl N-acetylneuraminic acid esterase and has been renamed NanS (Steenbergen et al, 2009). These results suggest that the operon containing the *nanc* gene is involved in the uptake and catabolism of alternative sialic acid molecules.

Structural data on the KdgM family

Structural knowledge on members of the KdgM family is limited. From cross-linking data and lipid bilayer experiments on KdgM, KdgN and NanC, it was suggested that proteins of that family are monomeric porins such as OmpG or Tsx (Blot et al, 2002; Condemine et al, 2005; Condemine & Ghazi, 2007). In a study based on cystein specific labelling experiments, a 14-stranded β-barrel topology was proposed for KdgM, similarly to OmpG (Blot et al, 2002).

Later, a study comparing electron microscopy projection maps of 2D-crystals of KdgM and KdgN from *D. dadantii* and NanC from *E. coli* with maps from other porins suggested that the three members of the KdgM family are presenting similar β-barrel structures with an oval shape in the membrane plane (Signorell et al, 2007). This was in accordance with sequence based transmembrane β-strands predictions. That study also gave confirmation that these proteins are monomeric. However, the secondary structure analysis concluded to a 12-stranded β-barrel pore contradicting the cystein labelling experiments (Signorell et al, 2007).

Goal and achievements of this project

The KdgM family is the first described that allows the uptake of oligomers of acidic sugars. This ability to facilitate diffusion of rather big molecules contrasts with the short length of the polypeptide chain of these proteins that is suggesting a probable small pore. However, structural data about this family is clearly lacking. The aim of this project was to obtain structural data of a member of the KdgM family in order to understand the mechanisms essential for the substrate translocation in this new family of specific porins. To achieve this, the work was carried out in parallel on three homologous proteins of the family: KdgM, KdgN from *D. dadantii* and NanC from *E. coli*. The following research reports will describe in a first part, the purification, crystallization and structure determination of NanC followed by a second part presenting the biological implications of the NanC structure we published (Wirth et al, 2009) as well as further additional comments and discussions. A third part will shortly summarize the results obtained by Roman Lehner during his Master practical under my and Dr. Caroline Peneffs' supervision and the ongoing attempts to solve the structure of KdgM.

Chapter I.1.

Solving the structure of NanC porin using membrane incorporated and inclusion bodies refolded protein

Christophe Wirth, Guy Condemine, Tilman Schirmer and Caroline M. Peneff.

I.1.1. Material and Methods

I.1.1.1. Expression, refolding, and purification of NanC produced in inclusion bodies

In order to obtain large amounts of pure NanC protein, the first production protocol took advantage of the previously set up methods (Condemine et al, 2005). NanC was expressed without secretion signal sequence leading to its accumulation in inclusion bodies. The cells were broken using a French press and the inclusion bodies were harvested and dissolved using 6M urea. NanC was then slowly refolded by overnight dialysis against 10mM Tris, 0.5% SDS. The refolded sample was then purified using a preparative SDS-PAGE followed by a preparative Tris-Tricine PAGE. Finally a size exclusion chromatography was performed in order to remove aggregated material as well as to change the buffer to the desired detergent for crystallization.

Later the refolding and purification protocols were modified by Dr. Caroline Peneff. Inclusion bodies were solubilized in 10mM Tris pH 8.0, 8M urea and refolding was performed by drop-wise fast dilution into a 10mM Tris pH 8.0, 5% N-dodecyl-N,N-dimethylamine-N-oxide (LDAO) solution at room-temperature and under constant mixing An ultracentrifugation (30 min at 200000g) was performed subsequently in order to remove protein aggregates. The supernatant was loaded on an anion exchange chromatography column equilibrated with 10mM Tris pH 8.0, 0.05% LDAO and the elution was carried out using a 0 to 1M NaCl gradient. Fraction containing NanC were concentrated and further purified using size exclusion chromatography in 10mM Tris pH 8.0, 150mM NaCl, 0.14% foscholine12. Prior to crystallization, the salt was removed by overnight dialysis.

I.1.1.2. Expression of NanC in the membrane, extraction and purification

In order to express NanC in the membrane, primers yjh-3 5'-ggtccctagcgattattcctgc-3' and yjh-4 5'-cgaccttgcgataattcacccg-3' were used to amplify a DNA fragment extending 1 kb upstream and downstream of the *nanC* coding sequence. The amplified fragment was cloned into plasmid pGEM-T (Promega). A NcoI-Acc65I fragment was cut from this plasmid and inserted into the same sites of plasmid pKSM717 (Maneewannakul et al, 1994). The insert was then shortened by a NcoI-AgeI deletion to give plasmid pKSMNanCΔ1. This plasmid

was introduced into strain BL21(DE3)omp8/pLys (F⁻ *ompT hsdS$_B$ gal dcm ΔlamB ompF*::Tn*5* Δ*ompA* Δ*ompC* (DE3)) (Prilipov et al, 1998). Transformants were grown in LB medium at 30°C. When OD reached 1.0, IPTG was added at a final concentration of 1mM and cells were grown overnight. Bacteria were collected by centrifugation at 6000*g* for 10 min, resuspended in 10mM Tris-HCl pH 8.0, 1mM EDTA, treated with benzonase (Merck) and disrupted using a French press. Unbroken cells were eliminated by centrifugation at 6000*g* for 20 minutes and the membrane fraction was collected by centrifugation at 100000*g* for 1h.

Total membrane pellets were resuspended in 50mM Tris pH 8.0. Addition of 1% lauroyl-sarcosine and 1 min vortexing allowed specific solubilization of the inner-membrane. The outer-membrane was pelleted by ultracentrifugation (100000*g* for 1h) and the solubilized inner-membrane proteins discarded with the supernatant. The outer-membrane pellet was resuspended in 50mM Tris pH 8.0. Addition of 0.5% octylpolyoxyethylene (OPOE) allowed solubilization of the outer-membrane proteins. After ultracentrifugation (100000*g* for 30 min), the remaining pellet was again resuspended in 50mM Tris pH 8.0. This procedure was done several times while increasing slowly the OPOE concentration until all the NanC protein was solubilised according to SDS-PAGE.

NanC was purified using anion exchange chromatography (MonoQ, GE Healthcare) in 20mM Tris pH 8.0, 0.6% OPOE and eluted with a 0 to 1M NaCl gradient. The fractions were analysed on SDS-PAGE and the ones containing NanC were collected, diluted 10 times in 25mM acetate pH 5.0, 0.6% OPOE and then loaded on a cation exchange chromatography column (MonoS, GE Healthcare). Bound proteins were eluted using a 0 to 1 M NaCl gradient. Prior to crystallization, a size exclusion chromatography (Superdex 200, GE Healthcare) was performed, allowing to estimate the monodispersity and to exchange the buffer and detergent.

I.1.1.3. NanC crystallization

All crystallization screens were done at room temperature using the sitting drop vapour diffusion method. Initial screens were done using the sparse-matrix sampling approach (Jancarik & Kim, 1991) with commercial screens. The typical reservoir volume was 80µl and drops were prepared mixing 0.5µl reservoir solution with 0.5µl of protein solution at various protein concentrations and in different detergents. Several crystallization conditions yielding

20

crystalline objects were identified. The best were optimised in 24-well plates with a reservoir volume of 500µl and drops made by mixing 0.5µl protein solution with 0.5µl reservoir solutions.

For heavy atom derivatization, both soaking and co-crystallization methods were tested. For each heavy atom used (Pt, Hg, Au, Sm, Eu), several compounds having various reactivities, were tested. In the co-crystallization experiments, the protein was mixed with the heavy atom solution (ranging from 1 to 10mM) several hours before the set up of the drops whereas in the soaking experiments, in solutions containing 0.1mM to 10mM heavy atoms for various times (10 min to 24h) were used.

I.1.1.4. Data collection

Prior data collection, crystals were quickly soaked into stabilizing solution containing cryoprotectant (15-20% glycerol or 15-20% ethylene glycol) when necessary, mounted in cryoloops and flash-frozen in liquid nitrogen. X-ray diffraction data collection was performed at 100K on beamline X06SA at the Swiss Light Source (SLS, Villigen, Switzerland) using a MARCCD for the native crystals and a PILATUS 6M detector for the heavy atoms derivitized crystals.

I.1.1.5. Structure solution

The diffraction data were processed using MOSFLM (Leslie & Powell, 2007) or XDS (Kabsch, 1993) and scaled with SCALA (Evans, 2006). Heavy atom sites were found using SHELXD (Schneider & Sheldrick, 2002) and refined with SHARP (Bricogne et al, 2003). Initial phases were calculated using SHARP. Solvent flattening was performed using the program SOLOMON (Abrahams & Leslie, 1996). To increase the phase quality, multi-crystal density averaging was applied using DMmulti (Cowtan & Main, 1993). The model was built using COOT (Emsley & Cowtan, 2004) and REFMAC (Winn et al, 2003) in interactive refinement. The model quality was assessed using MOLPROBITY (Davis et al, 2007). Crystal contact analysis was realized using the PISA server of the EBI (Krissinel & Henrick, 2007). Models were visualized and figures were done using Pymol (www.pymol.org).

I.1.2. Results and Discussion

NanC was expressed and purified in two ways. The first took advantage of the existing *nanc* construct that had been cloned by Dr. Guy Condemine without signal peptide resulting in NanC expression in inclusion bodies. Two different refolding methods (using SDS or LDAO as detergent) and purification protocols (preparative SDS PAGE or ion exchange chromatography) were used. In a second approach, NanC was expressed with its signal peptide resulting in a mature protein secreted to the outer-membrane. This protein was purified directly from the membrane using ion exchange chromatography methods.

I.1.2.1. Purification, crystallization and diffraction data analysis of NanC refolded from inclusion bodies in SDS.

NanC production from inclusion bodies was based on a protocol set up initially for its characterization (Condemine et al, 2005). It allowed producing large amounts of pure NanC protein. However, the several steps of purification (refolding using SDS as a detergent, preparative SDS-PAGE) were considered being too harsh compared to conventional purification methods. In the final gel filtration chromatography step, the detergent was exchanged and the monodispercity of the NanC peak was checked. Several detergents were tested such as octylglucoside (OG), N-dodecyl-N,N-dimethylamine-N-oxide (LDAO), octyltetraoxyethylene, octylpentaoxyethylene, foscholine 12 (Fos12) and foscholine 8. Samples with OG, LDAO and Fos12 were leading to monodisperse gel filtration peaks (data not shown). The corresponding size was confirming that NanC is in a monomeric state.

Several crystallization screens were done using KdgM in these three detergents and crystals could rapidly been obtained from a 0.14% Fos12 sample and at a protein concentration of 20-22mg/ml. Well shaped crystals could be obtained in a condition containing 0.1M MES pH 6.5 (or HEPES pH 7.5) and 40-45% PEG 400. During the crystallization optimization process, two different crystal morphologies were observed. The first crystals were very long rods of sometimes more than $100 \times 100 \times 800 \mu m$ whereas the second crystals were fine rhombohedric plates of about $10 \times 200 \times 300 \mu m$ (Fig. I.1.1).

The rod type crystals were easy to handle whereas the plate type crystals were very fragile and often breaking during fishing with the cryoloop. However, several datasets could be recorder from both crystal types. Typical diffraction images are shown in Fig. I.1.2. The best observed resolution was to about 3.4 Å but, unfortunately, the diffraction was strongly anisotropic (Fig. I.1.2).

Figure I.1.1: NanC crystals have two different shapes.

(left) Rod shaped NanC crystals and (right) rhombohedral plates of NanC.
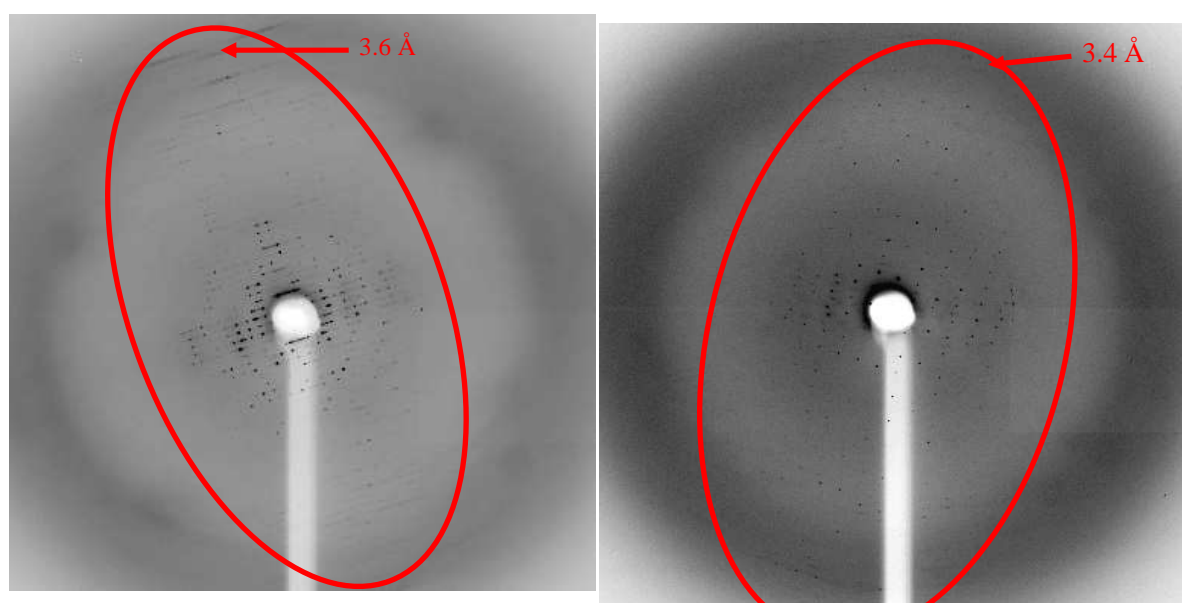
Figure I.1.2: Diffraction pattern from two types of NanC crystals.

Diffraction pattern of (a) the rod and (b) the rhombohedral crystal type showing the anisotropy in both cases (red oval) and the resolution limit (arrow).

The processing of the diffraction data was done using MOSFLM (Leslie & Powell, 2007) and XDS (Kabsch, 1993). In the case of the rod shaped crystals, the indexing led to an unambiguous trigonal or hexagonal space group solution. Self rotation function showed a clear peak at $\kappa=60°$ as well as at $\kappa=180°$ suggesting respectively a 6-fold axis and 2-fold axes. From the systematic extinctions (reflection condition for *00l : l=3n*) a 6_2 or 6_4 screw axis was concluded. This would suggest that the space group of the rod shaped crystals is $P6_222$ or $P6_422$ with huge cell dimensions. For the plate shaped crystals, the indexing was easier as the C2 space group was the only suggested with a low penalty.

Table I.1.3: Crystallographic data of the NanC crystals obtained from protein refolded in SDS

	Rod shaped crystals	Plate shaped crystals	
Space group	$P6_422 / P6_222$	C2	
Cell parameters (Å)	a = b = 278, c = 129	a = 247, b = 71, c = 199, β = 126	
Resolution (Å)	100 – 8.0 (8.43-8.0)	75 - 6.0 (6.33 - 6.0)	75 – 7.0 (7.38 – 7.0)
R_{merge}	12.1 (48.3)	10.2 (25.2)	8.9 (27.3)
$I/\sigma(I)$	10.0 (3.6)	10.7 (3.1)	12.8 (3.7)
Completeness	99.4 (99.8)	90.4 (64.4)	95.9 (98.2)

Values in brackets are for the high resolution shell of the data.

For both crystal types, data reduction with SCALA (Evans, 2006) led to reasonable overall statistics when the resolution was lowered in order to keep the R_{merge} values in the higher resolution shells acceptable (table I.1.3).

The precise number of molecules per asymmetric unit could not be clearly estimated using Matthews coefficient calculation (Matthews, 1968). The hexagonal crystals probably contains between 9 and 15 molecules par asymmetric unit (Vm between 2.00 Å^3/Da and 3.33 Å^3/Da) as well as the C2 crystals (Vm comprised between 1.96 Å^3/Da and 3.27 Å^3/Da for 15 and 9 monomers respectively). The data collected from rhombohedral plates were used in molecular replacement trials using several available models. As it was unclear if NanC had a 12 or 14 stranded barrel, models having both numbers of strands were used. OmpG (Subbarao & van den Berg, 2006; Yildiz et al, 2006), Tsx (Ye & van den Berg, 2004), NalP (Oomen et al, 2004), FadL (van den Berg et al, 2004) structures (respective PDB codes: 2IWV, 1TLY,

1UYN and 1T1L) were used as a search model. All side chains were removed except Cβ atoms in order to obtain poly-alanine models. The loops and turns were also removed in order to keep only the β-barrel scaffold for the molecular replacement models. Molecular replacement was tried using different programs but none could find a solution. This is most likely due to the low resolution and the severe anisotropy of the collected data or because the search models were too different from the NanC structure.

The tendency of NanC refolded in SDS to form low resolution diffracting crystals could eventually be explained by several problems in the protein production and purification. The crystals were diffracting with strong anisotropy which could be due to the intrinsic structure of the membrane protein that has hydrophilic areas (capable of strong and specific crystal contacts) along the pore axis direction and has hydrophobic areas (forming strong but unspecific crystal contacts) along the two other axes. Refolded NanC could also be badly folded when using SDS as a detergent. It might therefore be necessary to add a purification step in order to discriminate properly from badly refolded NanC. This can be achieved by comparing chromatograms of refolded and membrane purified NanC. The SDS might also interfere with a normal protein crystallization process as it is strongly negatively charged and is likely to be carried over the different purification steps as it could bind to the numerous positive residues present in NanC. If a SDS molecule is binding by its charged moiety to the charged residues of the protein, the hydrophobic tails of SDS could impair the formation of tight crystal contacts. To avoid these potential problems the expression and purification protocols were modified.

I.1.2.2. Expression, solubilisation and purification of the membrane targeted and LDAO refolded NanC

Expression was done using full length protein including the signal peptide to address NanC to the membrane and was yielding about 0.5 mg of pure protein per litre of culture. The use of BL21(DE3)omp8/pLys strains (Prilipov et al, 1998) as an expression system considerably diminished the number of contaminants, as the major porins are not expressed in this *E. coli* strain.

The membrane solubilisation protocol was designed such as to optimize the specific extraction of NanC and thereby avoid many contaminants. In a first step, *N*-lauroyl sarcosine was added. In these conditions, the detergent is unable to insert into the outer-membrane due to the presence of LPS whereas it can solubilise the inner-membrane and therefore specifically extract the inner-membrane proteins that will be removed from the sample by an ultracentrifugation step. The subsequent stepwise extraction of the outer-membrane proteins using OPOE allowed the fractionation of membrane proteins in function of their anchoring strength into the membrane. The first OPOE extraction removed loosely inserted proteins. NanC was mainly extracted in the 1, 2 and 3% OPOE steps and has already a high purity level as controlled on an SDS-PAGE (Fig. I.1.4) as the absence of major porins avoids the presence of these contaminants.
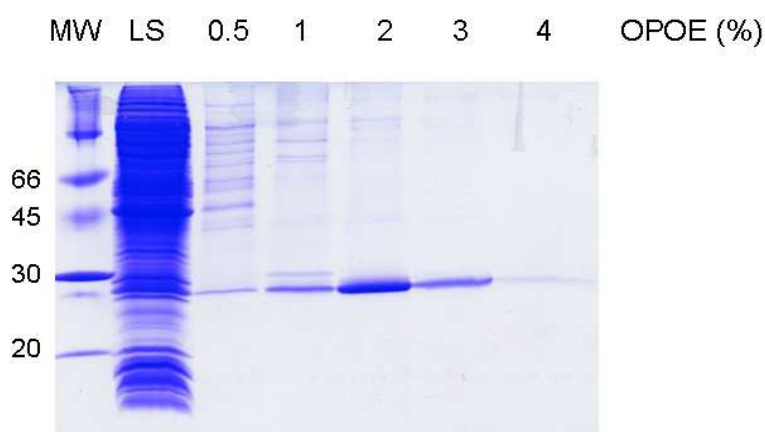
Figure I.1.4: SDS-PAGE of the NanC extracted samples.

MW stands for molecular weight marker in kDa, LS is corresponding to the 1% N-lauroyl sarcosine fraction, 0.5, 1, 2, 3 and 4 refer to the concentration of OPOE of the marked fractions.

Nevertheless, further purification was required to obtain a high amount of crystallization grade protein. NanC was expressed without affinity tag and therefore the purification was realized using ion exchange chromatography. The extracted samples were diluted to obtain a protein sample at about 20 mM Tris pH 8.0 and 0.8% OPOE and loaded on an anion exchange chromatography column. The elution with a 0 to 1M NaCl gradient led to several peaks. NanC eluted at a salt concentration of about 350-400 mM NaCl, together with a contaminant having a very similar molecular weight whereas some other unidentified contaminants were eluting at lower NaCl concentration (Fig. I.1.5).
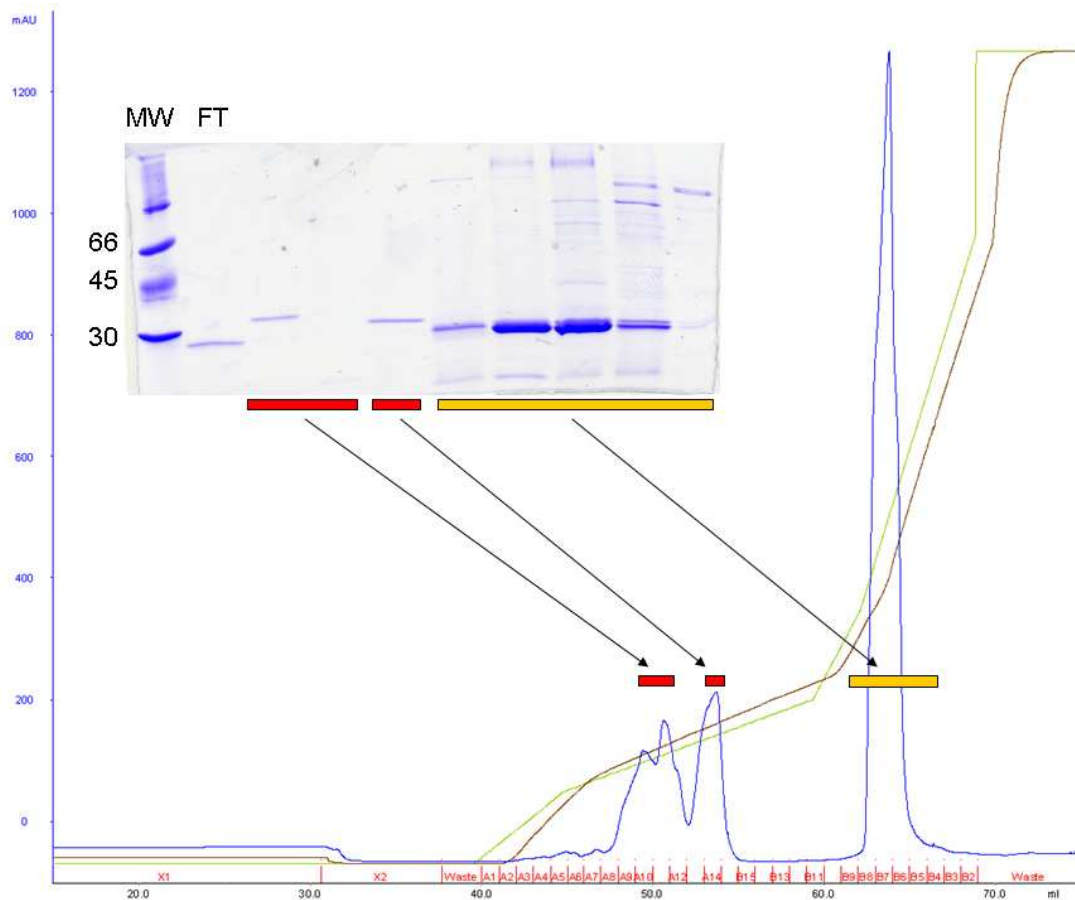
Figure I.1.5: Anion exchange chromatogram and corresponding SDS-PAGE.

Anion exchange chromatography using NanC extracted from the membrane. The blue trace represents the UV-absorption in mAU, the green curve is the NaCl gradient and the brown curve corresponds to the measured conductivity. Red bars represent contaminants and the yellow bar represents NanC containing protein fractions.

After analysis on SDS-PAGE, NanC containing fractions were pooled, diluted in 25mM acetate pH 5.5 and 0.6% OPOE, and subjected to cation exchange chromatography. Here NanC eluted at about 150 mM NaCl whereas the main other contaminant, that had a similar size than NanC, binds stronger to the columns and is eluted only later, at about 250mM NaCl (Fig. I.1.6).
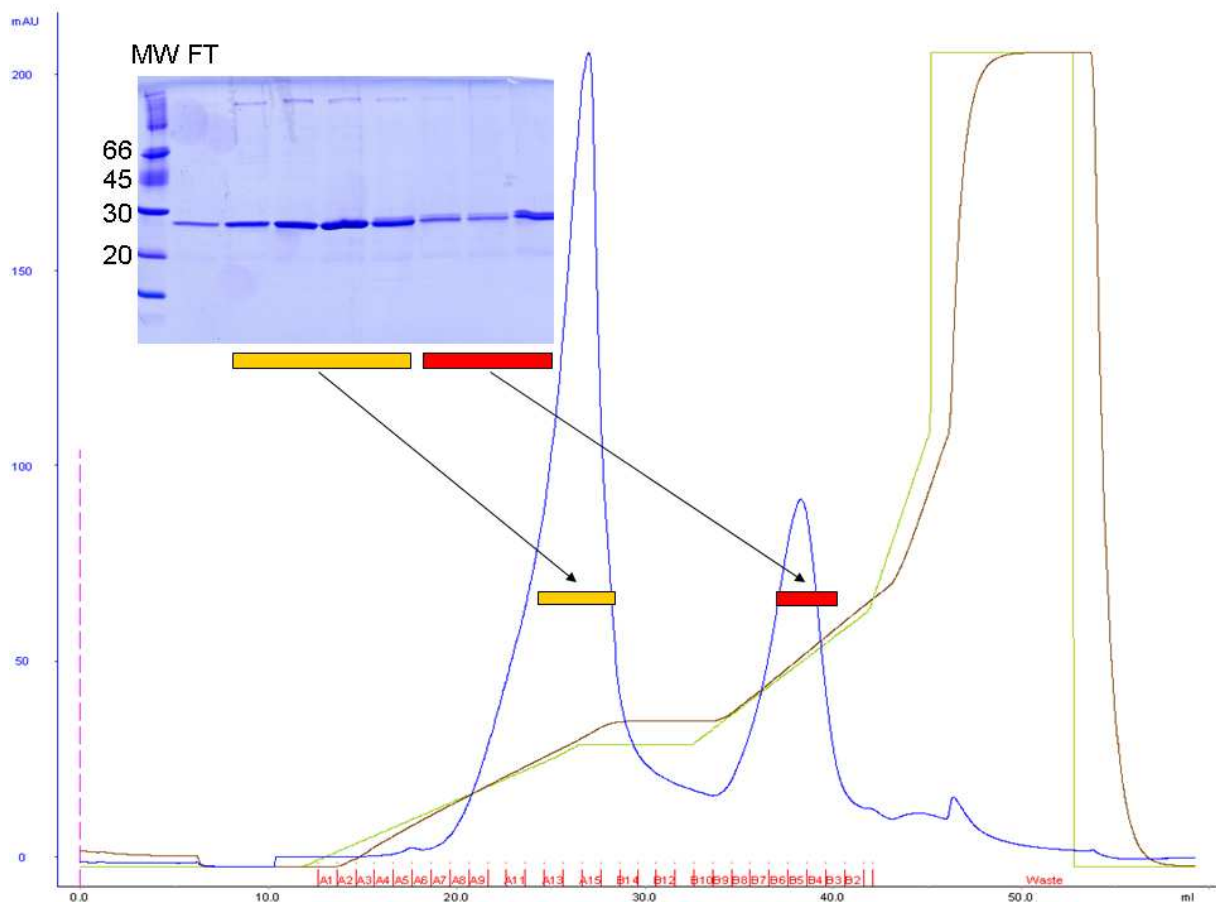
Figure I.1.6: Cation exchange chromatogram and corresponding SDS-PAGE.

Cation exchange chromatography of NanC containing samples allowing separation of the main contaminant. The blue trace represents the UV-absorption in mAU, the green is the NaCl gradient and the brown is the measured conductivity. Red bars represent contaminants and the yellow bar represents NanC containing protein fractions.

After these ion exchange chromatography steps, NanC is pure enough to allow crystallization experiments. However, to obtain monodispersity information of the sample and to exchange the detergent to the desired one used for crystallization, the fractions containing NanC were concentrated and submitted to size exclusion chromatography (Fig. I.1.7). In LDAO, NanC eluted as a monodisperse peak at a mass corresponding to about 60 kDa according to the retention time on the size exclusion chromatography. This is consistent with the molecular weight of NanC added to the mass of a detergent micelle.
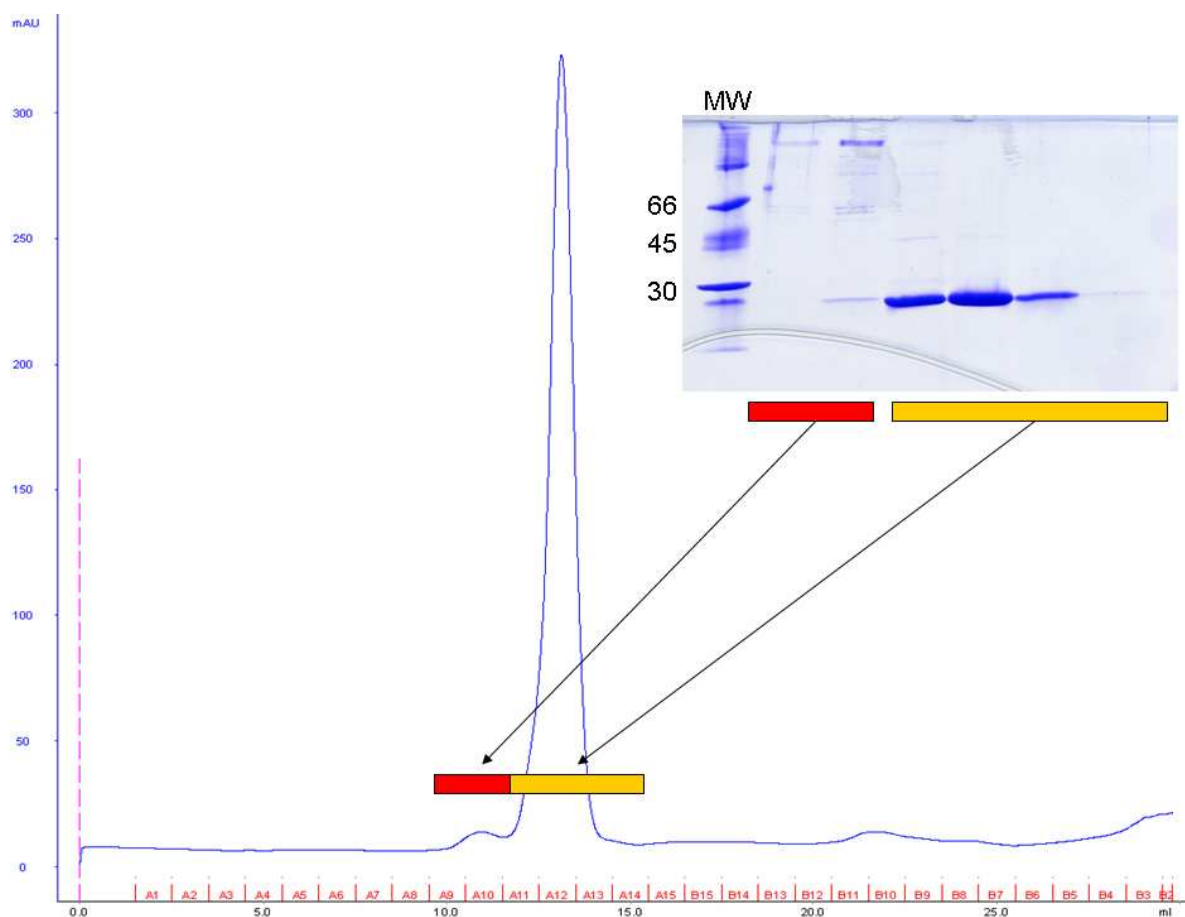
Figure I.1.7 Size exclusion chromatogram of NanC and corresponding SDS-PAGE.

Chromatogram realised in 0.05% LDAO as detergent. The blue trace corresponds to UV-absorption. Red bars represent contaminants and the yellow bar represents NanC containing protein fractions.

In parallel, refolding of NanC from inclusion bodies was realized using LDAO by Dr. Caroline Peneff as described in the material and method part. One challenge when working with refolded protein is to be able to discriminate between well and badly refolded protein in order to use, for subsequent crystallization, only protein in its native fold. Knowing the elution profile of the membrane extracted NanC on an anion exchange chromatography column, we decided to use this method in order to remove badly refolded protein. Except not well refolded NanC, no significant other contaminant was present and therefore, no further purification was required. However, a gel filtration chromatography was also done in order to exchange the detergent and to estimate the monodispercity of the NanC sample. Although the initial amount of protein was higher than for the membrane extracted NanC, the final yield was smaller in this case as a lot of protein was not refolded properly.

In both, inclusion body refolded and membrane expressed, cases the protein eluted during gel filtration was analysed using protein identification mass spectrometry in order to control that it was NanC (realized by Dr. Paul Jenö, Biozentrum, Basel).

Prior crystallization, NanC was concentrated using Amicon filtering devices. Interestingly the membrane purified NanC could be concentrated in a 50kD cut-off Amicon whereas the refolded NanC, when in the same detergent, was found in the flow through when using this type of material. This is suggesting that the refolded NanC has a smaller molecular weight than NanC purified from the membrane and thereby that the latter one might be associated to LPS molecules. This was assessed using a silver stained SDS-PAGE revealing that the membrane purified NanC sample is containing an additional, low-molecular weight, band whereas the refolded sample does not (Fig. I.1.8). This additional band can't be protein as it was not stained on the Coomasie gels. However, it could correspond to LPS that can be stained on silver gels. Although this is specific to LPS, further assessing using for instance mass spectrometry is required.
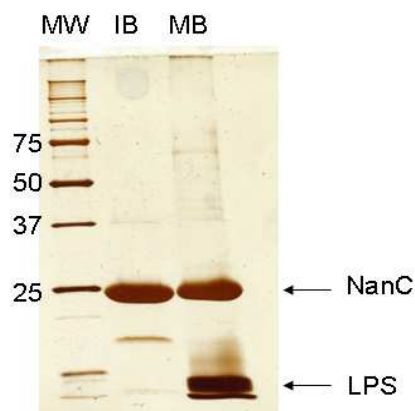
Figure I.1.8: Silver stained gels of NanC samples purified from the membrane and refolded from inclusion bodies.

The presence of LPS is confirmed in the membrane purified sample. MW stands for molecular weigh marker (in kDa), IB for inclusion bodies refolded NanC sample and MB for the membrane purified NanC protein.

I.1.2.3. Crystallization

Membrane extracted NanC

Initial crystallization trials were realized using the sparce matrix method (Jancarik & Kim, 1991). As the protein concentration and the detergent were considered as being critical parameters for crystallization, several protein concentrations as well as detergents were tried out during this initial screening. A first set of screens was done at 5 mg/ml protein concentration. These were leading to some needle shaped crystals in a few conditions when using LDAO as a detergent. These crystals were not looking suitable to be tested in diffraction experiments as they were clustered and very fine. Seeing that LDAO was yielding crystals it was decided to continue using this detergent. Further screens were realized using a 10 mg/ml protein concentration in 10 mM Tris pH 8.0, 150 mM NaCl and 0.05 % LDAO. This gave crystals in several conditions from the Classic Lite screen (Qiagen). The most promising was forming little hexagonal columns in 100 mM acetate pH 4.6, 10 % isopropanol and 200 mM $CaCl_2$ (Fig. I.1.9) These crystals were easily reproducible and improved in conditions with precipitant concentrations ranging from 100 to 200 mM $CaCl_2$ and 3 to 10 % isopropanol; also the crystal shape was often different. In many cases, the crystals grown in this condition were shaped as half-disks (Fig. I.1.9). These crystals had a typical size of about $50 \times 50 \times 50$ µm and were easily reproducible from one purification batch to the next.
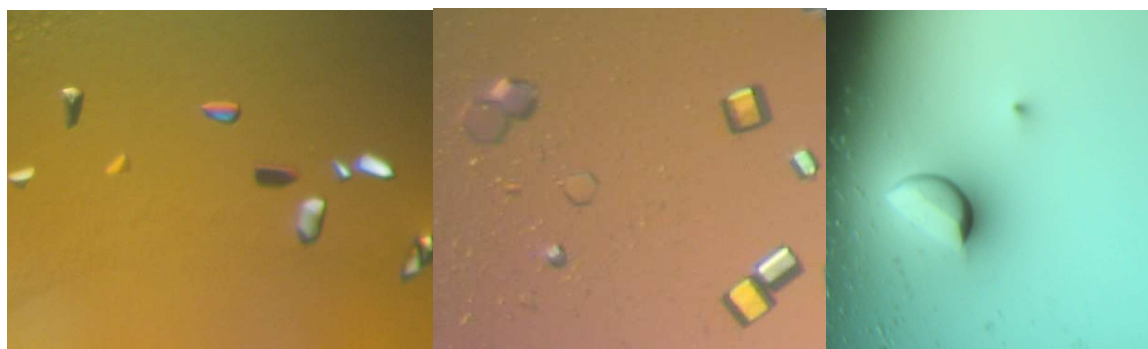
Figure I.1.9: Pictures of NanC crystals.

(left) Crystals obtained when refolding NanC in LDAO reaching a typical size of $50 \times 50 \times 50$ µm. *(middle and right)* Crystals produced with NanC purified from the membrane. The hexagonal crystals *(middle)* have a size of about $50 \times 50 \times 50$ µm and the half-disks *(right)* have a diameter of about 100 µm and a thickness of 30 µm.

Inclusion bodies refolded NanC

For the refolded NanC, crystals were obtained using protein at a concentration of 15 mg/ml in 10 mM Tris pH 8.0, 0.14 % Fos12. Here no screening was required because the protein was crystallizing in similar conditions than NanC refolded using SDS (see paragraph: I.1.2.1). However, the reservoir solution allowing these crystals to grow was slightly different with a higher PEG400 concentration (100 mM Hepes pH 7.5, 52% PEG 400). These crystals had a diamond shape and had an approximate size of about $50\times50\times50$ µm (Fig. I.1.9) and were very difficult to reproduce, as they could only been grown when using one given batch of refolded protein. When decreasing the PEG400 concentration, with the inclusion bodies refolded NanC, it was also possible to obtain the same, rod or plate shaped crystal form than previously observed with protein refolded in SDS (Fig. I.1.1).

Interestingly, it was impossible to obtain crystals with the membrane extracted protein using the condition leading to the crystals obtained from refolded NanC. Identically, refolded NanC did not yield any crystals when put in conditions were membrane extracted NanC did. This indicates that the two samples were not exactly identical, although the chromatograms seemed similar. Maybe the presence of LPS in the membrane extracted sample played an important role to form the hexagon shaped crystals observed with that protein.

I.1.2.4. Diffraction data collection on native protein crystals and processing

Prior freezing in liquid nitrogen, the crystals grown using membrane extracted NanC were quickly soaked in 15 to 20 % ethylene glycol or glycerol solutions, whereas the crystals grown from inclusion bodies don't need any additional cryoprotection as their condition was containing sufficient PEG400 concentration.

X-ray diffraction experiments were done at the SLS. The resolutions of the membrane purified NanC crystals and the inclusion bodies refolded NanC were going up to 2.0 Å and 1.8 Å respectively. Both crystal forms were presenting an isotropic diffraction pattern with well defined spots.

Data sets were collected using 1s exposure and 1° per image. The crystals did not suffer from radiation damage allowing complete data sets to be collected. Diffraction images were

indexed using MOSFLM (Leslie & Powell, 2007) and scaled with SCALA (Evans, 2006). The good data quality allowed fast processing. The crystals grown from refolded NanC are H3 whereas the ones grown from the membrane purified protein belong to space group $P6_322$. The data reduction statistics are satisfactory for both crystal forms and are shown in table I.1.10.

I.1.2.5. Heavy atom derivatization

Intensive molecular replacement trials were done using the two native data sets of NanC and several 12 and 14 β–stranded β-barrels such as OmpG (Subbarao & van den Berg, 2006; Yildiz et al, 2006), Tsx (Ye & van den Berg, 2004), NalP (Oomen et al, 2004) or FadL (van den Berg et al, 2004) in their full length model or in modified models in which only the β-barrel scaffold was used. Although these models strongly differ by their number of strands and their shape, none of these molecular replacement trials was successful and experimental phasing was necessary. Heavy atom derivatization was preferred to seleno-methionine methods because many crystals grown with membrane extracted NanC were readily available. Different heavy atom types as well as different compounds for each of them were used to increase the chance of success. Different soaking times, ranging from 10 min to 24 h, were also tried out as both quick and long soaks have proven to successfully allow derivatization (Boggon & Shapiro, 2000; Garman & Murray, 2003; Sun & Radaev, 2002; Sun et al, 2002). NanC doesn't contain cysteines, so mercury, the most popular heavy atom, could not be used efficiently. However, tests using ethyl-mercury were done because of its ability to bind to hydrophobic residues that are likely to be exposed at the surface of NanC. In the co-crystallization trials, the presence of heavy atoms either impaired the growth of crystals or the anomalous signal was too low when measured at the synchrotron. Most soaking tests led to cracks in the crystals. When tested at the SLS, these crystals were diffracting to a much lower resolution than the native ones and were also strongly suffering from radiation damage. However, two useful data sets showing clear incorporation of heavy atoms in the crystals according to the x-ray absorption scan, could be collected. They were recorded from two crystals soaked in stabilizing solution containing 20% ethylene glycol as cryoprotectant and 1mM K_2PtCl_4 and 10mM $SmCl_3$ during 22h and 10 min respectively. Statistics for these data sets can be found in table I.1.10.

