X-ray crystallographic studies on bacterial proteins involved in active membrane transport: $MalFGK_2$, MalE, ProX, and AcrB

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften

an der Universität Konstanz
Mathematisch- Naturwissenschaftliche
Sektion, Fachbereich Biologie

vorgelegt von
Dipl. Biochem. André Schiefner

Tag der mündlichen Prüfung: 22.07.04

Referent 1: Prof. Dr. Wolfram Welte
Referent 2: Prof. Dr. Winfried Boos
Referent 3: Prof. Dr. Reinhard Krämer

Contents

Abbreviations

CHES	2-(Cyclohexylamino)ethanesulfonic acid
CHM	Cyclohexylhexylmaltoside
DM	n-Decyl-β-D-maltopyranoside
DDAO	N,N-dimethyldecylamine-N-oxide
DDM	n-Dodecyl-β-D-maltopyranoside
DTT	1,4-Dithio-DL-threitol
EDTA	Ethylenediamine-tetraaceticacid
GB	N,N,N-trimethyl glycine (glycine betaine)
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
LDAO	N,N-dimethyldodecylamine-N-oxide
MES	2-Morpholinoethanesulfonic acid
OG	n-Octyl-β-D-glucopyranoside
OM	n-Octyl-β-D-maltopyranoside
PB	N,N-dimethyl-L-proline (proline betaine)
PEG	Polyethylene glycol
PIPES	1,4-Piperazine-bis-ethanesulfonic acid
PMSF	Phenylmethanesulfonyl fluoride
R.m.s.d.	Root mean square deviation
TM	Trimethylammonium
Tris	Tris(hydroxymethyl)aminomethane
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

1 Introduction

1.1 Biological membranes

One of the most important events that led to the emergence of life, as we know it today, was the invention of effective barriers to separate directed, metabolic reactions from random fluctuations of the environment. This barrier between a living cell and its surroundings is with approximately 5 nm very thin and consists largely of phospholipids and embedded proteins. Due to their amphipathic nature, the nonpolar chains of the phospholipids interact with each other whereas the hydrophilic head groups face the surrounding water. This results in the energetically most favorable arrangement, the lipid bilayer. Under *in vivo* conditions the hydrocarbon chains are not rigidly packed but are in a fluid state undergoing lateral diffusion. Therefore the bilayer behaves like a two-dimensional fluid, allowing embedded proteins to float around (Singer and Nicolson, 1972).

The lipid and protein composition of lipid bilayers varies depending on the task to be performed by the membrane and can be controlled by the cell. Beside forming a boundary to the surroundings by the plasma membrane other membranes within the cell are used to divide the space of the cytoplasm into compartments.

The hydrophobic interior of the bilayer creates a barrier to the passage of most hydrophilic molecules, depending on their concentration gradient across and their partition coefficient within the bilayer. Only small non-polar molecules such as O_2 and CO_2 rapidly permeate. In contrast, larger polar molecules such as glucose hardly permeate and for ions like H^+, and Na^+ the membrane is virtually impermeable. This effective insulating property of bilayers allows plasma membranes to establish an electrical potential difference across them, with the inside being usually negative and the outside positive.

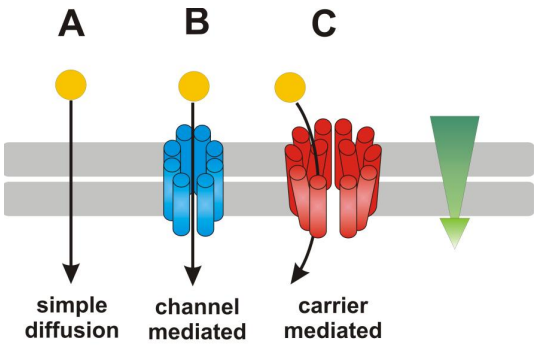

Figure 1.1: Passive transport Schematic representation of passive transport across biological membranes. The green arrow indicates the electrochemical gradient for the transported solute in yellow

1.2 Transport across biological membranes

In principle all ions and molecules pass the membrane down their concentration gradient (Figure 1.1A), even if the corresponding transport rates are very small. To speed up the transport rates of nutrients, metabolic waste products, ions and so forth, all types of biological membranes contain a certain set of different transport proteins. Each transport protein is specialized to transport a certain molecule or a restricted class of compounds. All transport proteins have been found to span the lipid bilayer multiple times to create a continuous protein pathway for the transported molecules. To penetrate the lipid surrounding of the bilayer the transmembrane polypeptide chain needs to have a hydrophobic surface formed by hydrophobic residues and the formation of α-helices or β-sheets within the membrane to satisfy all polar groups of the main chain by hydrogen bond formation.

Passive transport

Passive transport or facilitated diffusion of a molecule is mediated by channel proteins or carrier proteins (Figure 1.1B and C). If a molecule is not charged only its concentration difference across the membrane drives the transport and determines its direction. On the other hand, if the molecule carries a net charge the electrical potential of the membrane also matters. Thus facilitated diffusion and simple diffusion depend upon the electrochemical gradient of ions and molecules. Carrier proteins or facilitators specifically bind a molecule and undergo a series of conformational changes to transfer ions or molecules to the other side (Figure 1.1C). In contrast, channel proteins only interact weakly with the transported ions or molecules, they rather form aqueous pores and therefore act at much higher rates then facilitators (Figure 1.1B).

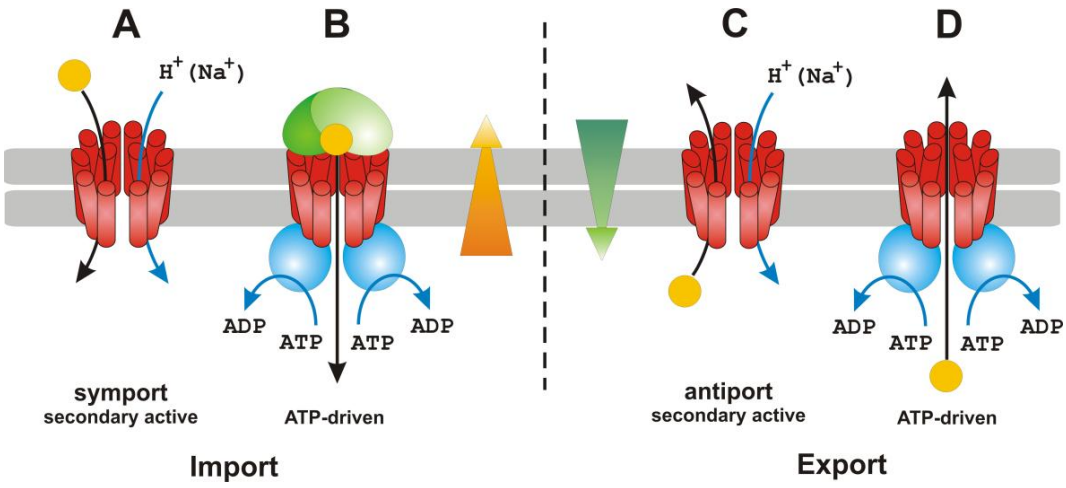

Figure 1.2: Active transport Schematic representation of active transport across biological membranes. Arrows indicate the electrochemical gradient (orange high inside, green high outside) for different solutes in yellow

Active transport

Many ions or molecules have to be transported against their electrochemical gradient to maintain cellular functions. Those transport processes necessarily have a positive Gibbs energy change and therefore must be coupled with a spontaneous exergonic reaction. In primary active transport there is direct coupling between the solute transport across the membrane and a reaction such as hydrolysis of ATP (Figure 1.2B and D). Secondary active transport utilizes the energy of the electrochemical gradient of a second solute, established by a primary process such as respiration (Figure 1.2A and C). Both primary and secondary active transport, like facilitated diffusion, depend upon conformational changes of the carrier protein. Coupled carriers are subdivided into two classes: a) symporters, if solute and driving ion are transported in the same direction and b) antiporters, if solute and driving ion are transported into opposite directions.