The K_2PtCl_4 and the $SmCl_3$ derivatized crystals were used together with the high-resolution native data set in order to solve the phase problem. Due to low-resolution of the heavy atom data sets, the working resolution was at 4Å. 4 Pt and 2 Sm sites were found using SHELX. Initial phases could be calculated using MIRAS. After solvent flattening, the electron density map was clearly showing β-strands forming a barrel but neither side chains nor connections between the strands were visible.

In absence of non-crystallographic symmetry, density modification by NCS averaging was impossible. A shape similarity was observed between the NanC β-barrel form observed in the experimental electron density obtained from the $P6_322$ crystal form and the β-barrel domain of the NalP autotransporter (Oomen et al, 2004). A NalP poly-alanine β-barrel model, from which the loops and turns were removed, was fitted to the electron density obtained from the $P6_322$ crystal form in order to get a first rough model of the NanC β-barrel. This modified NalP model was used as molecular replacement model in order to find the position of the NanC barrel in the H3 crystal form with Phaser (Z-score of 6.02 and log-likelihood gain of 257). Knowing the position of the NanC molecules in the H3 crystal form enabled the usage of multi-crystal electron density averaging in order to increase the quality of the electron density map. The multi-crystal averaging procedure was carried out using program DMMulti (Cowtan & Main, 1993) and experimental phase information for the $P6_322$ crystal form as well as calculated phase information from the H3 crystal form. The phases were extended from 4.0 Å resolution to 3.2 Å in 250 averaging cycles. The multi-crystal averaging allowed a remarkable increase of the electron density map quality (Fig. I.1.11). After the averaging, side chains were identifiable leading to an unambiguously assignment of the sequence in the electron density.

Trigonal / Hexagonal

Crystal data	Trigonal Native	Hexagonal Native	Hexagonal K$_2$PtCl$_4$	Hexagonal SmCl$_3$
Origin of the protein sample	Inclusion bodies	Membrane	Membrane	
Space group	H3	P6$_3$22	P6$_3$22	
Unit cell parameters (Å, °)	74.9, 74.9, 126,4 / 90.0, 90.0, 120.0	96.8, 96.8, 121.2 / 90.0, 90.0, 120.0	97.5, 97.5, 123.5 / 90.0, 90.0, 120.0	99.15, 99.15, 122.07 / 90.0, 90.0, 120.0
Molecule / A.U.	1	1	1	1

Diffraction data	Trigonal Native	Hexagonal Native	Hexagonal K$_2$PtCl$_4$	Hexagonal SmCl$_3$
Wavelength (Å)	1.0000	1.0000	1.0719	1.8353
Resolution (Å)	45.0 - 1.8 (1.9 - 1.8)	50.0 - 2.0 (2.1 - 2.0)	50.0 - 4.0 (4.2 – 4.0)	50.0 - 3.5 (3.5 - 3.5)
Total number of reflections	80794 (8699)	227563 (32852)	37605 (3682)	26908 (2079)
Unique reflections	24521 (3603)	22890 (3247)	5183 (607)	7289 (553)
Rmeas (%)	8.6 (31.5)	8.9 (41.0)	8.1 (12.4)	13.7 (23.5)
Completeness (%)	99.9 (99.9)	98.6 (97.9)	92.9 (80.1)	85.8 (80.8)
Multiplicity	3.3 (2.4)	9.9 (10.1)	6.9 (7.1)	3.7 (3.8)
I/σ(I)	12.5 (2.3)	19.2 (5.7)	21.5 (17.6)	10.7 (7.9)

Phasing data	K$_2$PtCl$_4$	SmCl$_3$
MIRAS phasing power		
Isomorphous (centric/acentric)	0.378 / 0.441	0.304 / 0.282
Anomalous (acentric)	0.528	0.299
MIRAS FOM after SHARP (centric/acentric)	0.38 / 0.31	

Refinement statistics	Trigonal	Hexagonal
Resolution range (Å)	45 - 1.8	50.0 - 2.0
R/R$_{free,\ 5\%}$	19.5 / 23.0	19.2 / 21.4
Number of reflections used for refinement	23265 (1753)	21713 (1178)
R.m.s.d. bond length (Å)	0.011	0.011
R.m.s.d. bond angle (°)	1.32	1.33
Number of protein atoms	1722	1714
Number of solvent/detergent atoms	149 / 32	138 / 136

Table I.1.10: Diffraction, phasing and refinement data of the NanC crystals and structures.
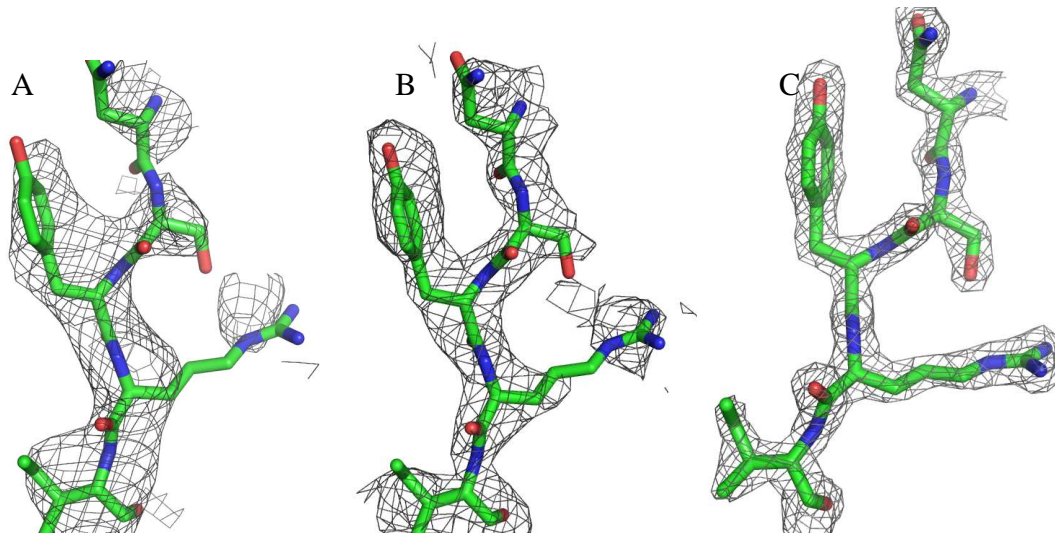
Figure I.1.11: Electron density maps of the hexagonal crystal form at several phasing steps showing the increasing quality of the map.

Electron density map represented at 1.5 σ after (A) the MIRAS phasing and solvent flattening (4.0 Å resolution), (B) the multi-crystal electron density averaging (3.8 Å resolution) and (C) the final refinement step (2.0 Å resolution). The structure model depicted is the final model.

The NanC model was then built in both electron density maps using COOT (Emsley & Cowtan, 2004) and REFMAC (Murshudov et al, 1997). The final R and Rfree values are respectively 19.5 and 23.0 for the H3 crystals and 19.2 and 21.4 for the P6$_3$22 crystals (Fig. I.2.10). Finally the quality of both NanC models were checked using MOLPROBITY (Davis et al, 2007) and deposited in the PDB with entries 2WJQ and 2WJR for the hexagonal and rhombohedral crystal forms respectively. 2WJQ showed 97% and 0.5% of the residues in the most favourable and outlier regions of the Ramachandran plot. No residues were showing bad rotamers. For 2WJR, 97% and 0.5% of the residues are in the most favourable and outlier regions of the Ramachandran plot, but only 1 residue (0.6%) has a bad rotamer.

NanC is folded as a hollow β-barrel composed of 12 anti-parallel strands. Long loops are connecting the β-strands on the extracellular side whereas short turns link them on the periplasmic side. NanC has an approximate height of 28 Å and its cross section is ellipsoidal (20 × 26 Å). The structure and its biological implications will be further discussed in section I.2, and in Wirth et al, 2009, also attached to this thesis.

I.1.2.6. Crystal contact analysis

The crystal contacts in the H3 and the $P6_322$ crystal forms have been analysed using the PISA web served of the EBI (Krissinel & Henrick, 2007).

H3 crystal form

In the H3 crystals, there is one NanC molecule in the asymmetric unit. Its crystal contacts are listed in table I.1.12. Along the c axis, NanC molecules are interacting head to tail, the extracellular loops of one molecule being in contact with the periplasmic turns of the other (Fig. I.1.13). The extracellular loop in this contact is almost entering the pore of the neighbour protein forming a large and tight contact surface (Table I.1.12., interface 1). The repetition of this contact along the c-axis creates continuous column-like arrangements of NanC monomers. Another, smaller interface (interface 3) allows contact between one NanC molecule and another one located in the next row of the neighbouring column-like arrangement of NanC monomers (Fig. I.1.13). In the a-b-plane, NanC molecules are forming very loose trimeric arrangements located on the crystallographic 3-fold axis and interacting *via* the base of the β-barrel in a V shape manner (Fig. I.1.13). This interface (Table I.1.12., interface 2) is almost exclusively composed of hydrophobic residues and no hydrogen bound or salt bridges could be detected.

Interface	Residues	Symmetry operations	Symmetry related residues	Interface area (Å²)
1	K111*, R114*, Q115, Q116***, D117, L118*, S119*, G120*, D121, M122, D125, K153, D156, Y157, R^{s}158, H162, K163, K164, D200	-y-1/3, x-y-2/3, z+1/3	A2, S23, E24^s, G25*, W26*, Q27*, G29, W30*, W31, A62, K64, D66, D67*, T70, R72*, D95, H137, K214*, A215	652.7
2	W94, A96, T97, L100, F102, V134, Y136, I138, W146,	-y-1, x-y-1, z	I63, L65, D66, W69, V71, L92, W94, W130	315,3
3	R192, R199^{s**}, N201*, L202	-x+y-1/3, -x-2/3, z+1/3	S140, D141^s, Y175, H176*, M177*, T178*, P179	168.5

Table I.1.12: List of residues involved in the H3 crystal interfaces.
^{s}: residues involved in a salt bridge. *, ** and ***: residues forming hydrogen bounds with one, two and three residues respectively.
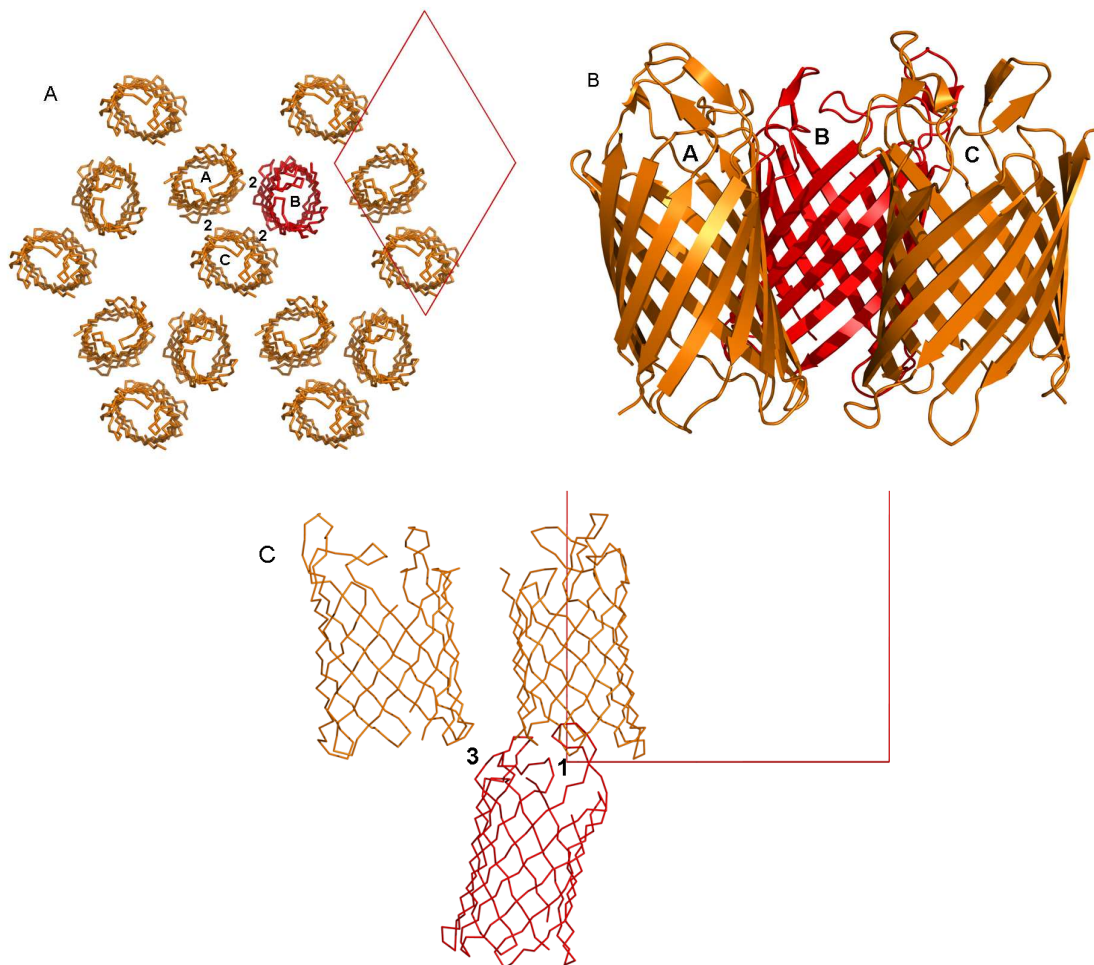
Figure I.1.13: H3 crystal packing

Interface numbers are referring to table I.2.12. (A) Packing in the a-b-plan showing very loose trimers (i.e. monomers A, B and C) along the 3-fold axis. In this plan the contacts are mediated only by the hydrophobic interface. (B) Close up view of the hydrophobic interaction zone within the loose NanC trimer (bottom). In this representation, the opened trimer is also visible. (C) Crystal contacts along the c axis presenting the two additional interfaces and in particular the large interface (number 1) showing the loop of one monomer almost entering the pore of the other monomer.

$P6_3 22$ crystal form

The crystal contact areas of the $P6_3 22$ crystal form are listed in Table I.1.14. This crystal form contains one monomer per asymmetric unit. The first striking feature of the packing is that NanC forms a tight trimeric arrangement located along the 3-fold symmetry axis generated in that hexagonal space group. This trimer, having no detergent molecules between the monomers, resembles to what could be expected if NanC would be a trimeric porin (Fig.

I.1.15). Within the trimer, the crystal contacts are mostly mediated by hydrophobic residues (Table I.1.14, interface 1) but hydrogen bounds are also present in order to allow that assembly.

In the a-b-plan, NanC trimers are forming 32-symmetry related arrangements resembling a honeycomb with trimers oriented in opposite directions (Fig. I.1.15). The crystal contacts between the trimers within the 6-fold related assemblies are mediated by two small interfaces (4 and 5 in Table I.1.14 and Fig. I.1.15) located on the 2-fold axis. Interestingly, such honeycomb like assembly is also present in several crystal structures of OmpF, the general porin of *E. coli* (see also Chapter II). Along the c axis, each NanC trimers is forming a tight head-to-head contact with another trimer (Fig. I.1.15). Each monomer of NanC forms a contact *via* its periplasmic loops with the loops of another one located on the top of it. This crystal contact (interface 2 in table I.1.14) involves several residues from different loops around the barrel. Each monomer also forms contacts with another monomer form the trimer on the top of it (interface 2 in Fig. I.1.15). All together, NanC forms tight contacts with its neighbours within the trimer and with another trimer forming hexameric assemblies (Fig. I.1.15).

Interface	Residues	Symmetry operation	Symmetry related molecules	Interface area (Å²)
1	A1, V22, E24, W26*, N28, W30, W31, A32, M34, V55, V57, V59, T77, F79, S80, S81***, N82, G83, T84, Y86*, L215	-y, x-y-1, z	N28*, W30, Y61, I63, K64, L65, V71, P73, G74, M75, Y86*, P88, Y89, V90, Y108, W110, A112, R114*, R124**	640.2
2	S81, N82, Q116, S119, G120, H162, N197, G198*, R199, D200*, N201	-y, -x, -z+1/2	S81, N82, Q116, S119, G120, H162, N197, G198*, R199, D200*, N201	303.3
3	R114*, Q116, D117, G120*, D121*, M122*, S123*, R124, N155	-x+y+1, y, -z+1/2	R114*, Q116, D117, G120*, D121*, M122*, S123*, R124, N155	233.0
4	Y16, Y175, M177, T178	-x, -x+y, -z+1	Y16, Y175, M177, T178	103.3
5	W94, T97, W146, L150, T165	x-y, -y, -z+1	W94, T97, W146, L150, T165	85.8

Table I.1.14: List of residues involved in the P6₃22 crystal contacts.
*, ** and *** is for residues forming one, two and three hydrogen bounds respectively.

The interface allowing the trimer formation is large (640 Å²) and is detected by PISA as being a potential biologically relevant protein-protein contact. However, numerous data, such as electrophysiology and cross-linking (Blot et al, 2002) as well as electron microscopy (Signorell et al, 2007) have shown that NanC is monomeric. The trimer is not observed in the H3 crystals but the gel filtration elution profile of both membrane extracted and inclusion bodies refolded proteins are similar. Therefore we can conclude that the trimer observed in the P6$_3$22 is not biologically relevant although it could be a reminiscence of a trimeric ancestor of NanC in a similar way than found in the general (or specific) classical trimeric porins such as OmpF, OmpC, or LamB (Basle et al, 2006; Cowan et al, 1992; Schirmer et al, 1995)
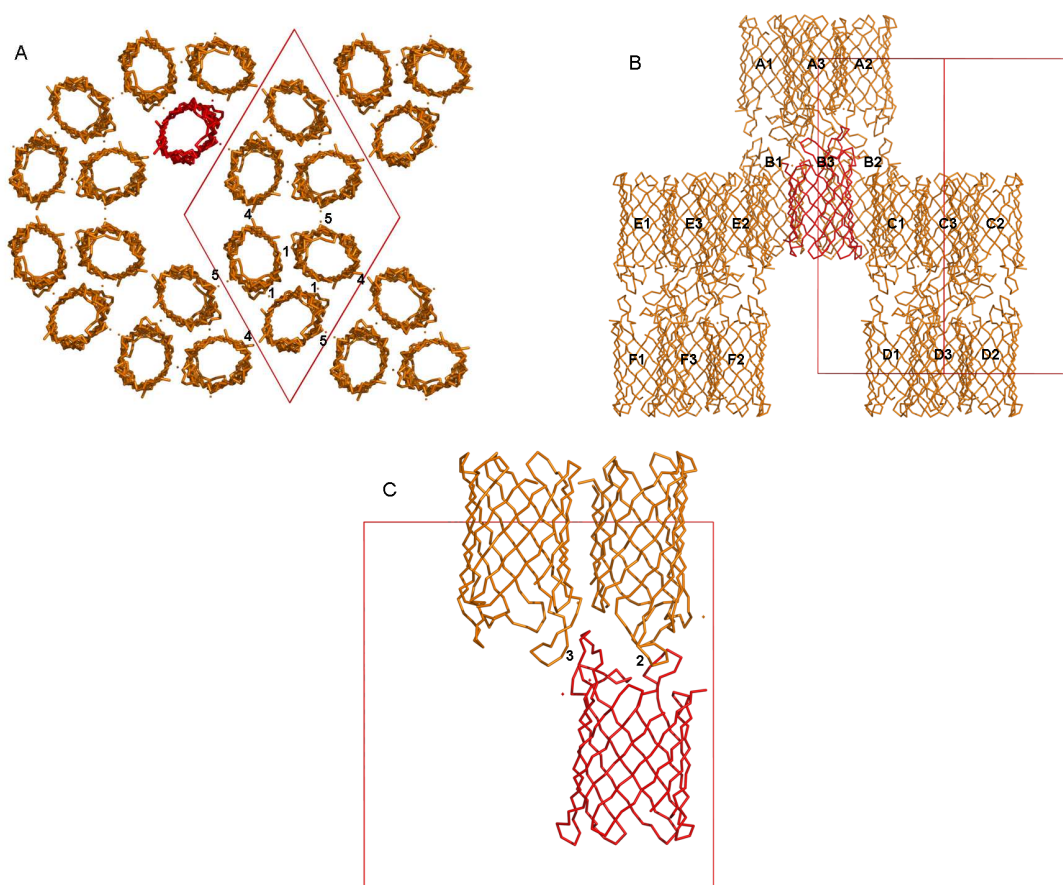
Figure I.1.15: crystal packing in P6$_3$22

(A) Representation in the a-b-plan. NanC molecules are forming loose trimers related by the 3-fold axis generated in the P6$_3$22 space group. These loose trimers are forming hexameric structures of trimers around the 6-fold axis of the space group. Interface numbers are referring to the number in Table I.1.14. (B) Crystal packing along the c-axis. The NanC loose trimers are interacting to form hexameric assemblies. (C) Close-up representation of the interactions along the c axis. Each monomer of NanC is interacting with two monomers of the neighbour trimer. Interface numbers are referring to the numbers in table I.1.14.

I.1.3. Conclusions

This chapter presented the purification of NanC protein expressed in inclusion bodies and in membrane using ion exchange chromatography. These protocols allowed obtaining high amounts of pure protein and subsequent crystallization of NanC. Two crystallization conditions, leading to well diffracting crystals, could be obtained when using NanC purified from the membrane or from inclusion bodies after refolding using LDAO. X-ray diffraction data were collected at the SLS using crystals obtained from both conditions leading to native data sets at 1.8 and 2 Å for crystals obtained from refolded and membrane purified proteins respectively. Using the numerous crystals available heavy atom derivatization was undertaken and anomalous data sets could be measured using platinum and samarium. Initial phases were calculated using the MIRAS method and the two derivatives and the native data sets. The electron density quality has been strongly improved thanks to multi-crystal density averaging allowing to solve the structure of NanC and to give the first high resolution structural information about a member of the KdgM porin family. The crystals obtained here are formed by two very different packing but in both cases the extracellular loops of the porin are playing a critical role for the formation of tight contacts. In the $P6_322$ crystal form, a symmetry related trimer can be observed but all other data, as well as the second crystal packing confirms that NanC is a monomeric protein *in vivo*. The structure and its biological relevance are discussed in section I.2. and in Wirth et al. 2009.

Chapter I.2

I.2.1. Research article:

NanC crystal structure, a Model for outer-membrane channels of the acidic sugar-specific KdgM porin family

Christophe Wirth, Guy Condemine, Céline Boiteux, Simon Bernèche, Tilman Schirmer and Caroline M. Peneff.

Note:

Christophe Wirth purified the protein from the membrane, obtained the hexagonal crystal form, processed the diffraction data, built and refined the NanC structures. Caroline Peneff designed the inclusion body purification protocol, obtained the rhombohedral crystal form and phased the crystallographic data. Guy Condemine made the molecular biology work and expressed the protein. Céline Boiteux and Simon Bernèche made the ligand binding modelling. Tilman Schirmer helped for phase improvement by multi-crystal averaging and for funding. The paper was mostly written by Caroline Peneff, with help from Christophe Wirth and Tilman Schirmer. The figures were realized by mostly Christophe Wirth with help of Caroline Peneff.

JMB

NanC Crystal Structure, a Model for Outer-Membrane Channels of the Acidic Sugar-Specific KdgM Porin Family

Christophe Wirth[1], Guy Condemine[2], Céline Boiteux[3], Simon Bernèche[3], Tilman Schirmer[1] and Caroline M. Peneff[1]*

[1]*Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

[2]*Université de Lyon, F-69003; Université Lyon 1, F-69622; INSA-Lyon, Villeurbanne, F-69621; CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, France*

[3]*Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

Sialic acids are acidic sugars present mostly on vertebrate cell surfaces, which can be metabolized by bacteria and act as an inflammation signal. *N*-Acetylneuraminic acid, the most abundant sialic acid, can enter into *Escherichia coli* K12 through NanC, an *N*-acetylneuraminic acid-inducible outer-membrane channel. With its 215 residues, NanC belongs to the family of small monomeric KdgM-related porins. KdgM homologues are found in gammaproteobacteria, including major plant and human pathogens, and together they define a large family of putative acidic sugar/oligosaccharide transporters, which are as yet poorly characterized. Here, we present the first high-resolution structure of a KdgM family member. NanC folds into a 28-Å-high, 12-stranded β-barrel, resembling the β-domain of autotransporter NalP and defining an open pore with an average radius of 3.3 Å. The channel is lined by two strings of basic residues facing each other across the pore, a feature that appears largely conserved within the KdgM family and is likely to facilitate the diffusion of acidic oligosaccharides.

© 2009 Elsevier Ltd. All rights reserved.

Introduction

Membranes constitute one of the most fundamental components of all living organisms, acting as protective barriers separating the cell from its environment and allowing it to carry out specific functions. Proteins that fold into channels are embedded in this lipid bilayer and are essential for

maintaining its selective permeability. In bacterial outer membranes, these proteins have been classified into three distinct categories: the general/nonspecific porins, solute-specific channels (also referred to as specific porins), and active transporters (for reviews, see Refs. 1 and 2). Proteins from the first group, exemplified by the well-characterized *Escherichia coli* OmpF porin, facilitate nonspecifically the diffusion of small solutes (molecular mass exclusion limit below 600 Da). The flux of molecules through these pores is driven by the concentration gradient, and transport is more efficient when the solute concentration difference between the extracellular and periplasmic space is high. For nutrients with a low external concentration, uptake occurs *via* specific channels that contain saturable binding sites for guided diffusion of molecules, or active transporters such as TonB-dependent FhuA.[3] The first outer-membrane passive transporters to be structurally

---

characterized included nonspecific porins from *Rhodobacter capsulatus*,[4] *Rhodobacter blasticus*,[5] *Paracoccus denitrificans*,[6] the *E. coli* general porins OmpF and PhoE,[7] the osmoporin (OmpK36) of *Klebsiella pneumoniae*,[8] and *Comamonas acidovorans* Omp32,[9] as well as the sucrose-specific *Salmonella typhimurium* ScrY channel,[10] the maltose-specific LamB porin of *E. coli*,[11] and *S. typhimurium*.[12] These proteins, most commonly referred to as classical porins, all form highly stable trimers with each subunit folded into a hollow barrel composed of 16 or 18 transmembrane β-strands. Recently, more structures have been reported, including that of the *Pseudomonas aeruginosa* specific channels OprD,[13] OprK,[14] and OprP;[15] the *E. coli* outer-membrane proteins OmpC,[16] Tsx (nucleoside-specific),[17] and OmpG (a putative oligosaccharide-specific channel);[18,19] and proteins involved in hydrophobic compound uptake: FadL from *E. coli*,[20] TodX from *Pseudomonas putida*, and TbuX from *Ralstonia pickettii*.[21] Interestingly, the latter five have defined a new structural group with features different from those of the classical porin family. In particular, these channels are monomeric and form smaller barrels, typically composed of 12 or 14 β-strands. According to transmembrane β-strand predictions, cysteine-labeling experiments, and additional biochemical characterization, the family of KdgM-related outer-membrane channels belongs to this class of recently characterized porins.[22]

KdgM is a major, 25-kDa, monomeric outer-membrane protein of *Dickeya dadantii* (formerly *Erwinia chrysanthemi*), a pathogen responsible for the soft rot disease that affects numerous plants of agricultural importance and causes severe economical losses every year. During the course of infection, oligogalacturonate (GAn) produced by the plant cell wall degradation process can cross the outer membrane *via* KdgM, which is essential and specific for transport of the largest molecules (GAn, *n*>3) that cannot pass through OmpF/OmpC. Inactivation of the KdgM gene impairs the virulence of the bacteria.[23] KdgM homologues have been identified in other gammaproteobacteria. Together, they define a large family of more than 100 members distributed in at least 50 species including major human and plant pathogens such as *Yersinia pestis*, *Yersinia enterocolitica*, *S. typhimurium*, *Pseudomonas*, *Vibrio*, *Shigella* spp., and uropathogenic and food poisoning strains of *E. coli*. These porins are the smallest currently characterized, with an average size for the mature protein of 220 amino acids. Present knowledge about this channel family is scarce and mostly relates to genomic information such as analysis of genetic context, which often corresponds to pathogenicity islands, and the possible modes of gene expression regulation.[24] In Enterobacteriaceae, typically two and up to four homologous sequences are found, underlying the biological advantage for the bacteria to maintain several paralogues. The reason for such redundancy is not clear, but initial data on KdgN, a close homologue of KdgM in *D. dadantii*, revealed differences in their mode of regulation.[25]

NanC, a KdgM orthologue identified in *Shigella* spp. and uncapsulated *E. coli* strains, has been further characterized and shown to allow translocation of *N*-acetylneuraminic acid (Neu5Ac), the most abundant of sialic acids.[26] Sialic acid refers to a family of nine-carbon, negatively charged sugar molecules prominently found at the terminal position of higher vertebrates' cell surface glycoconjugates but also present in some, mostly pathogenic, bacteria.[27] NanC-expressing strains of *E. coli* can use sialic acids as a carbon- and nitrogen-rich nutrient.[28] Recent studies also suggest a signaling role in balancing the host–pathogen relationship.[29,30] Neu5Ac strongly induces the expression of NanC, which is required for bacterial growth on Neu5Ac as the sole carbon source, in the absence of OmpF and OmpC.[26] Electrophysiological experiments with planar lipid bilayers showed that NanC forms high-conductance channels (typically 450pS) exhibiting a weak selectivity for anions,[26] similar to the results obtained with KdgM and KdgN.[23,25] However, the presence of a saturable binding site for Neu5Ac could not be shown, as the presence of up to 50 mM Neu5Ac did not affect ion conductivity, while, in similar experiments, blockage of the KdgM channel activity was observed when trigalacturonate was added.[23] Structural information for the KdgM family is limited to (i) a topology model of KdgM based on cysteine-labeling experimental data, which suggests a 14-stranded β-barrel fold,[22] and (ii) a 7-Å projection structure, obtained by 2D electron crystallography, which confirmed the monomeric β-barrel fold and estimated the dimensions of its elliptical cross section to about 24 Å×17 Å.[31]

In order to gain more insight into the structure and function of the KdgM family, we have determined the crystal structure of NanC at 1.8 Å. The protein folds into a barrel composed of 12-antiparallel β-strands. Inspection of the pore lining reveals the presence of aligned basic residues defining two positively charged tracks likely to assist solute translocation.

Results and Discussion

Overall structure

NanC was crystallized in two different crystal forms (Table 1). Hexagonal crystals were obtained using protein material extracted from the outer membrane (the protein was expressed with its signal sequence), whereas trigonal crystals were grown from refolded material purified from inclusion bodies (the signal sequence was removed). Both forms contain one molecule in the asymmetric unit. Heavy-atom derivatization of the hexagonal crystals with K_2PtCl_4 and $SmCl_3$ led to initial multiple isomorphous replacement with anomalous scattering (MIRAS) phase information. A preliminary model could be fitted into the electron density map and was subsequently used to solve the structure of the trigonal crystal form by molecular

Table 1. Data collection and refinement statistics

	Hexagonal			Trigonal
	K_2PtCl_4	$SmCl_3$	Native	Native
Data collection[a]				
Space group	$P6_322$	$P6_322$	$P6_322$	R3:H (H3)
Wavelength (Å)	1.0719	1.8353	1.0000	1.0000
Resolution limits (Å)				
Overall	50.0–4.0	50.0–3.5	50.0–2.0	45.0–1.8
Outer shell	4.2–4.0	3.6–3.5	2.1–2.0	1.9–1.8
Cell dimensions				
a, b, c (Å)	97.5, 97.5, 123.5	99.1, 99.1, 122.1	96.8, 96.8, 121.2	74.9, 74.9, 126.4
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
R_{meas}	8.1 (12.4)	13.7 (23.5)	8.9 (41.0)	8.6 (31.5)
$I/\sigma(I)$	21.5 (17.6)	10.7 (7.9)	19.2 (5.7)	12.5 (2.3)
Number of unique reflections	5183 (607)	7289 (553)	22,890 (3247)	24,521 (3603)
Multiplicity	6.9 (7.1)	3.7 (3.8)	9.9 (10.1)	3.3 (2.4)
Completeness (%)	92.9 (80.1)	85.8 (80.8)	98.6 (97.9)	99.9 (99.9)
MIRAS phasing power				
Iso (acentric/centric)	0.38/0.44	0.30/0.28		
Ano (acentric)	0.528	0.299		
Refinement statistics				
Resolution range (Å)			50.0–2.0	50.0–1.8
R/R_{free} (%)			19.2/21.4	19.5/23.0
Total number of reflections[b]			21,713 (1178)	23,265 (1753)
rmsd bond lengths (Å)			0.011	0.011
rmsd bond angles (°)			1.33	1.32
Total number of atoms (protein/detergent/solvent)			1714/136/138	1722/32/149

  [a] Numbers in parentheses refer to values for the highest-resolution (outer) shell.
  [b] Values in parentheses refer to the numbers of reflections used in the random set.

replacement. Multicrystal electron density averaging between the two forms was then applied, leading to an electron density map suitable for final model building. The hexagonal and trigonal crystal structures were refined independently to final R/R_{free} values of 19.2%/21.4% and 19.5%/23.0%, respectively (see Table 1 for complete statistics). The two models are very similar with an average rmsd of 0.68 Å for all C^{α} atoms. The complete amino acid sequence could be traced except for a 10-residue segment (His43-Leu52) located in the extracellular loop 2, an indication of the structural flexibility of this part of the molecule. An overview of the NanC structure and a schematic topology model are shown in Fig. 1.

Although the NanC molecules form a trimer in the hexagonal crystal (related by the crystallographic 3-fold symmetry), the small number of interactions between the monomers indicates a crystallization-driven association. This is consistent with the monomeric state of NanC observed in the trigonal crystal and deduced, in solution, from gel-filtration experiments (elution volumes for the membrane-extracted and refolded NanC protein are similar) as well as previously reported biochemical and biophysical data.[23,26,31] NanC folds into a hollow, 28-Å-high β-barrel, with ellipsoidal cross section of about 26 Å × 20 Å (Fig. 1b and c). The barrel is composed of 12 antiparallel β-strands and not 14 as has been suggested for its homologue KdgM.[22] The average angle between the β-strands and the barrel axis is 45°, and the shear number is 14.[32] The discrepancy between the crystal structure of NanC and the previously proposed topology model of KdgM is

explained mostly by the fact that KdgM Ser72 and Asn100 were predicted to lie in an extracellular loop by the cysteine-labeling experiments (cysteines at positions 72 and 100 could be labeled using intact cells). The structure, however, shows that the equivalent residues in NanC are found on the periplasmic side (turns T3 and T4). Since it is unlikely that KdgM and NanC would fold very differently, the reason for the misleading cysteine-labeling result is not clear and shows that such experimental data should be interpreted with caution.

The structure of NanC displays a number of structural features typical of outer-membrane channel proteins. In particular, the loops connecting the β-strands are long on the extracellular side, with the exception of L1 and L3, whereas short turns (T1–T5) are found facing the periplasm. Two lines of aromatic residues, separated by about 24 Å and also referred to as aromatic girdles, are encircling the barrel, delimiting the hydrophobic outer surface of NanC that would be embedded in the outer membrane *in vivo* (Fig. 1b). No loop is occluding the channel (Fig. 1c) in either of the two structures, which have been determined at different pH (4.6 and 7.8). This, together with the different crystal packing, argues against a pH-gating mechanism such as that reported for OmpG.[19]

The 12-stranded barrel structure and monomeric assembly directly relate NanC to the recently determined structures of Tsx[17] and OmpG.[18,19] However, the OmpG barrel has 14 β-strands with a wider and more circular geometry than NanC, and the 12-stranded barrel of Tsx appears significantly
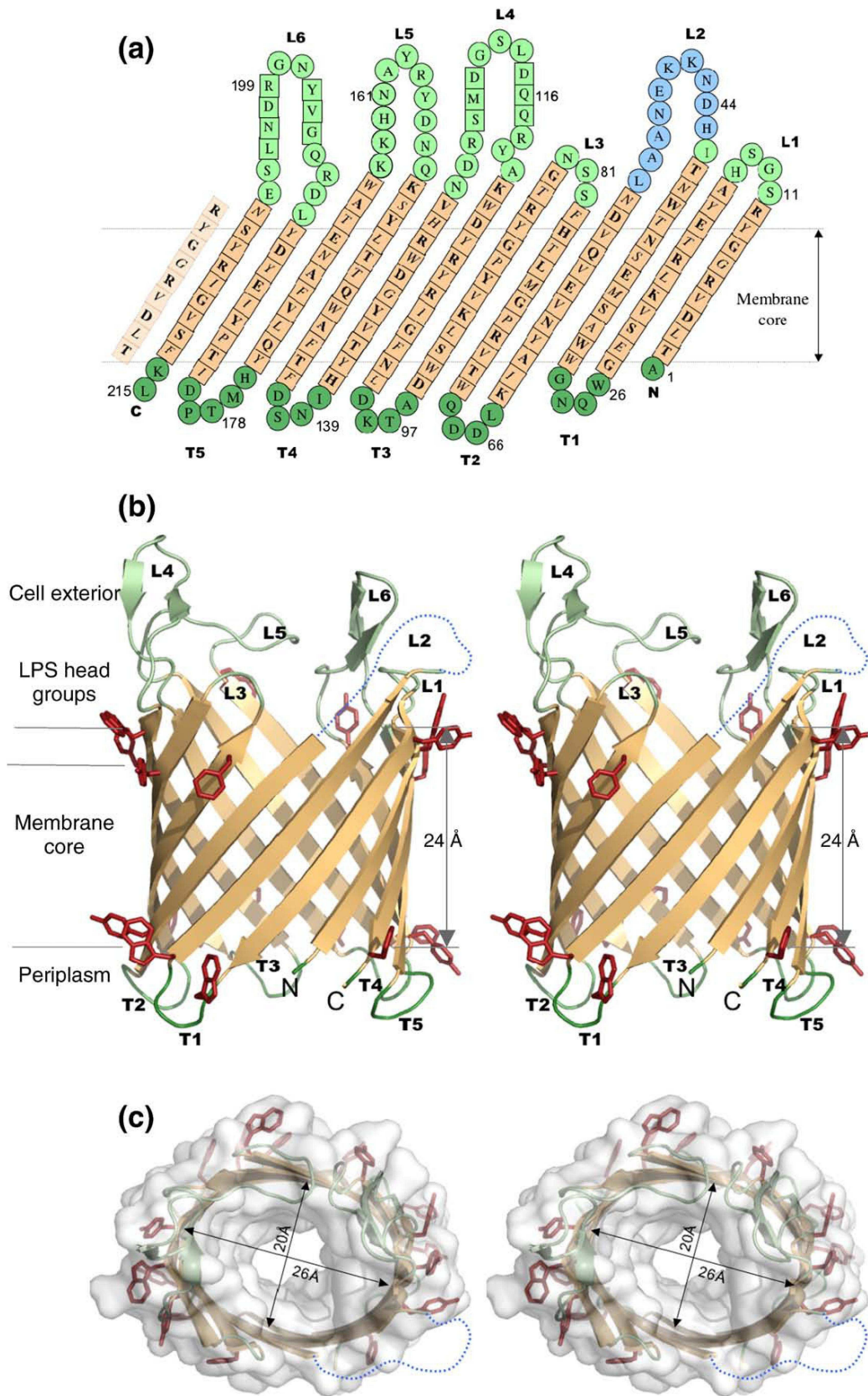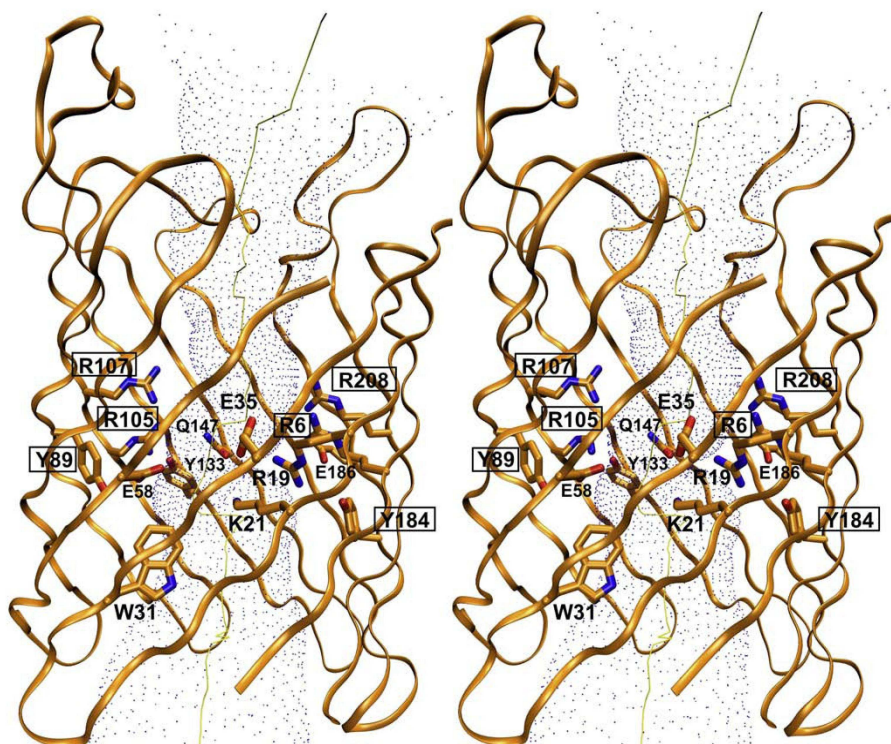
Fig. 1 *(legend on next page)*

Fig. 2. NanC channel architecture. Stereo view of a NanC ribbon model, together with the molecular surface of the pore generated in dot representation by the program HOLE.[34] Side chains constricting the channel at its center and all aromatic residues lining the channel are shown as sticks and labeled. Residues with boxed labels are highly conserved within the family.

more ellipsoidal. Unexpectedly, the closest structural homologues of NanC are not found among functionally related outer-membrane channels but in the group of bacterial autotransporters. Superimposition of the NanC structure onto that of the β-domain of autotransporter NalP[33] shows surprising similarities with an overall rmsd of 1.74 Å over all Cα atoms from transmembrane β-strands.