1.3 ATP binding cassette transporters

1.3.1 Overview

ATP binding cassette (ABC) transporters represent the largest family of membrane transport proteins, found in all studied bacterial, archaeal and eukaryotic species. They transport an enormous variety of substrates like ions, polysaccharides, peptides, amino acids, and so forth across the cytoplasmic membrane. The importance of this class of transporters becomes obvious by looking at the genome sequence of *E. coli* where 5 % of all genes encode for ABC transporters corresponding to 57 different transport systems

(Linton and Higgins, 1998).
In general ABC transporters are subdivided in the two classes, exporters and importers, with respect to the direction of substrate translocation. These two classes have been found to be segregated early during evolution. While most eukaryotic ABC transporters are exporters, the majority of bacterial ABC transporters import essential nutrients (Saurin et al., 1999).
ABC transporters also play an important role in many human diseases (Klein et al., 1999). a) The MDR1 (multi-drug resistance protein 1) is found on the apical surface of epithelial cells of a variety of tissues, *e.g.* in the intestine or the blood-brain barrier (Thiebaut et al., 1987; Cordon-Cardo et al., 1989). It is suggested that the physiological function is protection of the cells against toxic compounds. Special attention is paid to this protein due to the fact that multi-drug resistance of cancer cells has been found to be caused by this protein. b) The TAP (transporter associated with antigen processing), being active as a heterodimer of TAP1 and TAP2, transports peptides degraded by the proteasome into the lumen of the endoplasmic reticulum. Those peptides associate with the molecules of the major histocompatibility complex class I (MHC-I). The loaded MHC-I complexes are sent to the cellular surface via the Golgi apparatus and there they present the peptides to cytotoxic T cells. Humans with defective TAP1 protein are found to develop lung cancer (Chen et al., 1996). c) Cystic fibrosis is one of the most frequent inherited diseases. Every 25th person is carrier of an autosomal recessive mutation in the CFTR (cystic fibrosis transmembrane conductance regulator) gene. These mutations affect a variety of epithelial tissues including exocrine functions of the pancreas, intestinal glands, bronchial glands, sweat glands etc. (Klein et al., 1999). The intact CFTR protein has been found to form a cAMP activated chloride channel (Hart et al., 1996).

1.3.2 Architecture and mechanism of ABC transporters

Despite their large number and enormous substrate diversity, ABC transporters share a common modular architecture. Members of this family invariably consist of two membrane spanning domains (TMD) that form the translocation pathway through the membrane, and two cytoplasmic nucleotide binding (ABC) domains that energize the transport reaction through binding and hydrolysis of ATP. These four domains can be arranged in many different ways. In bacteria and archaea the four domains are usually separate polypeptides, whereas in higher organisms these polypeptides are fused in different ways. The membrane spanning domains are poorly conserved and vary considerably in the number of predicted α-helices This is likely to reflect the diversity of

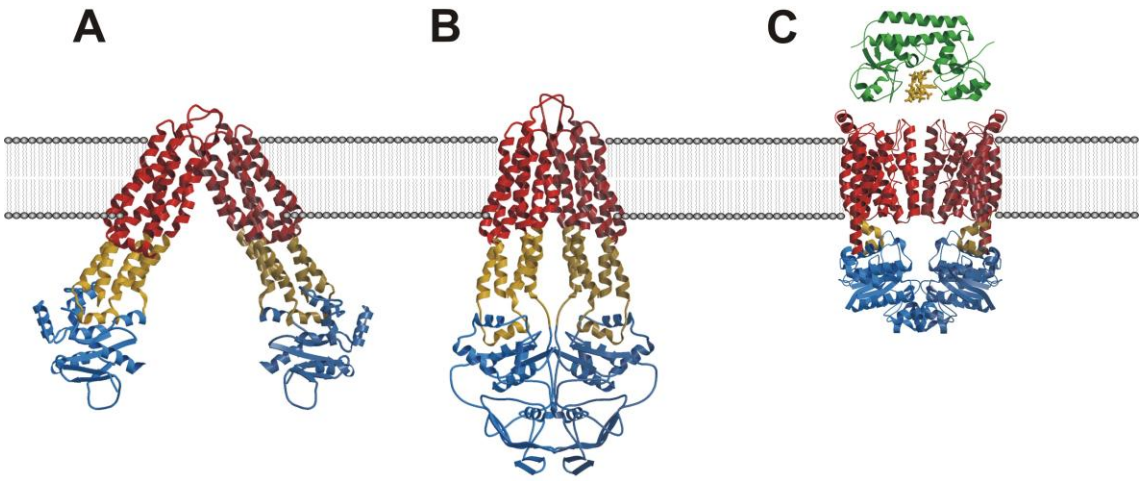

Figure 1.3: Structures of ABC transporters Transmembrane spanning domains (TMD) are shown in red, intracelluar domains (transmission interfaces) in yellow and ABC cassettes in blue. A) MsbA from *Escherichia coli*, B) MsbA from *Vibrio cholera*, C) Btu(CD)$_2$ from *Escherichia coli*, with binding protein BtuF in green.

transported substrates. What brings the different transporters together to a large family is the conserved ATP-binding cassette (ABC) engine that is common to all transporters of this class and shows highly conserved features. Among those are the Walker A and B motifs (Gx_4GKT and $Rx_{4\text{-}12}h_4D$, with h being any hydrophobic residue), signature motif or C-loop (LSGGQ/E), Q-loop, D-loop, and switch II. Because of this conservation, it is assumed that all ABC cassettes drive the transport through the membrane spanning domains by a common mechanism.

A first step in the determination of high resolution structures of ABC transport complexes was MsbA from *E. coli* (Chang and Roth, 2001) (Figure 1.3A). MsbA is closely related to the N- and C- terminal part of MDR1 and forms a homodimeric complex within the membrane that builds the functional unit of the transporter. In Gram-negative bacteria, where MsbA transports lipid A, a major component of the outer membrane, the loss of MsbA activity leads to a lethal accumulation of lipid A in the inner leaflet of the cytoplasmic membrane (Zhou et al., 1998). However, in the MsbA structure important parts in the ATPase domains are missing due to disorder of the crystals, and the large opening angle between the two monomers is likely to be artificially caused by the crystal lattice. A more suitable model of MsbA presents the structure of MsbA from *Vibrio cholera*. The model is much more complete and the ABC cassettes are close enough to interact with each other (Chang, 2003) (Figure 1.3B). The first structure of a binding protein dependent ABC transporter was that of the vitamin B_{12} transporter Btu(CD)$_2$ (Locher et al., 2002) (Figure 1.3C). This structure has been determined to a resolution of 3.2 Å and presents the most detailed structure of an ABC transporter available so

far.

From biochemical and structural studies it is known that ABC cassettes work in pairs of either homo- or heterodimers. Two ATPase domains form two composite ATP-binding active sites in their dimer interface. The ATP binding site is formed by the Walker A motif of one monomer and the signature motif of the other and vice versa (Hopfner et al., 2000; Chen et al., 2003). A monomeric ABC ATPase can be subdivided into two lobes: lobe I containing the P-loop that is the major ATP binding site formed by the Walker A motif and lobe II containing the signature motif. These two lobes are connected by a shared β-sheet that contains the Walker B motif. The signature motif is remote to the Walker A and B motifs of the same monomer. In the functional dimer the signature motif binds to the ATP γ-phosphate which results in the engagement and alignment of Walker A, B and the signature motif (see Figure 1.4) (Hopfner and Tainer, 2003). This may lead to a reorientation of lobe I and II within one monomer, thereby generating force. Furthermore, the D-loop (close to the Walker B motif) presents the attacking water via a main chain carbonyl to the other monomer, likely linking the ATP hydrolysis in one monomer to the other one. Additionally, the activation of the attacking water is likely to be controlled by the switch II including a conserved His (sometimes replaced by Q or S). Another important substructure of ABC cassettes is the Q-loop a loop with a conserved glutamine which binds the catalytic magnesium ion and the nucleophilic attacking water molecule. The residues adjacent to the glutamine of the Q-loop interact tightly with the L-loop (Locher et al., 2002) of the transmembrane subunits and form the transmission interface. It is supposed to be involved in the allosteric activation of the ABC ATPase by the TMD and the transmission of the ATP-dependent conformational changes back to the TMDs (Hopfner and Tainer, 2003). Mutations or deletions in the transmission interface cause severe transport defects as known from the deletion of F508 in CFTR, being responsible for 70 % of the cystic fibrosis cases.

Roughly, a transport cycle may involve the following course of events: 1) substrate binding or binding of a liganded binding protein induces conformational changes in the TMDs 2) these changes are transmitted via the L-loop to the Q-loop of the ATPase inducing engagement of the ATPase dimer 3) the conformational rearrangement caused by the engagement is now transmitted back to the L-loop of the TMDs via the Q-loop which then opens the channel for substrate translocation 4) ATP hydrolysis brings the system back to the ground state.

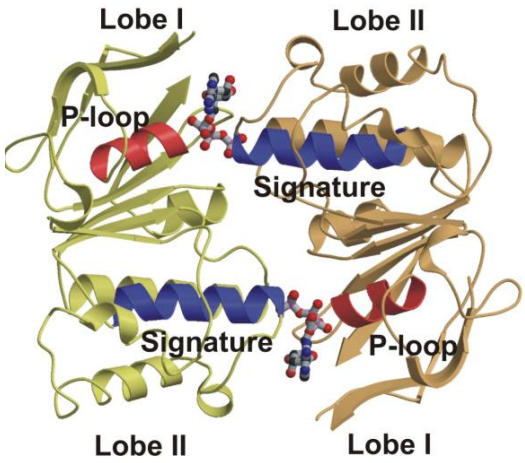

Figure 1.4: ABC ATPase dimer structure Structural model of MalK from *E. coli* (Chen et al., 2003). Two composite active sites are formed in the dimer interface where two ATP molecules are bound.

1.3.3 Binding protein dependent ABC transporters

In bacteria and archaea the large subfamily of ABC transporters that mediate nutrient uptake have an additional component, a periplasmic or membrane anchored binding protein that binds nutrients with high affinity prior to translocation (Boos and Lucht, 1996). The liganded binding protein binds to TMDs and stimulates via the TMDs the ATPase activity of the ABC cassettes in the cytoplasm. By this the binding protein becomes tightly bound to the TMDs to ensure that the substrate is passed through the transporter which refers to the transition state of the transport cycle (Chen et al., 2001). In the TMDs of binding protein dependent ABC transporters a local sequence similarity, the EAA-motif ($EAAAx_3Gx_9IxLP$), has been found which is referred to as the L-loop (Locher et al., 2002). Among the best studied transporters of this class are the maltose transporter $MalEFGK_2$ from *E. coli* and the histidine permease $HisJQMP_2$ from *Salmonella typhimurium*. Nevertheless, all attempts to determine the structure of the intact $MalFGK_2$ and $HisQMP_2$ transporters at atomic resolution failed until today. Only their water soluble components have been structurally characterized HisJ (Oh et al., 1994), HisP (Hung et al., 1998), MalE (Spurlino et al., 1991), and MalK (Chen et al., 2003). Instead, the first structure of a binding protein dependent ABC transporter was that of $Btu(CD)_2$ (see Figure 1.3C).

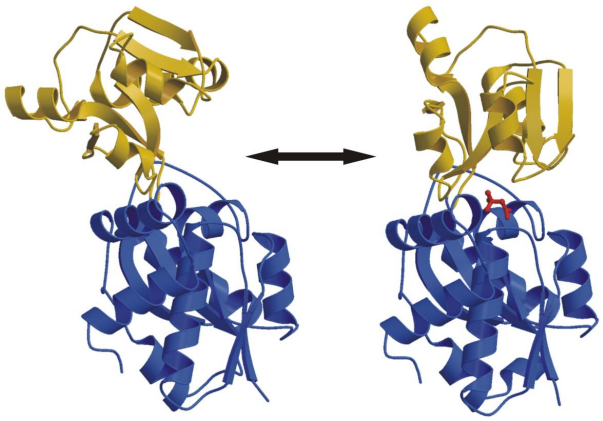

Figure 1.5: Binding proteins ProX from *Archaeoglobus fulgidus* in the open unliganded conformation (left) and the closed liganded conformation (right). The domains are colored blue and yellow and the ligand in red.

Binding proteins

Binding proteins bind their substrates with affinities in the low μM range. During evolution they have been optimized to bind a large variety of ligands like sugars, amino acids, vitamins, and so forth. Nevertheless, crystal structures of several binding proteins revealed a common blueprint. They consist of two globular domains or lobes which are flexibly linked to each other by one to three polypeptide switches between them, forming a hinge. The hinge region is able to undergo large conformational changes, allowing the two domains to move with respect to each other (Figure 1.5). The ligand binding site is located between the two domains in the hinge region. In the unliganded form several open states can be adopted by the binding protein (Björkman and Mowbray, 1998; Magnusson et al., 2002). Ligand binding induces a large conformational change that moves both domains toward each other, engulfing the ligand between both domains.

The maltose transporter MalEFGK$_2$

Maltose and maltodextrins are a major carbon source for microorganisms. The maltose transporter is responsible for the transport of these sugars across the cytoplasmic membrane. It consists of the periplasmic binding protein MalE, the two homologous TMDs MalF and MalG, and two copies of the ABC ATPases MalK. Additionally to its sugar transport function MalEFGK$_2$ is also involved in regulatory processes. This regulatory function is mediated by MalK that is not only an energizing module within the transport complex. The C-terminal 2/5 of MalK fold into a distinct barrel-like structure that is present only in a subset of all bacterial and archaeal ABC transporters. When no substrate is transported by MalFGK$_2$, MalT, the transcriptional activator of most of

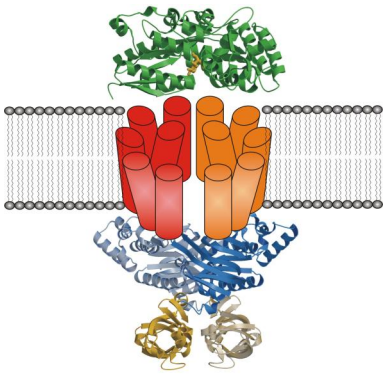

Figure 1.6: Maltose transporter from *Thermococcus litoralis* Structurally characterized components are shown as molecular model TMBP in green and the MalK dimer in blue (ATPase domain) and yellow (regulatory domain). The membrane spanning domains MalF and MalG are schematically drawn in red and orange

the *mal* genes, is bound to the C-terminal part of MalK and *mal* gene expression cannot occur. Furthermore, the dephosphorylated form of $EIIA^{Glc}$ binds to the C-terminal part of MalK as well. This inhibits non-PTS (phosphotransferases systems) sugar uptake systems like the maltose transporter as long as glucose is being phosphorylated during PTS transport, which leads in a series of phosphotransfer reactions to the dephosphorylation of $EIIA^{Glc}$ (Boos and Shuman, 1998). Both binding sites are located in two distinct areas in the barrel-like structure and do not interfere with each other (Böhm et al., 2002).

A very homologous system has been found in the hyperthermophilic archaeon *Thermococcus litoralis* (Xavier et al., 1996). This has some unique properties: it recognizes maltose and trehalose with equal affinities, but no maltodextrins, it exhibits high affinity transport at 85 °C, the optimal growth temperature of *T. litoralis*, and the binding protein is anchored in the cytoplasmic membrane. In our laboratory the structures of the ATP binding cassette MalK and the trehalose/maltose binding protein TMBP have been already structurally determined (Diederichs et al., 2000; Diez et al., 2001) (see Figure 1.6). Part of the present thesis is aimed at the crystallization and structural characterization of the transport complex $MalFGK_2$. Crystals were obtained which, however, did not diffract beyond 5 Å and no high resolution structure could be determined so far (Chapter 2). Currently, we are working on new constructs for the expression of $MalFGK_2$ from *Thermococcus litoralis* as well as *Thermus thermophilus* in *E. coli* in order to obtain more suitable crystals for structural analyses.

A homologue of $MalEFGK_2$ has also been found in the Gram-positive bacterium *Alicyclobacillus acidocaldarius*. This thermoacidophilic organism grows best at pH 3.6 and 57 °C. All its extracellular proteins, like the lipid anchored binding protein MalE, are

constantly exposed to the low pH and to high temperatures. Since structural information is available for mesophilic as well as for thermophilic MalE homologues (Sharff et al., 1992; Evdokimov et al., 2001; Diez et al., 2001), MalE from *A. acidocaldarius* is an excellent candidate to study acidostability. The structural analysis of MalE and its adaptation towards acidostability are discussed in Chapter 3.