Transmembrane channel

Pore shape and dimensions were analyzed using the program HOLE.[34] The NanC channel resembles a hollow cylinder with an average radius of 3.3 Å. Maximal pore constriction occurs close to its center (~13 Å away from the periplasmic outlet, with a minimum radius of 2.9 Å) and results from the presence of bulky side chains extending from the β-

strands (Fig. 2). Although the size of the NanC channel at the most constricted region is similar to that of OmpC (2.9 Å) or PhoE (3.1 Å), the tubular pore geometry contrasts with the hourglass architecture observed in classical porins resulting from the presence of a constriction loop. In that respect, NanC pore structure is most related to that of OmpG or Tsx.

The pore is lined predominantly with charged and polar residues, a common feature of bacterial porins. In NanC, a striking predominance of positive charges is clearly visible with the regular arrangement of basic amino acids facing each other across the pore. On one side of the channel, five arginines (Arg85, Arg107, Arg129, Arg105, and Arg72) and four lysines (Lys111, Lys153, Lys91, and Lys64) form a continuous electropositive surface (track 1), which spans the entire channel from the extracellular

Fig. 1. NanC porin structure. Residues composing the transmembrane β-strands are shown in beige. Extracellular loops (L1–L6) and periplasmic turns (T1–T6) are labeled and displayed in light and dark green, respectively. Blue residues or broken lines correspond to the missing part of the structure. N and C designate the termini of the polypeptide chain. (a) Topology diagram of NanC. The first β-strand is repeated to indicate interactions closing up the barrel. Circled residues belong to loop regions and squared residues belong to β-strands. Residues indicated in bold have their side chain pointing towards the pore interior and those in italic have their side chain pointing towards the membrane bilayer. Residue numbers correspond to the mature sequence. (b and c) Stereo cartoon representation of the β-barrel viewed from the membrane (b) and from the cell exterior together with its molecular surface (c). Horizontal lines indicate the approximate boundaries of the outer-membrane hydrophobic core and lipopolysaccharide (LPS) head groups. The side chains of residues forming the aromatic girdles are shown as red sticks.

vestibule to the periplasmic outlet (Fig. 3a). On the opposite side, four arginines and one lysine compose a second conspicuous succession of positively charged amino acids (track 2) connecting the extracellular face to the narrowest point of the channel constriction (Fig. 3b): Arg10 is found at the

exoplasmic tip of the β-barrel inside which Arg208, Arg6, Arg19, and Lys21 side chains are stacked on top of each other, defining an arc down one side of the pore. Similarly aligned basic residues, also referred to as basic "ladder" or "slide", have been described in OprP,[15] OprD,[13] and Omp32,[9] but to the
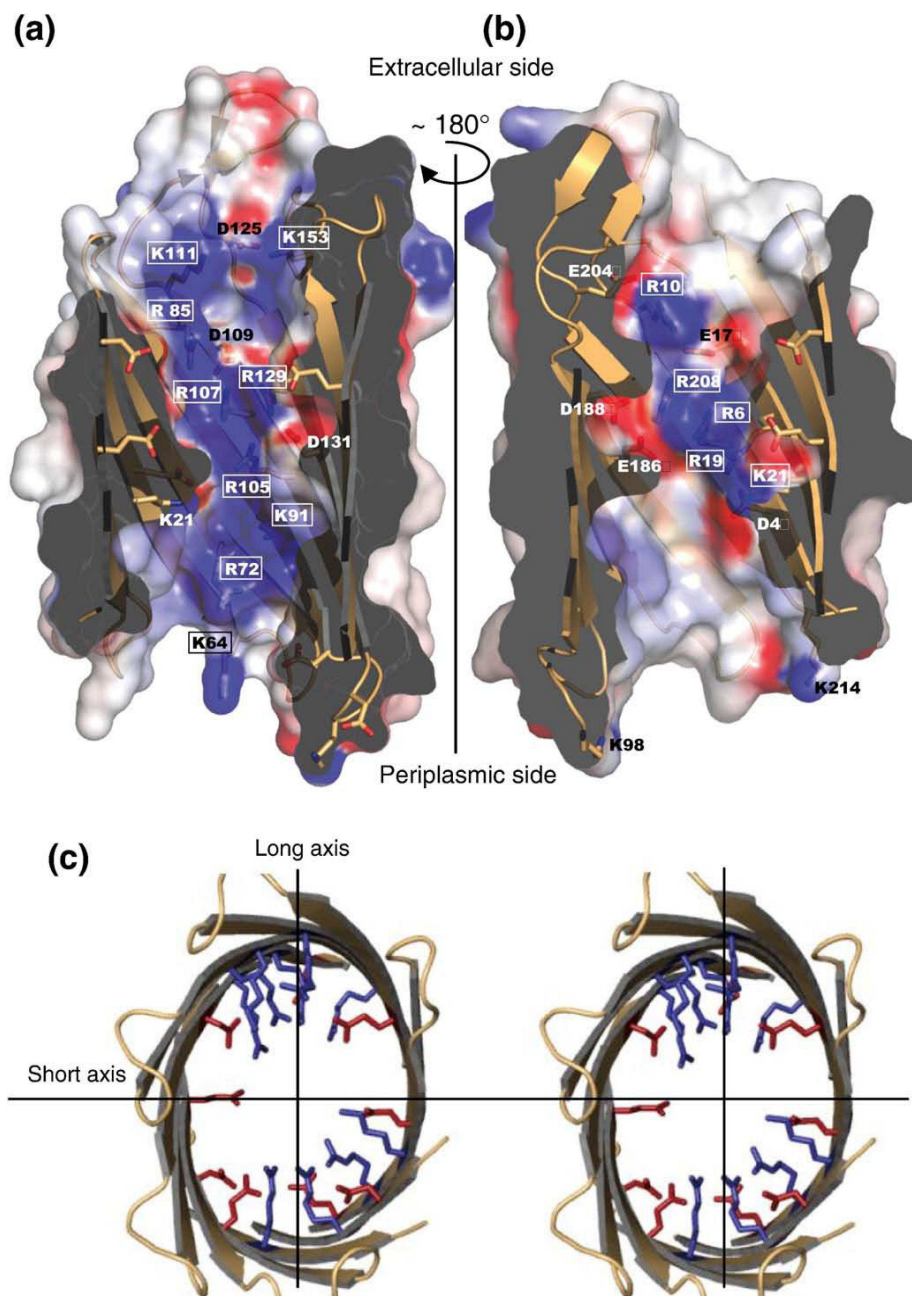
Fig. 3. Electrostatic properties. (a and b) Cut-open views of the channel are displayed with electrostatic potential mapped onto the molecular surface of the pore. Behind the semitransparent surface, charged side chains are shown in stick representation (carbon in light orange, nitrogen in blue, and oxygen in red). Boxed labels indicate basic residues from track 1 (a) and track 2 (b). Acidic residues in interaction with the tracks are also labeled. (c) Stereo representation of NanC porin viewed along the barrel axis from the cell exterior (for clarity, parts of the extracellular loops have been cut away). Charged side chains (blue, positive; red, negative) extending towards the pore interior are arranged as in a quadrupole.

best of our knowledge, this is the first time that two separate electropositive tracks are observed.

Positive charges within the pore are partly compensated by the presence of acidic residues. In particular, Asp154, Asp132, and Asp148 participate in direct electrostatic and H-bond interactions with basic residues from track 1 and similar interactions connect Glu204, Glu17, Asp188, Glu186, and Asp4 to Arg10, Arg19, and Arg208 from track 2 (Fig. 3a and b). Interestingly, the positioning of charged side chains, in particular at the channel center, suggests an electrostatic quadrupole (Fig. 3c), similar to that observed in the OmpG pore.[19] Again, this is different to the general porins OmpF, OmpC, or PhoE in which a dipolar arrangement of charged groups at the constriction zone has been described, a feature likely to facilitate the translocation of charged solutes while preventing that of apolar molecules.[35]

Specific solute transport

The NanC structure is consistent with a role in Neu5Ac transport, proposed on the basis of its genetic context, transcriptional upregulation in the presence of Neu5Ac, and complementation studies.[26] The channel appears large enough to accommodate Neu5Ac, and the preponderance of basic residues within the pore would facilitate transport of the negatively charged Neu5Ac molecule, in agreement with the reported anionic selectivity.[26] To understand the molecular mechanism of transport, we have attempted to determine the structure of NanC in complex with a monomer or homo-oligomer of Neu5Ac. So far, our efforts have been unsuccessful, possibly because of one or several of the following reasons. The absence of high-affinity binding sites for Neu5Ac, in particular under crystallization conditions, may prevent formation of the complex. Since trigonal crystals, growing at neutral pH, could not be reproduced, cocrystallization was only performed in the conditions of the hexagonal crystal form, that is, at pH 4.6. Unfortunately, at this acidic pH value, the Neu5Ac polymer is likely to be instable.[36] Soaking experiments were performed in various pH conditions and may have failed because of the non-accessibility of the channel in the crystal. One way to overcome the problem would be to subject NanC, together with its ligand, to a new broad crystallization screen for the identification of conditions favoring the complex formation and stability.

Nevertheless, a detailed analysis of the pore interior and comparison with structural and functional data from related specific outer-membrane channels can provide some indications as to the mode of solute translocation. First, aromatic side chains are often found in stacking interactions with sugar rings, and their contribution to the binding and transport of sugar molecules or related compounds has been described in the case of LamB,[37] ScrY,[10] and Tsx.[17] In NanC, four aromatic amino acids (Trp31, Tyr89, Tyr133, and Tyr184) are present inside the pore, all located close to the most constricted region (Fig. 2). In particular, aromatic

side chains at positions 89 and 184 are remarkably well conserved among enterobacterial members of the KdgM family, supporting a functional role for these residues. Second, and most importantly, we expect that the strong electropositive potential inside the pore constitutes the major driving force for translocation. The ligand carboxylate group could successively bind to the aligned Arg and Lys residues that pave the way from the extracellular space to the narrowest point of the channel, in a manner reminiscent to the proposed mechanism of phosphate translocation *via* OprP.[15] The presence of three lysines (at positions 98, 64, and 214) located at the periplasmic end of the pore, together with Arg72, may subsequently guide the solute towards the exit (Fig. 3a and b). However, what is unique to NanC is the presence of two electropositive tracks and it is not immediately clear how both tracks could participate in binding the Neu5Ac sole carboxylate group. One possible explanation is that the preferred ligand of NanC possesses several negative charges. Since NanC is closely related to GAn-specific KdgM, it is possible that, similarly, NanC transports preferentially oligomers of sialic acid. Computational simulations of the ligand binding and translocation have been performed using the program CHARMM.[38] The result shows that oligosialic acid could indeed orient the carboxylate groups of consecutive Neu5Ac units in alternating directions, thus fully taking advantage of the two tracks for guided diffusion through the pore (Fig. 4).

Unfortunately, the effect of precise and appropriate concentrations of oligomers of Neu5Ac on the channel conductance of NanC has not been assayed. Colominic acid, which is a mixture of homopolymers of Neu5Ac of different length, was tested in electrophysiological experiments and did not affect ion conductivity of the refolded NanC channel.[26] The refolding process reported in this study was performed in the presence of the negatively charged detergent SDS, which may have modified the ligand binding properties by interacting with the basic residues. Some of the tested polymer molecules may have been too large and the smaller ones not at a concentration high enough to induce blockage of the channel activity. In addition, the high salt concentration used in these experiments (0.8 to 1 M KCl) may not have been appropriate since it may have shielded considerably electrostatic interactions. Further investigation of the NanC channel conductivity in optimized experimental conditions is needed to clarify these points. Liposome swelling assays (i.e., the reconstitution of NanC into proteoliposomes and the measurement of solute diffusion rates derived from the osmotic swelling rates of the liposomes when placed in the presence of test solutes)[39] could also be performed to address this issue, although the use of this method for studying the diffusion of charged solutes has proven particularly difficult.[40]

Intestinal mucosal surfaces that are colonized by *E. coli* contain high levels of sialylated mucins, and
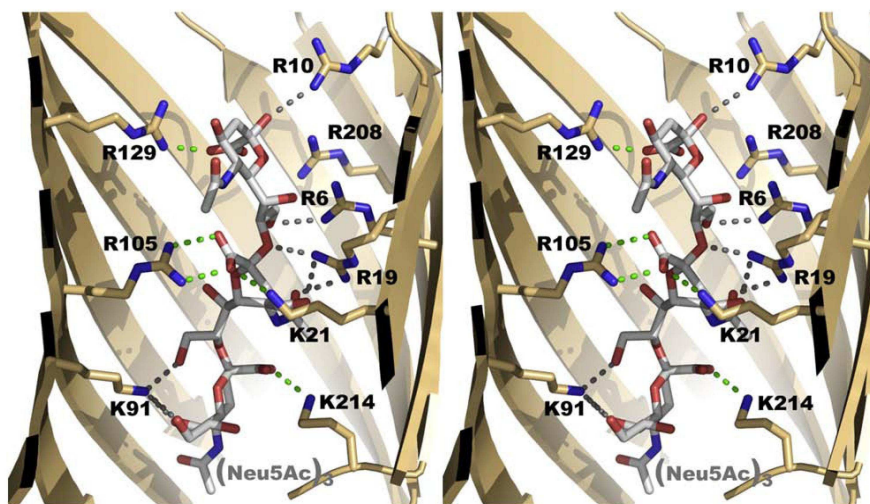
Fig. 4. Ligand binding model. Stereo view of a NanC ribbon model, together with a trimer of Neu5Ac, in one possible binding position suggested by computational simulations. The Neu5Ac trimer (white carbons) and basic residues from the pore linings (beige carbons) are displayed as sticks and labeled. Gray broken lines indicate possible H-bonds and green ones represent salt bridges between the sialic acid negatively charged carboxyl groups and residues from both track 1 and track 2.

monomers or oligomers of sialic acid can be made available to *E. coli* by the action of host sialidases, upregulated during the course of inflammation, or sialidases secreted by other bacteria present in the same niche. Moreover, in an environment where nutrient content is relatively low and competition is high, one would expect that the bacterium has acquired all possible ways to maximize specific nutrient uptake. However, in the absence of identified intracellular *E. coli* enzymes with sialidase activity, the reasons why the bacteria would have acquired a specific channel for translocation of sialic acid oligomers are unknown. *nanC* is the first gene of a three-gene operon, *nanC–yjhT–yjhS*. The recent characterization of YjhT/NanM,[41] a sialic acid mutarotase, shows that *E. coli* is capable of trans-forming, in the periplasm, the end product of polysialic acid degradation, that is, the α-anomer of Neu5Ac, into the β-anomer that can be trans-ported through the inner membrane *via* NanT. *yjhS*, the last gene of the NanC operon, codes for a 35-kDa periplasmic protein with unknown function. The functional characterization of this protein should shed more light on the specific role of NanC. In particular, it would be interesting to test whether YjhS could be the sialidase that cleaves oligosialic acids that have entered through NanC.

The KdgM porin family

KdgM homologues are found in gammaproteo-bacteria. Presently, 101 sequences distributed in more than 50 species have been identified† (family number: PF06178). The great majority of these KdgM homologues are found in Enterobacteriales.

If we exclude two distantly related genes from *Pectobacterium carotovorum* (formerly *Erwinia caroto-vora*), phylogenic sequence analysis of these family members reveals the existence of four main sub-families: the KdgM, NanC, and OmpL subgroups and a final set composed mainly of homologues from *Salmonella* spp., which we have named SkrPs for *Salmonella* KdgM-related porins (Fig. 5). Present functional data are scarce but analysis of genetic context and possible modes of transcriptional regulation suggest a role in acidic sugar/oligosac-charide transport.[23,24] In many species, two closely related sequences (typically sharing around 60% sequence identity) from all subfamilies except that of OmpL have been identified. These include the following protein pairs: (i) *D. dadantii* KdgM and KdgN; (ii) NanC and a related homologue named NanC2, both present in the pathogenic *E. coli* strains CFT073 and O6; and (iii) two uncharacterized homologues in *Salmonella* spp., named SkrP1 and SkrP2, whose gene location suggests a role in the transport of rhamnogalacturonate, a pectin degra-dation product.[24] To pinpoint possible conserved structural features within this family as well as to investigate further the reason(s) for genetic redun-dancy, we have used the structure of NanC to build 3D models of these homologues, using the web-based SWISS-MODEL workspace.[42] We restricted the analysis to the β-barrel domain that could be modeled with a high level of confidence, and we compared the number and location of charged (and particularly basic) pore residues between NanC and the modeled homologues.

First, comparison of the closely related homo-logues within each protein pair shows that signifi-cant differences are only observed between NanC and NanC2, in which four Arg and one Lys have been substituted to uncharged amino acids. Most of
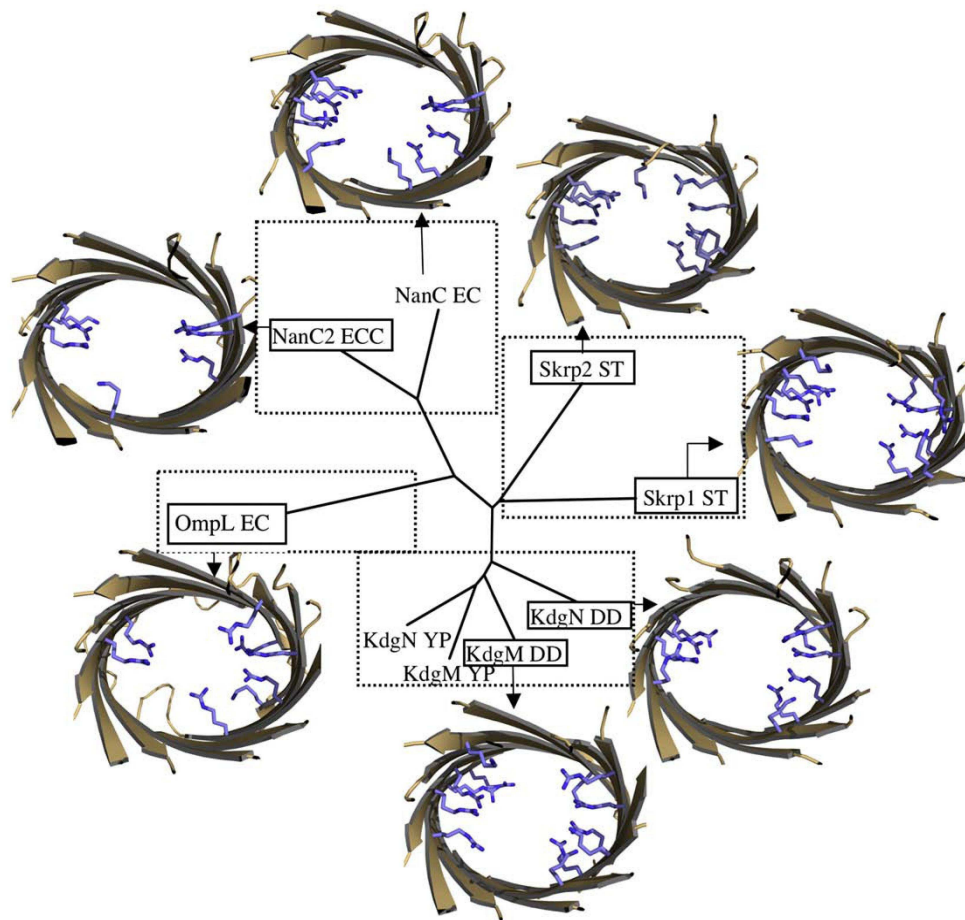
† http://pfam.janelia.org/site

Fig. 5. Homology models of KdgM family members. The phylogenic tree of selected members of the KdgM family is presented, together with the 3D homology models of the proteins with boxed labels. Dashed squares delimitate the NanC, KdgM, OmpL, and SkrP subfamilies. Models are displayed in a top view with the extracellular loops cut away for clarity and the side chains of basic residues lining the pore interior shown as blue sticks. EC, ECC, DD, YP, and ST stand for *E. coli*, *E. coli* CFT073, *D. dadantii*, *Y. pestis*, and *Salmonella typhi*, respectively. The accession codes for the proteins are as follows: NanC2_ECC, Q8CVR1; OmpL_EC, P76773; KdgM_DD, Q934G3; KdgN_DD, ABF-0015532 (https://asap.ahabs. wisc.edu/asap/logon.php); KdgM_YP, Q8ZF15; KdgN_YP, Q8ZA25; SkrP1_ST, P0A1U9; SkrP2_ST, Q8Z7M9.

the remaining basic residues lining the modeled NanC2 pore still cluster in two small groups facing each other, but this feature does not appear as prominent as in NanC. Electrostatic properties can vary significantly depending on osmotic conditions. This gene duplication might reflect the need for the bacteria to adapt to different environments in a manner reminiscent to what has been described for OmpF/OmpC.[43] On the contrary, no significant differences could be observed between the models of SkrP1 and SkrP2 and between those of KdgM and KdgN. This is consistent with previous studies revealing no major functional differences but a reciprocal gene expression regulation pattern, which suggests an upregulation of KdgM during the course of plant infection, whereas KdgN would be used under saprophytic conditions.[25]

Second, and most importantly, we were particularly interested to assess the level of conservation of the basic residues forming the two positively

charged tracks. Our modeling studies suggest a striking structural conservation of the two facing strings of arginines and lysines inside the pores of KdgM, KdgN, SkrP1, and SkrP2, with the exception of the OmpL model that exhibits one main track. Overall, about half of the relevant, positively charged amino acids are conserved in sequence. The four arginines Arg6, Arg105, Arg107, and Arg208, located in the center of the pore, are among the most conserved amino acids throughout the entire family, which is not surprising since conserved arginines at the constriction zone of porins are assumed to play an important role in the ion conductance and selectivity properties.[4,44] To compensate for the residues not conserved in sequence, the models suggest that arginines or lysines arising from another part of the sequence are likely to replace the missing basic residues and complement the cluster to form two opposite tracks resembling those observed in the structure of NanC

(Fig. 6). Such conservation supports a functional role for these residues in NanC and related porins and is consistent with the proposed common role for these KdgM family members in the transport of acidic oligosaccharides. For KdgM and several other homologues, the translocated solutes most probably result from pectin degradation occurring during the course of plant infection.[24] Thus, the structure of NanC, by revealing the architecture of KdgM family members, complements the extensive structural analyses performed on proteins from enterobacterial pectin degradation pathways.[45] It also provides valuable information for further characterization as well as for the design of KdgM channel activity inhibitors that could be used to fight *D. dadantii* plant infections.
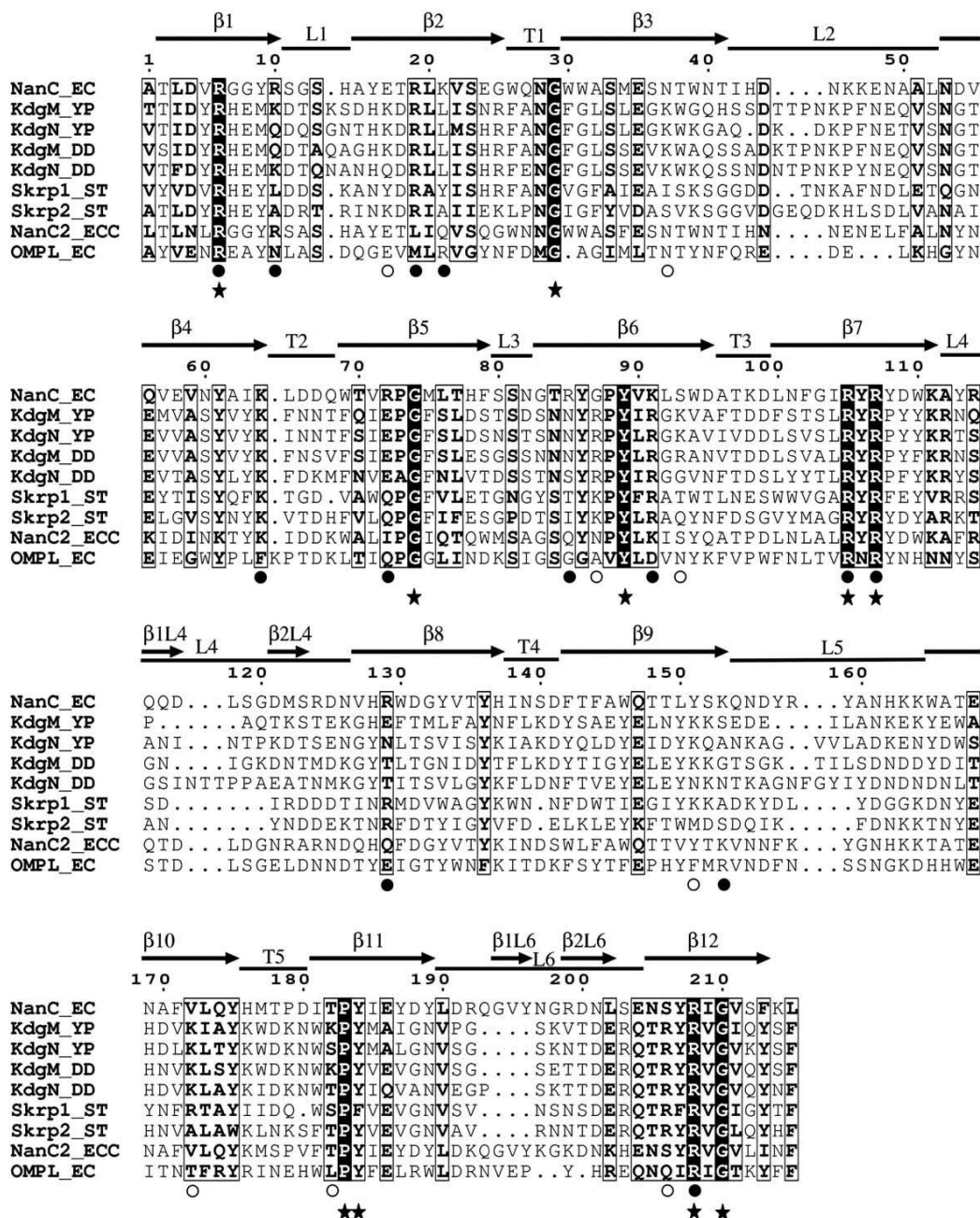
Fig. 6. Sequence alignment of KdgM family members. The alignment was generated by ClustalW and subsequently rectified to position insertions and deletions in loop regions. Secondary structural elements and residue numbers of NanC are indicated at the top. Identical residues are highlighted in black; semi-conserved ones are boxed. Stars designate residues that are highly conserved within the KdgM family. Filled circles indicate arginines or lysines forming the parallel tracks in the NanC pore, whereas unfilled circles mark the position of basic residues that could contribute to similar positively charged tracks in other family members, as suggested by the 3D homology models.

Biological significance

Sialic acids represent a valuable source of carbon and nitrogen, in particular for NanC-expressing *E. coli* strains that have acquired all the enzymatic machinery for growth on sialic acid as the only carbon source.[28] In the intestinal environment, the importance of that nutrient is supported by the significant overexpression of *nanAT* genes (essential for Neu5Ac transport through the inner membrane and cytoplasmic catabolism) detected during the initial phases of mouse colon colonization by *E. coli*.[46] In line with this, the same study reports a 1.9-fold upregulation of the *nanC* gene.

Furthermore, sialic acids could also serve as a regulatory signal to sense inflammation. Neu5Ac (the level of which raises during inflammation) has been shown to suppress the expression of type 1 fimbriae virulence factors, *via* the downregulation of FimB recombinase, possibly as a means for the bacteria to evade from the immune system.[29,30,47] Interestingly, the genes coding for FimB and NanC are adjacent in the *E. coli* genome and separated by a large intergenic region where complex transcriptional regulation takes place. As a result, gene expression regulation of FimB and NanC appears highly intertwined.[26,30] In this context, efficient uptake of monomers and oligomers of sialic acids *via* NanC might be important to level down host defenses.

In laboratory conditions and/or at high Neu5Ac concentration, Neu5Ac is efficiently translocated through OmpF/OmpC and the involvement of NanC may be minimal. However, this might not be the case under different conditions. A recent DNA-microarray analysis of temporal gene expression in *E. coli* K12 biofilms reports an overall 5-fold over-expression of NanC.[48] At low sialic acid concentrations such as in the intestinal tract, the specific features revealed by the structure of NanC and in particular the strong positive electrostatic potential inside the pore would make NanC more efficient for Neu5Ac transport than OmpF/OmpC. Moreover, the presence of two tracks of basic residues facing each other across the pore, which are likely to be conserved among acidic sugar-specific members of the family, appears well suited for the translocation of acidic oligosaccharides. These results, together with the mutarotase activity recently described for NanM, are consistent with the idea that primary functions of NanC and other genes within the operon (coding for NanM and YjhS) are probably not simply for Neu5Ac acquisition but rather for scavenging of host or microbial-derived sialic acid products.[41]

Materials and Methods

NanC expression and purification

Full-length *E. coli* NanC was cloned into a T7 promoter-containing expression vector pKSM717 either with or without the signal sequence and expressed into BL21(DE3) omp8/pLys *E. coli* cells.[49] Transformants were grown in LB medium at 30 °C, and overexpression was induced by addition of 1 mM isopropyl-D-thiogalactopyranoside. Bacteria producing outer membrane-targeted NanC (Mb-NanC, the protein synthesized with its signal sequence) were harvested after an overnight induction; resuspended in 10 mM Tris–HCl, pH 8.0, and 1 mM ethylenediaminetetraacetic acid; and disrupted in a French press. Two centrifugation steps ($6000g$ for 20 min followed by $100,000g$ for 1 h) allowed the elimination of unbroken cells and subsequent collection of the total membrane fraction that was resuspended in 50 mM Tris, pH 8.0. After solubilization of the inner-membrane fraction by addition of 1% lauroyl sarcosine (Sigma), the outer-membrane fraction was pelleted by ultracentrifugation ($100,000g$ for 1 h) and resuspended in 50 mM Tris, pH 8.0. NanC membrane extraction was achieved by stepwise solubilization of the outer membrane using increasing concentrations (from 0.5% to 3%) of octylpolyoxyethylene (OPOE; Bachem). Solubilized NanC was first purified by anion-exchange chromatography in the presence of 0.6% OPOE. NanC-containing elution fractions were pooled and diluted 10 times in 0.6% OPOE, 25 mM acetate, pH 5.0 prior to loading onto a cation-exchange chromatography column. The final purification step consisted in a gel filtration on a Superdex-75 column (GE Healthcare) in 0.05% lauryldimethylamine-*N*-oxide (LDAO; Anatrace), 150 mM NaCl, 50 mM Tris, pH 8.

Inclusion body-targeted NanC (IB-NanC, the protein synthesized without its signal sequence) was produced from cells harvested after a 2-h IPTG induction and disrupted in a French press. A 5-min, $2000g$ centrifugation step was performed to eliminate unbroken cells followed by a 10-min centrifugation at $8000g$ to pellet the inclusion bodies that were subsequently dissolved in 8 M urea, 10 mM Tris, pH 8. Refolding was carried out by rapid dropwise 8-fold dilution in 5% LDAO, 10 mM Tris, pH 8, at room temperature under thorough mixing. After a 30-min, $150,000g$ ultracentrifugation step, the supernatant was loaded onto an anion-exchange chromatography column pre-equilibrated with 0.05% LDAO, 10 mM Tris, pH 8.5. NanC-containing fractions were pooled, concentrated, and subjected to size-exclusion chromatography in 0.14% Fos-Choline 12 (Anatrace), 150 mM NaCl, 10 mM Tris, pH 8. Eluted protein fractions were finally dialyzed against 0.14% Fos-Choline 12, 10 mM Tris, pH 8. This last chromatography step was performed in the presence of various detergents, and the eluted protein was subsequently subjected to crystallization screening, but only the presence of Fos-Choline 12 led to diffracting crystals of the refolded protein.

NanC crystallization

Protein samples were concentrated on an Amicon filtration device with a molecular cutoff of 30 kDa with final concentrations of 10 mg/ml. Crystals were grown using the vapor diffusion technique at room temperature. Sitting drops were prepared by mixing 0.5 μl of the protein sample with 0.5 μl of the reservoir solution that consisted of 5–10% isopropanol, 150 mM CaCl$_2$, 100 mM acetate, pH 4.6, and 52% polyethylene glycol 400, 100 mM Hepes, pH 7.5, for Mb-NanC and IB-NanC samples, respectively. Mb-NanC crystals appeared within 10 days; they belong to the hexagonal space group $P6_322$ with cell dimensions $a = b = 96.8$ Å and $c = 121.2$ Å. IB-NanC crystallized in the trigonal space group R3:H (also referred to as H3) with cell constants $a = b = 74.9$ Å and $c = 126.4$ Å. Crystals were solubilized and analyzed by SDS-PAGE and mass spectrometry, confirming the nature of the crystallized

protein and that no proteolysis had occurred. The two crystal forms have one molecule in the asymmetric unit and solvent contents of 62% and 54% for the hexagonal and trigonal forms, respectively. The trigonal crystals were not reproducible from protein batch to batch and replaced by thin crystalline plates that diffracted aniso-tropically to 6 Å only. Consequently, only the hexagonal Mb-NanC crystals were used for heavy-atom derivatiza-tion, which was successfully achieved with a 22-h soak in the presence of 1 mM K_2PtCl_4 and a 10-min soak in a solution containing 10 mM $SmCl_3$. Both soaks were carried out in crystallization conditions supplemented with 20% ethylene glycol.

Data collection and structure determination

Prior to data collection, native Mb-NanC crystals were soaked a few minutes into the crystallization solution supplemented with 20% glycerol or ethylene glycol, whereas IB-NanC and derivatized Mb-NanC crystals could be directly flash-frozen in liquid nitrogen. X-ray diffraction data collection was performed at 100 K on beamline X06SA at the Swiss Light Source (Villigen, Switzerland) using either a MARCCD (for native crystals) or a PILATUS 6M detector (for derivatized crystals). Native data were processed using MOSFLM[50] and scaled with SCALA.[51] Diffraction data from derivatized hexag-onal crystals were processed and scaled using XDS.[52] Positions of heavy-atom sites (four Pt and two Sm sites) were determined using SHELXD[53] and refined using SHARP,[54] which was used to calculate initial phases. After solvent flattening using the program SOLOMON,[55] the resulting experimental MIRAS electron density showed the presence of a 12-stranded β-barrel but the map quality was insufficient for full interpretation. Nevertheless, the structural similarity between NanC and the β-domain of NalP was revealed and an alanine-substituted model of the latter could be used to solve the trigonal crystal structure by molecular replacement using Phaser.[56] Subsequently, the electron density map was decisively improved by applying multi-crystal density averaging using the program DMmulti.[57] Model building of the two crystal forms was performed with Coot,[58] and refinement was carried out by REFMAC.[53] The stereo-chemistry of the refined models was checked by the program MolProbity.[59] Ninety-seven percent of all residues are found in the favored regions of the Ramachandran plot. One outlier, Gly87, is found in the disallowed regions. Since the electron density corresponding to this residue is very well defined, this conformation is likely to result from the presence of a proline at position 88. Figures 1b and c, 3, 4, and 5 were generated with PyMOL‡, Fig. 2 with HOLE[34] and VMD,[60] and Fig. 6 with ESPRIT.[61] Surface electrostatic potential representations shown in Fig. 3a and b were created using GRASP.[62]

Modeling of the ligand binding

The simulation system consisted of the NanC structure together with a Neu5Ac trimer. Several initial configura-tions were used, each corresponding to a different position of the trimer center of mass along the pore axis. All the calculations were performed using the programs CHARMM[38] and the CHARMM27[63,64] all-atom force

‡ www.pymol.org

field with a GBSW (Generalized Born with a simple SWitching) implicit solvent model. The protocol of simulation consisted in 200 steps of minimization (SD procedure in alternation with ABNR method) followed by 1 ns of Langevin dynamics in implicit solvent using the GBSW[65] module of CHARMM (time step of 0.002 ps, temperature set at 300 K).

The membrane was modeled as a solvent-inaccessible infinite planar low-dielectric slab of 28 Å thickness.[66] The nonpolar solvation energy was estimated from the solvent-exposed surface area using a phenomenological surface tension coefficient of 0.04 kcal/mol/$Å^2$. All bonds involving hydrogen atoms were fixed using SHAKE algorithm.[67]

Accession numbers

Structure factor amplitudes and coordinates of the hexagonal and trigonal NanC porin structures have been deposited in the Protein Data Bank (respective accession codes: 2WJQ and 2WJR).

References

1. Delcour, A. H. (2003). Solute uptake through general porins. *Front. Biosci.* **8**, 1055–1071.
2. Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656.
3. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K. & Welte, W. (1998). Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science*, **282**, 2215–2220.
4. Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E. & Schulz, G. E. (1991). Molecular architec-ture and electrostatic properties of a bacterial porin. *Science*, **254**, 1627–1630.
5. Kreusch, A. & Schulz, G. E. (1994). Refined structure of the porin from *Rhodopseudomonas blastica*. Comparison with the porin from *Rhodobacter capsulatus*. *J. Mol. Biol.* **243**, 891–905.
6. Hirsch, A., Breed, J., Saxena, K., Richter, O. M., Ludwig, B., Diederichs, K. & Welte, W. (1997). The structure of porin from *Paracoccus denitrificans* at 3.1 Å resolution. *FEBS Lett.* **404**, 208–210.
7. Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A. *et al.* (1992). Crystal structures explain functional properties of two *E. coli* porins. *Nature*, **358**, 727–733.

8. Dutzler, R., Rummel, G., Alberti, S., Hernandez-Alles, S., Phale, P., Rosenbusch, J. *et al.* (1999). Crystal structure and functional characterization of OmpK36, the osmoporin of *Klebsiella pneumoniae*. *Structure*, **7**, 425–434.

9. Zeth, K., Diederichs, K., Welte, W. & Engelhardt, H. (2000). Crystal structure of Omp32, the anion-selective porin from *Comamonas acidovorans*, in complex with a periplasmic peptide at 2.1 Å resolution. *Structure*, **8**, 981–992.

10. Forst, D., Welte, W., Wacker, T. & Diederichs, K. (1998). Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Biol.* **5**, 37–46.

11. Schirmer, T., Keller, T. A., Wang, Y. F. & Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science*, **267**, 512–514.

12. Meyer, J. E., Hofnung, M. & Schulz, G. E. (1997). Structure of maltoporin from *Salmonella typhimurium* ligated with a nitrophenyl-maltotrioside. *J. Mol. Biol.* **266**, 761–775.

13. Biswas, S., Mohammad, M. M., Patel, D. R., Movileanu, L. & van den Berg, B. (2007). Structural insight into OprD substrate specificity. *Nat. Struct. Mol. Biol.* **14**, 1108–1109.

14. Biswas, S., Mohammad, M. M., Movileanu, L. & van den Berg, B. (2008). Crystal structure of the outer membrane protein OpdK from *Pseudomonas aeruginosa*. *Structure*, **16**, 1027–1035.

15. Moraes, T. F., Bains, M., Hancock, R. E. & Strynadka, N. C. (2007). An arginine ladder in OprP mediates phosphate-specific transfer across the outer membrane. *Nat. Struct. Mol. Biol.* **14**, 85–87.

16. Basle, A., Rummel, G., Storici, P., Rosenbusch, J. P. & Schirmer, T. (2006). Crystal structure of osmoporin OmpC from *E. coli* at 2.0 Å. *J. Mol. Biol.* **362**, 933–942.

17. Ye, J. & van den Berg, B. (2004). Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO J.* **23**, 3187–3195.

18. Subbarao, G. V. & van den Berg, B. (2006). Crystal structure of the monomeric porin OmpG. *J. Mol. Biol.* **360**, 750–759.

19. Yildiz, O., Vinothkumar, K. R., Goswami, P. & Kuhlbrandt, W. (2006). Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J.* **25**, 3702–3713.

20. van den Berg, B., Black, P. N., Clemons, W. M., Jr. & Rapoport, T. A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. *Science*, **304**, 1506–1509.

21. Hearn, E. M., Patel, D. R. & van den Berg, B. (2008). Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. *Proc. Natl Acad. Sci. USA*, **105**, 8601–8606.

22. Pellinen, T., Ahlfors, H., Blot, N. & Condemine, G. (2003). Topology of the *Erwinia chrysanthemi* oligogalacturonate porin KdgM. *Biochem. J.* **372**, 329–334.

23. Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A. & Condemine, G. (2002). The oligogalacturonate-specific porin KdgM of *Erwinia chrysanthemi* belongs to a new porin family. *J. Biol. Chem.* **277**, 7936–7944.

24. Rodionov, D. A., Gelfand, M. S. & Hugouvieux-Cotte-Pattat, N. (2004). Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology*, **150**, 3571–3590.

25. Condemine, G. & Ghazi, A. (2007). Differential regulation of two oligogalacturonate outer membrane channels, KdgN and KdgM, of *Dickeya dadantii* (*Erwinia chrysanthemi*). *J. Bacteriol.* **189**, 5955–5962.

26. Condemine, G., Berrier, C., Plumbridge, J. & Ghazi, A. (2005). Function and expression of an *N*-acetylneuraminic acid-inducible outer membrane channel in *Escherichia coli*. *J. Bacteriol.* **187**, 1959–1965.

27. Angata, T. & Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* **102**, 439–469.

28. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. *Microbiol. Mol. Biol. Rev.* **68**, 132–153.

29. Sohanpal, B. K., El-Labany, S., Lahooti, M., Plumbridge, J. A. & Blomfield, I. C. (2004). Integrated regulatory responses of fimB to *N*-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12. *Proc. Natl Acad. Sci. USA*, **101**, 16322–16327.

30. Sohanpal, B. K., Friar, S., Roobol, J., Plumbridge, J. A. & Blomfield, I. C. (2007). Multiple co-regulatory elements and IHF are necessary for the control of fimB expression in response to sialic acid and *N*-acetylglucosamine in *Escherichia coli* K-12. *Mol. Microbiol.* **63**, 1223–1236.

31. Signorell, G. A., Chami, M., Condemine, G., Schenk, A. D., Philippsen, A., Engel, A. & Remigy, H. W. (2007). Projection maps of three members of the KdgM outer membrane protein family. *J. Struct. Biol.* **160**, 395–403.

32. Schulz, G. E. (2002). The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta*, **1565**, 308–317.

33. Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. & Gros, P. (2004). Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**, 1257–1266.

34. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. (1993). The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460.

35. Schulz, G. E. (1993). Bacterial porins: structure and function. *Curr. Opin. Cell Biol.* **5**, 701–707.

36. Manzi, A. E., Higa, H. H., Diaz, S. & Varki, A. (1994). Intramolecular self-cleavage of polysialic acid. *J. Biol. Chem.* **269**, 23617–23624.

37. Dutzler, R., Wang, Y. F., Rizkallah, P., Rosenbusch, J. P. & Schirmer, T. (1996). Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure*, **4**, 127–134.

38. Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B. *et al.* (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614.

39. Nikaido, H. & Rosenberg, E. Y. (1981). Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of *Escherichia coli*. *J. Gen. Physiol.* **77**, 121–135.

40. Nikaido, H. & Rosenberg, E. Y. (1983). Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. *J. Bacteriol.* **153**, 241–252.

41. Severi, E., Muller, A., Potts, J. R., Leech, A., Williamson, D., Wilson, K. S. & Thomas, G. H. (2008). Sialic acid mutarotation is catalyzed by the *Escherichia coli* beta-propeller protein YjhT. *J. Biol. Chem.* **283**, 4841–4849.

42. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.

43. Pratt, L. A., Hsing, W., Gibson, K. E. & Silhavy, T. J. (1996). From acids to osmZ: multiple factors influence

synthesis of the OmpF and OmpC porins in *Escherichia coli*. *Mol. Microbiol.* **20**, 911–917.

44. Schirmer, T. & Phale, P. S. (1999). Brownian dynamics simulation of ion flow through porin channels. *J. Mol. Biol.* **294**, 1159–1167.

45. Abbott, D. W. & Boraston, A. B. (2008). Structural biology of pectin degradation by Enterobacteriaceae. *Microbiol. Mol. Biol. Rev.* **72**, 301–316.

46. Chang, D. E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J. *et al.* (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl Acad. Sci. USA*, **101**, 7427–7432.

47. El-Labany, S., Sohanpal, B. K., Lahooti, M., Akerman, R. & Blomfield, I. C. (2003). Distant cis-active sequences and sialic acid control the expression of fimB in *Escherichia coli* K-12. *Mol. Microbiol.* **49**, 1109–1118.

48. Domka, J., Lee, J., Bansal, T. & Wood, T. K. (2007). Temporal gene-expression in *Escherichia coli* K-12 biofilms. *Environ. Microbiol.* **9**, 332–346.

49. Prilipov, A., Phale, P. S., Van Gelder, P., Rosenbusch, J. P. & Koebnik, R. (1998). Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from *E. coli*. *FEMS Microbiol. Lett.* **163**, 65–72.

50. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **27**, 30.

51. Evens, P. R. (2005). Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 72–82.

52. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.

53. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 122–133.

54. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **59**, 2023–2030.

55. Abrahams, J. P. & Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial

F1 ATPase. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **52**, 30–42.

56. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.

57. Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.

58. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.

59. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, 375–383.

60. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.

61. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.

62. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Funct., Genet.* **11**, 281–296.

63. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

64. Mackerell, A. D., Jr., Feig, M. & Brooks, C. L., III (2004). Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415.

65. Im, W., Lee, M. S. & Brooks, C. L., III (2003). Generalized born model with a simple smoothing function. *J. Comput. Chem.* **24**, 1691–1702.

66. Im, W., Feig, M. & Brooks, C. L., III (2003). An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **85**, 2900–2918.

67. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints—molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341.

I.2.2. Further discussions, conclusions and perspectives

I.2.2.1. Recent characterisation of NanM and NanS

Bacteria genomes are generally clustered in co-transcribed operons in which all genes are belonging to the same pathway. In *E. coli*, the *nanC* gene belongs to an operon composed of two additional genes called *nanM* (or *YjhT*) and *nanS* (or *YjhS*) coding for two proteins involved in the sialic acid catabolism. NanM has been recently characterized as being a periplasmic sialic acid mutarotase (Severi et al, 2008). This protein is yet the only one known to have this function in bacteria. *E. coli* colonises mucosal surfaces in which sialic acid is mostly present in glycoconjugate under α-anomeric form. When cleaved away from these glycoconjugates by host sialidases or other sialidase competent bacteria (*E. coli* K12 is unable to cleave sialic acid away from glycoconjugates), sialic acid α-anomers are spontaneously, but very slowly, converted in β-anomers. NanM is folding in a 6-bladed unclosed β-propeller, catalyzing this otherwise very slow mutarotation reaction and considerably accelerating it. As NanT, the inner-mebrane transporter responsible for the sialic acid entrance in the cytoplasm, is specifically translocating the β-anomer, the presence of NanM gives *E. coli* an advantage in comparison to other bacteria for the acquisition of sialic acid (Severi et al, 2008), a good carbon and nitrogen source in the colonized host. Another advantage for *E. coli* could be the rapid depletion of the sialic acid released by host sialidases in the extracellular environment during the course of inflammation, slowing the host immune response down (Severi et al, 2008; Sohanpal et al, 2004).

NanS, the third gene in the *nanCMS* operon, has been recently identified as being a potential 9-O-acetyl N-acetylneuraminic acid (Neu5,9Ac$_2$) esterase (Steenbergen et al, 2009). NanS allows the growth of *E. coli* on minimal medium containing Neu5,9Ac$_2$ as sole carbon source. This enzyme would allow the catalysis of the Neu5,9Ac$_2$ into Neu5Ac that can be further acquired and processed by the bacteria as carbon and nitrogen source. "Sialic acids" is a general term for a family of nine-carbon sugar acids. Neu5Ac is the most frequently found and the most important in metazoans. However, several other sialic acids are present in animals including Neu5,9Ac$_2$ which is the second most abundant (Steenbergen et al, 2009). The acquisition of alternative sialic acid molecules could give *E. coli* a substantial advantage in colonising the host intestine compared to other bacteria reflecting its success in colonizing the lowest intestine.

It was shown that neither NanC nor NanM or NanS are essential to allow the growth of *E. coli* on Neu5Ac clearly indicating that these proteins are not directly involved in Neu5Ac acquisition (Condemine et al, 2005; Severi et al, 2008; Steenbergen et al, 2009). However, their regulation clearly suggests an implication in, probably alternative, sialic acid metabolism. Data obtained on NanM and NanS suggest that the proteins belonging to this operon could be involved in the processing of sialic acid molecules others than Neu5Ac, the most common one. On the basis of structural data and comparison with its homologues, we proposed that NanC is implied in the acquisition of sialic acid oligomers. Combined with the mutarotase activity of NanM, this proposal would imply the existence of a, yet unknown, periplasmic enzyme that would be able to cleave oligomers of sialic acid. Although our structure of NanC supports its putative activity in translocating oligomeric chains of acidic sugars to the periplasm (Wirth et al, 2009), the recent study on NanS (Steenbergen et al, 2009) suggests that it could specifically translocate alternative sialic acid molecules such as $Neu5,9Ac_2$ or directly released sialic acid containing glycoconjugate molecules. However, the hypothesis of alternative sialic acid uptake would not explain the function of NanM and why it is in this operon. Alternatively, the glycoconjugate uptake would require an additional, not discovered yet, periplasmic protein able to cleave sialic acid (or other sugars) away of these molecules.

The limited biological information on NanC does not allow, yet, to determine the exact function of the *NanCMS* operon. This clearly indicates that further experimental data on NanC is required. Therefore, a collaboration has been started with the group of Prof. Eisenberg, (Chicago) in order to determine the basal electrophysiological properties of NanC as well as in presence of several potential substrates. However, data using $Neu5,9Ac_2$ or glycoconjugates were not recorded yet.

I.2.2.2. Two HEPES molecules are present in the rhombohedral NanC structure

The rhombohedral crystals of NanC were grown in a crystallization condition containing 100 mM HEPES as buffer. After refinement of the structure obtained using these crystals, an unassigned electron density was visible at the entrance and the exit of the pore. First resembling a sulfate ion electron density, the refinement of the fitted ion model led to additional electron density. Placing the sulfate moiety of the HEPES molecule into the initial

large electron density, a complete HEPES molecules could be fitted in these densities and could be refined properly when the occupancy was set to 0.7 (Fig. I.2.1).
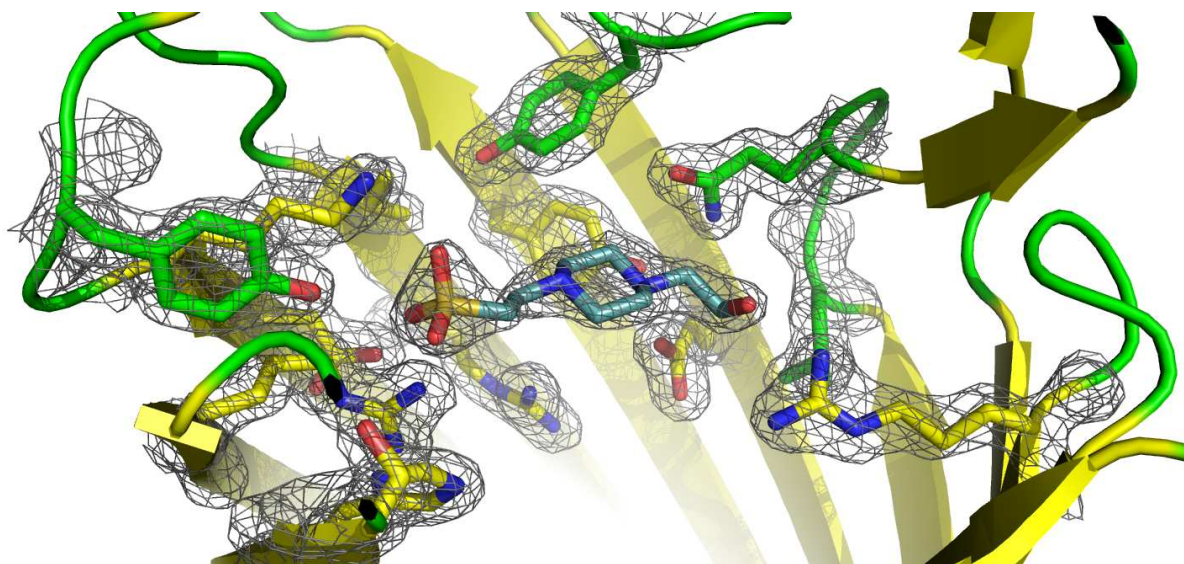
Figure I.2.1: HEPES molecule located at the extracellular entrance of the pore.

Also shown are adjacent residues as well as the electron density at 1.4 σ

The presence of two ordered HEPES molecules in the NanC pore is surprising. The binding determinants can be clearly identified. In both cases, the sulfate moiety of the HEPES molecule is bound to positively charged residues located inside the β-barrel and belonging to the two positively charged tracks that have been identified in the NanC structure (Fig. I.2.2). The hydroxyl "tail" of the HEPES molecule located at the extracellular pore entrance, is forming a hydrogen-bound with residue R10 of the other side of the pore. For the HEPES at the periplasmic side, a loose hydrogen bound between the hydroxyl of the HEPES and T135 is possible (3.75Å). However, except the sulfate moiety, the electron density for this molecule is weakly defined and the HEPES molecule might only tightly bind to NanC *via* its sulfate.
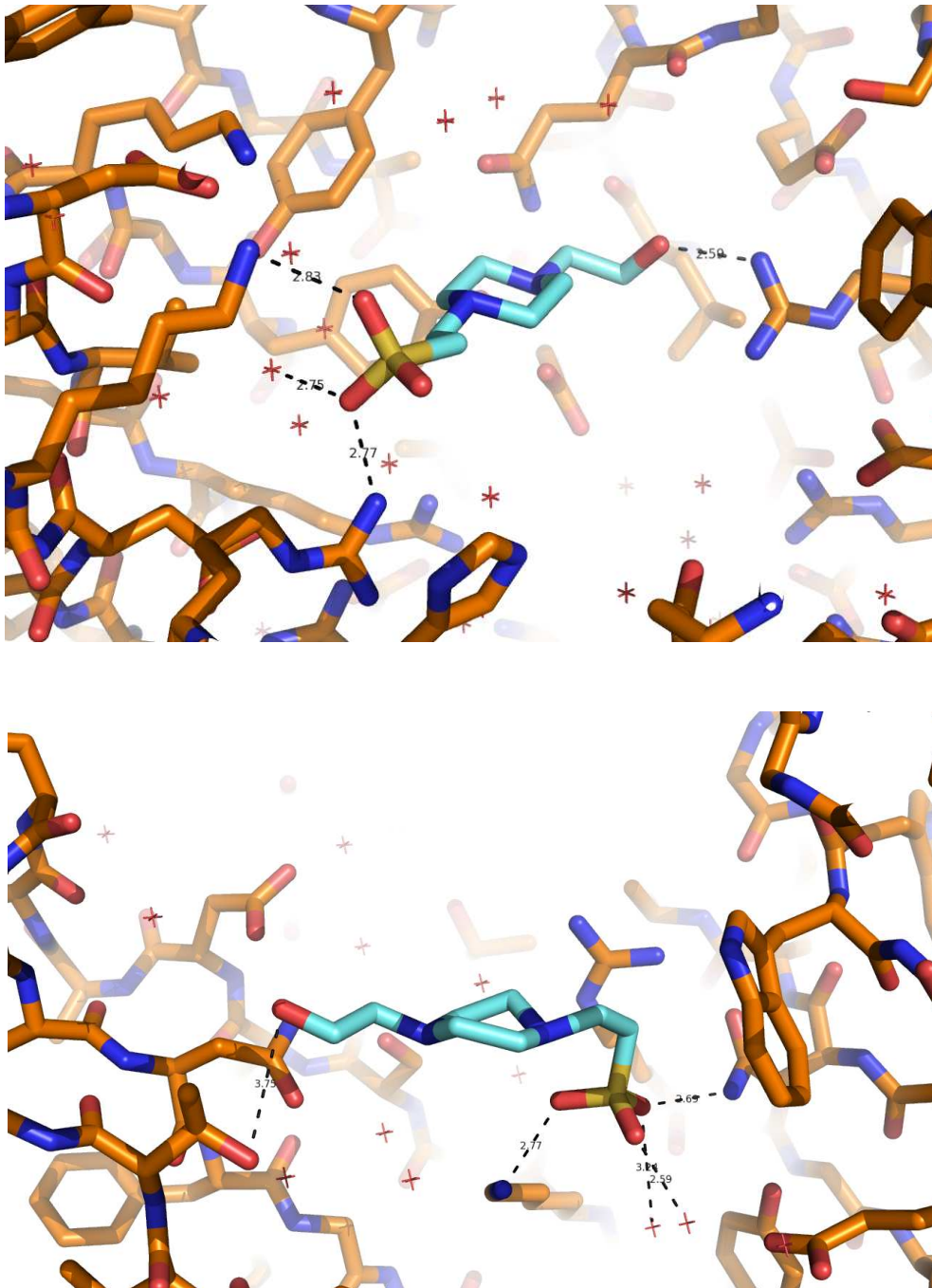
Figure I.2.2: The two HEPES molecules are bound to basic residues.

Possible ionic or hydrogen bounds are represented in black. (Top) HEPES molecule located close to the extracellular exit of the pore. The sulfate moiety is interacting with R85 and K111. (Bottom) HEPES molecule located at the periplasmic side of NanC. The sulfate moiety is interacting with K91. In both cases additional hydrogen bounding is provided by water molecules.

I.2.2.3. Implications of HEPES bound inside the pore

The HEPES molecules binding into the NanC pore are likely to have important implications on both biological and experimental aspects.

The binding of the HEPES molecules is mostly mediated by its sulfate moiety which is hydrogen bounded to basic residues of the most striking feature of NanC, its two positively charged tracks. This, before never observed feature, is believed to be involved in the translocation of N-acetylneuraminic acid oligomers (oligo-Neu5Ac) by orienting and guiding the substrate through NanC (Wirth et al, 2009). Also the HEPES molecule is a bit similar to Neu5Ac (it has nearly the same size and an acidic group on one side) its binding to the pore raises the question of the substrate specificity of NanC. It is likely that NanC is able to attract and bind, not only (oligo-)Neu5Ac or sialic acid derivatives but also a vast variety of negatively charged molecules such as other acidic sugars or smaller negatively charged ions. This will have to be determines in lipid bilayer experiments in collaboration with Prof. Eisenberg, Chicago.

A second question rising from the binding of HEPES into the NanC pore is if HEPES can be translocated through the channel or is it only binding to it? This question will have to be addressed by future experiments as it has important implications. The fact that the electron density for the HEPES molecules is relatively well defined is suggesting a rather tight binding to these two sites. This would imply that the pore might be partially blocked by these molecules.
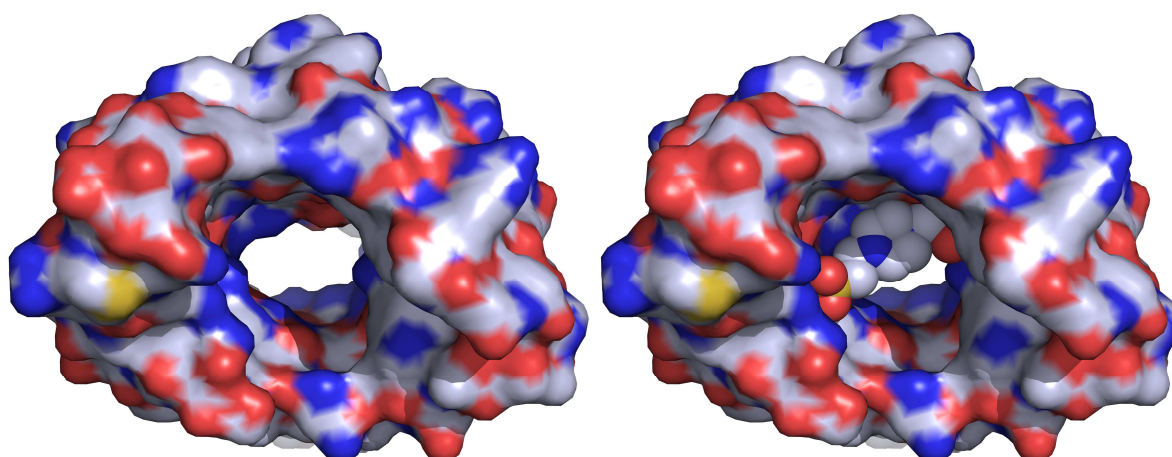
Figure I.2.3: Surface representation of NanC with and without HEPES bound.

(left) surface representation of NanC when the HEPES molecules has been removed. In this case the pore in wide open whereas when HEPES is present (represented in balls) it is partially occluded (right).

The two molecules are located at both the entrance and exit of the channel but are not completely closing the channel. However, the radius of the pore at these positions is significantly diminished and therefore HEPES might act as a competitive inhibitor for translocation of Neu5Ac through NanC (Fig. I.2.3). HEPES was used as buffer for the electrophysiology experiments on NanC and the other KdgM family members (Blot et al, 2002; Condemine et al, 2005; Condemine & Ghazi, 2007). The concentration was only 10 mM compared to the 100 mM used for crystallization, but the presence of HEPES could have influenced these measurements. To investigate the HEPES effect and to eventually reproduce the measurements with Neu5Ac and oligo-NeuAc using another buffer, electrophysiology experiments are being done in collaboration with the group of Pr. Eisenberg in Chicago using NanC purified from the membrane in order to avoid eventual problems due to the refolding in SDS method (see I.1.2.1). Initial results clearly indicate that HEPES is significantly diminishing the channel conductance already at 5 mM (Fig. I.2.4). With 100 mM only 50-60 % of the conductance is remaining compared to a sample without HEPES. In the future, other buffers will have to be used for electrophysiology experiments and their iniquity assessed before the experiments.
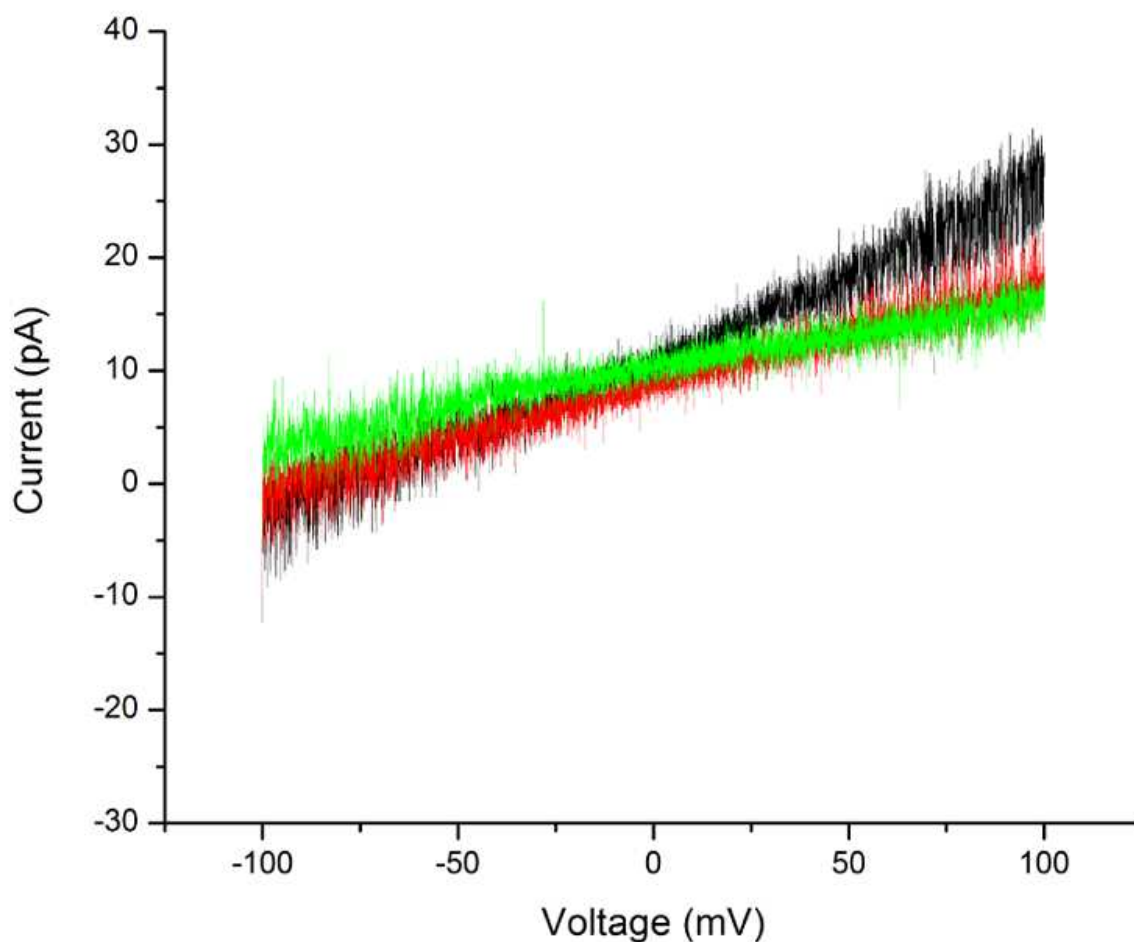
Figure I.3.4: Current-voltage traces of NanC inserted in artificial lipid bilayer;

All measurements were done using 250 mM KCl. The experiments have been performed without HEPES (black trace) with 5 mM (red) and 100 mM (green). This figure is provided by J. Giri, Department of Bioengineering, University of Illinois, Chicago.

I.2.2.4. Potential influence of salt on substrate translocation

As deduced from the crystal structure of NanC and from the models of other members of the KdgM family, the translocation through these porins is most probably driven by the two conserved alignments of basic residues facing each other inside the aqueous pore. The charges probably play a fundamental role in long range attraction and translocation of the substrate similarly as in Omp32 (Zachariae et al, 2006; Zachariae et al, 2002). Therefore it is likely that the presence of numerous anions employed in the standard electrophysiology experiments (800 mM KCl) (Condemine et al, 2005) can shield partially the two positively tracks and

diminish the translocation efficiency. However, experimental data of trigalacturonate translocation could be measured for KdgM in similar salt conditions as for NanC. Therefore, the ligand translocation is probably not completely impaired, but may be less efficient than in *in vivo* conditions.

I.2.2.5. NanC-ligand complexes

Many attempts were made to try to obtain the crystal structure of a complex of NanC with Neu5Ac or oligomers of sialic acid. No binding constant could be measured for Neu5Ac in the NanC characterization study (Condemine et al, 2005). However, data could be obtained for KdgM and trigalacturonate and the K_D was estimated to be about 30 mM (Blot et al, 2002) and it could be speculated that NanC has a similar affinity for its substrate. This low affinity value is necessary to allow the proper function of the specific porin. In order to efficiently translocate its substrate, a specific porin should not bind it too tightly as it would lead to a difficult release of the substrate. In the case of the *E. coli* maltoporin LamB, it had been shown that the affinity of a specific porin for its sugar substrate increases with the length of the sugar and the protein-substrate complex took advantage from that (Dutzler et al, 1996). Furthermore from the NanC structure we deduced that it may be able to translocate oligomeric sugar acids. Therefore, several molecules have been tested up to a high substrate concentration: Neu5Ac (up to 100 mM), Neu5Ac α-2,8-linked dimer (up to 40 mM), trimer (up to 25 mM) and tetramer (up to 10 mM).

As the rhombohedral crystals were not reproducible, soaking and co-crystallization experiments were tried out using the hexagonal crystal form. For co-crystallization, the ligand was added to the NanC protein solution 12-18 hours before the set up of the crystallization drops. Soaking experiments were tried out with various ligand concentrations and soaking times ranging from several minutes to 18 hours. However, all co-crystallization and soaking experiments failed as no crystals could be obtained in the co-crystallization experiments, and no electron density for sialic acid was visible the electron density obtained from soaked crystals. This can be caused by several parameters such as too low binding site affinity (as discussed before). Sialic acid oligomers have been shown to be unstable at acidic pH (Manzi et al, 1994). The hexagonal crystal form condition is at pH 4.6 and may results in spontaneous cleavage of the oligomers of sialic acid leading to monomeric sialic acid. This would

unfortunately lead to the loss of the benefit of using oligomers in order to increase the affinity of NanC for Neu5Ac. Taken together, these facts probably explain why the co-crystallization and soaking attempts failed.

Additionally, the recent conclusions of the study of NanS (Steenbergen et al, 2009) imply that NanC may be a porin specific for modified sialic acid molecules. Therefore, intensive translocation data using electrophysiology lipidic bilayer experiments using Neu5,9Ac$_2$, glycoconjugates and (oligo)Neu5Ac are needed before further co-crystallization or soaking experiments can be undertaken.

I.2.2.6. Mutations proposed to probe translocation through NanC

NanC and the related homologues belonging to the KdgM family have been classified as specific porins. This implies, that they have properties similar to enzymes and especially, the presence of a specific binding site allowing to filter the solutes and to limit the translocation to a given (class of) molecules. Once the substrate of NanC will be identified by lipidic bilayer experiments and as the NanC structure in complex with ligands could not been obtained, the mutation of specific residues impairing the translocation could give experimental evidences into the transport mechanism.

The inside of the NanC pore is mostly formed by two consecutions of basic residues forming a continuous positively charged patch from the extra-cellular environment to the periplasm. This feature is likely to be important for substrate guidance and translocation through the pore and is probably conserved in NanC homologues. In order to confirm the role of these two lines of positive charges, mutations to neutral or even negatively charged residues could be done. *In vivo* studies, electrophysiology experiments as well as *in silico* modelling on these mutants, could allow understanding the transport mechanism at a molecular level. These experiments could be done in collaboration with the groups of Dr. Condemine, CNRS, Université de Lyon (Villeurbanne, France), Prof. Eisenberg University of Illinois (Chicago, USA) and Prof. Bernèche, Biozentrum, University of Basel.
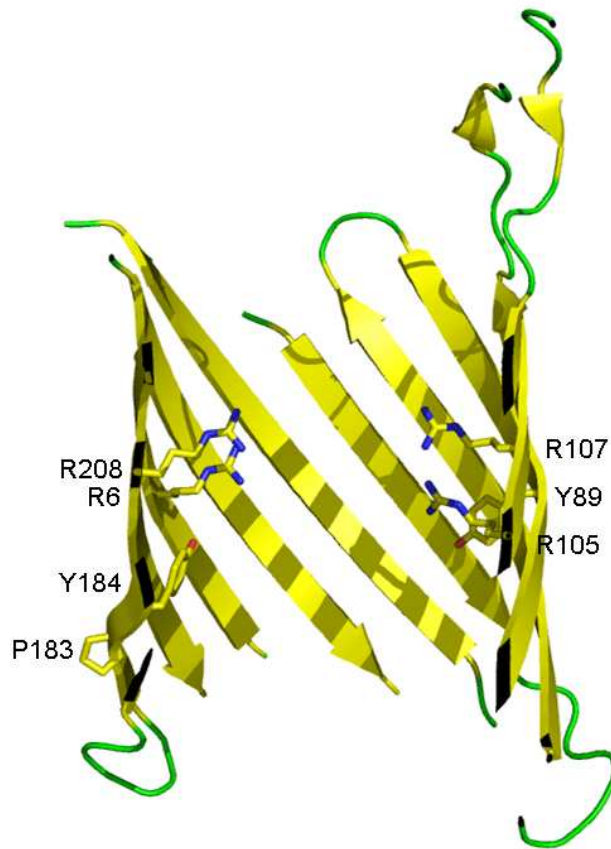
Figure I.2.5: Cut away representation of the conserved residues NanC pore.

Residues (except glycins) conserved amongst the members of the KdgM family are reprensented in sticks. They all cluster in the central, most constricted, region of the pore.

Although the two charged lines seem to be structurally well conserved in the KdgM family, only four arginine residues (R6, R105, R107 and R208 in NanC) are highly conserved in all the sequences of KdgM family members (Wirth et al, 2009). These four residues are likely to play a key role in substrate translocation as all four are grouped together at the centre of the pore, close to the constriction zone (Fig. I.2.5). Mutation of these residues could probably give insight into their importance for substrate translocation.

Chapter I.3

Crystallization of KdgM using
microseeding matrix screening and surface entropy
reduction mutants

**Roman Lehner*, Christophe Wirth*, Guy Condemine, Caroline M Peneff and
Tilman Schirmer**

*: Contributed equally to the project

The following section is describing the expression, purification and crystallization of KdgM that was done in parallel to the work on NanC. This project can be divided in three main phases: (I) initial crystallization assays using KdgM protein expressed in inclusion bodies and refolded and purified using SDS as a detergent, (II) expression and purification of 6xHisTagged KdgM from the membrane as well as its crystallization with and without microseeding matrix screening and (III) design and surface entropy mutant generation, its expression and initial crystallization results.

Dr. Caroline Peneff and I worked together on the first phase and obtained the initial KdgM crystals that were unsuitable for structure determination. The second phase was done by Roman Lehner during his Master practical under my and Dr. Caroline Peneffs' supervision. In addition, Roman Lehner designed the surface entropy mutants used during the third phase of the project but could not, during his practical, obtain the variants. Our collaborator, Dr. Guy Condemine, obtained a first mutant (K124 to serine) that I purified and crystallized in several conditions. The crystallographic analysis was done by me.

As a large part of this section is redundant with the Master Thesis of Roman Lehner, only the first and the third phase of the project will be discussed in details. The second part will be very shortly summarized but details can be found in the Master Thesis of Roman Lehner.

I.3.1. Material and methods

I.3.1.1. Protein expression

Expression in inclusion bodies

Over-expression of KdgM and KdgN was done in inclusion bodies by Dr. Guy Condemine (Unité de Microbiologie et Génétique Villleurbanne, France) using the *kdgm* (or or *kdgn*) gene without the membrane addressing signal sequence. The gene was cloned into a pKSM717 plasmid containing a T7 promoter (Maneewannakul et al, 1994). The expression of this protein was following the protocol established for the *in vivo* and electrophysiology studied of NanC (Condemine et al, 2005).

Expression in membranes

In order to express KdgM in the membrane, the gene was recloned in pKSM717 (Maneewannakul et al, 1994) with its signal sequence addressing the protein to the outer-membrane. A 6-histidine tag (6xHis-tag) was also added, as other monomeric porins could be expressed, purified and crystallized with a such an affinity tag. The protein was then expressed in *E. coli* BL21(DE3)omp8/pLys (F$^-$ *ompT hsdS$_B$ gal dcm ΔlamB ompF*::Tn*5 ΔompA ΔompC* (DE3)) (Prilipov et al, 1998) grown in LB medium at 30°C. The expression was induced by addition of 1 mM IPTG when OD reached 1.0. The cells were harvested after overnight expression at 30°C and stored at -80°C.

Mutations

The mutation work was realized by Guy Condemine using Quickchange cloning protocol (Stratagen) and various primers depending which mutation has to be introduced.

I.3.1.2. Protein purification

Inclusion bodies refolding and purification

The first productions and purifications of KdgM and KdgN were realized by Dr. Guy Condemine using protein expressed in inclusion bodies. After production, the inclusion bodies were harvested and dissolved in 6M urea. The solubilised material was refolded by an over

night dialysis against a 10mM Tris, 0.5% SDS solution. The refolded protein sample was purified using a preparative SDS-PAGE and a Tris-tricine PAGE. Subsequently, gel filtration chromatography was performed in order to exchange buffer and detergent to the one used for crystallization.

Purification of membrane produced KdgM-His and KdgM-His K124S

Cells were resuspended in 10mM Tris pH 8.0, 1mM EDTA and lysed using a French press. Unbroken cells were discarded after a 15 min centrifugation at 6000g. The total membrane fraction was harvested by centrifugation at 100000g for 1h. Membranes were resuspended in 50mM Tris pH 8.0, 1% N-lauroyl sarcosine (LS) in order to specifically solubilise the inner-membrane. The outer-membrane was collected by centrifugation at 100000g for 45 min and resuspended in 50mM Tris. The protein was then extracted from the outer-membrane by addition of increasing octylpolyoxyethylene (OPOE) concentrations. In the first step, 0.5% OPOE was added, the sample mixed and centrifuged for 30 minutes at 100000g. The pellet was then resuspended and an increased concentration of OPOE added. This was done until the whole KdgM–His proteins were extracted according to control SDS-PAGE.

The purification took advantage of the presence of the 6xHis-tag. Extracted KdgM-His protein samples were pooled and bound to a pre-equilibrated (50mM Tris pH 8.0, 0.6% OPOE) HisTrap column (GE healthcare) and eluted with a 0 to 500 mM imidazol gradient. The fractions containing KdgM-His were pooled, concentrated and subjected to a gel filtration chromatography (Superdex 200, GE healthcare) equilibrated with 10 mM Tris pH 8.0, 100 mM NaCl and the desired detergent for crystallization.

I.3.1.3. Crystallization

All crystallization trials were realized at room temperature. Initially, crystallization assays were performed using the sparse matrix approach and commercial screens (Jancarik & Kim, 1991). The reservoir volume was 80µl and the drop was done by mixing 0.5µl of the reservoir and 0.5µl of the KdgM solution at concentrations ranging from 5 to 20 mg/ml. This was done using 96-well plates and pipetted either manually or using a Phenix crystallization robot from Art Robin (in that case the drop was composed of 0.2µl reservoir and 0.2µl of protein solution). The found crystallization conditions were refined in 24-well sitting drop plates with

a reservoir volume of 500µl and a drop composed of 0.5µl protein solution and 0.5µl reservoir solution.

The microseeding matrix screening used in the second phase of this project is an extrapolation of the classical microseeding method in which the microseeds are used to sow a whole crystallization plate. This part is described in details in the Master Thesis of Roman Lehner.

I.3.1.4. Diffraction data and processing

Although some diffraction data were collected using the in-house x-ray generator, most data collections were realized at the Swiss Light Source (SLS) on beamlines PXI and PXIII. Before data collection, crystals requiring cryoprotection were soaked in stabilizing solution containing either 15 % glycerol or 15 % ethylene glycol and flash frozen in liquid nitrogen. Most diffraction data collections were done at 100 K but some measurements at room temperature were performed in order to control the crystal quality before freezing. The typical oscillation range was 1° and the exposure time was varying from 1 to 3 s per frame at the SLS. Data processing and integration was done with MOSFLM (Leslie & Powell, 2007) and XDS (Kabsch, 1993), scaling was realized using either SCALA (Evans, 2006) or XSCALE (Kabsch, 1993). Phasing by molecular replacement was tried out using PHASER (Mccoy et al, 2007) and various models.

I.3.2. Results and Discussions

I.3.2.1. Crystallization and diffraction data of proteins refolded using SDS

The goal of this project was to understand the molecular mechanisms of the GAn uptake by *D. dadantii.* In order to increase the success probability, work was done in parallel on both KdgM and KdgN porins. These proteins were first produced by Dr. Guy Condemine with a protocol adapted from previous studies (Condemine et al, 2005; Condemine & Ghazi, 2007). KdgM and KdgN were both expressed in inclusion bodies and purified using a method based on preparative PAGE. The proteins were then subjected to gel filtration chromatography in

order to exchange the detergent and to check the monodispercity of the sample in that detergent. Many detergents have been tested (Table I.3.1). Depending which protein was used, detergents yielding monodisperse size exclusion chromatography peaks could be identified (Table I.3.1).

Table I.3.1: List of detergents used in gel filtration and to crystallize KdgM and KdgN.

Detergent	KdgM			KdgN		
	Tested	Monodisperse	crystals	Tested	Monodisperse	crystals
LDAO	√	√				
C_8E_4	√	√	√	√	√	
C_8E_5	√	√	√			
OG	√	√	√	√		
OTG	√					
NG	√	√				
DM	√					
OM	√	√				
Fos12	√			√		
Cymal 5	√					

Knowing the best detergents for each protein, crystallization screens were set up using commercial screens, such as the CrystalScreen from Hampton as well as the Wizard, MemSys and MemStart screens from Molecular Dimensions.

Crystals were obtained for KdgM but not for KdgN. These conditions were all leading to long and very thin needle shaped crystals except the most promising crystallization condition that was obtained using octylglucoside with nonylmaltoside as additive and that is described in the following table.

Protein	Concentration	Detergent	Precipitant
KdgM	17 mg/ml	0.6% OG 0.28% NM	0.1M MES pH 6.5 18-23% PEG 600 50-100mM zinc acetate

This crystallization condition was intensively optimized by varying the precipitant concentration and the pH and by using additive screens. However, most of the obtained conditions led to similar crystals than the original condition. The KdgM crystals were small rods of about 30×30×150 µm (Fig. I.3.1).

Figure I.3.1: Rod shaped crystals of KdgM obtained after protein refolding in SDS

Prior testing these crystals for diffraction, they were frozen in liquid nitrogen without further addition of cryoprotectant. X-ray diffraction data were collected at the PXI beamline of the SLS. These crystals showed unfortunately very reproducible low resolution diffraction (10-15Å). No quality improvement could be observed when using crystals of various sizes, grown with additives or treated with a dehydration method. In house room temperature data collection did not improve the diffraction quality compared to data collection at 100K, indicating that the crystals were intrinsically of poor quality and were not suffering from freezing in liquid nitrogen. The work on these crystals was not continued further.

Similarly to NanC when purified from inclusion bodies and refolded in SDS, the tendency to form crystals leading to low resolution data could be explained by several deficiencies in the protein production and purification. As no control was possible, the protein might have been improperly refolded when using SDS as a detergent. Another possible reason could have been the presence of remaining traces of SDS that could interfere with a normal protein crystallization process. SDS is strongly negatively charged and is likely to bind to the positive charges of the proteins (KdgM and KdgN have numerous R and K residues). If SDS

molecules were binding by their negatively charged moieties to the positively charged residues of the protein, the floppy hydrophobic tails of SDS could have impaired the formation of tight crystal contacts. However, membrane proteins are known to be difficult to crystallize as their huge hydrophobic parts can't create strong specific crystal contacts that are required to obtain good diffracting. In order to avoid potential problems rising from the refolding in SDS, the rest of this project was carried out using KdgM purified form the membrane.

I.3.2.2. Purification and crystallization of the membrane expressed 6xHisTagged KdgM.

As the crystallization of KdgM and KdgN expressed in inclusion bodies and purified with preparative PAGE was unsuccessful, a new KdgM expression and purification protocol, using a 6xHisTag, was designed by Roman Lehner during his Master training (for the details, see the Master Thesis of Roman Lehner).

Structures of small monomeric β-barrel proteins, in particular OmpG and Tsx (Subbarao & van den Berg, 2006; Ye & van den Berg, 2004; Yildiz et al, 2006), have been reported in which the crystallization as well as the folding of the protein was not disturbed by the presence of a C-terminal 6xHis-tag. This is in contrast with the case of the general porins where the N- and C-termini are connected by a salt bridge in the middle of a strand (Cowan et al, 1992). Therefore, the use of tags was believed to be not applicable for outer-membrane proteins. However, considering the similarity of KdgM with OmpG and Tsx, expression of KdgM in BL21/Omp8 cells with a C-terminal His-Tag and the signal sequence targeting the protein to the membrane (KdgM-His) was undertaken.

KdgM-His membrane extraction took advantage of the experience gained with NanC using OPOE as a detergent (Wirth et al, 2009) and yielded in amounts suitable for structural studies. In contrast to NanC, the presence of the 6xHis-tag allowed an easier purification using HisTrap affinity columns. A subsequent gel filtration was used as final purification step as well as to exchange the detergent and buffer. Crystallization assays were done with KdgM in several detergents, but only LDAO yielded crystalline objects. However, after optimisation, these P222 crystals turned out to diffract only to 9Å and, therefore, were unsuitable to solve the KdgM structure.