The compatible solute transporter ProU

Microorganisms lack systems for active water transport. They adjust their cellular water content and turgor by controlling the pool of osmolytically active substances in the cytoplasm. As a first response to an osmotic upshift, bacterial cells transport potassium ions into the cytoplasm. Since ions influence protein stability and function, this is only a transient solution to cope with high osmolarity. After a while, the ions are replaced by the synthesis or accumulation of compounds known as compatible solutes. Compatible solutes are defined as compounds that can be accumulated up to molar concentrations without disturbing cellular functions. According to the preferential exclusion model, compatible solutes are excluded from the immediate hydration shell of proteins due to unfavorable interactions. The resulting non-uniform distribution of these compounds within the cytoplasm forces proteins to occupy a smaller volume by reducing the amount of hydration water. These effects are not only compatible with cellular functions but also stabilize the tertiary structures of proteins.
An important system for the uptake of compatible solutes is the osmoregulated binding protein dependent ABC transporter ProU from *E. coli*. It consists of two copies of the ABC ATPase ProV, two TMDs ProW and the periplasmic binding protein ProX. ProU shows a rather broad substrate specificity but has a clear preference for N,N,N-trimethyl glycine (glycine betaine) and N,N-dimethyl-L-proline (proline betaine) (Gouesbet et al., 1994; Haardt et al., 1995). As deduced from genome sequence, a homologous system has been found in the hyperthermophilic archaeon *Archaeoglobus fulgidus* which consists of two distinct TMD copies ProW1 and ProW2, two copies of ProV and the lipid anchored binding protein ProX. This shows the same ligand preferences as ProU from *E. coli*.
Both systems require a binding protein, ProX, that is able to overcome the usual exclusion of the ligand within its binding site. Crystallization and structural analysis of both ProX proteins has been achieved in the thesis presented here. The structural prerequisites for high affinity binding of glycine betaine and proline betaine is discussed in Chapter 4 for ProX from *E. coli* and in Chapter 5 for ProX from *A. fulgidus*.

1.4 Secondary active transporters

1.4.1 Overview

Secondary active transporters are responsible for the transport of most organic solutes across membranes, especially in eukaryotes that lack uptake systems of the ABC type. They are known to transport almost any inorganic ion of biological importance, as well as all classes of organic molecules. In contrast only a few secondary active transporters are believed to function as export systems for biological macromolecules (Saier Jr., 2000). Among the secondary transporters the major facilitator superfamily (MFS) represents the largest family, which is in fact the second largest family after the ABC family (Paulsen et al., 2000).
Some of the secondary transporters play a key role in the mammalian metabolism a) a Na^+/glucose symporter in the apical domain of intestinal cells transports glucose into the cells. b) a Na^+/P_i symporter is responsible for the reabsorption of inorganic phosphate at the brush border membrane, which is a key process in P_i -homeostasis (Hernando et al., 2000). So far only a few transport defects of secondary transporters have been described. Their clinical importance rather comes from the fact that quite a number of them is involved in drug efflux from pathogenic bacteria (van Bambeke et al., 2000).

1.4.2 Architecture and mechanism

The functional unit of secondary transporters usually is the product of a single gene. Generally, the architecture of such facilitators is a symmetrical structure with 6 + 6 helices as the functional unit. The sequences of loops between those helices may vary considerably, particularly the one between helix 6 and 7, which is in many cases responsible for substrate recognition.
Secondary transporters are low-affinity but high transport rate systems. They couple the diffusion of an ion (usually H^+ and Na^+) down its electrochemical gradient to the translocation of another ion or solute up its chemical gradient. An electrochemical gradient is established across all cytoplasmic membranes by the action of reversible proton or ion pumps. It consists of the two components pH-gradient (ΔpH) and membrane potential ($\Delta\psi$). Since it depends on the pH of the environment, alkalophile bacteria mainly have Na^+-coupled secondary active transporters whereas most systems of bacteria living in neutral or acidic environments are H^+-driven. The larger the electrochemical gradient is, the higher is the transport rate. Secondary active transporters are reversible machines, they are the only active transport systems which can operate by facilitated diffusion when the cell is de-energized. This general behavior and the fact that a single

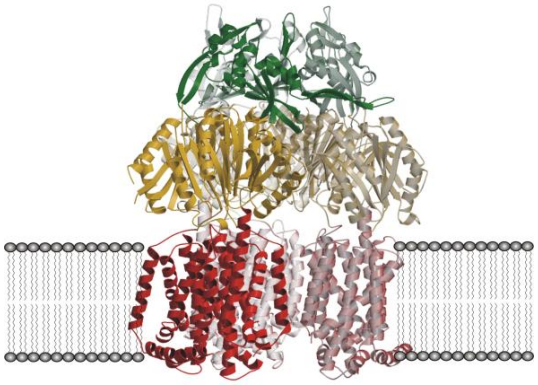

Figure 1.7: AcrB Structural model of an AcrB trimer. Membrane spanning domain is shown in red, periplasmic domain 1 in yellow and periplamic domain 2 in green.

mutation in the lactose permease can convert it into an uniporter, only depending on the concentration gradient of the substrate, supports the idea that those systems evolved from passive carriers. Coupled carriers are subdivided into two classes: a) symporters, if solute and driving ion are transported in the same direction and b) antiporters, if those are transported in opposite directions.

The multi-drug efflux pump AcrB

Multi-drug efflux pumps are widespread among microbes and to date five families are known in prokaryotes that provide resistance to clinically significant drugs. Besides the already introduced ABC transporters, there are four different classes of secondary active transporters involved in drug efflux: 1) the major facilitator superfamily (MFS), 2) small multidrug resistance (SMR), 3) multidrug and toxic compounds efflux (MATE), 4) resistance/nodulation/cell division (RND) (Paulsen, 2003).
Among the 37 known drug efflux pumps found in *E. coli*, AcrB is the most important one. AcrB belongs to the RND family of transporters. It displays an extremely wide substrate specificity, ranging from simple solvents, over dyes and detergents to practically all amphiphilic and lipophilic antibiotics (Elkins and Nikaido, 2003). The high efficiency of AcrB in generating resistance is due to the fact that it forms a tripartite complex with the outer membrane protein TolC and the membrane fusion protein AcrA that ties AcrB and TolC together. Thereby a continuous efflux-channel is formed that crosses the entire periplasm and the outer membrane, allowing the transport of drugs directly to the exterior without any periplasmic intermediates. Apparently, the system transports substrates from the periplasm or from the outer leaflet of the cytoplasmic membrane (where amphiphilic compounds become accumulated) but not from the cytoplasm (Nikaido et al., 1998). A first structure of AcrB has been determined by

Murakami et al. (2002). Like TolC, the functional unit of AcrB is a homotrimer, each consisting of 12 transmembrane helices and a large periplasmic domain (see Figure 1.7). The AcrB trimer is tied together by a long β-hairpin that is exchanged between adjacent monomers. Within the trimer a 35 Å wide central cavity is spanning the membrane and partially the periplasmic part, being accessible by three vestibules from the periplasm that allow drug entry.

2 $MalFGK_2$ from Thermococcus litoralis

2.1 Abstract

Trehalose and maltose uptake in the hyperthermophilic archaeon *Thermococcus litoralis* is mediated by an ABC transport system. The heterotetrameric transport complex $MalFGK_2$ consisting of two membrane spanning subunits and two copies of an ATP binding cassette protein has been crystallized. The crystals belong to the monoclinic space group C2 with unit cell parameters of a = 106.5 Å; b = 150.5 Å; c = 170.1 Å; β = 107.8 °. A native dataset has been obtained at a resolution of 5 Å.

2.2 Introduction

The trehalose/maltose transport complex $MalFGK_2$ from *Thermococcus litoralis* belongs to the family of ATP binding cassette (ABC) proteins. Members of this large superfamily of proteins are ABC transporters which are found in all eubacterial, archaeal and eukaryotic species (Saurin et al., 1999). They share a common architecture usually consisting of four distinct units, two nucleotide binding ATPase subunits attached cytoplasmatically to two membrane spanning domains (Holland and Blight, 1999). Several subclasses can be defined according to the direction of substrate translocation, specificity and subunit organization. Among many others there are the mammalian P-glycoprotein (MDR) involved in multi-drug resistance, the gated ion channel cystic fibrosis transmembrane conductance regulator (CFTR), the transporter associated with antigen processing (TAP) and bacterial binding protein-dependent transport systems. One of the best studied systems of binding protein dependent ABC transporters is the maltose/maltodextrin system of *Escherichia coli* (Boos and Shuman, 1998). In the hyperthermophilic archaeon *T. litoralis* a highly homologous system for maltose uptake has been discovered (Xavier et al., 1996). The trehalose/maltose transport complex of *T. litoralis* consists of the two transmembrane spanning parts MalF (41.6 kDa) and

Table 2.1: Detergents

detergent		solubilization [%(w/v)]	final concentration [%(w/v)]
n-Dodecyl-β-D-maltopyranoside	(DDM)	1.0	0.05
n-Decyl-β-D-maltopyranoside	(DM)	1.0*	0.2
n-Octyl-β-D-maltopyranoside	(OM)	1.0*	1.2
Cyclohexylhexylmaltoside	(CHM)	1.0*	0.1
n-Octyl-β-D-glucopyranoside	(OG)	2.0	1.0
N,N-dimethyldodecylamine-N-oxide	(LDAO)	1.0	0.1
N,N-dimethyldecylamine-N-oxide	(DDAO)	1.0	0.5

Detergents used for solubilization and purification of the MalFGK$_2$ complex with the final concentration in the protein solution used for crystallization. For purification in DM, OM, and CHM the membranes were solubilized using DDM, marked by *.