To increase the number of successful crystallization conditions, microseeding matrix screening (D'Arcy et al, 2007; Ireton & Stoddard, 2004) was performed using the crystals obtained before that were crashed in order to sow complete commercial screens. Using these seeds, 35 crystallization conditions could be obtained out of 288 tested (success rate of 12.5 %). In contrast, without seeds, only 2 conditions could be obtained (success rate of 0.7 %) when using the same screens. This clearly shows the advantage of overcoming the nucleation phase of crystallization by introducing a regular growth matrix into the crystallization condition and is the first statistics about this method when using a membrane protein.

However, after optimization of the most promising conditions, the diffraction was still limited to 6Å in the best case. Interestingly, at least one of the conditions obtained after microseeding matrix screening, resulted in crystals having a different space group (P622) than the seeds (P222) used in the screen. This demonstrates that the microseeding matrix screening method not only allows obtaining more crystallization conditions but also allows getting new crystal forms. It seems that, as previously suggested (D'Arcy et al, 2007; Ireton & Stoddard, 2004), the seed is only a regular matrix required to grow a new crystal form.

I.3.2.3. Design of the surface entropy reduced KdgM mutants.

As neither classical screening, nor microseeding matrix screening yielded crystals suitable to solve the KdgM structure, surface entropy reduced mutant were designed. Using the NanC structure, sequence alignment of KdgM with NanC, the available KdgM secondary structure information and SwissModeller (Arnold et al, 2006), a homology model of KdgM was obtained (See the Master Thesis of Roman Lehner for more details as well as for the model analysis).

The modelling of KdgM gave insights into its overall structure but could not be used to reveal the fine details of its structure. However it was used to design KdgM mutants using the surface entropy reduction approach (Derewenda, 2004; Derewenda & Vekilov, 2006; Vekilov, 2003; Vekilov et al, 2002). In order to enhance the crystallization ability of a protein, bulky solvent exposed residues (such as Lys or Glu) can be mutated in shorter residues (Ala, Ser or

Asn) in order to allow easier crystal contacts formation. Usually Lys or Glu residues are found at the surface of soluble protein. In the case of a porin, however, many charged residues are located in the channel but only mutations in the extracellular loops and the periplasmic turns can eventually significantly help the increase the crystallization ability of these proteins. Using the homology model, six possible mutations sites were found in KdgM (K124 and K130 in loop L4, K156 and K161 in loop L5 and E199 and E 203 in loop L6). These residues are not conserved in the KdgM family and sequence alignment with homologues was used to determine the residue type to replace the selected mutation targets. This led to the identification of six mutations (K124S, K130N, K156N, K161N, E199G and E203S).

I.3.2.4. Expression and crystallization of a "surface entropy reduction" mutant.

Although all KdgM single mutations have been tried in parallel, only replacement of K124 to serine was conclusive yet (KdgM K124S). The difficulties to obtain the other mutants might be due to the low C/G content of the KdgM gene and other members of the KdgM family (Blot et al, 2002; Ilatovskiy & Petukhov, 2009).

Table I.3.2: Composition of the crystallization conditions obtained with KdgM K124S

	Buffer	Precipitant	Salt
1	-	1 M Na/K phosphate pH 7.5	0.7 M ammonium sulfate
2	0.1 M Na phosphate pH 6.8	15 % PEG 2000	0.5 M NaCl
3	0.1 M Tris pH 8.0	20 % PEG 2000, 2% MPD	0.3 M Mg nitrate
4	0.1 M tri-sodium citrate pH 5.6	10 % PEG 2000 MME 3 % PEG 200 20 % glycerol	-
5	0.05 M tricine pH 8.0	22 % PEG 4000	-
6	0.1 M tri-sodium citrate pH 5.6	10 % PEG 2000 MME	100 mM Mg Acetate
7	0.1 M tri-sodium citrate pH 5.6	12 % PEG 6000	100 mM Lithium sulphate
8	0.05 % Na succinate pH 6.5	15 % PEG 6000	-
9	0.1 M HEPES pH 7.5	1 M tri-sodium citrate	-

KdgM K124S was expressed and purified as wild-type KdgM-His in similar amounts. The protein sample was then concentrated to 4 and 6 mg/ml and first crystallization screens were done using 0.05 % LDAO as detergent. However, although LDAO had proven to be the detergent of choice in order to obtain KdgM crystals, the crystallization assays of the K124S mutant led to almost only clear drops at these concentrations whereas the wild type KdgM had resulted in mostly precipitated drops or even crystallizing at similar concentrations. This suggests that the K124S mutant has a higher solubility than the wild type KdgM or a decrease in its aggregation tendency.

In order to confirm these observations, a second set of crystallization screens was done at 8 and 18 mg/ml leading to 24 conditions (out of 384) resulting in crystalline objects from which the best are listed in table I.3.2. Some of these conditions (number 3 and 4) have been refined using 24-well plates. These crystals (Fig. I.3.2) will be tested for diffraction at the SLS in the near future.

Figure I.3.2: crystal obtained after refinement of conditions 3 (left) and 4 (right) using KdgM K124S.

The crystals from condition 3 are rods having a size of about $250 \times 20 \times 20$ µm whereas the crystals of condition 4 form imperfect triangles of about $150 \times 150 \times 20$ µm.

I.3.3. Conclusions and outlook

Crystallization with refolded in SDS material

The use of KdgM protein refolded and purified with SDS led to poorly diffracting crystals. The crystallization was not problematic as crystals could be obtained easily. However these crystals were invariably of poor quality. Although the KdgM sample had been intensively washed by dialysis or gel filtration chromatography, the presence of SDS in the protein buffer can't be excluded. This would have a negative effect on the crystallization of KdgM because SDS is an ionic detergent able to bind strongly to the positive charges of KdgM. This could create further problems in the crystallization step as the detergent molecules, bound to the protein *via* ionic interactions are not labile and can't easily be displaced to allow crystal contacts.

Purification and crystallization of KdgM-His

The refolding and purification with SDS having proven to be a problem in the case of NanC (see I.1.), we decided to use KdgM in as native as possible conditions. It was decided to reclone the *kdgm* gene in order to express the protein in the membrane with addition of a His-Tag. The protein could be expressed and purified in reasonable amount and to a high purity grade. Crystallization gave first hits of, unfortunately, bad diffracting crystals but thanks to microseeding matrix screening, several other crystallization conditions could be obtained. This is the first time that this method was applied for membrane proteins with a partial success as it helped obtaining more crystallization conditions and crystals of different space groups using one single seed type. Unfortunately, although the method seemed promising, the microseeding matrix screening did not allow obtaining crystals of good enough quality to solve the KdgM structure.

Surface entropy reduction

6 mutants were designed with the help of a KdgM homology model in order to increase its crystallization ability but only one has been successfully cloned yet. However, initial expression, purification and crystallization trials are very promising as many crystallization conditions have been obtained from this mutant. This could never been obtained directly with the wild type KdgM. The crystals have not been tested for diffraction yet; however, the high

number of crystals and the clear various morphologies allows expecting a positive result in the close future.

Although the NanC structure has given the first structural data of the KdgM family the amount of available microbiological and biochemical data on KdgM is more important. Therefore the KdgM structure could probably allow getting a more complete view of the acidic oligosaccharides acquisition mechanisms.

Chapter I

Conclusions and perspectives

The aim of this project was to determine the structure of a member of the recently characterized acidic (oligo)saccharide specific KdgM porin family. Members of this outer-membrane proteins family are mostly present in plant and animal pathogens and supposed to be important for virulence.

Work was done in parallel on three members of the KdgM family (KdgM and KdgN from *D. dadantii* and NanC from *E. coli*). Expression and purification protocols were set up for all three proteins, but rapidly, the project was focused on KdgM and NanC. Finally, the NanC structure was solved in two different crystal forms at 1.8 and 2.0 Å resolution giving the first structural information on members of the KdgM outer-membrane protein family.

Despite many attempts to solve the KdgM structure and the numerous crystals obtained with this protein, no good diffraction data could be collected yet. However, the surface entropy reduction method clearly showed significant increase in the crystallization ability of KdgM but it might be necessary to further mutate other long and bulky amino-acids in order to obtain well diffracting crystals. To solve the phase problem for KdgM, molecular replacement using the NanC structure as model could be used. The KdgM structure remains of primary interest as in its case the translocated substrate has been clearly identified as being oligogalacturonate. If co-crystallization or soaking with ligand is possible, the substrate-KdgM structure could help revealing the details of the molecular mechanisms involved in substrate translocation.

The structure of NanC raises several biological questions as mentioned in Section I.2. However, although structural data are giving hints for its solute translocation capabilities, the subtrate of NanC is not clearly identified yet. From genetic data and from the function of the other genes in its operon, NanC seems clearly involved in sialic acid or sialic acid derivatives uptake. The structure suggests that NanC could, as KdgM, be involved in the uptake of oligomers of acidic sugars or sialic acid containing glycoconjugates. This hypothesis will have to be confirmed in the future using *in vitro* assays. Due to the importance of sialic acid in host-pathogen interactions it is important to determine the exact substrate of NanC and how its translocation occurs.

References

Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52:** 30-42

Angata T, Varki A (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem Rev* **102:** 439-469

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22:** 195-201

Basle A, Rummel G, Storici P, Rosenbusch JP, Schirmer T (2006) Crystal structure of osmoporin OmpC from E. coli at 2.0 A. *J Mol Biol* **362:** 933-942

Batchelor E, Walthers D, Kenney LJ, Goulian M (2005) The Escherichia coli CpxA-CpxR envelope stress response system regulates expression of the porins ompF and ompC. *J Bacteriol* **187:** 5723-5731

Benz R, Schmid A, Vos-Scheperkeuter GH (1987) Mechanism of sugar transport through the sugar-specific LamB channel of Escherichia coli outer membrane. *J Membr Biol* **100:** 21-29

Biswas S, Mohammad MM, Movileanu L, van den Berg B (2008) Crystal structure of the outer membrane protein OpdK from Pseudomonas aeruginosa. *Structure* **16:** 1027-1035

Biswas S, Mohammad MM, Patel DR, Movileanu L, van den Berg B (2007) Structural insight into OprD substrate specificity. *Nat Struct Mol Biol* **14:** 1108-1109

Blot N, Berrier C, Hugouvieux-Cotte-Pattat N, Ghazi A, Condemine G (2002) The oligogalacturonate-specific porin KdgM of Erwinia chrysanthemi belongs to a new porin family. *J Biol Chem* **277:** 7936-7944

Boggon TJ, Shapiro L (2000) Screening for phasing atoms in protein crystallography. *Structure* **8:** R143-149

Braun V, Endriss F (2007) Energy-coupled outer membrane transport proteins and regulatory proteins. *Biometals* **20:** 219-231

Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* **59:** 2023-2030

Brillet K, Journet L, Celia H, Paulus L, Stahl A, Pattus F, Cobessi D (2007) A beta strand lock exchange for signal transduction in TonB-dependent transducers on the basis of a common structural motif. *Structure* **15:** 1383-1391

Cobessi D, Celia H, Folschweiller N, Schalk IJ, Abdallah MA, Pattus F (2005) The crystal structure of the pyoverdine outer membrane receptor FpvA from Pseudomonas aeruginosa at 3.6 angstroms resolution. *J Mol Biol* **347:** 121-134

Condemine G, Berrier C, Plumbridge J, Ghazi A (2005) Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *J Bacteriol* **187:** 1959-1965

Condemine G, Dorel C, Hugouvieux-Cotte-Pattat N, Robert-Baudouy J (1992) Some of the out genes involved in the secretion of pectate lyases in Erwinia chrysanthemi are regulated by kdgR. *Mol Microbiol* **6:** 3199-3211

Condemine G, Ghazi A (2007) Differential regulation of two oligogalacturonate outer membrane channels, KdgN and KdgM, of Dickeya dadantii (Erwinia chrysanthemi). *J Bacteriol* **189:** 5955-5962

Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP (1992) Crystal structures explain functional properties of two E. coli porins. *Nature* **358:** 727-733

Cowtan KD, Main P (1993) Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr D Biol Crystallogr* **49:** 148-157

D'Arcy A, Villard F, Marsh M (2007) An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* **63:** 550-554

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375-383

Delcour AH (2003) Solute uptake through general porins. *Front Biosci* **8:** d1055-1071

Derewenda ZS (2004) Rational protein crystallization by mutational surface engineering. *Structure* **12:** 529-535

Derewenda ZS, Vekilov PG (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr* **62:** 116-124

Driessen AJ, Nouwen N (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77:** 643-667

Dutzler R, Rummel G, Alberti S, Hernandez-Alles S, Phale P, Rosenbusch J, Benedi V, Schirmer T (1999) Crystal structure and functional characterization of OmpK36, the osmoporin of Klebsiella pneumoniae. *Structure* **7:** 425-434

Dutzler R, Wang YF, Rizkallah P, Rosenbusch JP, Schirmer T (1996) Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* **4:** 127-134

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126-2132

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* **62:** 72-82

Ferguson AD, Deisenhofer J (2002) TonB-dependent receptors-structural perspectives. *Biochim Biophys Acta* **1565:** 318-332

Ferguson AD, Hofmann E, Coulton JW, Diederichs K, Welte W (1998) Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282:** 2215-2220

Forst D, Welte W, Wacker T, Diederichs K (1998) Structure of the sucrose-specific porin ScrY from Salmonella typhimurium and its complex with sucrose. *Nat Struct Biol* **5:** 37-46

Garman E, Murray JW (2003) Heavy-atom derivatization. *Acta Crystallogr D Biol Crystallogr* **59:** 1903-1913

Hearn EM, Patel DR, van den Berg B (2008) Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. *Proc Natl Acad Sci U S A* **105:** 8601-8606

Hirsch A, Breed J, Saxena K, Richter OM, Ludwig B, Diederichs K, Welte W (1997) The structure of porin from Paracoccus denitrificans at 3.1 A resolution. *FEBS Lett* **404:** 208-210

Hugouvieux-Cotte-Pattat N, Condemine G, Nasser W, Reverchon S (1996) Regulation of pectinolysis in Erwinia chrysanthemi. *Annu Rev Microbiol* **50:** 213-257

Hugouvieux-Cotte-Pattat N, Reverchon S (2001) Two transporters, TogT and TogMNAB, are responsible for oligogalacturonide uptake in Erwinia chrysanthemi 3937. *Mol Microbiol* **41:** 1125-1132

Ilatovskiy A, Petukhov M (2009) Genome-wide search for local DNA segments with anomalous GC-content. *J Comput Biol* **16:** 555-564

Ireton GC, Stoddard BL (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* **60:** 601-605

Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallisation of proteins. *J Appl Cryst* **24:** 409-411

Kabsch W (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* **26:** 795-800

Kalivoda KA, Steenbergen SM, Vimr ER, Plumbridge J (2003) Regulation of sialic acid catabolism by the DNA binding protein NanR in Escherichia coli. *J Bacteriol* **185:** 4806-4815

Kreusch A, Schulz GE (1994) Refined structure of the porin from Rhodopseudomonas blastica. Comparison with the porin from Rhodobacter capsulatus. *J Mol Biol* **243:** 891-905

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372:** 774-797

Leslie AGW, Powell HR (2007) Processing diffraction data with MOSFLM. *Nato Sci Ser Ii Math* **245:** 41-51

191

Locher KP, Rees B, Koebnik R, Mitschler A, Moulinier L, Rosenbusch JP, Moras D (1998) Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* **95:** 771-778

Maneewannakul S, Maneewannakul K, Ippen-Ihler K (1994) The pKSM710 vector cassette provides tightly regulated lac and T7lac promoters and strategies for manipulating N-terminal protein sequences. *Plasmid* **31:** 300-307

Manzi AE, Higa HH, Diaz S, Varki A (1994) Intramolecular self-cleavage of polysialic acid. *J Biol Chem* **269:** 23617-23624

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491-497

Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658-674

Moraes TF, Bains M, Hancock RE, Strynadka NC (2007) An arginine ladder in OprP mediates phosphate-specific transfer across the outer membrane. *Nat Struct Mol Biol* **14:** 85-87

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240-255

Nikaido H (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67:** 593-656

Nikaido H, Vaara M (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49:** 1-32

Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J, Gros P (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23:** 1257-1266

Plumbridge J, Vimr E (1999) Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. *J Bacteriol* **181:** 47-54

Prilipov A, Phale PS, Van Gelder P, Rosenbusch JP, Koebnik R (1998) Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli. *FEMS Microbiol Lett* **163:** 65-72

Pugsley AP (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57:** 50-108

Reverchon S, Expert D, Robert-Baudouy J, Nasser W (1997) The cyclic AMP receptor protein is the main activator of pectinolysis genes in Erwinia chrysanthemi. *J Bacteriol* **179:** 3500-3508

Rouanet C, Reverchon S, Rodionov DA, Nasser W (2004) Definition of a consensus DNA-binding site for PecS, a global regulator of virulence gene expression in Erwinia chrysanthemi and identification of new members of the PecS regulon. *J Biol Chem* **279:** 30158-30167

Schirmer T (1998) General and specific porins from bacterial outer membranes. *J Struct Biol* **121:** 101-109

Schirmer T, Keller TA, Wang YF, Rosenbusch JP (1995) Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267:** 512-514

Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58:** 1772-1779

Schulz GE (2002) The structure of bacterial outer membrane proteins. *Biochim Biophys Acta* **1565:** 308-317

Severi E, Hood DW, Thomas GH (2007) Sialic acid utilization by bacterial pathogens. *Microbiology* **153:** 2817-2822

Severi E, Muller A, Potts JR, Leech A, Williamson D, Wilson KS, Thomas GH (2008) Sialic acid mutarotation is catalyzed by the Escherichia coli beta-propeller protein YjhT. *J Biol Chem* **283:** 4841-4849

Shevchik VE, Condemine G, Robert-Baudouy J, Hugouvieux-Cotte-Pattat N (1999a) The exopolygalacturonate lyase PelW and the oligogalacturonate lyase Ogl, two cytoplasmic enzymes of pectin catabolism in Erwinia chrysanthemi 3937. *J Bacteriol* **181:** 3912-3919

Shevchik VE, Kester HC, Benen JA, Visser J, Robert-Baudouy J, Hugouvieux-Cotte-Pattat N (1999b) Characterization of the exopolygalacturonate lyase PelX of Erwinia chrysanthemi 3937. *J Bacteriol* **181:** 1652-1663

Signorell GA, Chami M, Condemine G, Schenk AD, Philippsen A, Engel A, Remigy HW (2007) Projection maps of three members of the KdgM outer membrane protein family. *J Struct Biol* **160:** 395-403

Sohanpal BK, El-Labany S, Lahooti M, Plumbridge JA, Blomfield IC (2004) Integrated regulatory responses of fimB to N-acetylneuraminic (sialic) acid and GlcNAc in Escherichia coli K-12. *Proc Natl Acad Sci U S A* **101:** 16322-16327

Sohanpal BK, Friar S, Roobol J, Plumbridge JA, Blomfield IC (2007) Multiple co-regulatory elements and IHF are necessary for the control of fimB expression in response to sialic acid and N-acetylglucosamine in Escherichia coli K-12. *Mol Microbiol* **63:** 1223-1236

Steenbergen SM, Jirik JL, Vimr ER (2009) YjhS (NanS) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *J Bacteriol* **191:** 7134-7139

Subbarao GV, van den Berg B (2006) Crystal structure of the monomeric porin OmpG. *J Mol Biol* **360:** 750-759

Sun PD, Radaev S (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures. *Acta Crystallogr D Biol Crystallogr* **58:** 1099-1103

Sun PD, Radaev S, Kattah M (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr D Biol Crystallogr* **58:** 1092-1098

van den Berg B, Black PN, Clemons WM, Jr., Rapoport TA (2004) Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304:** 1506-1509

Vekilov PG (2003) Solvent entropy effects in the formation of protein solid phases. *Methods Enzymol* **368:** 84-105

Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D (2002) Solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallogr D Biol Crystallogr* **58:** 1611-1616

Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM (2004) Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68:** 132-153

Vimr ER, Troy FA (1985) Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* **164:** 845-853

Weiss MS, Abele U, Weckesser J, Welte W, Schiltz E, Schulz GE (1991) Molecular architecture and electrostatic properties of a bacterial porin. *Science* **254:** 1627-1630

Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374:** 300-321

Wirth C, Condemine G, Boiteux C, Berneche S, Schirmer T, Peneff CM (2009) NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM porin family. *J Mol Biol* **394:** 718-731

Wirth C, Meyer-Klaucke W, Pattus F, Cobessi D (2007) From the periplasmic signaling domain to the extracellular face of an outer membrane signal transducer of Pseudomonas aeruginosa: crystal structure of the ferric pyoverdine outer membrane receptor. *J Mol Biol* **368:** 398-406

Ye J, van den Berg B (2004) Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO J* **23:** 3187-3195

Yildiz O, Vinothkumar KR, Goswami P, Kuhlbrandt W (2006) Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J* **25:** 3702-3713

Zachariae U, Kluhspies T, De S, Engelhardt H, Zeth K (2006) High resolution crystal structures and molecular dynamics studies reveal substrate binding in the porin Omp32. *J Biol Chem* **281:** 7413-7420

Zachariae U, Koumanov A, Engelhardt H, Karshikoff A (2002) Electrostatic properties of the anion selective porin Omp32 from Delftia acidovorans and of the arginine cluster of bacterial porins. *Protein Sci* **11:** 1309-1319

Zeth K, Diederichs K, Welte W, Engelhardt H (2000) Crystal structure of Omp32, the anion-selective porin from Comamonas acidovorans, in complex with a periplasmic peptide at 2.1 A resolution. *Structure* **8:** 981-992

Chapter II

Structure of OmpF,

the major *Escherichia coli* porin,

in an orthorhombic crystal form

Christophe Wirth, Matthias Barone[1], Arnaud Baslé[2] and Tilman Schirmer[#]

Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, 4056-Basel, Switzerland.

#. Corresponding author: tilman.schirmer@unibas.ch

1. Present address: Allschwillerstrasse 65, 4056-Basel, Switzerland

2. Present address: Institut for cell and molecular biosciences, Medical School, University of Newcastle, Framlington Place, Newcastle upon Tyne, NE2 1HH, UK

Abstract

OmpF, the trimeric general porin from *Escherichia coli*, is one of the most studied bacterial outer-membrane proteins. Its structure has been determined in 1992 by Cowan et al. in a trigonal crystal form in which OmpF trimers make lateral contacts via their hydrophobic surfaces, generating layers of 32-symmetry. We have adapted the OmpF crystallization to vapor diffusion methods and could reproduce these trigonal crystals. Data were collected to 2.1Å and the structure was further refined. In addition, a new OmpF crystal form diffracting to 2.8Å resolution and belonging to space group $C222_1$ was obtained. This new crystal form is formed by the same 32-symmetry related layers than observed in other OmpF crystal packings.

II.1. Introduction

The outer-membrane of Gram-negative bacteria is composed of phospholipids and lipopolysaccharides and is spanned by trimeric channel forming proteins called porins that can be classified in two main groups. A first group comprises the non-specific or general porins, having very little selectivity for the solute they enable the translocation of various small hydrophilic molecules (MW ≤ 600 Da) such as nutrients (Nikaido, 2003). The second group, called specific porins, regroups proteins with a specific binding site for a given (class of) molecule (Schirmer et al, 1995). In both cases, the solute translocation is driven by its concentration gradient (Schirmer, 1998).

OmpF, the general porin of *Escherichia coli*, has been intensively biochemically and structurally characterized over the last 30 years. It forms stable trimers and presents weak selectivity for cations (Benz, 1988; Nikaido, 2003). The first OmpF structure was solved to 2.4Å resolution using trigonal crystals (Cowan et al, 1995; Pauptit et al, 1991). However, these were not the first OmpF crystals obtained as a tetragonal crystals form could be grown earlier. These were the first well diffracting membrane protein crystals grown (Garavito et al, 1983) but their structure was solved only later and to a lower resolution than the trigonal crystal form (Cowan et al, 1995). More recently a high resolution structure (1.6Å) of OmpF has been reported in a similar trigonal crystal form but using an other condition for crystallization (Yamashita et al, 2008). Additionally a new hexagonal ($P6_3$) crystal form has also been described (Dhakshnamoorthy et al, 2009; Yamashita et al, 2008).

Like all structurally characterized general porins, OmpF folds in a 16-stranded hollow β-barrel structure encompassing a water-filled channel (Nikaido, 2003). The β-strands are connected by long loops on the extracellular side and short periplasmic turns. The transmembrane aqueous pore extending roughly along the barrel axis is almost perpendicular to the membrane and is forming an hour glass shaped channel. Its most constricted zone is delimited by barrel residues on one side and residues of the L3 loop that is folding into the pore on the other side. Strong hydrophobic and hydrophilic

(mostly mediated by loop L2) interactions within the monomers are explaining the tight heat stable trimers formed by OmpF (Cowan et al, 1992).

OmpF also remains a good model protein to study ion transport through general porins (Karshikoff et al, 1994; Miedema et al, 2006; Pezeshki et al, 2009) as well as interactions with antimicrobials molecules (Yamashita et al, 2008).

From a methodological point of view and as only few membrane proteins crystallize easily, OmpF is an interesting protein to try obtaining information about the crystallization of β-barrel membrane proteins and to develop specific crystallization protocols. In order to adapt the micro-dialysis OmpF crystallization protocol used in the past, to state-of-the-art methodology, we crystallized OmpF using the sitting drop method. Here we present a 2.1Å resolution structure of OmpF obtained from trigonal crystals as well as a new orthorhombic form. X-ray diffraction data were collected and the structure solved to 2.7Å resolution. The lattice of this form has pseudohexagonal symmetry and forms 32-symmetry related layers, in which the interactions are mediated by the same residues than in previously reported OmpF crystal forms.

II.2. Material and methods

II.2.1. Protein expression and purification

Wild-type OmpF expression and purification was carried out as described previously (Cowan et al, 1992). In short, wild type OmpF was expressed in a BL21(DE3)omp8/pLys strain (Prilipov et al, 1998). Cells were harvested, broken and OmpF was extracted using 3% n-octylpolyoxyethylene (OPOE). Subsequently, the protein sample was purified using chromatofocusing (PBE94, GE healthcare) followed by size exclusion chromatography (Superdex 200, GE healthcare) in order to exchange buffer and detergent.

II.2.2. Crystallization

All crystallization experiments were performed at room temperature using the sitting-drop vapor-diffusion method. The protein concentration was 12-14 mg/ml in a buffer containing 50 mM Tris pH 8.5, 0.1% OPOE and 0.8% 2-hydroxyethyloctylsulfoxide (C_8HESO). To screen for new crystallization conditions, the sparse matrix sampling approach (Jancarik & Kim, 1991) with two commercial screens (Wizzard from Molecular Dimensions and MbClass Suite from Qiagen) and Intelliplates 96-well plates (Hampton Research) was used. The plates were set up using a Phoenix robot (Art Robins) with a reservoir volume of 80 µl and drops composed of 0.2 µl protein solution mixed with 0.2 µl reservoir solution. Refinement of promising crystallization conditions was done in Cryschem plates (Hampton Research) using 500 µl as reservoir volume and 0.5-1.5 µl protein and 0.5-1.5 µl reservoir solutions as the drop.

II.2.3. Data acquisition and processing

Prior to data collection, OmpF crystals were soaked few seconds in 10 % ethylene glycol to ensure cryoprotection when required and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the X06SA and X06DA beamlines of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). Data were processed using XDS (Kabsch, 1993) or MOSFLM (Leslie & Powell, 2007) and scaled using XSCALE or SCALA (Evans, 2006). Program Pointless (Evans, 2006) was used to confirm the Laue group.

II.2.4. Phasing, refinement and structure validation

The 2.4Å OmpF model (PDB code 2omf) and rigid body refinement with Refmac (Winn et al, 2003) were used to calculate the phases for the trigonal crystal form. The phase problem of the orthorhombic crystal form was solved by molecular replacement using

PHASER (McCoy et al, 2007) and the OmpF model obtained from the new trigonal crystals as search model. The models were refined using several cycles of rebuilding with COOT (Emsley & Cowtan, 2004) and refinement with REFMAC (Winn et al, 2001). The final model quality was assessed using MOLPROBITY (Davis et al, 2007). The crystal contact analysis was carried out using the PISA server (Krissinel & Henrick, 2007).

II.3. Result and discussion

II.3.1. Refinement of the standard procedure for obtaining trigonal OmpF crystals

Our initial aim was to adapt, with state of the art methods, the crystallization protocol of OmpF allowing the growth of the high resolution diffracting trigonal crystal form. The OmpF structure has been solved several years ago (Cowan et al, 1992) using crystals grown with the micro-dialysis method (Table II.1). The crystallization condition was adapted to vapor diffusion by screening around the original condition. After few trials, hexagonal plates could be obtained within few days using 50 mM Tris pH 9.8, 10 % PEG 6000, 700 mM $MgCl_2$, 0.1 % OPOE and 0.8 % C_8HESO (Fig. II.1). X-ray diffraction data were collected up to 2.1Å resolution. The crystals were isomorphous to the already known trigonal crystal form (cell parameter a = b = 117.1 Å, c = 51.6 Å) (Table II.1). The OmpF model (PDB code 2omf) was fitted to these data using few rigid body cycles and further refined using REFMAC and rebuilt using COOT. The final model having respective R and Rfree of 16.1 % and 19.8 % will be used to replace the older, lower resolution, 2omf in the PDB. This new OmpF model for the trigonal crystal form superposes on the high resolution structure of OmpF (Yamashita et al, 2008) with a r.m.s.d of only 0.120 Å for the Cα (0.371 Å for all protein atoms) with the maximal deviations located in the extra-cellular loops indicating that the different crystallization condition has almost no impact on the protein conformation.

II.3.2. OmpF in a related orthorhombic crystal form

During the attempts to reproduce the trigonal crystal form, several complete commercial sparse matrix screens were also done. Micro- or needle shaped crystals that were not suitable for structural studies were obtained in various conditions and not further optimized. Other conditions were all containing high NaCl concentrations (higher than 2.5 M) but had various buffers ranging from pH 6.2 to 8.0. These conditions were leading to thin hexagonal plate crystals resembling the trigonal crystals and were further optimized (Fig. II.1). The best crystals were grown from 0.1 M HEPES pH 7.5 and 3.2 M NaCl (Table II.2). These new OmpF crystals appeared after 10 to 15 days and were reaching their maximum size (up to $400 \times 400 \times 20\,\mu m$) about 2 days later.

These crystals diffracted up to 2.7Å resolution at the Swiss Light Source. Automatic processing with MOSFLM and XDS revealed a hexagonal lattice with a = b = 116.7 Å and c = 145.4 Å. When processed in P3, however, the statistics where unsatisfactory. Therefore, the data were reprocessed in P1 and the Laue group determined using Pointless (Evans, 2006) as being Cmmm. Systematic absences were observed for $00l \neq 2n$ suggesting a 2_1 screw axis along c. After data processing with XDS and reduction with XSCALE, the statistics in $C222_1$ were acceptable (table II.2). Matthews coefficient calculation (Matthews, 1968) suggested the presence of one trimer in the asymmetric unit ($V_M = 3.96$ $Å^3$/Da). The phases were obtained using molecular replacement, program PHASER and the new trigonal OmpF model at 2.1Å resolution. A significant molecular replacement solution was found with a rotation and translation functions z-scores of 20.3 and 74.7, respectively, and a final Log Likelihood Gain of 9150. Few cycles of rigid body refinement were followed by cycles of rebuilding/refinement using COOT and REFMAC resulting in a final model at 2.7Å resolution with R and R_{free} values of 21.3 % and 24.4 % respectively.

The overall structure of OmpF in the orthorhombic crystal form is almost identical to previously solved OmpF structures (2ZFG, 2OMF and 1OPF). The monomeric structure is composed of a 16-stranded β-barrel superimposing with an r.m.s.d for all atoms of

0.402 Å, 0.424 Å and 0.329 Å for monomer A, B and C, respectively, with the high resolution structure of OmpF (PDB code 2ZFG).

The OmpF trimer observed in the asymmetric unit of the orthorhombic crystal form (Fig. II.2) is identical to the one found in the trigonal or tetragonal OmpF crystals (Cowan et al, 1995; Cowan et al, 1992). However, some variation can be observed at the level of the loops, especially at the tip of loop L6 from monomer B (residues 243 to 246) where the electron density is not interpretable. Intriguingly this region is very close to a crystal contact mediated by loop L7 with a 2-fold related symmetric mate of monomer B. If loop 6 of monomer B would have the same structure as in monomers A and C, the side chains of this loop would clash with the 2-fold related equivalents in monomer B. It is possible that this position is occupied by only one of the two symmetry-related loops at a given time. This would explain the disorder of the tip of loop L6.

II.3.3. The reoccurring packing of OmpF in layers of 32-symmetry.

OmpF is naturally forming tight trimers *in vivo*. In the a-b-plane of the orthorhombic crystal form, OmpF trimers are arranged in layers of 32-symmetry resembling honeycombs. Within a honeycomb, the OmpF trimers are alternatively oriented up and down along the 3-fold symmetry (Fig. II.2). This honeycomb-like arrangement in the a-b-plane can be described by a unit cell with the following calculated dimensions: a = 117.5 Å, b = 116.9 Å and γ = 120.2°. These values are very close to the requirements for a trigonal or hexagonal space group (a = b; α = β = 90° and γ = 120°) and are probably the explanation of the difficulty to index these data with both MOSFLM and XDS as well as of the characteristic 622 pattern of the self-rotation.

In the a-b-plane, the OmpF trimers are exclusively forming crystal contacts *via* isologous hydrophobic areas. Two major interaction areas (1 and 2) can be identified (Table II.3 and Fig. II.3). Interestingly, these crystal contact (interfaces 1-1 and 2-2) are identical to those found in two other OmpF crystal forms reported in the literature (P321 and P6$_3$)

(Cowan et al, 1992; Yamashita et al, 2008) (Tab II.1) but are not observed in the tetragonal crystals (Cowan et al, 1995).

Membrane protein crystals can be classified in two different classes (Ostermeier & Michel, 1997). Type I membrane protein crystals are composed of layers of proteins forming hydrophobic contacts. The layers are superposed in the third dimension where contacts are formed by the hydrophilic parts of the protein. In type II membrane protein crystals, proteins are exclusively interacting *via* their hydrophilic parts. Interfaces 1-1 and 2-2 helping the formation of the $C222_1$ crystals are favoring the OmpF crystallization in type I membrane protein crystals as they allow the formation of planar honeycombs in which proteins are interaction exclusively via their hydrophobic interfaces. These two-dimensional architectural elements can then stack in layers in order to form 3D-crystals of various space groups.

Along c, the crystal contacts are unrelated betwenn the three crystal forms (Fig. II.4). In the orthorhombic crystal form, the contacts in this area are only formed by the interaction of loop L7 from monomer B with its symmetry related neighbor whereas in the trigonal crystal form the interaction areas are larger, each β-barrel interacting in a "head to tail" fashion with its neighbor.

Conclusions

In this report we present the protocol allowing obtaining the "classical" trigonal crystal form of OmpF using modern sitting drop techniques. During our attempts of reproducing these well diffracting trigonal crystals, we found a new orthorhombic crystal form growing in high NaCl concentrations and presenting the same hexagonal shape than the trigonal crystals. These new crystals are diffracting to 2.7 Å resolution and are belonging to space group $C222_1$. In the a-b-plan, the OmpF packing is identical to the previously reported trigonal crystals as well as the recently reported hexagonal crystals suggesting that the two interfaces involved are highly favorable in order to form tight type I membrane protein crystals.

References

Benz R (1988) Structure and function of porins from Gram-negative bacteria. *Ann Rev Microbiol* **42:** 359-393

Cowan SW, Garavito RM, Jansonius JN, Jenkins JA, Karlsson R, Konig N, Pai EF, Pauptit RA, Rizkallah PJ, Rosenbusch JP, et al. (1995) The structure of OmpF porin in a tetragonal crystal form. *Structure* **3:** 1041-1050

Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP (1992) Crystal structures explain functional properties of two E. coli porins. *Nature* **358:** 727-733

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375-383

Dhakshnamoorthy B, Raychaudhury S, Blachowicz L, Roux B (2009) Cation-selective Pathway of OmpF Porin Revealed by Anomalous X-ray Diffraction. *J Mol Biol*

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60:** 2126-2132

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* **62:** 72-82

Garavito RM, Jenkins J, Jansonius JN, Karlsson R, Rosenbusch JP (1983) X-ray diffraction analysis of matrix porin, an integral membrane protein from Escherichia coli outer membranes. *J Mol Biol* **164:** 313-327

Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallisation of proteins. *J Appl Cryst* **24:** 409-411

Kabsch W (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *Journal of Applied Crystallography* **26:** 795-800

Karshikoff A, Spassov V, Cowan SW, Ladenstein R, Schirmer T (1994) Electrostatic properties of two porin channels from Escherichia coli. *J Mol Biol* **240:** 372-384

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372:** 774-797

Leslie AGW, Powell HR (2007) Processing diffraction data with MOSFLM. *Evolving Methods for Macromolecular Crystallography* **245:** 41-51

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491-497

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658-674

Miedema H, Vrouenraets M, Wierenga J, Eisenberg B, Schirmer T, Basle A, Meijberg W (2006) Conductance and selectivity fluctuations in D127 mutants of the bacterial porin OmpF. *Eur Biophys J* **36:** 13-22

Nikaido H (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67:** 593-656

Ostermeier C, Michel H (1997) Crystallization of membrane proteins. *Curr Opin Struct Biol* **7:** 697-701

Pauptit RA, Zhang H, Rummel G, Schirmer T, Jansonius JN, Rosenbusch JP (1991) Trigonal crystals of porin from Escherichia coli. *J Mol Biol* **218:** 505-507

Pezeshki S, Chimerel C, Bessonov AN, Winterhalter M, Kleinekathofer U (2009) Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF. *Biophys J* **97:** 1898-1906

Prilipov A, Phale PS, Van Gelder P, Rosenbusch JP, Koebnik R (1998) Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli. *FEMS Microbiol Lett* **163:** 65-72

Schirmer T (1998) General and specific porins from bacterial outer membranes. *J Struct Biol* **121:** 101-109

Schirmer T, Keller TA, Wang YF, Rosenbusch JP (1995) Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267:** 512-514

Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* **57:** 122-133

Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374:** 300-321

Yamashita E, Zhalnina MV, Zakharov SD, Sharma O, Cramer WA (2008) Crystal structures of the OmpF porin: function in a colicin translocon. *EMBO J* **27:** 2171-2180

Figures

Table II.1: Comparison of the different crystallization conditions leading to OmpF crystals.

PDB code	Space group	Resolution (Å)	Protein concentration (mg/ml)	Detergent	Buffer	Salt	Precipitant	Reference
2omf (old)*	P321	2.4	10.0	0.1 % C8E4, 0.6 % C8HESO	50 mM Tris pH 9.8	0.7 M $MgCl_2$	10.5 % PEG 2K	Cowan et al, 1992
1opf*	$P4_2$	3.1	10.0	0.5 % OG	100 mM phosphate pH 7.0	0.5 M NaCl	25 % PEG 4K	Cowan et al, 1995
2zld[#]	$P6_3$	3.0	5.0	0.8 % OPOE	100 mM MES pH 6.5	-	15 % PEG 6K 5 % MPD	Yamashita et al, 2008
2zfg	P321	1.59	5.0	0.8 % OPOE	100 mM MES pH 6.5	1M $MgCl_2$	15 % PEG 6K	This work
2omf (new)	P321	2.1	12.5	0.1 % C8E4, 0.4 % C8HESO	50 mM Tris pH 9.8	0.7 M $MgCl_2$	10 % PEG 6K	This work
ortho	$C222_1$	2.7	14.0	0.1 % C8E4, 0.4 % C8HESO	100 mM HEPES pH 7.5	3.2M NaCl	-	This work

*: obtained using microdialysis, all other crystal forms were obtained using vapour diffusion method

#: crystallized with colicin peptide

116

Table II.2: Diffraction and refinement statistics

Diffraction data

Space group	P321	C222$_1$
Wavelength	0.9000 Å	1.0000 Å
Resolution limits #	58.0 – 2.1 (2.21 – 2.1)	50.0 – 2.7 (2.9 – 2.7)
Molecules per asymmetric unit	1	3
Cell dimensions (Å, °)	117.1, 117.1, 51.6, 90.0, 90.0, 120.0	117.4, 202.1, 148.3, 90.0, 90.0, 90.0
R$_{merge}$ #	8.2% (27.7%)	13.3% (43.3%)
I/σ(I) #	24.0 (8.4)	15.5 (5.0)
Number of unique reflections #	23977 (3466)	47103 (8585)
Multiplicity #	11.0 (10.8)	7.47 (7.59)
Completeness #	99.8% (100.0%)	96.7% (92.7%)

Refinement statistics

Resolution range (Å)	58.0 – 2.1	50 – 2.7
R / Rfree	16.1% / 19.8%	21.3% / 24.4%
Rmsd bound length	0.009 Å	0.011 Å
Rmsd bound angle	1.224°	1.235°
Total number of atoms (protein/detergent/solvent)	2627 / 75 / 227	7877 / 40 / 70

#: Numbers in brackets are values for the high resolution shells.

Table II.3: Residues involved in the lateral apolar van der Waals contacts between OmpF trimers.

Interface 1
Asp92*
Val93
Tyr139
Phe144
Phe145
Leu147
Val148
Leu151
Phe153
Ala154
Val155
Tyr157*
Val174
Ile178
Tyr191

Interface 2
Phe265
Phe267
Leu269
Ile273
Val297
Tyr313
Ile315

* may form a hydrogen bound between the main chain oxygen of Asp92 and the hydroxyl group of the Tyr157 side chain.