MalG (33.6 kDa), two copies of the nucleotide binding domain MalK (41.6 kDa) and an extracellular, lipid anchored binding protein TMBP (50.4 kDa). The number of structures of ABC ATPases is constantly growing, and currently known members are HisP (Hung et al., 1998), MalK (Diederichs et al., 2000), MJ1267 (Karpowich et al., 2001), MJ0796 (Yuan et al., 2001), and TAP1 (Gaudet and Wiley, 2001). On the other hand, little is known about the mechanochemical coupling mechanism between ATPases and membrane spanning parts. The first structure of a complete ABC transport complex is the MsbA molecule of *E. coli* determined by Chang and Roth (2001). MsbA is an export complex that forms a homodimer with a V-like quarternary arrangement. The structure was solved at 4.5 Å resolution and a part of the ATPase domain is disordered. Recently, Locher et al. (2002) published the vitamin B_{12} transporter Btu(CD)$_2$ from *E. coli* at 3.2 Å resolution. This is the first structure of a complete binding protein dependent ABC import complex consisting of four polypeptide chains: two membrane spanning parts BtuC and two ATP binding cassettes BtuD forming a homodimer. The overall structure shows a relatively compact arrangement with the ATPases in close contact presenting most likely one stage of the transport cycle of substrate import. In this report, we describe the crystallization and preliminary X-ray analysis of MalFGK$_2$, the trehalose/maltose ABC import complex of *T. litoralis* which consists of a heterodimeric membrane spanning dimer MalFG and a dimer of the C-terminally elongated ATPase MalK.

2.3 Materials and Methods

2.3.1 Expression and Purification

For overexpression of the $MalFGK_2$ complex in *E. coli* strain TG-1, a two vector system was used, one coding for the membrane domains and one for the ABC domain, as described by Greller et al. (2001). Cells were grown in batch culture with NZA medium, induced with 0.1 mM isopropyl thio-β-D-galactopyranosid (IPTG) at $OD_{600} = 0.8$ and harvested after 4 hours. The preparation of membranes was performed according to the procedure described by Greller et al. (2001). In order to find optimal conditions for crystallization, the $MalFGK_2$ complex was purified in seven different detergents. Solubilization was carried out using 1 %(w/v) N,N-dimethyldodecylamine-N-oxide (LDAO), N,N-dimethyldecylamine-N-oxide (DDAO), n-Dodecyl-β-D-maltopyranoside (DDM), or 2 %(w/v) n-Octyl-β-D-glucopyranoside (OG). In case of n-Decyl-β-D-maltopyranoside (DM), n-Octyl-β-D-maltopyranoside (OM) or Cyclohexylhexylmaltoside (CHM) the membranes were solubilized with 1 %(w/v) DDM. After 1h solubilization the solution was heated up to 323 K for 20 min. This led to the precipitation of most of the *E. coli* proteins which were removed by centrifugation at 100000 x g for 20 min. The yellow supernatant was loaded onto a Ni-NTA column, washed with either 80 ml of buffer A: 50 mM Tris-HCl pH 7.5, 500 mM sodium chloride, 5 mM magnesium chloride, 20 mM imidazole with DDM, LDAO, DDAO or OG at the final concentration (see Table 2.1), or washed with 160 ml of buffer A containing the detergents DM, OM or CHM at the final concentration (see Table 2.1). In the latter case the flow rate was set to ≤ 0.2 ml/min in order to replace DDM by another detergent. The concentrations of all detergents used for solubilization and purification are summarized in Table 2.1. Afterwards the protein was eluted with 100 mM imidazole in buffer A with the detergent used for washing. In a further purification step the transporter was loaded onto a red sepharose column with affinity to the ATPase domains and was eluted with a linear gradient (0.5 - 2.0 M sodium chloride). Fractions containing $MalFGK_2$ were pooled (98 % pure as judged by SDS-Page), concentrated and dialyzed in buffer B: 50 mM Tris pH 7.5, 500 mM sodium chloride, 5 mM magnesium chloride to lower the sodium chloride concentration for crystallization. ATPase activity was not tested during purification because transport activity of the $MalFGK_2$ complex in detergent solution cannot be measured, as discussed by Greller et al. (2001). The protein concentration was determined by absorption at 280 nm using a theoretically calculated extinction coefficient $\epsilon_{calc} = 129700\ M^{-1}cm^{-1}$. Finally, the crystallization solution contained 2.5 - 3 mg/ml $MalFGK_2$ in buffer B with one of the detergents at the final concentration listed in Table 2.1.

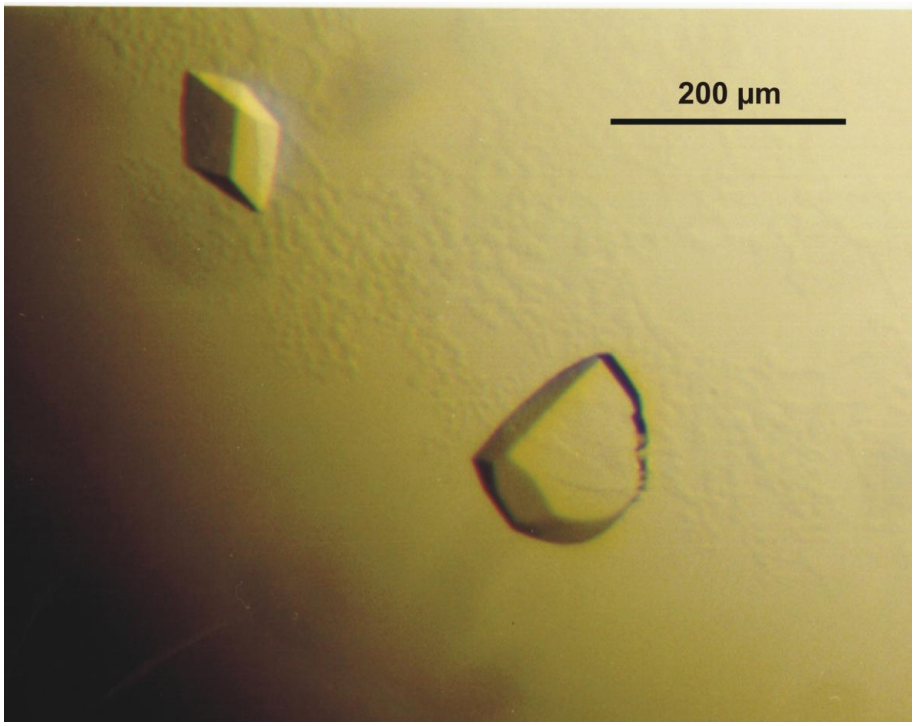

Figure 2.1: Crystals of MalFGK$_2$ Two single crystals grown in 0.1 M sodium cacodylate pH 6.5, 0.4 M calcium acetate, 13-14 %(w/v) PEG4000, 20 %(v/v) glycerol and DDM.

2.3.2 Crystallization and data collection

Initial screening was performed using Screen I (Jancarik and Kim, 1991), Screen II and Membfac (Hampton Research) at 291 K. First small crystals were obtained in Screen I conditions 42 and 46. Further fine screening yielded regular but very thin platelets, which grew thicker in the third dimension after further improvement. No significant differences were observed between crystallization conditions, crystal shape and properties upon the purification with different detergents. For cryocooling of the crystals the following cryoprotectants were tested: glycerol, ethylene glycol, PEG200, PEG400, sucrose, trehalose in buffer B (containing detergent) or dried paraffin oil. All of these solutions damaged the crystals if they were directly transferred to them. Better results were achieved when the cryoprotectant was added to the crystallization solution. The final presaturation condition for crystallization was 0.1 M sodium cacodylate pH 6.5, 0.4 M calcium acetate, 13-14 %(w/v) PEG4000, 20 %(v/v) glycerol mixed with the MalFGK$_2$ complex in buffer B containing DDM. The only visible differences in case of the crystals grown in 20 %(v/v) glycerol were the slightly round edges. The crystals shown in Figure 2.1 had a maximum size of 250 x 150 x 100 μm^3 and grew within 5 days. Two hours before flashcooling of the crystals 100 μl of glycerol (100 %) were added to the reservoir (1 ml).