The same residues are involved in all the crystal forms (P321, C222$_1$, P6$_3$) composed of layers of 32 symmetry.
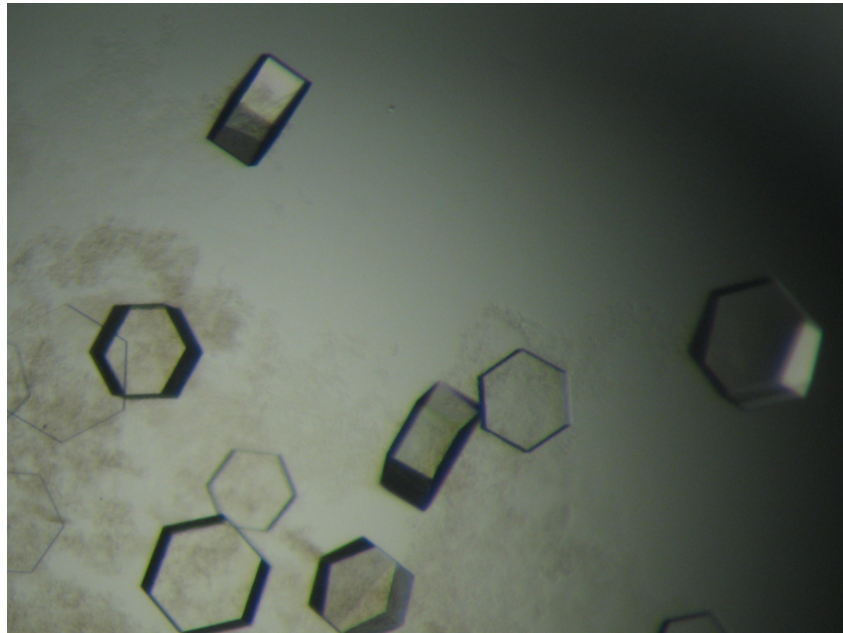
Figure II.1: OmpF crystal with hexagonal morphology belonging to space-group P321 (top) and C222₁ (bottom)
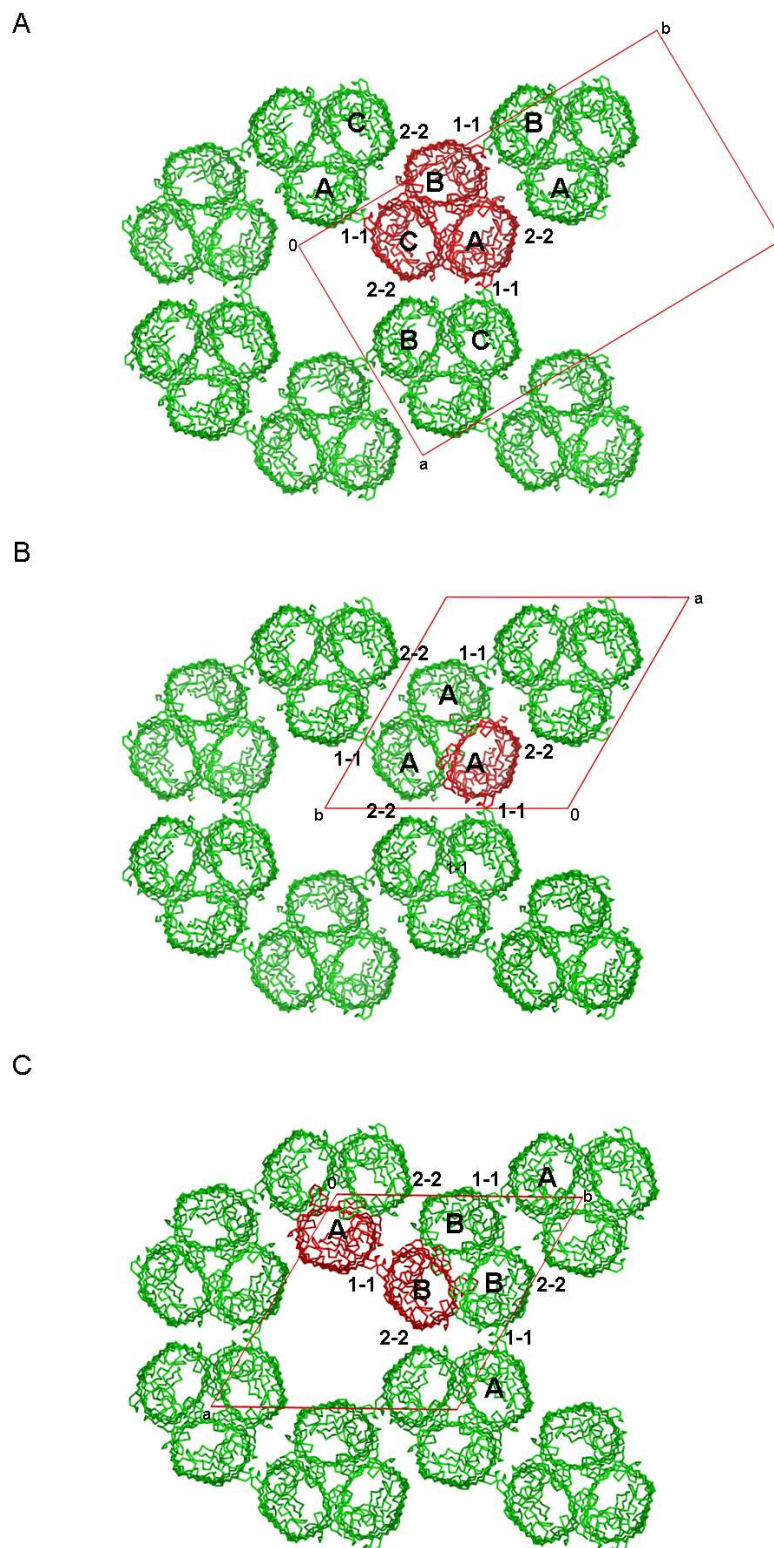
Figure II.2: Crystal contacts in the a-b-plane and comparison of the orthorhombic form with the trigonal and hexagonal crystal forms of OmpF.

(A) Ribbon representation of the OmpF trimer arrangement in the a-b-plane of the orthorhombic crystal, (B) of the trigonal crystal form and (C) of the hexagonal crystal form. OmpF coloured in red are representing the molecules present in the asymmetric unit.
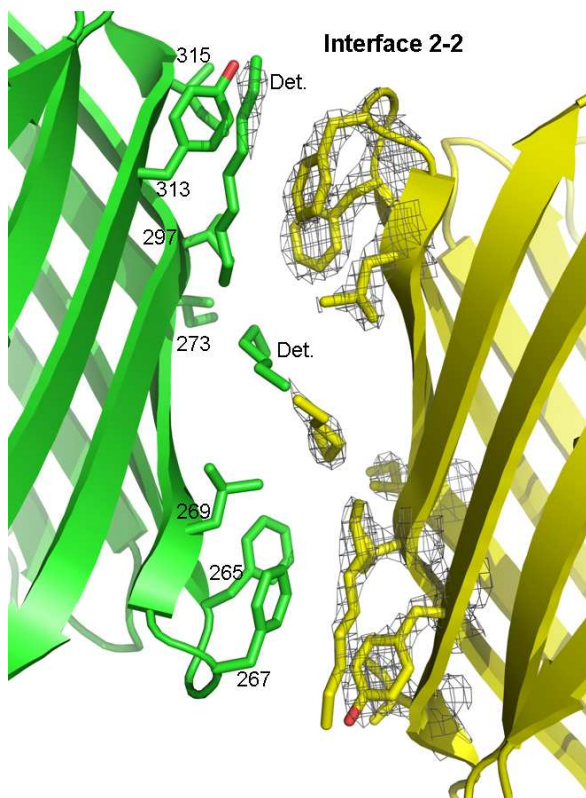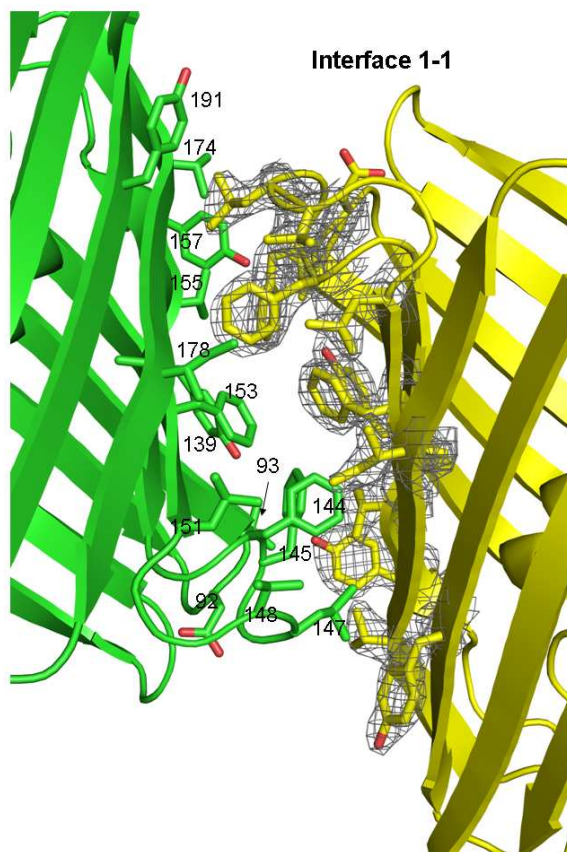
120

Figure II.3: Isologous hydrophobic crystal contacts located on 2-fold axes.

These contacts are observed in the orthorhombic, trigonal and hexagonal crystal forms. Interface numbering is referring to table 2. Note the presence of a detergent or lipidic alkyl chain in the interface 2-2 (marked det.).
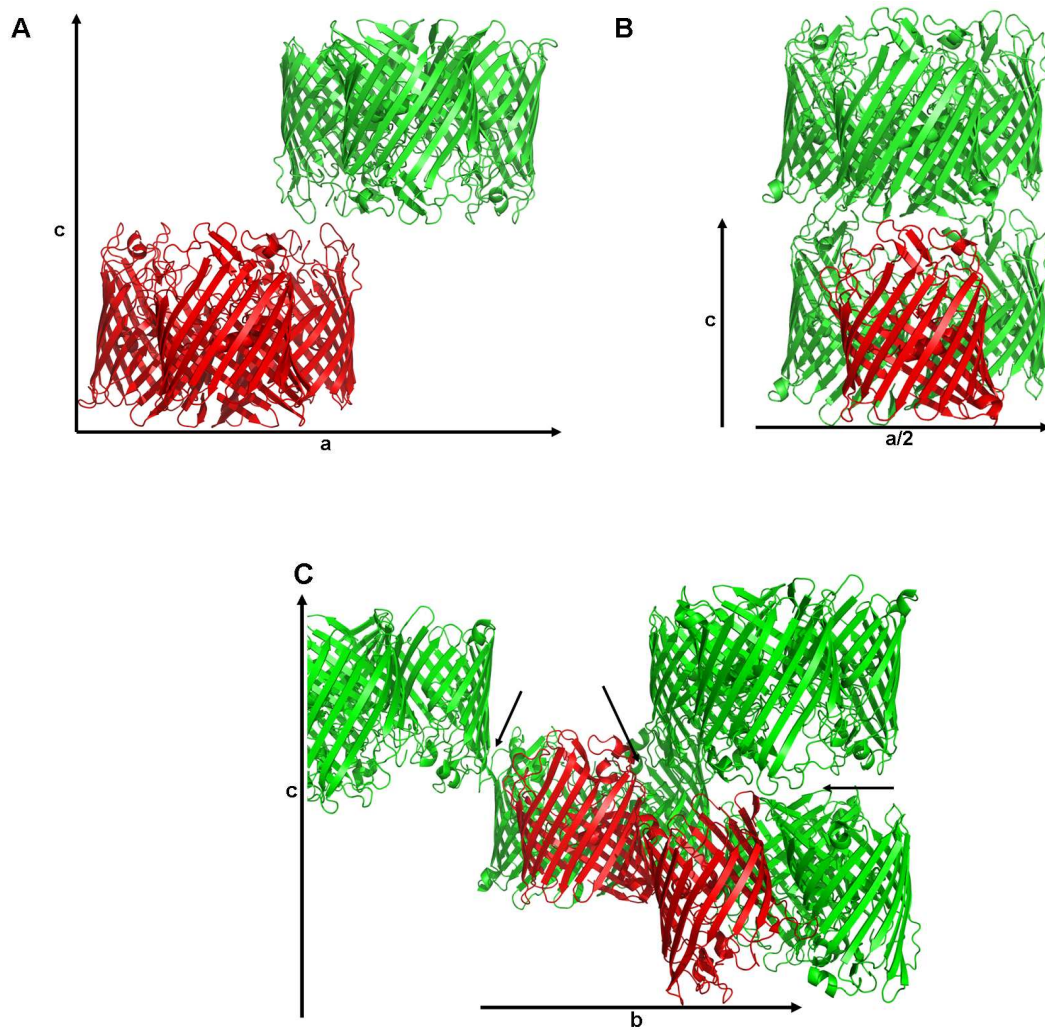
Figure II.4: Crystal contacts along the c axis

Comparison between crystal arrangement along the c axis in the (A) orthorhombic, (B) trigonal and (C) hexagonal crystal forms (arrows indicate the contact sites).

Chapter III

Expression, purification and crystallization assays of the sucrose permease IIBC and its complex with a Fab fragment

Christophe Wirth, Caroline Peneff, Jun-ichi Saito[1] and Tilman Schirmer[#]

Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, 4056-Basel, Switzerland.

1. Present address: Research Division, Kyowa Hakko Kirin Co., Ltd., 1188 Shimatogari Nagaizume, Sunto-gun, Shizuoka 411-8731, Japan.

#: Corresponding author tilman.schirmer@unibas.ch

Abstract

The inner-membrane of Gram-negative bacteria is a selective barrier allowing the specific uptake of nutrients or signalling molecules and release of waste molecules. Amongst the existing nutrient uptake systems, the phosphoenolpyruvate-dependent carbohydrate transport system (PTS) is the structurally least well-characterized. The PTS complex is belonging to the group translocation super-family of permeases. The transmembrane moiety (IIC protein or domain) of the PTS is allowing specific translocation of a sugar molecule and its concomitant phosphorylation by the IIB protein or domain. The PTS is also involved in the bacterial virulence and in sensing the energy state of the cell and therefore regulating the general metabolism. Here we report the expression and purification of the sucrose specific IIBC^{suc} permease (sharing domain B and C on a single polypeptide chain). In order to enhance the crystallization ability of the permease and to rigidify its structure, Fab fragments were produced and purified before IIBC^{suc}-Fab complexes were formed. Crystallization assays were realized using both IIBC^{suc} and IIBC^{suc}-Fab complexes and crystals were obtained. The diffraction quality was promising but no optimisation could be done as the raised crystals were not reproducible. Topology models will also be discussed.

III.1. Introduction

Bacteria are protected from the environment by membranes that are also playing an essential role in sensing signals coming from the surrounding environment and in the uptake of all nutrients essential for bacterial survival and development. They play a permeability barrier role in order to form discrete compartments in the bacteria. Gram-negative bacteria are surrounded by two membranes delimiting the periplasm.

The outer-membrane is an asymmetric selective barrier composed of a layer of phospholipids on the periplasmic side and a layer of lipopolysaccharide turned to the outer environment. Small molecules of a typical size of less than 600 Da can freely diffuse through the outer membrane thanks to β-barrel proteins called porins (Nikaido, 2003). These porins can be specific or non-specific to one particular molecule (Schirmer, 1998). Bigger molecules can be transported into the periplasm thanks to energy dependant transporters (Ferguson & Deisenhofer, 2002).

The inner-membrane (or cytoplasmic membrane) is composed of a phospholipid bilayer. In contrast to the outer-membrane where substrate translocation is done by wide opened porins, at the inner-membrane level, the transport is done by numerous channels and transporters displaying a high specificity and a very different behaviour. The cytoplasmic membrane is used to store energy by forming a proton gradient between the periplasm compartment and the cytoplasm. To keep this proton gradient, wide open, hole forming, proteins cannot exist in the inner-membrane as they would result in proton leakage. Therefore the channel and transporter proteins located in the cytoplasmic membrane are usually shaped as switches that are able to tightly control the substrate translocation without unwanted release of protons. All the transmembrane domains of inner-membrane proteins structurally characterized yet are α-helical. This probably confers a high flexibility to the transmembrane protein allowing it to change conformation during substrate uptake.

Depending which type of energy they use to translocate their substrate, inner-membrane transporters can be classified in three different super-families (Saier et al, 2009). First, proteins that can be called channels as they are acting as passive transporter allow the translocation of ions or molecules downhill their concentration gradient. These proteins do not

require additional energy to function. A second group, also called transporters, needs energy in order to translocate molecules or ions, against their concentration gradients. The energy can be provided by ATP hydrolysis, light or by concomitant transport of another molecule downhill its concentrations gradient such as in syn- or antiporters. A third super-family of inner-membrane transporters is called group translocators. This family is coupling the translocation of a given molecule with its chemical modification (i.e. phosphorylation).

The transport of carbohydrates is mediated by the phosphoenolpyruvate (PEP): sugar phosphotransferase system (PTS) belonging to the group translocation family (Saier et al, 2009). The PTS are ubiquitous in eubacteria but are not found in archaebacteria or eukaryotes (Robillard & Broos, 1999) and typically, up to 20 different PTS proteins are found in a bacteria. The PTS is also implied in several sensing and signal transduction phenomena (Lengeler & Jahreis, 2009) as well as regulation of virulence (Tsvetanova et al, 2007). The PTS allows sensing of saccharides in the growth environment as well as the energy state of the cell.

The PTS is a multi-protein system transporting sugars and hexitols from the periplasm to the cytoplasmic with their concomitant phosphorylation. The system consists in a membrane-spanning transporter and in several soluble proteins implied in a phosphate transfer cascade leading finally to sugar phosphorylation (figure III.1). There are two general PTS proteins called enzyme I (EI) and histidine-containing phosphocarrier protein (Hpr) and a variable number of sugar specific transporting and phosphorylating complexes called enzymes II. The enzyme II complex is composed of three to four functional parts (IIA, IIB, IIC and sometimes IID), which can be found as a single peptidic chain forming three domains or as individual proteins forming a heteromultimer complex (Postma et al, 1993). This is the result of evolutionary fusion and splitting mechanisms that do not reflect mechanistic differences.

A phosphoryl group is transferred from PEP to His15 of the phosphocarrier protein Hpr by enzyme I (transiently phosphorylated during the reaction on His189). Hpr transfers the phophoryl group to the different sugar specific transporting complexes. The phosphoryl group is first transferred to the IIA subunit which, then, phosphorylates the IIB subunit. When a sugar molecule is translocated into the cytoplasm by the IIC domain, it is immediately phosphorylated by the IIB subunit (figure III.1). The enzyme II complexes (EII) are also able

to realize the so-called exchange reaction, in which a phosphorylated carbohydrate is used as phosphoryl donor to phosphorylate another carbohydrate (Robillard & Broos, 1999).
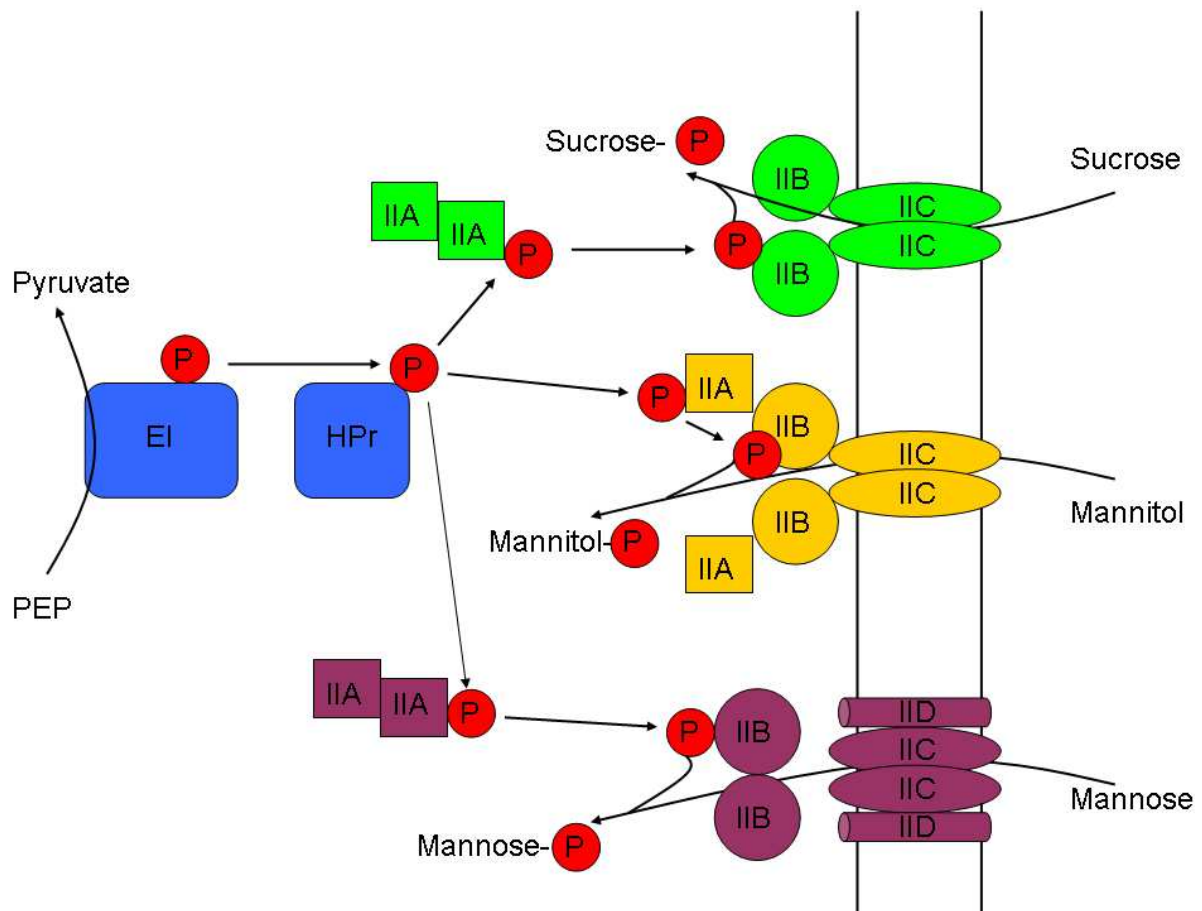

Figure III.1: Schematic representation of the mannitol, mannose and glucose specific PTS.

The enzymes common to all sugars are represented in blue and each sugar-specific system is represented in a different colour.

Experiences on the mannitol specific EII complex have shown that there is a coupling between the transport and the phosphorylation activities (Elferink et al, 1990) as the transport of the sugar is much more efficient when EII is phosphorylated. EII complex is also able to transport the sugar without being phosphorylated. In that case, the transport efficiency is two orders of magnitude lower that in the phosphorylated form of the complex. The transport is not necessary to allow phosphorylation of the carbohydrate by the IIB domain as shown using transport deficient mutants (Manayan et al, 1988).

Apart HPr and EI that are common for all sugars, EII can be classified in five different families according to their sequences similarities. These are the glucose-sucrose, the mannose-sorbose, the mannitol-fructose, the lactose-cellobiose and the glucitol-galactitol families (Lengeler et al, 1994; Postma et al, 1993; Robillard & Broos, 1999; Tchieu et al, 2001). Within a family, EII proteins are sharing more than 25% overall sequence identity and domains can often be exchanged and complemented. This is not the case between members of different families for which the high sequence identity is only limited to some specific regions and therefore, no domain exchange can be done. All members of these families are believed to function as dimers (Erni et al, 1989; Lolkema et al, 1992; Ruijter et al, 1992).

In the glucose-sucrose family, IIB and IIC are always fused in either IIBC (sucrose) or IICB (glucose) polypeptide chains. IIA domain can be either a free, distinct protein or can be linked to the IICB protein, forming a IIBCA protein. Intermolecular complementation has been described between several members this family (Hummel et al, 1992; Schnetz et al, 1990; Vogler et al, 1988) suggesting a modular evolution of this family from single domain proteins to a more complex multi-domain polypeptide. This family is phosphorylating glucose containing sugars at the C6 position (Postma et al, 1993) and has a high affinity: IICB^{glu} has a Kd of 1.5μM for glucose (Ruijter et al, 1992). According to hydropathy and PhoA and LacZ protein fusion analysis, 8- and 6-transmembrane helices folds have been proposed for IICB^{glu} and a IIBC^{suc} respectively (Buhr & Erni, 1993; Titgemeyer et al, 1996).

The mannitol enzyme II system of *E. coli* consists in a single IICBA^{mtl} polypeptide chain that is having two binding sites of high and low affinities for mannitol (Lolkema et al, 1992). Protein fusion experiments suggest that IICBA^{mtl} folds in a 6-transmembrane helices structure (Sugiyama et al, 1991).

In contrast with the other families, mannose specific enzyme II is composed of two membrane bound subunits (IIC and IID) forming a IIABCD^{man} complex. The mannose binding side is located on the IIC domain, but IID is required to form a fully active enzyme (Mao et al, 1995; Rhiel et al, 1994). Stoechiometry analysis suggests a complex composed of two IIAB^{man}, one IIC^{man} and two IID^{man} protein domains (Erni et al, 1987). The IIC domain has been predicted as having six α-helices whereas the IID domain is anchored into the membrane by a single α-helix (Huber & Erni, 1996).

132

Over the last 15 years, structural information on the EI, HPr and EII proteins of the PTS has been accumulated (figure III.2) (Siebold et al, 2001). Full length structures of EI have been determined in both phosphorylated and dephosphorylated states (Marquez et al, 2006; Oberholzer et al, 2009; Teplyakov et al, 2006). EI is composed of two domains: the N-terminal (EIN) binding the HPr and the C-terminal domain binding the PEP. The first HPr structure was characterized in 1992 (Herzberg et al, 1992) using X-ray crystallography and several others, including NMR structures have been obtained to intensively study this protein (van Nuland et al, 1994). Complexes between EIN and HPr could also been obtained (Garrett et al, 1999; Liao et al, 1996) giving insight into the phosphate transfer mechanisms.
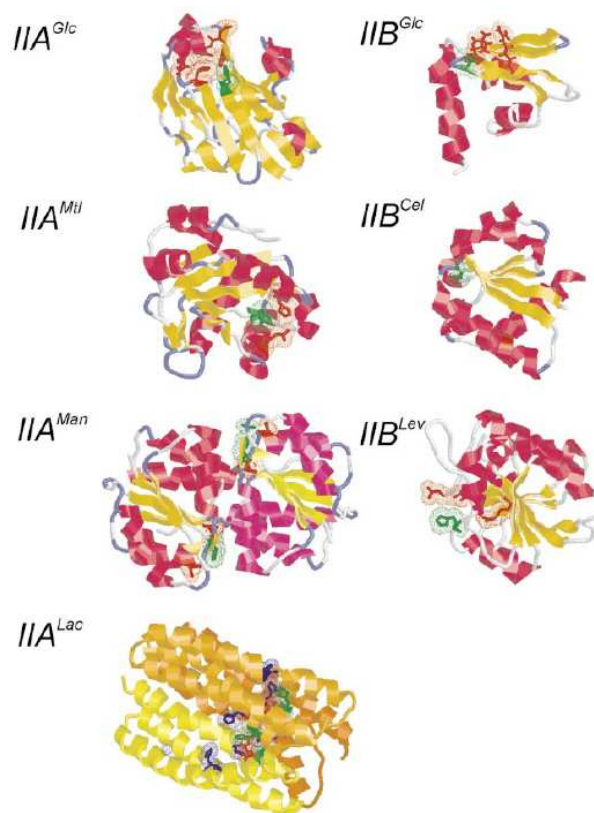

Figure III.2: Cartoon representation of IIA and IIB domains belonging to different PTS families. Figure taken from Siebold et al, 2001.

Phosphorylated residues (green) and other active site residues (red and blue) are represented.

Many structures of EIIA and EIIB have also been solved. Both are folding differently according to their family (figure III.2) (Robillard & Broos, 1999; Siebold et al, 2001). EIIA representatives from all four PTS classes have been characterized. IIA^{glu} of *E. coli* is folding in a sandwich of two six-stranded β-sheets and two α-helices (Worthylake et al, 1991). EIIA^{man} of *E. coli* is forming a homodimer of α/β subunits stabilized by a strand exchange

between the two involved monomers (Nunn et al, 1996). IIA^{mtl} is a 5-stranded β-sheets surrounded by α-helices (van Montfort et al, 1998) whereas IIA^{lac} is forming a homotrimer composed of three-helix-bundles subunits (Sliz et al, 1997). EIIB are also folding in different ways according their family. IIB^{glu} is folding in a four-stranded β-sheet flanked by three helices on one side (Eberstadt et al, 1996) whereas IIB^{cel} and IIB^{man} are folding in open-twisted four- and seven- stranded β-sheets respectively (Schauder et al, 1998; van Montfort et al, 1997). More recently additional structural information on the less studied glucitol-galcatitol family, has been obtained, completing the overview on these proteins (Volpon et al, 2006). Structure of EIIA-EIIB complexes have also been obtained recently (Cai et al, 2003) giving further insight in Enzyme II structures and interactions. However, both EIIA and EIIB have variable folds according to their family, therefore the atomic details of EIIA and EIIB complex formation are still unclear for most of the EII families as well as the molecular mechanisms of the non-specific interaction between HPr and the IIA enzymes of each sugar-specific family.

Natural plasmids are small circular DNA fragments present in bacteria. They are generally carrying a particular function or ability that can be transmitted from one bacterium or specie to the other by conjugation. The 70kb pUR400 plasmid is conferring the ability to use sucrose as sole carbon source as well as tetracycline resistance to its host. Examples have been isolated from *Salmonella* and *E. coli* strains (Schmid et al, 1982; Wohlhieter et al, 1975) indicating that this plasmid is transmissible from one specie to the other (Wohlhieter et al, 1975). The pUR400 plasmid is carrying an operon allowing the uptake and catabolism of sucrose (Cowan et al, 1991). This operon is composed of five genes: (I) *ScrK* encoding for an ATP-dependent fructokinase, (II) *ScrY* for a sucrose specific porin, (III) *ScrA*, for a sucrose specific PTS enzyme IIBC^{suc}, (IV) *ScrB* encoding for β-D-fructofuranoside fructohydrolase cleaving sucrose 6-phosphate into β-D-fructose and a-D-glucose 6-phosphate and (V) the negative *Scr* operon regulator encoded by *ScrR*. The ScrA gene product, called IIBC^{suc}, is belonging to the glucose-sucrose family of the PTS (Postma et al, 1993).

Obtaining x-ray grade crystals remains the principal bottleneck of membrane proteins structural elucidation. This is mostly due to the high flexibility of these proteins as well as, the limited size of the hydrophilic areas on the protein surface susceptible to form strong and specific crystal contacts. In order to overcome this problem, several attempts were done using

134

monoclonal antibodies (mAb) Fab or Fv fragments (Hunte et al, 2000; Hunte & Michel, 2002; Ostermeier et al, 1997; Rasmussen et al, 2007; Uysal et al, 2009; Zhou et al, 2001). The tight binding of the Fab or Fv is believed to stabilize the structure of the membrane protein and increasing substantially the size of its hydrophilic part. This can eventually allow formation of tighter crystal contacts. In addition, when using a Fab targeting a structural epitope (i.e. an element of the folded protein) the membrane protein can be trapped in a certain conformation that could limit its flexibility and increase its crystallization ability (Hunte & Michel, 2002). In our study of IIBC^{suc}, we used this approach using different Fab fragments.

Because of its ubiquitous presence in bacteria and its absence in archaebacteria and eukaryotes, the PTS is a potential anti-microbial drug target. In order to complete the structural knowledge of the PTS, structures of the transmembrane domain of EII from the different families have to be obtained. Moreover, group translocation proteins are one of the few inner-membrane family class for which no structural data is available. Here we present our over-expression and purification results on the IIBC^{suc} from *S. typhimurium* as well as production and purification of anti-IIBC^{suc} Fab antibody fragments. Both protein and Fab were to form IIBC^{suc}-Fab complexes that were utilized for crystallization experiments. IIBC^{suc} *in silico* study will also be discussed.

III.2. Material and methods

III.2.1. Bioinformatic analysis

In order to try predicting the IIBC^{suc} secondary structure, several programs, such as TopPred (Claros & von Heijne, 1994), TMHMM (Sonnhammer et al, 1998), SOSui (Hirokawa et al, 1998) or HMMTOP (Tusnady & Simon, 2001) were used with the amino acid sequence of the C-terminal domain of IIBC^{suc}. Homology modelling was also tried out using both HHpred (Soding et al, 2005) and SwissModel (Arnold et al, 2006).

III.2.2. IIBC^{suc} cloning and expression

Cloning

The project was started by Dr. Jun-ichi Saito who made the following IIBC^{suc} constructs. The *scra* gene was amplified from plasmid pUR400 and inserted in a pET15b vector using primers including 6-histidine tag (6xHisTag) to generate tagged proteins either at the N- or C-terminus. The pET15b plasmid, however, contains already a 6-histidine tag at the N-terminus of the inserted sequence. Therefore this cloning led to (I) a His-His-IIBC^{suc} construct having two 6xHisTags at the N-terminus and (II) to a His-IIBC^{suc}-His construct having a 6xHisTag on both termini. The plasmid was then transformed into an *Escherichia coli* C41 strain. The positive transformants were selected using the ampicillin resistance conferred by the pET15b plasmid.

In order to correct these cloning mistakes, new constructs were designed and standard cloning protocols were applied to generate constructs of IIBC^{suc} with only one 6xHisTag as well as IIC^{suc} and IIB^{suc} (constructs of the C- and N-terminal domains of IIBC^{suc}). Using primers corresponding to the N- or C-terminal nucleotide sequence of the IIBC^{suc} gene, an additional 6xHisTag and including an NcoI or NdeI, the IIBC^{suc} gene was amplified using a standard PCR protocol (28 cycles of 30s denaturation at 95°C, 60s of annealing at 53°C and 120s elongation at 68°C) and Pfu polymerase. Primers were designed in order to have a NdeI restriction site on the N-terminal side of the construct and a NcoI on the C-terminal side. PCR product was purified using a preparative agarose gel and amplified fragments were

subsequently subjected to NcoI or NdeI digestion. In parallel, a MidiPrep (Qiagen) allowed production of pRUN plasmid that were also digested with both NcoI and NdeI and subsequently dephosphorylated. After over-night ligation at 4°C, competent XL1-Blue cells (Stratagene) were transformed and grown over-night on a LB plate containing 100 µM ampicillin. Few transformant colonies were grown over-night and a MiniPrep (Qiagen) was done to obtain the plasmid which was subsequently checked by sequencing. Once the plasmid was controlled, it was transformed in BL21(DE3)/pLys for protein expression.

Expression

Initial expression tests were done at various temperatures and IPTG concentrations for induction in order to optimize the amount of protein obtained. Typical tested temperatures were 37, 30 and 25°C. Bacteria were grown to an OD of 0.6 and IPTG was added to final concentrations of 0.1 or 1 mM. Samples were collected at several time points and the amount of protein obtained was checked using Western blot and an anti-HisTag antibody. For over expression, several litres of LB containing 100 µM ampicillin and 20 µM chloramphenicol were used with over-expression protocols designed previously. After various induction times, the cells were harvested by centrifugation (15 min, 10000 g) and frozen at -80°C.

III.2.3. Membrane preparation and IIBC^{suc} extraction

Cells were resuspended in lysis buffer (50 mM HEPES, 400 mM NaCl, 5 mM β-mercaptoethanol, 4 mM EDTA) and disrupted using 3 cycles of French Press (1200 psi). Lyzate was centrifuged at low speed (20 min, 10000 g) to remove cell debris followed by an ultracentrifugation (45 min at 100000 g) to collect the membranes. The total-membrane pellet was resolubilised in 50 mM Tris, 150 mM NaCl. In order to separate the inner-membrane from the outer-membrane, a sucrose gradient was done. The lower third of the tube was filled with a 50 % sucrose solution in 50 mM Tris, 150 mM NaCl, 5 mM β-mercaptoethanol, and 3 mM EDTA on the top of which a 25 % sucrose solution was slowly pipetted. Finally, the tube was filled with the resuspended membrane solution. Tubes were centrifuged over-night at 200000g. Next day, the fraction containing the inner-membrane was collected, diluted to reduce the sucrose concentration and centrifuged again (45 min at 100000 g) in order to pellet the inner-membrane and to remove eventual remaining soluble proteins. At this point, the

inner-membrane pellet was resupended in 50 mM Tris pH 8.0, 150 mM NaCl and frozen or directly used for subsequent membrane protein extraction.

The IIBC^{suc} extraction was achieved mixing either a 7 % DM or a 0.5 % DDM in 25 mM Tris pH 8.0, 150 mM NaCl, 10 mM sucrose, 5 mM β-mercaptoethanol solution with the same volume of inner-membrane solution to obtain a final solution at 3.5 % DM or 0.25% DDM. The extraction was performed over 2 hours at 4°C. The sample was then ultracentrifuged during 45 min at 100000 g in order to discard the remaining inner-membrane pellet.

III.2.4. IIBC^{suc} purification

The purification of IIBC^{suc} took advantage of the 6xHisTag of the construct. Purification was realized "in batch" using NiNTA beads (Qiagen) allowing a better binding of the IIBC^{suc} to the affinity chromatography material. The extracted IIBC^{suc} protein sample was mixed with 2 ml of NiNTA beads pre-equilibrated in 25 mM Tris pH 8.5, 150 mM NaCl, 10 mM sucrose, 5 mM β-mercaptoethanol (TNS buffer) supplemented with the desired detergent at 3 times the critical micellar concentration. The mixture was incubated for 1 hour at 4°C with gentle mixing. After a washing step with TNS buffer, elution was performed with several increasing steps of imidazol (60, 90, 150, 250 mM). The fraction and purity of IIBC^{suc} was controlled by SDS-PAGE. Subsequently, IIBC^{suc} fraction was concentrated and loaded on a size exclusion chromatography (Superdex 200 or Superose 12 from GE healthcare) in order to exchange the buffer for 10 mM Tris pH 8.0, 150 mM NaCl, 10 mM sucrose and the desired detergent at the desired concentration.

III.2.5. IIB^{suc} purification

Bacteria overexpressing IIB^{suc} were resuspended in 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol and 10 mM imidazol and subjected to French Press (3 times 1200 psi). The cell lysate was centrifuged (1h at 100000 g) and the supernatant loaded on a HisTrap column. The presence of 10 mM imidazol in the binding buffer avoided non-specific binding of contaminants. Elution was done with a linear gradient of 10 to 750 mM imidazol. IIB^{suc}

containing fractions were identified by SDS-PAGE analysis, concentrated and loaded on a superdex 75 gel filtration column (GE healthcare) equilibrated in 10 mM Tris, 100 mM NaCl.

III.2.6. Monoclonal antibody purification

Production

Antibodies were produced by our collaborator Dr. René Fischer (Institute of Organic Chemistry, ETH, Zurich, Switzerland). Purified $\mathrm{IIBC^{suc}}$ solubilized in octyl-glucoside was injected in mice together with adjuvants in order to raise anti-$\mathrm{IIBC^{suc}}$ antibodies. To produce large amounts of monoclonal antibodies, the mouse hybridoma cells production system (Schwaber & Cohen, 1973) was used. Antibodies were produced in cell culture medium containing or not serum with similar efficiency. After production, the cells were removed by centrifugation and the supernatant stored at 4°C. Anti-$\mathrm{IIBC^{suc}}$ monoclonal IgG antibodies (mAb) were selected on their ability to recognize $\mathrm{IIBC^{suc}}$ in a ELISA test but not on a Western blot, allowing the selection of antibodies binding only to folded $\mathrm{IIBC^{suc}}$. An additional antibody able to bind to $\mathrm{IIBC^{suc}}$ in both ELISA and western blot was also selected and produced in order to be used as labeling tool for $\mathrm{IIBC^{suc}}$.

Purification

After 0.22µm filtration, hybridoma cells growth medium containing the mAb was loaded on a HiTrap Protein G HP affinity column (GE healthcare) after column equilibration in 40 mM phosphate pH 7.0. To wash the column, 20 ml of the same phosphate buffer was used allowing removing unspecific binding. The elution was performed using a 0.1 M solution of glycine buffer at pH 2.7. Fractions of 0.5 ml were collected in collection tubes containing 100 µl of 1 M Tris pH 9.0 in order to neutralize the pH avoiding unfolding or precipitation of the IgG. The purity of the sample was checked using a 15% SDS-PAGE electrophoresis in reducing conditions. The samples containing IgG were dialyzed over night and at 4°C against a 100 mM acetate pH 5.5, 2 mM EDTA solution to exchange the buffer in a 12 kDa cut-off dialysis tube (dilution 1:100)

III.2.7. Fab fragment production and purification

Papain digestion

To generate Fab fragment, IgGs were cleaved using papain digestion as previously reported (Padavattan et al, 2007). To set up the protocol and find out ideal digestion conditions, the mAb solution in a 100 mM acetate pH 5.5, 2mM EDTA was mixed with L-cystein acting as reducing agent and activator of papain at a 2.5 or 5mM final concentration. The papain was added at different ratios (typically 1 papain/50 IgG, 1/100, 1/250 and 1/500) and the sample was incubated for several hours at 37°C under agitation. Every 30 minutes an aliquot was taken out of the reaction mix and a 50-fold excess, compared to the papain concentration, of fresh E64 papain inhibitor was added in order to instantaneously stop the reaction. The different aliquots were then analyzed using a 15% SDS-PAGE. Once the best conditions had been set up for a given antibody at a given concentration, the IgG was cleaved routinely using the same protocol on larger amounts. To remove slowly the L-cystein, an over-night dialysis was performed against 25mM acetate buffer at pH 5.5 using a 12 kDa cut-off dialysis tube.

Purification

In order to purify Fab fragments from contaminants, a strong cation exchange chromatography was performed. The dialyzed sample containing the cleaved IgG was loaded on a Mono S HR 5/5 column (GE healthcare) that had previously been equilibrated in 25 mM acetate pH 5.5. The bound proteins were eluted with a very shallow 0 to 500 mM NaCl gradient. To perform optimal separation, the gradient was hold as soon as the OD monitoring shows a raising peak and was started again only when the absorption had returned to the baseline. The eluted peaks were collected in 1 ml fractions and analyzed using a 15 % SDS-PAGE electrophoresis in reducing or non-reducing conditions (no β-mercaptethanol in the sample buffer). Subsequently, a size exclusion chromatography was realized on the Fab fragments purified with the cation exchange chromatography to remove aggregates and to change the Fab buffer. This was performed using a Superdex 75 10/30 column (GE healthcare) equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, 10 mM sucrose. Fractions containing Fab could be stored at 4°C for several weeks.

III.2.8. IIBC-Fab complex formation and purification

To prepare the IIBC-Fab complexes, the solutions containing IIBC^{suc} and the Fab were mixed. A 2.2 to 2.5-fold molar excess of Fab was used and both proteins were incubated together for 30 minutes at room temperature. Prior mixing, the detergent used for IIBC^{suc} purification was added to the Fab solution at a concentration between 2 and 3 times the CMC. Then the sample was concentrated in a 50 kDa Amicon (Millipore) and loaded on a Superdex 200 HR 10/300 size exclusion column (GE healthcare). This final purification step allowed the removal of the Fab excess.

III.2.9. Dot blot experiments

In order to determine to which domain the Fab fragments are binding, dot blot experiments were realized. Few µl of IIBC^{suc} or IIB^{suc} were adsorbed on nitrocellulose paper. The paper was saturated during 2h with a 10% milk solution in 50 mM Tris pH 8.8, 150 mM NaCl (TBS-milk). After few washing steps, the nitrocellulose membrane is incubated 1h with the desired Fab fragment or antibody in TBS-milk with 1:1000 and 1:2000 dilution factors respectively. Anti-Fab antibodies linked to alkaline phosphatase was used as secondary antibody and incubated with the membrane for 1h. Dots were revealed using the alkaline phosphatase activity of the secondary antibody and FAST tablets (Sigma).

III.2.10. Crystallization

Prior crystallization protein samples (IIBC^{suc}, IIB^{suc} and IIBC^{suc}-Fab) were concentrated using Amicon filtering devices (Millipore) of corresponding molecular weight cut-off. The protein concentration for crystallization was ranging from 5 to 20 mg/ml. All crystallization screens were realized at 20°C. Initial crystallization assays were done in 96-well plates using the sitting drop technique (vapour diffusion method). The following commercial crystallization screens were carried out: Crystal screens I and II (Hampton Research), MbClass I, Classic Lite (Qiagen) or Wizard and MemSys/MemStart (Molecular dimensions). Once initial crystallization conditions had been obtained from these screenings, optimization was realized in 24-well plates.

III.2.11. X-ray data collection and processing

Crystals were frozen in liquid nitrogen after a quick soak in a cryoprotection solution containing either 15% ethylene glycol or 15% glycerol. All X-ray diffraction data were collected on beamline X06SA of the Swiss Light Source (Villigen, Switzerland) using typical exposure time of 1s and a MAR CCD detector. The data where processed using MOSFLM (Leslie & Powell, 2007) and scaled with Scala (Evans, 2006). The solvent content of the found solutions was estimated using the Matthews coefficient calculation (Matthews, 1968).

III.3. Results and discussions

Our aim was to solve the crystal structure of IIBC^{suc} using specific Fab fragments in order to enhance its crystallization ability. Before the beginning of this project, IgG monoclonal antibodies (mAb) were raised in mice by Dr. René Fischer (ETH Zurich) using purified IIBC^{suc} and the hybridoma cells production system. Three clones, termed G13, G22 and F24, were selected based on their ability to bind IIBC^{suc} on ELISA or on Western Blot.

G22 and F24 were binding IIBC^{suc} in its native state as both IgG were recognizing IIBC^{suc} in an ELISA test but unable to bind to unfolded IIBC^{suc} in a Western blot. In contrast, G13 was able to recognize IIBC^{suc} in both ELISA test and Western blot. Therefore, G22 and F24 are recognizing a structural epitope whereas G13 is probably recognizing a linear sequence of IIBC^{suc}. *A priori*, G22 and F24 are supposed to bind tightly to the IIBC^{suc} protein enhancing its hydrophilic surface and, as they recognize a structural epitope, stabilizing a given conformation. G13 is binding a linear epitope and is therefore probably less efficient in stabilizing the structure of IIBC^{suc} for crystallization but can be used to detect it in Western Blot for example.

Prior to crystallization trials, mAb purification followed by Fab production and purification as well as IIBC^{suc} purification was required. Both IIBCsuc (with a single or with two Histags) constructs could be used without significant differences.

III.3.1. IIBC^{suc} expression, membrane extraction and purification

Expression

The ideal growth conditions for the *E. coli* strain and the IIBC^{suc} expression protocol were well established after several systematic expression tests. For large scale expression, typical volumes were 10 flasks of 500 ml of LB medium each. The induction was done at OD = 0.6 with 1 mM IPTG and IIBC^{suc} was expressed for 4 hours at 30°. This protocol allowed obtaining routinely an amount of nearly 1.5 grams of dry cells resulting in about 0.5 mg of IIBC^{suc} per litre of culture.

Membrane preparation and protein extraction

Dr. Jun-ichi Saito and Dr. Caroline Peneff, in their initial attempts to crystallize IIBC^{suc}, could obtain crystals diffracting to a resolution at about 4Å. However, after structure solving, it turned out that these where crystals of OmpF, the major porin of the outer-membrane from *E. coli*. A deeper analysis of the reasons of this contamination resulted in the following conclusions. When extracting the membrane proteins directly after cell lysis, OmpF, the major general porin *E. coli* was representing the major contaminant. OmpF contaminants could not be completely removed using Ni-NTA chromatography. To avoid OmpF contamination, an additional sucrose gradient step was included to the membrane preparation protocol. After cell disruption using a French Press, unbroken cells were removed by centrifugation and an ultracentrifugation step allowed the total membrane fraction to be collected. The resolubilized membrane pellet was loaded on a 25 % / 50 % sucrose gradient and ultracentrifuged for several hours. This allowed separation of the inner- from the outer-membrane. The outer-membrane was pelleted at the bottom of the tube whereas the less dense inner-membrane fraction which is unable to enter the 50 % sucrose solution and remained at the 25 % - 50 % interface. The inner-membrane fraction (the 25 % sucrose and about half of the 50 % sucrose solution) was collected, diluted and ultracentrifuged in order to obtain an inner-membrane pellet. The pellet was resuspended in 50 mM Tris pH 8.0, 150 mM NaCl and could be readily used for protein extraction or frozen for several weeks at -80°C.

For protein extraction, the protocols were first tested in small scale experiments using different concentration (ranging from 2 to 100 times the critical micellar concentration) of various detergents such as decyl-maltoside (DM), dodecyl-maltoside (DDM), or thesit (C12E9), the most used for inner-membrane protein extraction. Extraction efficiency was controlled with Western blots and anti-His as primary antibodies (figure III.3). One condition (0.25 % DDM) allows almost complete solubilisation of IIBC^{suc} in a homogenous manner whereas increased DDM concentration or other detergents were leading to inhomogeneous samples (figure III.3). At 0.25 % DDM, the protein was extracted in a single band that could, according to it's size, correspond to the monomeric form of IIBC^{suc}. Other detergents were yielding two bands: the same, likely monomeric, form of IIBC^{suc} and an additional band of higher molecular weight that was likely to consist in IIBC^{suc} dimers. In some cases (0.5 % C$_9$E$_{12}$ or 2.5 % DDM) higher molecular weight aggregates were also visible. Once optimized, the IIBC^{suc} extraction conditions were used routinely with larger amounts of proteins.
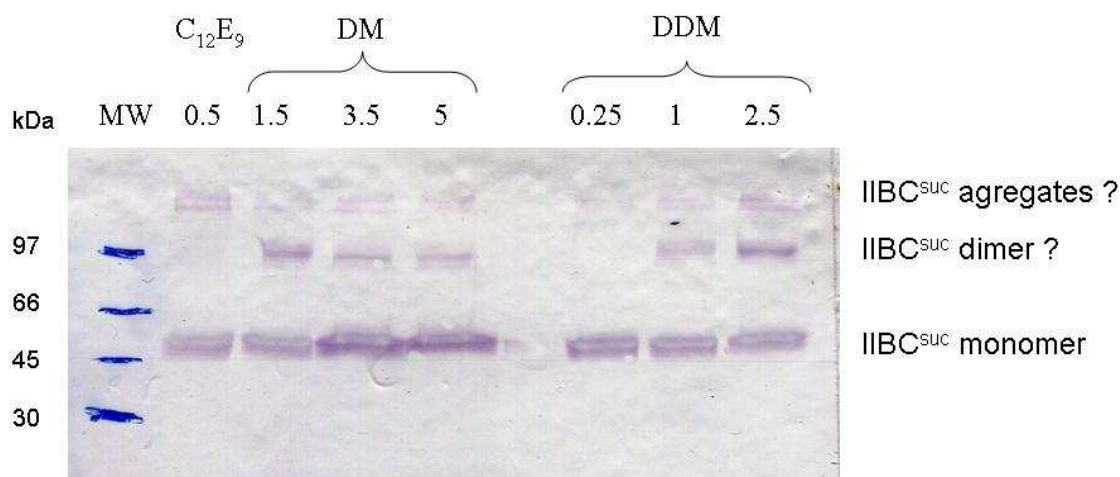
Figure III.3: Western blot of IIBC^{suc} extraction tests using 3 different detergents.

Detergent concentrations are in percentage. Potential IIBC^{suc} oligomeric states are mentioned.

III.3.2. IIBC^{suc} purification

IIBC^{suc} purification took advantage of the 6xHisTag present in the construct. The binding on HisTrap columns (GE Healthcare) was weak and only a small fraction of the total IIBC^{suc} protein could be attached and purified using this column. In order to allow better binding, IIBC^{suc} samples were mixed with Ni-NTA beads and kept at 4°C for 2 hours under gentle agitation. Using beads allowed binding of virtually all IIBC^{suc} to the Ni-NTA material. This is likely to be due to the longer binding time when using beads. However, there was also a substantial difference in binding efficiency between the double- and mono-HisTagged IIBC^{suc}. The almost non-binding of IIBC-His can be explained by an eventual buried HisTag at the C-terminus side of the protein. Elution was realized in steps using increasing imidazole concentration (Figure III.4). Almost all IIBC^{suc} is eluted between 60 and 150mM imidazol.
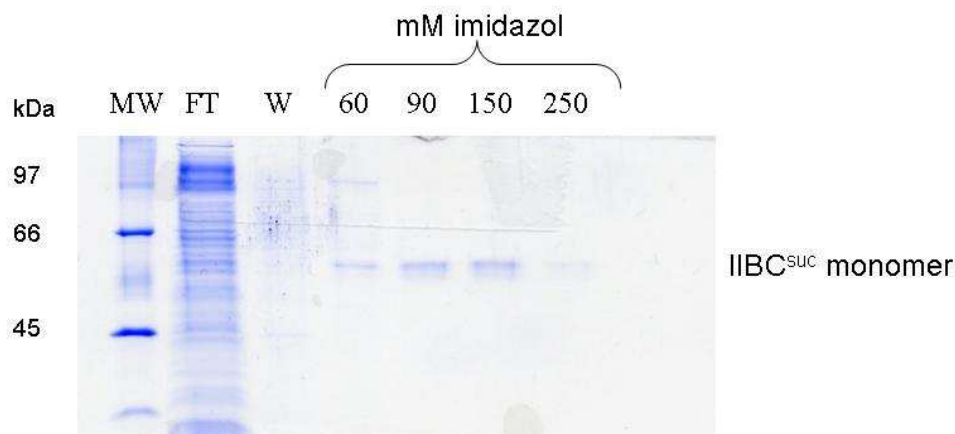

Figure III.4: SDS-PAGE of IIBC^{suc} Ni-NTA affinity purification

Fractions containing IIBC^{suc} are pooled and subjected to gel filtration chromatography in order to exchange the buffer. Detergents are considered as critical for the protein monodispersity. Therefore, only detergents allowing elution of IIBC^{suc} as a monodisperse peak and showing only minor aggregation were conserved for crystallization trials (Figure III.5).
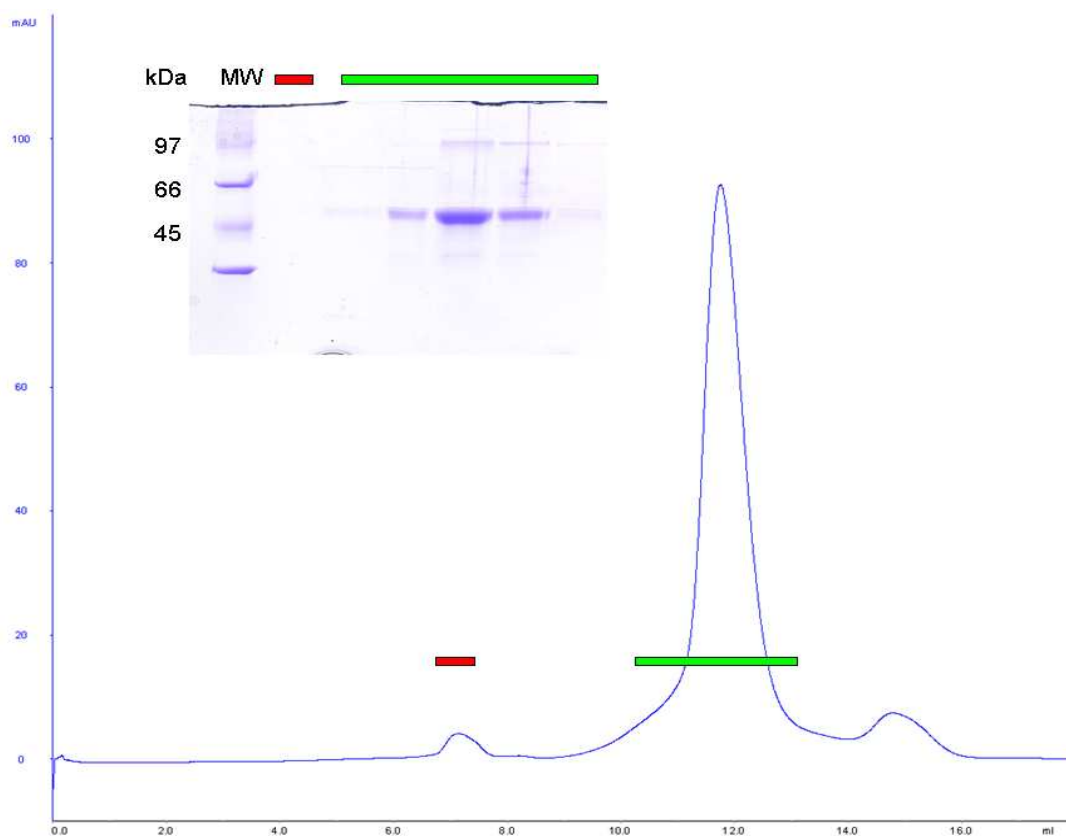

Figure III.5: Gel filtration chromatogram of IIBC^{suc} in DM detergent and corresponding SDS-PAGE

The blue trace represents the UV absorption. Green and red bars correspond to the fractions loaded on the gel.

According to the gel filtration elution peak at a volume of 12.75 ml, IIBC^{suc} might be a dimer in solution (estimated molecular weight of about 100 kDa). In order to confirm this, blue native polyacrylamide gels (BN-PAGE) were done allowing running the protein samples according to their molecular mass. IIBC^{suc} migrates in a rather smeary line, but forms a predominant band at roughly a similar size than the 140 kDa band of the MW ladder.

III.3.3. Antibody purification

The IgG purification is realized using a Protein G affinity chromatography column (HiTrap Protein G HP, GE Healthcare). Protein G is a cell surface type III Fc receptor of the Group G *streptococci* able to bind the Fc region of IgG by a non-immune mechanism similar to that of the protein A from *Staphylococcus aureus* (Akerstrom et al, 1985).

The purification of the different IgG used in that study led to various results. Both G22 and G13 could be easily purified and these IgG are eluting in a single peak, after 1.5 to 2 column volumes (figure III.7). G13 presented a good solubility during the elution contrariwise to G22 that had a tendency to precipitate during the elution. As the acid pH used to elute the IgG was neutralized immediately after elution, this aggregation process must be the result of the very high G22 concentration during the elution of the protein G column. Diluting ten times the samples containing the eluted G22 could solve the aggregation problem. The analysis on SDS-PAGE (figure III.7) revealed, for both G13 and G22 antibodies, the presence of sharp bands corresponding to the heavy and light chains of the purified IgG. A smeary band of nearly the same molecular weight surrounds these bands. This corresponds certainly to the contaminant IgG coming from the serum added to the medium used to raise the hybridoma cells. In the flow through, one major band is visible certainly corresponding to serum albumin present in the medium.
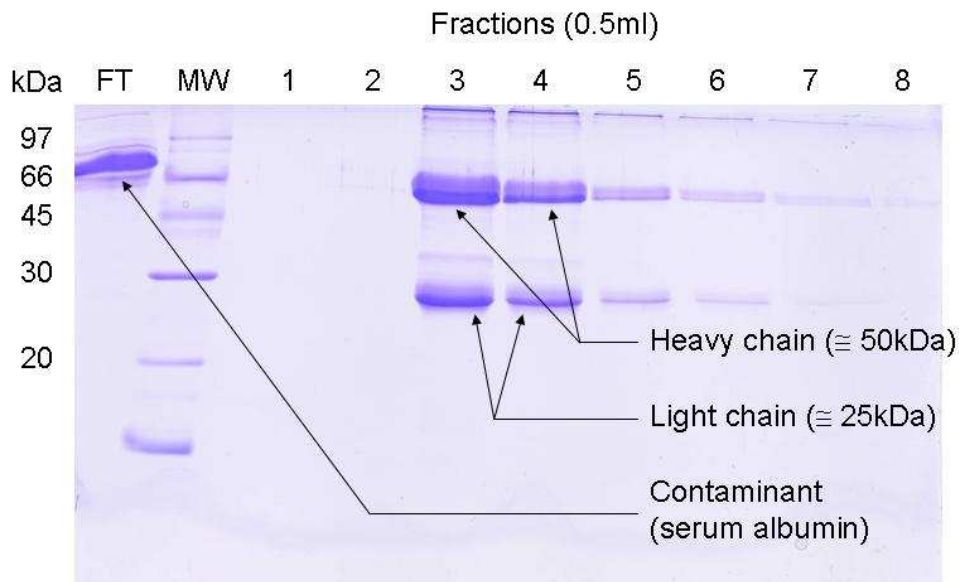

Figure III.7: SDS-PAGE after monoclonal antibody purification

Example of SDS-PAGE in reducing conditions after HiTrap Protein G column purification of G22. FT stands for flow through. Elution with 100 mM glycine pH 2.7 allowed quick release of the IgG. The two characteristic heavy and light chains are visible.

In the case of F24, there was also one single peak on the absorption chromatogram, but no sharp band on the SDS-PAGE. Only the smeary bands were visible corresponding to heavy and light chains of antibodies but not from one single species. As the ELISA tests after expression were positive and as there was no corresponding band neither on the loaded medium sample nor in the flow through sample, we conclude that the expression of F24 must be very low and therefore we decided to not continue working with this IgG.
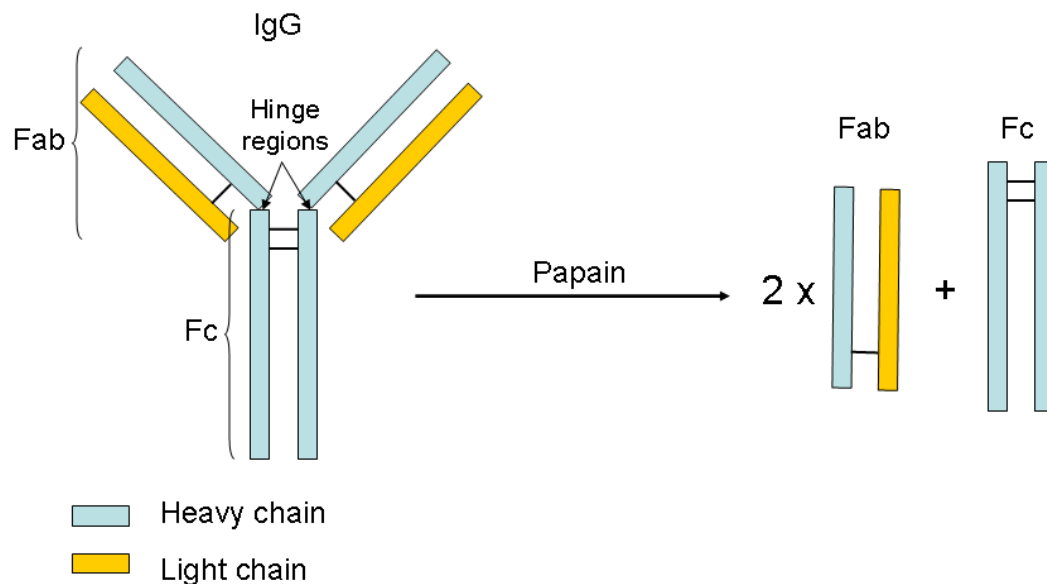
	Bands on SDS-PAGE		
	IgG	Fab	Fc
Reducing conditions	55 kDa (heavy chain) 25 kDa (light chain)	Double band at ~25 kDa	30 kDa
Non-reducing conditions	150 kDa	50 kDa	60 kDa

Figure III.8: Schematic representation of IgG, its papain cleavage product and description of the bands visible on SDS-PAGE in reducing and non-reducing conditions.

III.3.4. Fab production by antibody cleavage

The Fab production protocol was adapted from a previous protocol developed in the lab (Padavattan et al, 2007). Papain is cleaving IgG at the so-called "hinge region", a region that is particularly easy to access for the protease (figure III.8). In order to determine the optimal

cleavage time and conditions, small scale tests were done while varying parameters such as papain concentration compared to the IgG concentration, L-cystein concentration and digestion time (figure III.9). The optimal cleavage condition for G22 was a low concentration of papain (1 papain per 250 IgG molecules) and low L-cystein concentration (2.5 mM). The incubation time for this IgG was 90 minutes. For G13, the best cleavage condition was with a higher concentration of papain (molar ration of 1:50 IgG molecules) and higher L-cystein concentration (5mM). In this case, the optimal incubation time was 45 minutes.
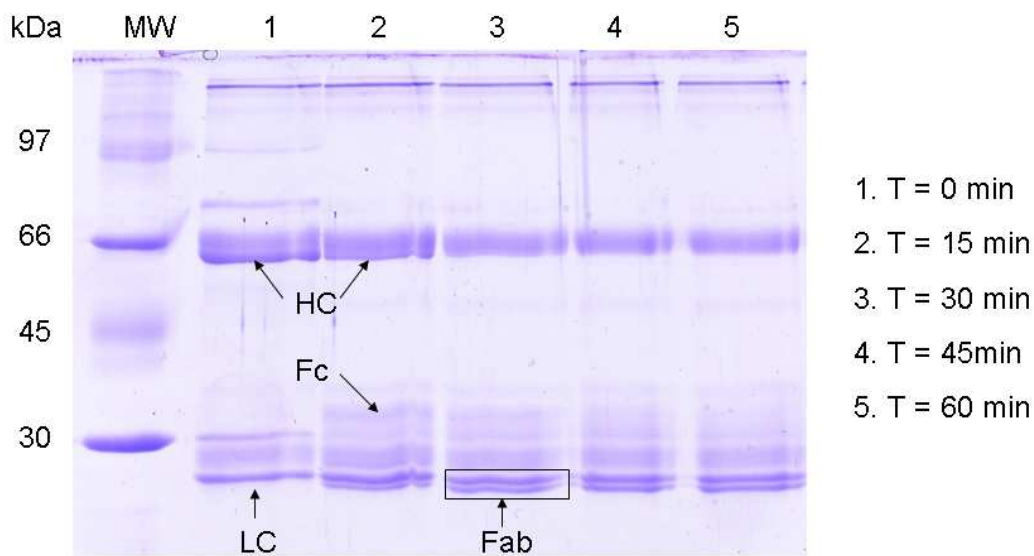

Figure III.9: SDS-PAGE of papain digestion tests of G13.

SDS-PAGE of the optimal G13 cleavage conditions showing the increase of the Fab specific double band at about 25 kDa. The papain to IgG ratio was 1:50 and the L-cystein concentration was 5 mM. The heavy-chain (HC), light chain (LC) are disappearing while the Fc and Fab bands are appearing.

According to Adamczyk *et al.* (Adamczyk et al, 2000) this suggests that G22 could be from the IgG2a or IgG3 subtypes and G13 could be from the IgG1 subclass. As protein G columns are able to bind strongly all these different subclasses of IgG, it was not possible to discriminate these subclasses at the previous purification step.

The cleavage led to a characteristic observation on the SDS-PAGE. For an uncleaved IgG, in reducing conditions, two main bands were visible at nearly 27 and 50-55 kDa. These bands correspond respectively to the light and heavy chain of the IgG (figure III.8). Once the IgG

was cleaved, the SDS-PAGE in reducing conditions was presenting two additional bands. One was very close to the size of the one at 27 kDa and the second band was about 30 kDa in size. In the mean time, the heavy chain band was disappearing on the SDS-PAGE. In fact, the heavy chain band was cleaved into two parts resulting in a Fab composed of the light chain and of the fragment of the heavy chain that is very similar in size compared to the light chain forming a characteristic double band on SDS-PAGE under reducing conditions. The 30 kDa corresponds to the Fc fragment obtained after digestion.

III.3.5. Fab purification

Various proteins and peptide chains were contaminating the Fab samples after IgG cleavage. This includes mostly Fc fragments and uncleaved antibodies but probably also different Fab forms (as the papain cleavage is not sequence specific). Cation exchange chromatography (MonoS, GE Healthcare) was used with a very shallow NaCl gradient (and its eventual "hold" as soon as a protein peak was appearing on the chromatogram), in order to separate different forms of Fab and to conserve only the peaks containing the major Fab form (figure III.10). This procedure turned out to be really efficient in differentiating forms of Fab fragments thanks to their surface charge differences. This purification step was also allowing the removal of Fc fragments that were eluting at higher NaCl concentration than the Fab. According to SDS-PAGE, the fractions belonging to the major Fab peak were mixed and further purified.
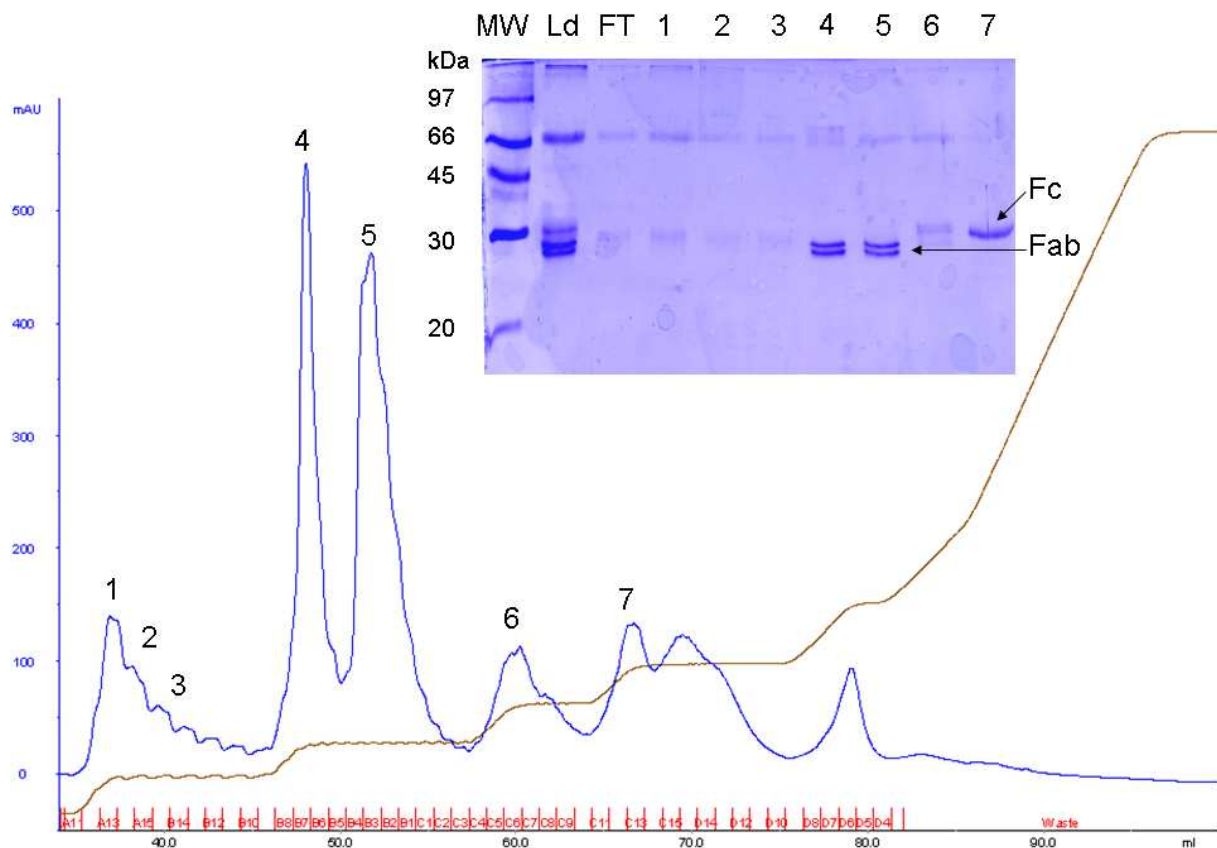
Figure III.10: Example of cation exchange chromatogram of Fab fragments from G22 purification and corresponding SDS-PAGE.

The blue trace represents the 280nm UV absorption. Correspondences between SDS-PAGE lines and chromatography peaks are marked with numbers. Ld stands for "load" and corresponds to the protein solution that was loaded on the column. FT stands for flow through.

In order to finalize the Fab purification, the sample was subjected to gel filtration chromatography (Superdex 75, GE Healthcare). In both G22 and G13 cases, the chromatogram was very similar allowing the separation of the remaining uncleaved mAb contaminants, found in the exclusion volume, from the Fab. Eluted from the column in a monodisperse peak, G22 and G13 Fab fragments were unambiguously identified on SDS-PAGE (figure III.11). Once purified, the Fab fragments could be stored for several weeks at 4°C. However, to avoid aggregation problems, the concentration had to be kept low (below 1 mg/ml) for long term storage.
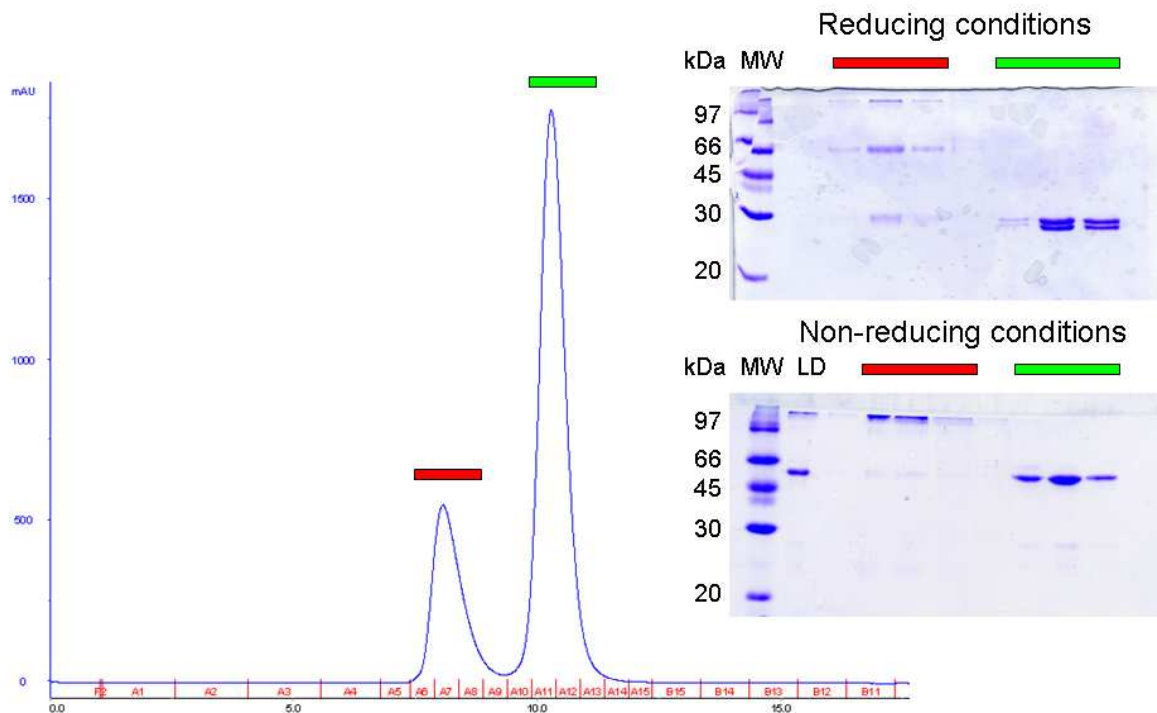
Figure III.11: Example of gel filtration chromatogram of G22 Fab fragment and the corresponding SDS-PAGE.

Gel filtration chromatogram of the UV (280nm) absorption (blue trace). The fractions marked by the red and green bars are representing contaminants (probably non-digested IgG) and purified Fab fragments respectively. The bars are also marking the corresponding lines on the SDS-PAGE in both reducing and non reducing conditions.

III.3.6. IIBC-Fab complex formation and purification

IIBC^{suc}-Fab(G22)

The IIBC^{suc}-Fab complex formation was obtained by mixing both components of the complex and by incubating it under gentle mixing. The protein sample was then concentrated and purified using a Superdex 200 gel filtration column. Two elution peaks were visible on the chromatogram (figure III. 12).
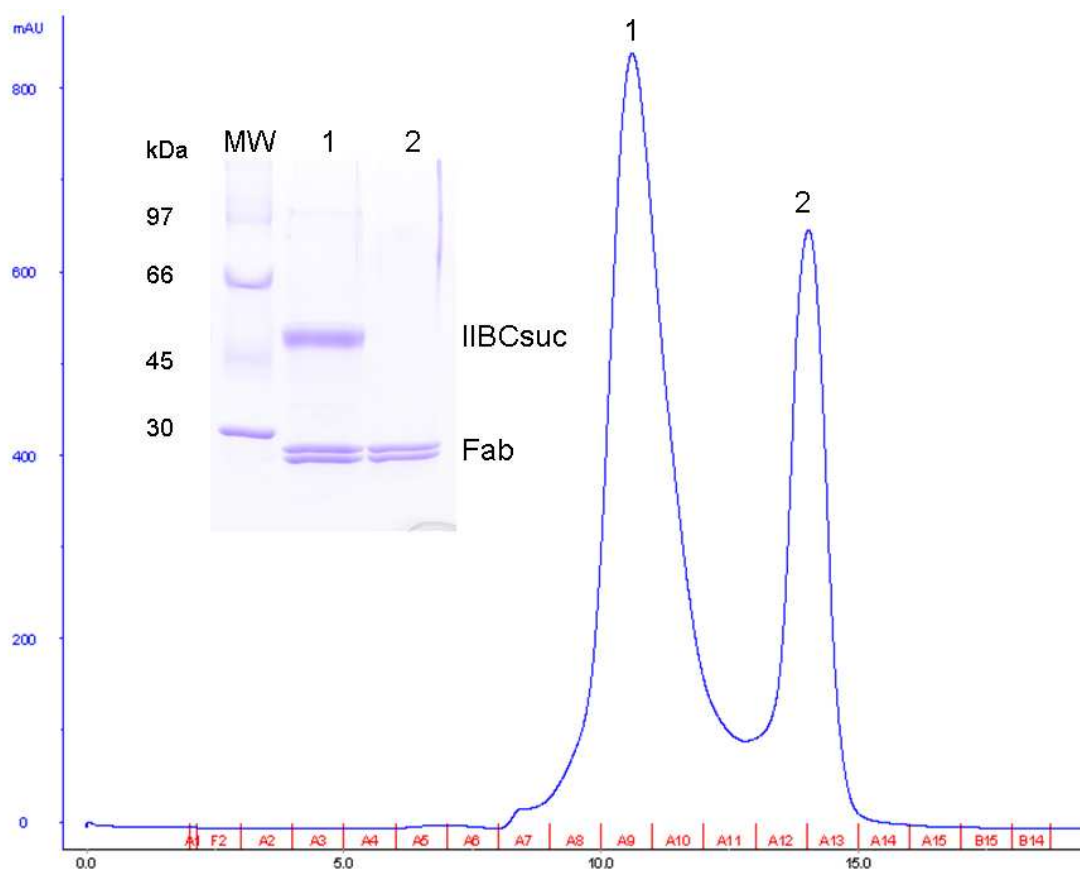
Figure III.12: Gel filtration chromatogram of IIBC^{suc}-Fab(G22) and its corresponding SDS-PAGE.

The blue trace represents the UV absorption. According to the SDS-PAGE, peak 1 was containing IIBC^{suc} and Fab, whereas peak 2 was containing only G22 Fab fragments.

According to the SDS-PAGE, the peak 1, eluting at a molecular weight of about 10.5 ml (about 280 kDa of MW), was probably corresponding to IIBC^{suc}-Fab(G22) complex. The mass determined by the elution volume of this complex corresponded roughly to the sum of a IIBC^{suc} dimer and two Fab(G22) molecular weight (~200kDa) embedded in a detergent micelle. No significant difference in band intensity was visible on the gel suggesting a 1:1 ratio of IIBC^{suc} and Fab(G22). The second peak, corresponding to a lower molecular weight of around 50 kD, wass eluted later. This peak contains the excess of Fab (G22).

IIBC^{suc}-Fab(G13)

G13 is a mAb directed against a linear epitope of IIBC^{suc}. As already observed during papain digestion, G13 mAb belongs probably to the IgG1 family. The difference in MW of the Fab

153

two chains is also smaller than in the G22 case resulting in a single band on a SDS-PAGE in reducing conditions. During the IIBC^{suc}-Fab(G13) complex formation analysis, an interesting feature concerning Fab(G13) has been observed. Compared to IIBC^{suc}-Fab(G22), the complex with Fab(G13) showed a complete different behaviour when subjected to gel filtration chromatography (figure III.12 and III.13). The chromatogram was showing a double peak of high molecular weight in addition to the Fab(G13) excess peak. On SDS-PAGE, these two peaks were consisting in a single identical band of high molecular weight.
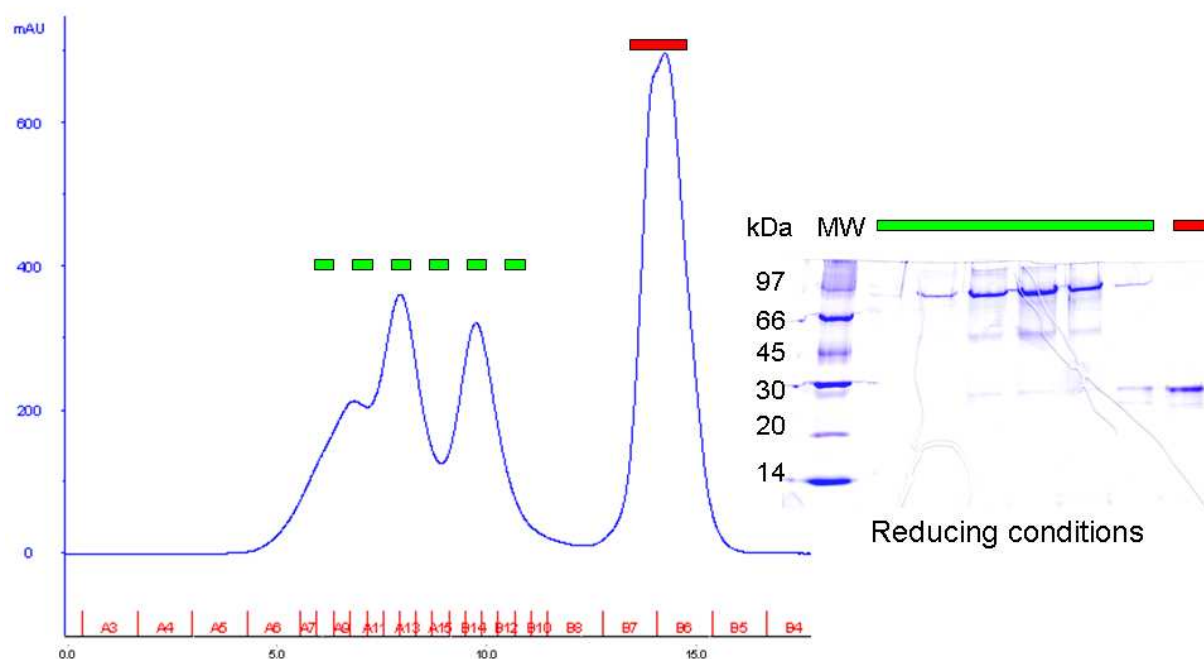

Figure III.13: Gel filtration chromatogram of IIBC^{suc}-Fab(G13) and the corresponding SDS-PAGE.

The blue trace corresponds to the UV-absorption. Green and red bars represent fractions in which IIBC^{suc}-Fab(G13) and Fab(G13) are present respectively.