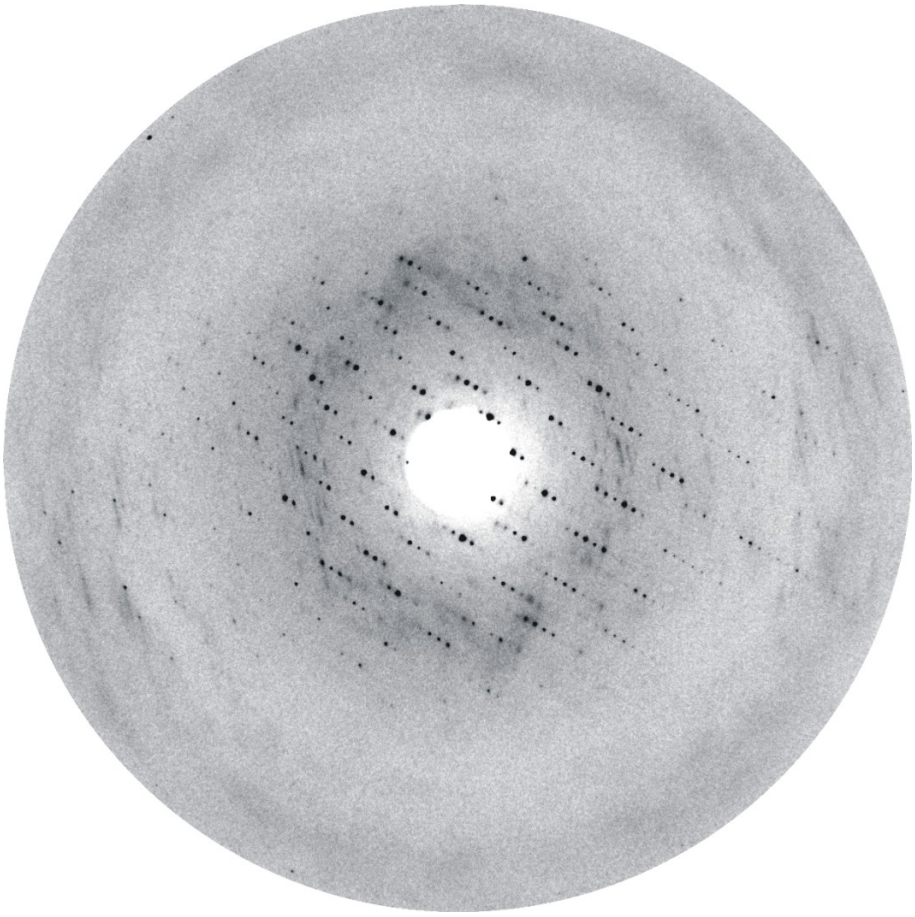

Figure 2.2: Diffraction pattern Data collected at BW7B at the EMBL outstation DESY Hamburg. The resolution at the edge is at 3.8 Å, some single reflections could be observed to 4 Å.

Then the crystals were directly transferred into liquid nitrogen. Data collection was carried out at the EMBL outstation DESY at the beamlines X13 and BW7B equipped with MarCCD and Mar345 detectors, respectively. Crystals diffracted to a resolution of 5 Å; some single reflections could be observed up to 4 Å. Figure 2.2 shows a diffraction pattern collected at BW7B. The data were processed using the program package XDS (Kabsch, 1993).

2.4 Results and discussion

Initial conditions for crystallization were found by screening experiments. Further improvement indicated that a pH range between 6.0 to 7.0 is suitable for crystallization with no significant difference in crystal quality. Crystal growth was more sensitive to the PEG chain length. PEG4000 was found to be optimal and also mixtures of PEG8000

Table 2.2: Data collection statistics

Space group	C2
Unit cell parameters	a = 106.5 Å; b = 150.5 Å; c = 170.1 Å; β = 107.8 °
Wavelength [Å]	0.8456
Resolution [Å]	5
Reflections observed	67990
Unique reflections	11093
Completeness [%]	97.3 (96.8)
I/σI	15.5 (3.54)
R_{sym} [%]	14.7 (53.8)

Data collection statistics for a MalFGK$_2$ crystal. Values in parentheses refer to resolution shell 6 - 5 Å.

and PEG1000 worked well. Addition of nucleotides and their analogs, e.g. ATP, ADP, ADPNP, and ADP-V_i and various detergents seemed to have no strong influence on crystal quality. Crystallization in the presence of 20 %(v/v) glycerol in the reservoir solution allowed direct cryocooling of the crystals without chemical stress during transfer into another solution. Dissolved crystals showed the same pattern on SDS gels as the protein solution used in the crystallization setups. A dataset of 97.3 % completeness could be collected to 5 Å resolution, with some reflections extending to 4 Å at a wavelength of λ = 0.8456 Å and an oscillation range of 1 °. The results of the data collection are summarized in Table 2.2. Data analysis showed that MalFGK$_2$ crystallizes in the monoclinic space group C2 with cell parameters of a = 106.5 Å; b = 150.5 Å; c = 170.1 Å; β = 107.8 °. Assuming a molecular weight of 147.6 kDa and one molecule per asymmetric unit the VM value is 4.4 $Å^3Da^{-1}$ with a solvent content of 72 %, which is in the normal range for a membrane protein surrounded by a detergent micelle. The diffraction image in Figure 2.2 shows that the crystals diffract anisotropically; the lattice order appears to be much better along the c-axis than along the a- and b-axes. We are now working on the improvement of the crystal quality and the phase determination by experimental methods.

3 MalE from Alicyclobacillus acidocaldarius

3.1 Abstract

Maltose binding proteins act as primary receptors in bacterial transport and chemotaxis systems. We report here crystal structures of the thermoacidostable maltose-binding protein from *Alicyclobacillus acidocaldarius*, and explore its modes of binding to maltose and maltotriose. Further, comparison with the structures of related proteins from *Escherichia coli* (a mesophile), and two hyperthermophiles (*Pyrococcus furiosus* and *Thermococcus litoralis*) allows an investigation of the basis of thermo- and acido-stability in this family of proteins.

The thermoacidophilic protein has fewer charged residues than the other three structures, which is compensated by an increase in the number of polar residues. Although the content of acidic and basic residues is approximately equal, more basic residues are exposed on its surface whereas most acidic residues are buried in the interior. As a consequence, this protein has a highly positive surface charge. Fewer salt bridges are buried than in the other MBP structures, but the number exposed on its surface does not appear to be unusual. These features appear to be correlated with the acidostability of the *A. acidocaldarius* protein rather than its thermostability.

An analysis of cavities within the proteins shows that the extremophile proteins are more closely packed than the mesophilic one. Proline content is slightly higher in the hyperthermophiles and thermoacidophiles than in mesophiles, and this amino acid is more common at the second position of β-turns, properties that are also probably related to thermostabilty. Secondary structural content does not vary greatly in the different structures, and so is not a contributing factor.