In order to get insight into the composition of this, SDS-PAGE stable, high molecular weight complex, Western blot experiments were undertaken using several antibodies. The high molecular weight band was recognized by anti-HisTag antibodies as well as by anti-Fab antibodies, but not by anti-Fc antibodies indicating that both IIBC^{suc} and Fab(G13) were present in this single high molecular weight band. This suggests that Fab(G13) is binding to IIBC^{suc} tightly enough to not be disturbed in SDS-PAGE conditions. Increased concentration of SDS or of β-mercaptoethanol did not help to separate the Fab(G13) from its IIBC^{suc} epitope.

154

Only heating of the sample, inducing aggregation of IIBC^{suc} allows obtaining a Fab band situated at the expected Fab size (data not shown). As β-mercaptoethanol had no effect, a disulfide bound between IIBC and Fab(G13) is probably excluded to explain the strong binding of Fab(G13). No effect of SDS was detected as well, suggesting that detergents do not affect the binding of Fab(G13) and that this strong binding may be due to hydrophobic interactions.

The presence of a double peak indicates polydispersity of the complex. The elution volumes of about 11 and 12.2 ml indicate that the two species have a MW of about 250 and 125 kDa respectively. As both peaks contain IIBC^{suc} and Fab(G13) it is likely that the first peak is corresponding to the IIBC^{suc} dimer (100 kDa) in complex with two Fab(G13) molecules (50 kDa each) surrounded by detergents. The second peak could correspond to a IIBC^{suc} monomer with one Fab(G13) bound. If this assumption is true, this would mean that Fab(G13) is binding to an epitope located close to the dimerization interface and that it is disturbing the quaternary structure of IIBC^{suc}.

In order to further characterize the binding of Fab(G13) on IIBC^{suc}, its N-terminal domain (IIB^{suc}) was expressed in *E. coli* and purified using Ni-NTA affinity chromatography and subsequent gel filtration (data not shown). The protein was strongly expressed and easily purified in high amounts (50 – 100 mg per litre of culture). IIBC^{suc} and IIB^{suc} were used for dot-blot experiments with G13 and G22 Fab fragments. No signal could be detected for Fab(G22) probably because it is a conformational epitope and that during the dot blot experiment, IIBC^{suc} is unfolding because of the absence of detergent in the solutions. In contrast, Fab(G13) was binding only to IIBC^{suc} but not to IIB^{suc}, suggesting that Fab(G13) is specifically binding to the transmembrane domain IIC^{suc}. The binding on transmembrane domain corresponds to a major selection criterion for using Fab fragments as it allows the increase of the insoluble part of IIBC^{suc}. The tight binding of G13, as well as its binding location on the IIC^{suc} domain indicates that it might be a good candidate to continue structural investigation. However, the gel filtration chromatogram also suggests that this Fab may disturb the oligomeric state of IIBC^{suc}, eventually leading to problems during crystallization.

III.3.7. Crystallization of IIBC^{suc} and IIBC^{suc}-Fab

IIBC^{suc}

As the IIBC^{suc} crystallization ability had already been intensively investigated, most of the screens realized here were to study the behaviour of IIBC^{suc} when it was extracted and purified using a new detergent or buffer. This was mainly in order to test if the detergent concentration was acceptable and to have a reference for the comparison between IIBC^{suc} and IIBC^{suc}-Fab crystallization screens. Screens designed for membrane specific proteins were used in that case (MemSys/MemStart or MbClassI) at usually rather low protein concentration (3-8 mg/ml).

The crystallization behaviour of IIBC^{suc} was tested in 0.3 % DM at up to 8 mg/ml, 0.025-0.03 % DDM at up to 10 mg/ml and 0.05 % C12E9 at 3mg/ml of IIBC^{suc}. In all cases the screens presented clear and precipitated drops as expected for a protein in suitable concentration for crystallization. In addition, DM screens often led to many drops with phase separation that was often turning into round solid objects. Although they were containing protein, as assessed using the micro-spectrophotometer, these "crystal" were presenting only powder diffraction.

Different crystals coming from different crystallization conditions were obtained during these assays. All IIBC^{suc} crystals obtained were grown when using DDM for extracting, purifying and crystallizing the protein. The crystals were all obtained at a IIBC^{suc} concentration of 8 mg/ml and using 0.025% as a DDM concentration in the protein buffer prior to crystallization. All conditions obtained with commercial screens are listed in Table III.14.

The typical growing time of these crystals was ranging between 2 and 4 weeks. Crystals were usually very small (number 1 was the biggest with an approximate size of $100 \times 50 \times 30$ µm). These crystals, unfortunately, could never been reproduced and optimized when using another IIBC^{suc} batch.
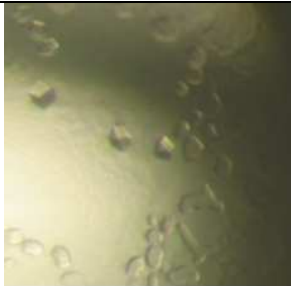
#	Protein buffer	Crystallization condition	Crystal picture
1	20 mM Tris pH 8.0, 10 mM sucrose, 200 mM NaCl, 0.025% DDM	0.1 M Hepes pH 7.5 30 % PEG 400	
2	20 mM Tris pH 8.0, 10 mM sucrose, 200 mM NaCl, 0.025% DDM	0.1 M MOPS pH 7.0 0.1M NaCl 12% PEG 4K	
3	20 mM Tris pH 8.0, 10 mM sucrose, 100 mM NaCl, 5% glycerol, 0.025% DDM	0.2 M imidazole/malate pH 8.5 17.5 % PEG 10K	

Table III.14: List of IIBC^suc crystallization conditions and crystal pictures.

IIBC^suc-Fab

Both IIBC^suc-Fab complexes (with G22 and G13) obtained have been used in crystallization screens. The typical threshold above which the complex was starting to strongly precipitate in the screens was around 15 mg/ml. Crystals have been obtained in different conditions using the IIBC^suc-Fab(G22) complex. The protein solution used to obtain these crystals was coming from a single purification realized using 3.5 % DM for the extraction followed by 0.3 % DM during the IIBC^suc purification and the Fab binding procedure. During the final purification step (to remove the excess of Fab) a detergent exchange was performed for 0.025 % DDM in 20 mM Tris pH 8.0, 100 mM NaCl, 5 % glycerol, 10 mM sucrose. All crystals were grown at a concentration of 18 mg/ml. These very tiny crystals could not be reproduced and optimized yet. The conditions in which they appear are summarized in the Table III.15.
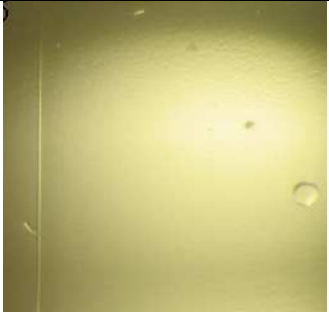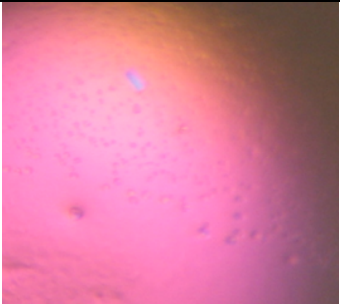
#	Crystallization condition	Crystal pictures
4	0.1 M Hepes pH 7.5, 0.1 M NaCl, 0.1 M LiSO4, 12% PEG 4000	
5	0.1 M Tris pH 8.5, 0.2 M Tri sodium citrate, 30% PEG 400	
6	0.2 M imidazole/malate pH 7.0, 10% PEG 4000	

Table III.15: List of IIBC^{suc}-Fab(G22) crystallization conditions and corresponding crystal pictures.

III.3.8. Diffraction experiments and data analysis of IIBC^{suc} and IIBC^{suc}-Fab

Although they were very fragile, few crystals obtained from each conditions could be fished using cryoloops and frozen in liquid nitrogen. Prior to freezing, crystals from conditions 2 and 3 were quickly soaked into a cryoprotecting solution. The x-ray diffraction experiments were done at 100K on beamline X06SA at the SLS.

IIBC^{suc}

Crystals obtained from conditions 2 and 3 were showing only very weak diffraction with only a few reflections on the diffraction pattern whereas the crystal obtained in condition 1 was showing a clear pattern at a muchbetter resolution (12 Å) (figure III.16). Few diffraction images could be collected allowing indexing with MOSFLM (Leslie & Powell, 2007). These crystals were belonging to space group P2 or P21 and had a = 117 Å, b = 131 Å, c = 170 Å and β = 101° as cell constants. Assuming IIBC^{suc} is a constitutive dimmer, Matthews coefficient (Matthews, 1968) analysis of these crystals suggested a number of 3, 4 or 5 IIBC^{suc} dimers corresponding to 72 %, 62 %, 53 % solvent content, respectively.
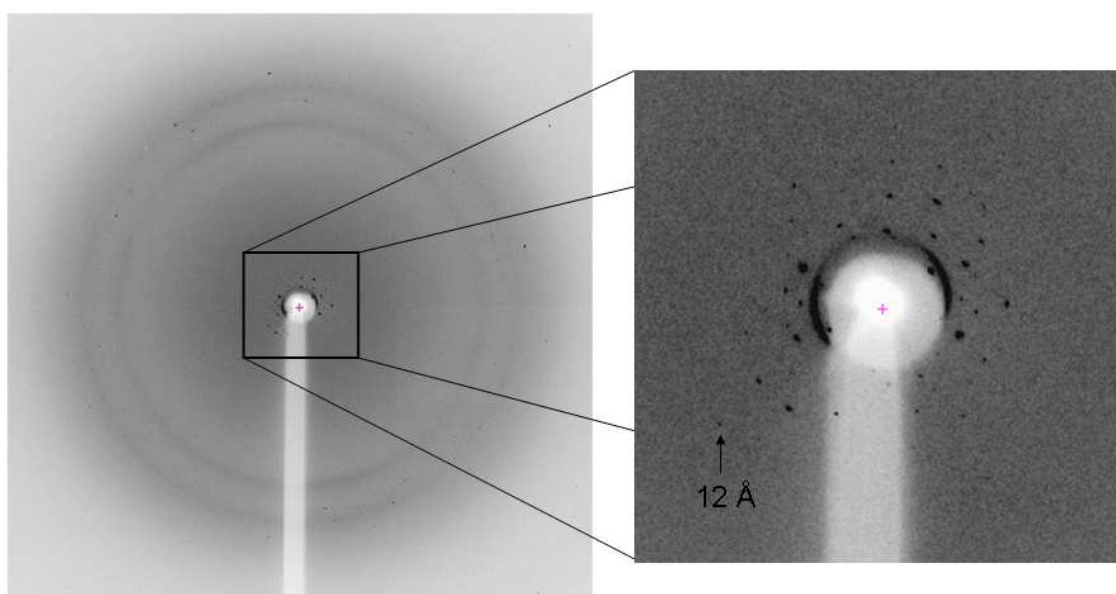
Figure III.16: IIBC^{suc} X-ray diffraction pattern

The right panel shows a zoomed in image presenting the pattern and the best resolution spot (arrow)

IIBC^{suc}-Fab

The biggest crystals obtained with the IIBC-Fab(G22) complex were tested for diffraction on beamline X06SA at the SLS. The crystals obtained from conditions 4 and 5 showed very week diffraction but a clear protein pattern (figure III.17). The resolution in this case was very low at nearly 30 Å. This can be partially due to the very small size of the crystals (about 10 × 10 × 30 µm).

A IIBC^{suc}-Fab(G22) crystal grown in condition 6 was also tested. This was the only crystal in that drop but had grown slightly bigger than in crystals from other conditions (around $20 \times 20 \times 40$ µm). The diffraction was at a resolution around 12 to 15 Å with a rather clear pattern (figure III.17).
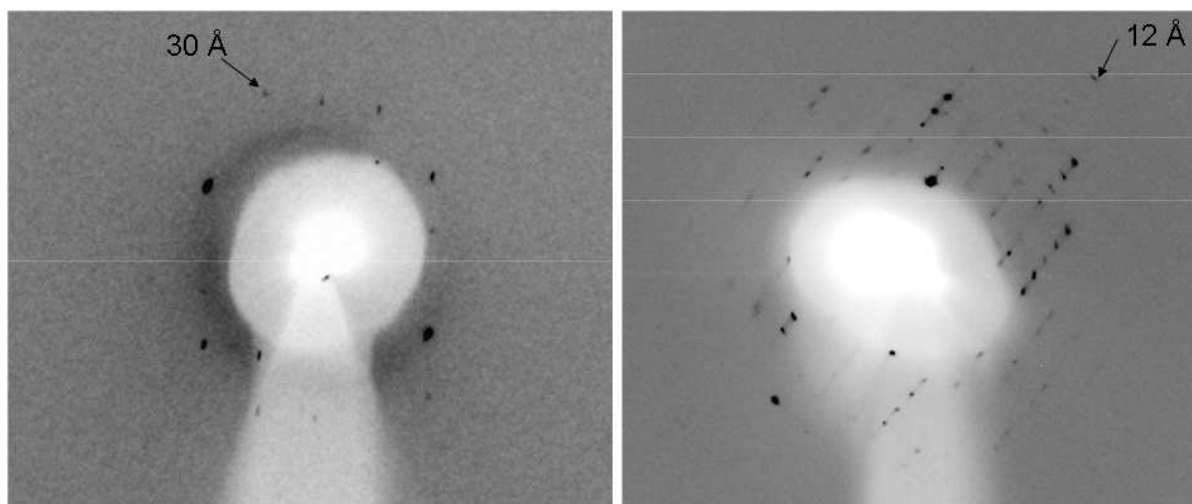

Figure III.17: Diffraction pattern of IIBC^{suc}-Fab(G22)

Examples of diffraction patterns obtained with IIBC^{suc}-Fab(G22) crystals grown (left) from condition 4 and (right) from condition 6. Maximal resolutions are represented with an arrow.

In this later case the indexing of the crystal could be achieved using MOSFLM. The space group of these crystals was P2(?)2(?)2(?) with unit cell parameters equal to a = 79.4 Å, b = 119.8 Å and c = 189.3 Å. Solvent content analysis using Matthews calculation revealed an unlikely low solvent content (45%) if the asymmetric unit would contain one dimer of IIBC^{suc} in complex with two Fab molecules (total MW = 200 kDa). However, if one assumes that one of the crystallographic two fold axis is located on a local two fold axe of the putative dimer of IIBC^{suc}, each IIBC^{suc} monomer in complex with one Fab could fit into the asymmetric unit with a solvent content of 73 %. This solvent content would be in better agreement with typical solvent content values found in crystals of membrane protein that are embedded in a detergent micelle.

III.3.9. Bioinformatic analysis of IIBC^{suc}

Transmembrane helices prediction

The transmembrane part of the bacterial inner-membrane proteins is composed of α-helices. These helices have an overall high hydrophobicity and a characteristic amphipathic pattern and therefore it is possible to directly predict the number and location of transmembrane helices from the protein sequance. Several servers were used to predict the transmembrane helices of the IIBC^{suc} C-terminal domain (figure III.18). All but one program predicted a ten transmebrane helices topology and all predicted helices are located at almost identical positions in the sequence. Except the SOSui program, all would suggest a topology in which IIBC^{suc} would have its both N- and C- termini in the cytoplasm. The topology predicted by the SOSui program would lead to a N-terminus located in the periplasm which is inconsistent with the location of the IIB domain that has to be located in the cytosol for functional reasons.

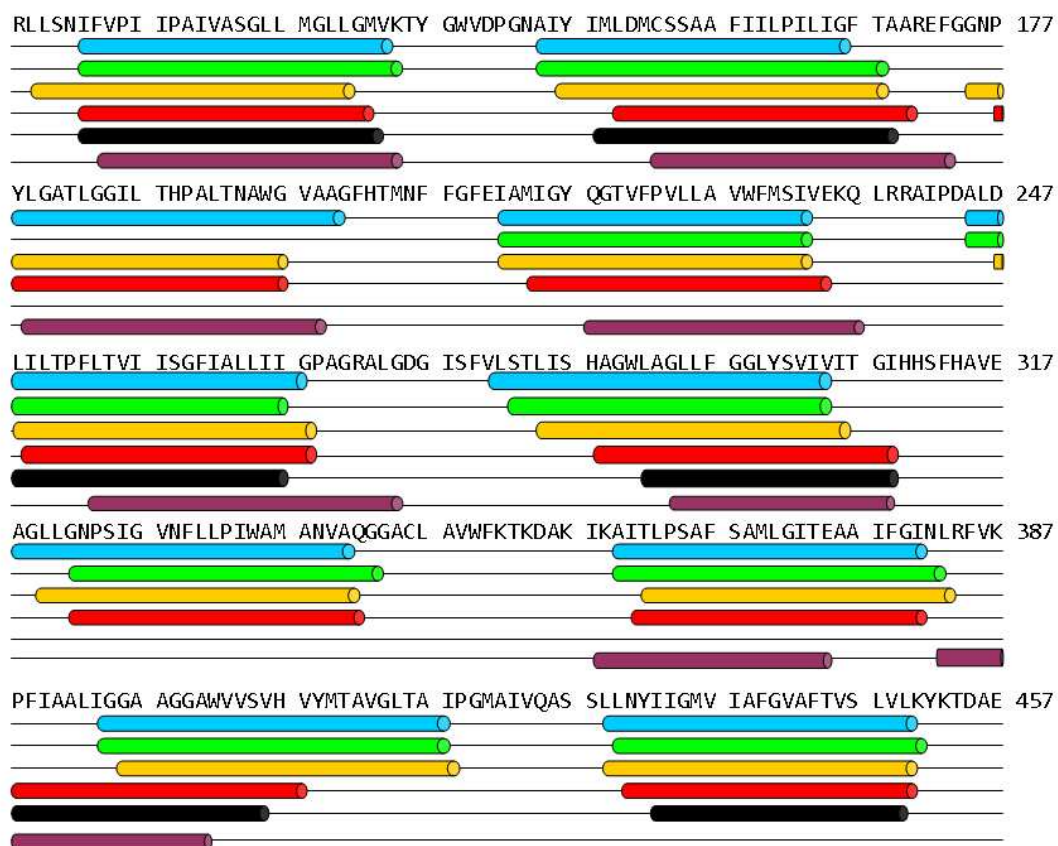**Figure III.18: Compilation of transmembrane helices prediction of the C-terminal domain of IIBC^{suc}**

Putative transmembrane helices predicted by HMMTOP (blue), SOSui (green), TMHHM (yellow) and TopPred (red) programs are marked. Black and purple helices represent the prediction made by Titgemeyer et al. (1996) and Buhr et al. (1993) respectively. Residue numbers refer to the full length IIBC^{suc}.

These results are in contrast with Titgemeyer et al. (Titgemeyer et al, 1996) which proposed a 6 helices model for the IIBC^{suc} close homologue ScrA from *K. pneumoniae.* The helices predicted in that study were also predicted *in silico* by the programs (figure III.18). However, in Titgemeyer et al., the protein was predicted to have two long loops in the cytoplasm that are detected by the prediction programs as being putative transmembrane helices. The 4 additional detected helices are presenting a majority of hydrophobic residues. In order to form the amphipatic helices, the hydrophilic residues present in these helices should be located in intervals of multiples of 3 to 4 residues. In the sequence, no long patches of hydrophilic residues are visible at the localization of the additional computer proposed helices, giving a supplementary indication that these helices might span the membrane. Another model based on fusion protein experiments and hydropathy analysis was proposed for the related IICB^{glu} (Buhr & Erni, 1993) and consists in an eight transmembrane helices bundle.

Taken together these data suggest that the transmembrane domain of IIBC^{suc} has an unclear number of transmembrane helices. Experimental data suggesting 8 helices, it is possible that the two addition "transmembrane" helices detected by the programs could be membrane associated and not transmembrane. In order to obtain more insights, further structural knowledge is clearly required.

Homology modelling

Sequence based homology modelling was carried out in parallel with SwissModel (Arnold et al, 2006) and with HHpred (Soding et al, 2005). In both cases, the programs identified structurally characterized homologues in the N-terminal part of the IIBC^{suc} sequence, corresponding to the IIB domain of the protein. Two homologous proteins (IIB domain of the IIABC of *clostridium difficile* (PDB code: 3ipj) (unpublished data) and IIB^{glu} from *E. coli* (PDB code: 1iba and 3bp3) (Eberstadt et al, 1996; Nam et al, 2008) with high sequence identities (respectively 42, 30 and 29 %), were used as templates in order to compute a structural model of the IIB domain from IIBC^{suc} using Modeller (Sali & Blundell, 1993).

The computed IIB^{suc} model strongly resembles the identified homologues used as template. IIB^{suc} is forming a globular α/β protein domain composed of an antiparallel β-sheet flanked by three α-helices (figure III.19).
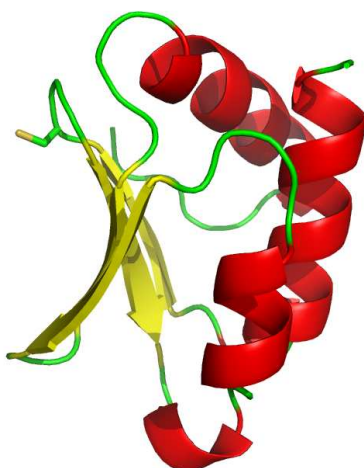
Figure III.19: Ribbon representation of the HHPred generated model of IIB^{suc} domain of IIBC^{suc}.

Its active site (Cys 26) is represented in sticks.

However, no proteins sharing enough similarity with the transmembrane domain of IIBC^{suc} and for which the structure is known was detected by homology modelling programs. Therefore, obtaining structural data is still necessary to understand the molecular mechanisms underlying the sugar uptake and its concomitant phosphorylation and more generally to determine the fold of the IIC domain.

Conclusions

This study aimed to solve the crystal structure of the IIBC^{suc} protein from *S. typhimurium* in order to obtain the first structural insights into the sugar uptake at the inner-membrane level by the PTS system. In order to achieve this, IIBC^{suc} was expressed and purified in *E. coli*. In addition, in order to increase the crystallization ability of IIBC^{suc}, specific antibody Fab-fragments were produced, purified and used to increase the soluble part and to rigidify IIBC^{suc}. Although initial crystallization conditions were also obtained with IIBC^{suc} without Fab, crystals could also been grown for the IIBC^{suc}-Fab(G22) complex confirming the potential of this method. However, the crystals were never reproducible from one IIBC^{suc} batch to the other suggesting that an uncontrolled parameter influences the crystallization. Cases have been reported were the amount of lipids present in the protein sample was critical for crystallization and obtaining well diffracting crystals (Lemieux et al, 2003). It is likely that such, yet uncontrolled, parameters are critical for crystal reproduction. In the past years other methods have been used to solve structures of membrane proteins. DARPINS, that are binding very tightly to their target and are not flexible in contrast to Fab fragments, have proven to allow easier crystallization of membrane proteins (Sennhauser et al, 2007). More drastic solutions have even arisen in the last few years, such as inserting an easy crystallizing protein (i.e. lysozyme) in a loop of a G-protein coupled receptor (Cherezov et al, 2007; Hanson et al, 2008). This approach could also be successful with IIBC^{suc} if replacement of a soluble part/domain (such as the IIB domain) of IIBC^{suc} is not disturbing the protein structure.

References

Adamczyk M, Gebler JC, Wu J (2000) Papain digestion of different mouse IgG subclasses as studied by electrospray mass spectrometry. *J Immunol Methods* **237:** 95-104

Akerstrom B, Brodin T, Reis K, Bjorck L (1985) Protein G: a powerful tool for binding and detection of monoclonal and polyclonal antibodies. *J Immunol* **135:** 2589-2592

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22:** 195-201

Buhr A, Erni B (1993) Membrane topology of the glucose transporter of Escherichia coli. *J Biol Chem* **268:** 11599-11603

Cai M, Williams DC, Jr., Wang G, Lee BR, Peterkofsky A, Clore GM (2003) Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIAGlucose and the cytoplasmic domain of the glucose transporter IICBGlucose of the Escherichia coli glucose phosphotransferase system. *J Biol Chem* **278:** 25191-25206

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318:** 1258-1265

Claros MG, von Heijne G (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10:** 685-686

Cowan PJ, Nagesha H, Leonard L, Howard JL, Pittard AJ (1991) Characterization of the major promoter for the plasmid-encoded sucrose genes scrY, scrA, and scrB. *J Bacteriol* **173:** 7464-7470

Eberstadt M, Grdadolnik SG, Gemmecker G, Kessler H, Buhr A, Erni B (1996) Solution structure of the IIB domain of the glucose transporter of Escherichia coli. *Biochemistry* **35:** 11286-11292

Elferink MG, Driessen AJ, Robillard GT (1990) Functional reconstitution of the purified phosphoenolpyruvate-dependent mannitol-specific transport system of Escherichia coli in phospholipid vesicles: coupling between transport and phosphorylation. *J Bacteriol* **172:** 7119-7125

Erni B, Zanolari B, Graff P, Kocher HP (1989) Mannose permease of Escherichia coli. Domain structure and function of the phosphorylating subunit. *J Biol Chem* **264:** 18733-18741

Erni B, Zanolari B, Kocher HP (1987) The mannose permease of Escherichia coli consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage lambda DNA. *J Biol Chem* **262:** 5238-5247

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* **62:** 72-82

Ferguson AD, Deisenhofer J (2002) TonB-dependent receptors-structural perspectives. *Biochim Biophys Acta* **1565:** 318-332

Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM (1999) Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat Struct Biol* **6:** 166-173

Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, Velasquez J, Kuhn P, Stevens RC (2008) A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure* **16:** 897-905

Herzberg O, Reddy P, Sutrina S, Saier MH, Jr., Reizer J, Kapadia G (1992) Structure of the histidine-containing phosphocarrier protein HPr from Bacillus subtilis at 2.0-A resolution. *Proc Natl Acad Sci U S A* **89:** 2499-2503

Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14:** 378-379

Huber F, Erni B (1996) Membrane topology of the mannose transporter of Escherichia coli K12. *Eur J Biochem* **239:** 810-817

Hummel U, Nuoffer C, Zanolari B, Erni B (1992) A functional protein hybrid between the glucose transporter and the N-acetylglucosamine transporter of Escherichia coli. *Protein Sci* **1:** 356-362

Hunte C, Koepke J, Lange C, Rossmanith T, Michel H (2000) Structure at 2.3 A resolution of the cytochrome bc(1) complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure* **8:** 669-684

Hunte C, Michel H (2002) Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol* **12:** 503-508

Lemieux MJ, Song J, Kim MJ, Huang Y, Villa A, Auer M, Li XD, Wang DN (2003) Three-dimensional crystallization of the Escherichia coli glycerol-3-phosphate transporter: a member of the major facilitator superfamily. *Protein Sci* **12:** 2748-2756

Lengeler JW, Jahreis K (2009) Bacterial PEP-dependent carbohydrate: phosphotransferase systems couple sensing and global control mechanisms. *Contrib Microbiol* **16:** 65-87

Lengeler JW, Jahreis K, Wehmeier UF (1994) Enzymes II of the phospho enol pyruvate-dependent phosphotransferase systems: their structure and function in carbohydrate transport. *Biochim Biophys Acta* **1188:** 1-28

Leslie AGW, Powell HR (2007) Processing diffraction data with MOSFLM. *Nato Sci Ser Ii Math* **245:** 41-51
191

Liao DI, Silverton E, Seok YJ, Lee BR, Peterkofsky A, Davies DR (1996) The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the E. coli PEP:

sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure* **4:** 861-872

Lolkema JS, Dijkstra DS, Robillard GT (1992) Mechanics of solute translocation catalyzed by enzyme IImtl of the phosphoenolpyruvate-dependent phosphotransferase system of Escherichia coli. *Biochemistry* **31:** 5514-5521

Manayan R, Tenn G, Yee HB, Desai JD, Yamada M, Saier MH, Jr. (1988) Genetic analyses of the mannitol permease of Escherichia coli: isolation and characterization of a transport-deficient mutant which retains phosphorylation activity. *J Bacteriol* **170:** 1290-1296

Mao Q, Schunk T, Flukiger K, Erni B (1995) Functional reconstitution of the purified mannose phosphotransferase system of Escherichia coli into phospholipid vesicles. *J Biol Chem* **270:** 5258-5265

Marquez J, Reinelt S, Koch B, Engelmann R, Hengstenberg W, Scheffzek K (2006) Structure of the full-length enzyme I of the phosphoenolpyruvate-dependent sugar phosphotransferase system. *J Biol Chem* **281:** 32508-32515

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491-497

Nam TW, Jung HI, An YJ, Park YH, Lee SH, Seok YJ, Cha SS (2008) Analyses of Mlc-IIBGlc interaction and a plausible molecular mechanism of Mlc inactivation by membrane sequestration. *Proc Natl Acad Sci U S A* **105:** 3751-3756

Nikaido H (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67:** 593-656

Nunn RS, Markovic-Housley Z, Genovesio-Taverne JC, Flukiger K, Rizkallah PJ, Jansonius JN, Schirmer T, Erni B (1996) Structure of the IIA domain of the mannose transporter from Escherichia coli at 1.7 angstroms resolution. *J Mol Biol* **259:** 502-511

Oberholzer AE, Schneider P, Siebold C, Baumann U, Erni B (2009) Crystal structure of enzyme I of the phosphoenolpyruvate sugar phosphotransferase system in the dephosphorylated state. *J Biol Chem* **284:** 33169-33176

Ostermeier C, Harrenga A, Ermler U, Michel H (1997) Structure at 2.7 A resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody FV fragment. *Proc Natl Acad Sci U S A* **94:** 10547-10553

Padavattan S, Schirmer T, Schmidt M, Akdis C, Valenta R, Mittermann I, Soldatova L, Slater J, Mueller U, Markovic-Housley Z (2007) Identification of a B-cell epitope of hyaluronidase, a major bee venom allergen, from its crystal structure in complex with a specific Fab. *J Mol Biol* **368:** 742-752

Postma PW, Lengeler JW, Jacobson GR (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57:** 543-594

Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, Kobilka BK (2007)

Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450:** 383-387

Rhiel E, Flukiger K, Wehrli C, Erni B (1994) The mannose transporter of Escherichia coli K12: oligomeric structure, and function of two conserved cysteines. *Biol Chem Hoppe Seyler* **375:** 551-559

Robillard GT, Broos J (1999) Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim Biophys Acta* **1422:** 73-104

Ruijter GJ, van Meurs G, Verwey MA, Postma PW, van Dam K (1992) Analysis of mutations that uncouple transport from phosphorylation in enzyme IIGlc of the Escherichia coli phosphoenolpyruvate-dependent phosphotransferase system. *J Bacteriol* **174:** 2843-2850

Saier MH, Jr., Yen MR, Noto K, Tamang DG, Elkan C (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37:** D274-278

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234:** 779-815

Schauder S, Nunn RS, Lanz R, Erni B, Schirmer T (1998) Crystal structure of the IIB subunit of a fructose permease (IIBLev) from Bacillus subtilis. *J Mol Biol* **276:** 591-602

Schirmer T (1998) General and specific porins from bacterial outer membranes. *J Struct Biol* **121:** 101-109

Schmid K, Schupfner M, Schmitt R (1982) Plasmid-mediated uptake and metabolism of sucrose by Escherichia coli K-12. *J Bacteriol* **151:** 68-76

Schnetz K, Sutrina SL, Saier MH, Jr., Rak B (1990) Identification of catalytic residues in the beta-glucoside permease of Escherichia coli by site-specific mutagenesis and demonstration of interdomain cross-reactivity between the beta-glucoside and glucose systems. *J Biol Chem* **265:** 13464-13471

Schwaber J, Cohen EP (1973) Human x mouse somatic cell hybrid clone secreting immunoglobulins of both parental types. *Nature* **244:** 444-447

Sennhauser G, Amstutz P, Briand C, Storchenegger O, Grutter MG (2007) Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol* **5:** e7

Siebold C, Flukiger K, Beutler R, Erni B (2001) Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). *FEBS Lett* **504:** 104-111

Sliz P, Engelmann R, Hengstenberg W, Pai EF (1997) The structure of enzyme IIAlactose from Lactococcus lactis reveals a new fold and points to possible interactions of a multicomponent system. *Structure* **5:** 775-788

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33:** W244-248

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6:** 175-182

Sugiyama JE, Mahmoodian S, Jacobson GR (1991) Membrane topology analysis of Escherichia coli mannitol permease by using a nested-deletion method to create mtlA-phoA fusions. *Proc Natl Acad Sci U S A* **88:** 9603-9607

Tchieu JH, Norris V, Edwards JS, Saier MH, Jr. (2001) The complete phosphotranferase system in Escherichia coli. *J Mol Microbiol Biotechnol* **3:** 329-346

Teplyakov A, Lim K, Zhu PP, Kapadia G, Chen CC, Schwartz J, Howard A, Reddy PT, Peterkofsky A, Herzberg O (2006) Structure of phosphorylated enzyme I, the phosphoenolpyruvate:sugar phosphotransferase system sugar translocation signal protein. *Proc Natl Acad Sci U S A* **103:** 16218-16223

Titgemeyer F, Jahreis K, Ebner R, Lengeler JW (1996) Molecular analysis of the scrA and scrB genes from Klebsiella pneumoniae and plasmid pUR400, which encode the sucrose transport protein Enzyme II Scr of the phosphotransferase system and a sucrose-6-phosphate invertase. *Mol Gen Genet* **250:** 197-206

Tsvetanova B, Wilson AC, Bongiorni C, Chiang C, Hoch JA, Perego M (2007) Opposing effects of histidine phosphorylation regulate the AtxA virulence transcription factor in Bacillus anthracis. *Mol Microbiol* **63:** 644-655

Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17:** 849-850

Uysal S, Vasquez V, Tereshko V, Esaki K, Fellouse FA, Sidhu SS, Koide S, Perozo E, Kossiakoff A (2009) Crystal structure of full-length KcsA in its closed conformation. *Proc Natl Acad Sci U S A* **106:** 6644-6649

van Montfort RL, Pijning T, Kalk KH, Hangyi I, Kouwijzer ML, Robillard GT, Dijkstra BW (1998) The structure of the Escherichia coli phosphotransferase IIAmannitol reveals a novel fold with two conformations of the active site. *Structure* **6:** 377-388

van Montfort RL, Pijning T, Kalk KH, Reizer J, Saier MH, Jr., Thunnissen MM, Robillard GT, Dijkstra BW (1997) The structure of an energy-coupling protein from bacteria, IIBcellobiose, reveals similarity to eukaryotic protein tyrosine phosphatases. *Structure* **5:** 217-225

van Nuland NA, Hangyi IW, van Schaik RC, Berendsen HJ, van Gunsteren WF, Scheek RM, Robillard GT (1994) The high-resolution structure of the histidine-containing phosphocarrier protein HPr from Escherichia coli determined by restrained molecular dynamics from nuclear magnetic resonance nuclear Overhauser effect data. *J Mol Biol* **237:** 544-559

Vogler AP, Broekhuizen CP, Schuitema A, Lengeler JW, Postma PW (1988) Suppression of IIIGlc-defects by enzymes IINag and IIBgl of the PEP:carbohydrate phosphotransferase system. *Mol Microbiol* **2:** 719-726

Volpon L, Young CR, Matte A, Gehring K (2006) NMR structure of the enzyme GatB of the galactitol-specific phosphoenolpyruvate-dependent phosphotransferase system and its interaction with GatA. *Protein Sci* **15:** 2435-2441

Wohlhieter JA, Lazere JR, Snellings NJ, Johnson EM, Synenki RM, Baron LS (1975) Characterization of transmissible genetic elements from sucrose-fermenting Salmonella strains. *J Bacteriol* **122:** 401-406

Worthylake D, Meadow ND, Roseman S, Liao DI, Herzberg O, Remington SJ (1991) Three-dimensional structure of the Escherichia coli phosphocarrier protein IIIglc. *Proc Natl Acad Sci U S A* **88:** 10382-10386

Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R (2001) Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution. *Nature* **414:** 43-48

Appendix:

Curriculum Vitae

Christophe WIRTH

Personal details

Address : 25B avenue de la liberté, 68000 Colmar – France
E-mail : w.christop@orange.fr
Nationality: French
29 years old, married

Work experience:

<u>December 2005-March 2010</u>: PhD in the group of Prof. T. Schirmer at the Biozentrum - University of Basel on the project: **Structural investigation on membrane proteins: the case of KdgM porin family, OmpF and IIBC^{suc} sucrose permease.** During this period, I also worked on few soluble proteins, acquired decisional skills and supervised two master students.

<u>September 2004-October 2005</u>: training period followed by a short term contract in the membrane protein department at the Ecole Supérieure de Biotechnologies de Strasbourg (ESBS) where I **purified, crystallized and solved the structure of a siderophore membrane receptor** of *Pseudomonas aeruginosa* called FpvA under the supervision of Dr. David Cobessi.

<u>March-July 2004</u>: training period in the team of Prof. M. Primig at the Biozentrum where I worked on a **yeast meiosis regulation factor** by making yeast mutants and genechip experiments.

<u>March 2003</u>: training period in the group of Pr. D. Kern at the IBMC - Strasbourg where I purified an aminoacyl-tRNA-synthetase.

Education

<u>December 2005-March 2010</u>: **PhD in structural biology** in the group of Pr. Tilman Schirmer, Biozentrum University of Basel, Switzerland.

<u>2004-2005</u>: **Master - M2** (DEA) in Molecular and Cellular Biology, with a specialisation in Structural Biology at the Université Louis Pasteur (ULP), Strasbourg, France.

<u>2003-2004</u>: **Master - M1** (Maîtrise) in Bioinformatics at the ULP - Strasbourg (last year of the Professional University Institute -IUP- specialised in Advanced Technologies of Life Science) obtained with honours and with the "Ingénieur Maître" diploma.

<u>2002-2003</u>: **Licence**, equivalent to a Bachelor's degree, in Bioinformatics at the IUP of the ULP.

<u>2000-2002</u>: **DEUG** (Diplôme d'Etudes Universitaire Générales) in Bioinformatics, degree award after a two years university program, with honours, at the ULP.

<u>1998-2000</u>: First year of medical school in the ULP.

<u>1998</u>: **Baccalauréat** scientifique, French secondary school level diploma, with honours.

Scientific methods:

- **<u>Bioinformatic skills</u>**: data mining, sequence alignment, homology modelling, etc…

- **<u>Protein expression, purification</u>** using affinity, ion exchange, size exclusion chromatography and **<u>characterization</u>** using static light scattering and western blot

- **<u>Crystallization</u>**, crystal handling, heavy atom derivatization

- X-ray diffraction **<u>data collection</u>** at several synchrotron beamlines and home sources

- Diffraction **<u>data processing</u>** and solving the phase problem using **<u>molecular replacement</u>** and **<u>experimental phasing</u>**

- Electron density **<u>map interpretation</u>**, model building and **<u>structure analysis</u>**

- Knowledge in C, Python and Perl as well as in SQL data bases **<u>programming</u>** but little training

Supervision and teaching:

<u>Supervision</u> of two master students during my PhD (Roman Lehner and Matthias Barone);

<u>Teaching</u> during the practical and exercise sessions of the "structural biology blockkurs" (5th semester students): 30 h / year;

In charge of the structural biology exercise sessions preparation and teaching to students in the 4th semester.

Languages:

<u>French</u>: native speaker

<u>English</u>: fluent, written and spoken (860 points at the TOEIC in 2004)

<u>German</u>: fluent, written and spoken

Scientific Publications

<u>Wirth, C.</u>, Barone, M., Baslé, A., Schirmer, T., (2009). Structure of OmpF, the major *Escherichia coli* porin, in orthorhombic crystal form. *In preparation*

<u>Wirth, C.,</u> Condemine, G., C., Boiteux, C., Bernèche, S., Schirmer, T., Peneff, CM., (2009) NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM porin family. *J. Mol. Biol.*, **394**, 718-731.

<u>Wirth, C.</u>, Meyer-Klaucke, W. Pattus, F. & Cobessi, D. (2007). From the periplasmic signaling domain to the extracellular face of an outer membrane signal transducer of *Pseudomonas aeruginosa:* crystal structure of the ferric pyoverdine outer membrane receptor. *J. Mol. Biol.*, **368**, *398-406.*

<u>Wirth, C.</u>, Hoegy, F., Pattus, F. & Cobessi, D. (2006). Preliminary X-ray investigations of several crystal forms of the ferripyoverdine FpvA outer membrane receptor from *Pseudomonas aeruginosa* bound to ferripyoverdine. *Acta Cryst. F*, **62**, 460-463.

176

Scientific communications

Oral presentations:

The structure of NanC porin provides further insights into sialic acid update by *Escherichia coli*. 22nd Rhine-Knee Regiomeeting on the structural biology of biological macromolecules, Freudenstadt, Germany (2008).

Regular group and floor seminars (about 5 a year).

Posters:

Wirth, C., Condemine, G., Schirmer, T., Peneff, CM. (2009) Structure of the NanC porin from *Escherichia coli*, 13th European symposium of the Protein Society, Zurich, Switzerland.

Wirth, C., Condemine, G., Schirmer, T., Peneff, CM. (2009) Structure of the NanC porin from *Escherichia coli*, USGEB 2009, Interlaken, Switzerland.

Cobessi, D., Celia, H., Wirth, C., Schalk, I. and Pattus F. (2005) Structures of iron-siderophore outer membrane receptor from *P. aeruginosa*. 15th IUPAB & 5th EBSA International Biophysicis Congress, Montpellier, France.

Other interests and activities

- Swimming and hiking

- Reading (especially historical, police and science fiction novels)

- Informatics

References

Prof. Tilman Schirmer: Structural Biology Department, Biozentrum, University of Basel
Kligelbergstrasse 50/70, CH-4056, Basel, Switzerland
+41 (0) 61 267 20 89
tilman.schirmer@unibas.ch

Dr. Caroline Peneff   Structural Biology Department, Biozentrum, University of Basel
Kligelbergstrasse 50/70, CH-4056, Basel, Switzerland
+41 (0) 61 267 20 88
caroline.peneff@unibas.ch

Dr David Cobessi   Institut de Biologie Structurale
41 rue Jules Horowitz, F-38027, Grenoble Cedex-1, France
+33 (0) 4 38 78 96 13
david.cobessi@ibs.fr