3.2 Introduction

Starch is one of the major sources of carbon and energy available to heterotrophic bacteria and archaea. Since polysaccharides cannot penetrate the cell membrane, many microorganisms secrete amylases that produce maltose and maltodextrins (oligosaccharides with up to seven α-1,4 linked glucose units) as major degradation products of starch. The subsequent uptake of maltodextrins is usually mediated by a member of the superfamily of ATP-binding-cassette (ABC) transport systems. Maltose ABC transporters are composed of a membrane-bound complex comprising the two hydrophobic permease subunits (MalF and MalG) and two copies of the ATPase subunit (MalK). In addition, a cognate receptor, the maltose binding protein (MBP) is essential for this activity (Schneider, 2003). The binding proteins interact with the membrane permeases that actually transport the solute into the cell. The ATPase component provides the energy both for transport itself, and for overcoming the binding protein's tight interaction with the ligand. In Gram-negative bacteria, the binding proteins are situated in the periplasm between the outer and inner membrane, where they bind their ligands with high affinity. Gram-positive bacteria and archaea, which do not have a periplasm, instead anchor the binding proteins to the outer surface of the cell membrane via an N-terminal lipid moiety (Sutcliffe and Russell, 1995), or alternatively, as observed for some archaea, an N-terminal hydrophobic helix (Albers et al., 1999).
Like other essential systems, binding-protein dependent ABC transporters are widespread in organisms that thrive in extreme environmental niches. The principles that allow some proteins to remain stable and active under harsh conditions, such as high temperature or extreme pH, are of great interest for academic reasons as well as for potential industrial applications. For thermophilic organisms, there is no escape from their hostile environment, and all proteins in the cell must be able to function at high temperature. Therefore, many structures of thermostable proteins have been solved, and a number of features have been identified that are correlated with this property. In contrast, acidophilic microorganisms maintain their cytoplasmic pH close to neutrality by the means of respiratory chain activity or ATP-driven proton pumps (Bakker, 1990). Only proteins that are secreted or otherwise exposed to an acidic environment actually need to be stable under these conditions, and so there are currently very few subjects suitable for studies of the molecular basis of acidostability. Slightly more is known about proteins from thermoacidophilic organisms such as *Alicyclobacillus acidocaldarius*. This Gram-positive bacterium has a pH optimum of 3.6 and a temperature optimum of 57 °C (Darland and Brock, 1971; Wisotzkey et al., 1992). Comparison of the core $(\beta\alpha)_8$-barrel domain of the secreted α-amylase (amylopullulanase) (Matzke et al., 2000) from

A. acidocaldarius with the available structures of homologous proteins from mesophilic organisms indicated that charged amino-acid residues were often replaced with polar but uncharged ones in the thermoacidophile's protein, especially on the molecular surface (Schwermann et al., 1994). This trend had not been identified in thermostable proteins, and so was believed to reflect the enzymes's acidostablity. At the time, it was not clear whether this reflected a more general phenomenon.
We report here the ligand-bound structure of maltose-binding protein of *A. acidocaldarius* (*Acy*MBP) at high resolution, and in multiple crystal packing environments. This protein binds maltose with high affinity (K_D 1.5 μM) over a wide pH range (2.5 to 7) and at temperatures up to 80 °C, making it an excellent candidate for studies of thermoacidostability (Hülsmann et al., 2000). Furthermore, the availability of the crystal structures for *Escherichia coli* MBP (*Eco*MBP (Sharff et al., 1992)) as well as two hyperthermostable MBPs from *Pyrococcus furiosus*, (*Pfu*MBP (Evdokimov et al., 2001)) and from *Thermococcus litoralis*, (*Tli*MBP (Diez et al., 2001)) allow an excellent opportunity to investigate the molecular adaptations related to acidostability.

3.3 Materials and Methods

3.3.1 Protein preparation and analysis

The N-terminally truncated version of *Acy*MBP was prepared from the culture supernatant of maltose-grown cells of *A. acidocaldarius* as described previously (Herrmann et al., 1996). Isoelectric focusing experiments were carried out using Ampholine PAG-plate gels (pH range 3.5 - 9.5, Amersham Biosciences) on a PHAST electrophoresis system. Dynamic light scattering experiments were performed at a protein concentration of 6 mg/ml, in a solution of 20 mM sodium acetate pH 3.5, 9.8 mM ammonium sulfate and 10 mM maltose (the normal storage buffer).
A full-length version of *Acy*MBP was purified from the cytosolic fraction of *E. coli* strain JM109(pRF1) by ion exchange chromatography and Ni-NTA affinity chromatography. Plasmid pFR1 is a derivative of expression vector pQE9 (Qiagen, Germany) that contains the *malE* gene lacking its signal sequence and with the wild type initiation codon TGT (Cys) replaced by GCG (Ala). The resulting translated protein has the sequence $MRGSH_6GS$ fused to the N-terminal alanine residue. Cells were grown in Luria-Bertani broth at 30 °C; gene expression was induced with 0.5 mM IPTG at OD_{650} = 0.4 and growth continued for 4 h. Cell disruption was carried out with a French pressure cell in a solution containing 50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 20 %(v/v) glycerol, 2 mM DTT, 0.1 mM PMSF, 10 mM maltose, and followed by ultracentrifugation. The

resulting supernatant was applied to an SP-Sepharose column equilibrated with 50 mM Tris-HCl pH 7.2, 5 %(v/v) glycerol, 0.5 mM PMSF, 10 mM maltose (buffer A). *Acy*MBP was eluted with buffer A containing 100 mM sodium chloride and subsequently incubated with Ni-NTA matrix in the same buffer for 20 min at room temperature. The mixture was then transferred to a column and the resin was extensively washed with buffer A containing 20 mM imidazole. *Acy*MBP was eluted with the same buffer containing 150 mM imidazole, concentrated, dialyzed against buffer A without imidazole and stored at 4 °C until use.

3.3.2 Crystallization and data collection

The crystals for the *Acy*MBP-1 structure (intact mature protein) were obtained by the sitting-drop vapor diffusion method: 1 μl protein solution (9.6 mg/ml in 50 mM Tris-HCl pH 7.2, 5 %(v/v) glycerol, 10 mM maltose) was mixed with 1 μl of reservoir solution containing 20 %(w/v) PEG8000 and 100 mM CHES buffer pH 9.5, and equilibrated against the reservoir solution at 18 °C. Data were collected in a glass capillary at room temperature using a rotating anode X-ray source equipped with a Mar345 detector, and processed using XDS (Kabsch, 1993). Data statistics are summarized in Table 3.1.
Crystals of the intact mature protein in a larger unit cell (*Acy*MBP-2) were obtained using the hanging-drop vapor diffusion method: 2 μl protein solution (15 mg/ml in 50 mM in 50 mM Tris-HCl pH 7.2, 5 %(v/v) glycerol, 10 mM maltose) were mixed with 2 μl of reservoir solution (18 %(w/v) PEG8000, 100 mM CHES pH 9.5, 10 %(v/v) glycerol) and equilibrated against the latter. Prior to freezing, crystals were transferred to a cryoprotection solution containing 20 %(w/v) PEG8000, the mixed CHES/Tris-HCl buffer, 20 %(v/v) glycerol, 10 mM maltose. Data were collected at the Swiss-Light-Source (beamline X06SA), as high and low-resolution sets that were processed and merged using XDS (Table 3.1).
Crystals of the truncated protein in the larger unit cell (*Acy*MBP-3) grew from hanging drops after three weeks at 37 °C: 2 μl of protein solution (10 mg/ml in 10 mM sodium acetate pH 3.8, 9.8 mM ammonium sulfate, 10 mM maltose) was mixed with 2 μl of reservoir solution (20 %(w/v) PEG5000, 100 mM Tris-HCl pH 8.0, 80 mM ammonium sulfate). Prior to freezing in liquid nitrogen, crystals were soaked in mother liquor containing 20 %(v/v) glycerol. Two diffraction data sets were collected, at 7II at MAX-lab, Lund, Sweden and ID14-4, Grenoble, France. Data were processed using DENZO and SCALEPACK (Otwinowski and Minor, 1997) and merged using programs of the CCP4 package (CCP4, 1994) (Table 3.1).

3.3.3 Structure solution and refinement

The structure of *Acy*MBP-1 was solved using molecular replacement with the program package CNS (Brünger et al., 1998) including data in the resolution range 15 - 4 Å. The search model consisted of all protein atoms of the ligand-bound *E. coli* protein (PDB entry 1ANF, (Berman et al., 2000)). Refinement was also carried out using CNS including all data to 1.8 Å. Combined energy minimization/simulated annealing, and alternating cycles of rebuilding with the graphics program 'O' (Jones et al., 1991), resulted in an R-factor of 19.4 % and a R_{free} of 21.7 % for the final model (summarized in Table 3.1). The structures *Acy*MBP-2 and -3 were solved by molecular replacement using the final structure of *Acy*MBP-1, with all protein atoms as a search model. Two molecules were located in the asymmetric unit of both datasets using AMoRe (Navaza and Saludjian, 1997) (resolution range 8 - 4 Å). This was followed by rigid body refinement (resolution range 40 to 3 Å) that allowed the two domains of each molecule to move independently (carried out using the program REFMAC (Murshudov et al., 1997) as implemented in CCP4 (CCP4, 1994). Further cycles of restrained refinement using all data to the maximum observed resolution were alternated with manual rebuilding. Water molecules were added using CNS (Brünger et al., 1998). Statistics relating to the final refined models are summarized in Table 3.1.

3.3.4 Structural analysis, sequence and structural comparisons

Coordinate sets used for the comparisons with the new MBP structures were obtained from the Protein Data Bank (Bernstein et al., 1977; Berman et al., 2000) as follows: 1ANF (*Eco*MBP with maltose (Quiocho et al., 1997)), 3MBP (*Eco*MBP with maltotriose (Quiocho et al., 1997)), 4MBP (*Eco*MBP with maltotetraose (Quiocho et al., 1997)), 1ELJ (*Pfu*MBP (Evdokimov et al., 2001)) and 1EU8 (*Tli*MBP (Diez et al., 2001)). Structures were compared using 'O', Swiss-PdbViewer (Guex and Peitsch, 1997) LSQMAN (Kleywegt and Jones, 1997) VOIDOO (Kleywegt and Jones, 1994b) and FIT (Lu, 1998). Secondary structural content was assessed using DSSP (Kabsch and Sander, 1983). Surface accessibility was assessed using a per-residue cutoff of 30 $Å^2$ in Swiss-PdbViewer (accessibility calculated using a 1.5 Å probe). Structure-based and other sequence alignments were carried out using LSQMAN (Kleywegt, 1996) and Indonesia (Madsen, Kleywegt and Johansson, Uppsala University). Similar sequences were located using BLAST (Altschul et al., 1990). Sequences were obtained from GenBank (Benson et al., 2003). Signal sequence cleavage sites were predicted using SignalP version 1.1 (Nielsen et al., 1997). Calculation of isoelectric points and amino acid compositions was performed using the server at http://www.expasy.org/tools/protparam.html. Fig-

Table 3.1: Data collection and refinement statistics

	***Acy*MBP-1**	***Acy*MBP-2**	***Acy*MBP-3**
Data collection			
Environment	Rotating anode Mar345 (T = 291 K)	Synchrotron MarCCD (T = 100 K)	Synchrotron Mar345 (T = 100 K)
Wavelength [Å]	1.5418	0.9196	1.0201
Space group	$P2_1$	$P2_1$	$P2_1$
Cell [Å and °]	50.11, 72.16, 57.78, $\beta = 109.63$	49.18, 70.53, 104.06, $\beta = 96.98$	49.23, 70.78, 104.67, $\beta = 96.58$
Resolution range [Å]	50.0 - 1.80 (1.90 - 1.80)	40.14 - 1.45 (1.50 - 1.45)	40.0 - 1.53 (1.56 - 1.53)
No. of unique reflections	35,989	137,804	99,996
Average multiplicity	3.9 (3.3)	6.2 (3.7)	5.2 (2.8)
R_{merge} [%]	15.5 (52.4)	8.4 (50.9)	8.3 (18.7)
Completeness [%]	99.5 (91.4)	99.6 (98.7)	98.8 (88.3)
I/σI	7.5 (2.1)	12.7 (2.4)	11.9 (5.8)
Refinement			
Resolution range [Å]	50 - 1.8	40.0 - 1.45	40.0 - 1.53
No. of reflections total	34,421	137,798	99,996
R, R_{free} values	19.4, 21.7	21.0, 23.2	20.6, 23.1
No. of reflections R_{free} set	1722	3450	3965
No. of non-hydrogen atoms	3065	5705	5755
No. of solvent waters	171	501	549
B-factor of protein atoms [$Å^2$]	23.3	17.3, 19.0	11.6, 12.9
B-factor of ligand atoms [$Å^2$]	12.8	15.6	10.1
B-factor of solvent atoms [$Å^2$]	32.0	27.3	21.2
Ramachandran plot outliers* [%]	0.6	1.2	1.5
R.m.s.d. bond length [Å]	0.005	0.005	0.005
R.m.s.d. bond angle [Å]	1.23	1.21	1.19

Numbers in parentheses refer to the highest resolution shell. *Using a stringent boundary Ramachandran plot (Kleywegt and Jones, 1996).

ures were prepared with the programs 'O', Molray (Harris and Jones, 2001), MolScript (Kraulis, 1991), LIGPLOT (Wallace et al., 1995) and Canvas (Deneba Systems, Inc.).

PDB accession codes

Coordinates and structure factors have been deposited with the Protein Data Bank with entry codes 1URG, 1URS and 1URD.

3.4 Results

3.4.1 Protein expression and purification

A truncated variant of *Acy*MBP was isolated from the culture supernatant of maltose-grown cells of *A. acidocaldarius* and purified by affinity chromatography on Sepharose-linked amylose (Hülsmann et al., 2000; Herrmann et al., 1996). N-terminal sequencing (SALPKGQTIT) and mass spectrometry (MW 41,250 Da) confirmed earlier conclusions (Herrmann et al., 1996) that the truncated protein begins at residue 24 of the mature sequence. Cleavage of the protein during the preparation is believed to be due to the action of an extracellular protease (Schwermann et al., 1994). Dynamic light scattering experiments showed this sample represented a monomer of molecular weight 42 kDa in the temperature range +5 to +50 °C. A full-length version of the protein was purified from *E. coli* strain JM109 harboring plasmid pRF1 as an N-terminal fusion to six consecutive histidine residues by Ni-NTA chromatography (see Materials and Methods for details).

3.4.2 Crystallization, data collection, structure solution and refinement

The structure of one molecule of closed, ligand-bound *Acy*MBP (full-length protein, referred to as *Acy*MBP-1 in further discussion) was obtained by molecular replacement using the structure of ligand-bound *Eco*MBP (Quiocho et al., 1997), and refined to 1.80 Å resolution. The first 28 amino acids of the mature protein (out of a total of 402) were not visible in the electron density and so were not included in the model. Molecular replacement with this structure located two molecules of the full-length protein in a different unit cell; these structures (*Acy*MBP-2a and -2b; residues 30 to 395 visible in both molecules) were refined to 1.45 Å resolution. An additional data set represented two copies of the truncated protein in a very similar crystal form. These were refined to 1.53 Å resolution (residues 30 to 399, and 31 to 399, respectively), and will be referred

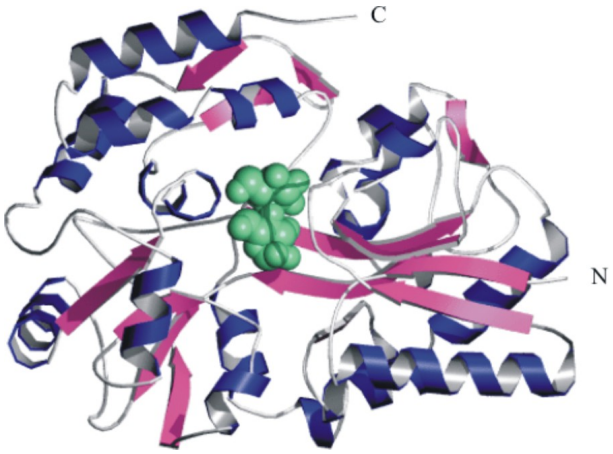

Figure 3.1: Overall structure The structure of *Acy*MBP is shown as a ribbon cartoon, with β-strands colored magenta, and α-helices in blue. The N-terminal and C-terminal ends are marked. The N-terminal domain consists of residues 30-140 and 287-340, while the C-terminal domain consists of residues 141-286 and 341-399. Bound maltose is shown as a space-filling model in green.

to as *Acy*MBP-3a and -3b. Statistics for data collection and final refined structures are presented in Table 3.1.

3.4.3 Overall structures

As typically found in periplasmic binding proteins, the structure of *Acy*MBP is comprised of two domains, each containing a central β-sheet flanked by α-helices (Figure 3.1). The two domains are linked by three hinge segments; rotations of bonds in the hinge region allow the opening and closing of the protein required for entry and exit of ligand during its acquisition and transport (Sharff et al., 1992; Shilton et al., 1996). The five *Acy*MBP structures are very similar, with small differences in the relationship between the two domains ($\sim 2°$, calculated in FIT (Lu, 1998)) accounting for much of the observed r.m.s. differences (0.21 to 0.43 Å when comparing all Cα atoms). The different conformations of the closed forms are probably the result of crystal packing effects, as they are correlated with differences in the environment of the various molecules. Changes in side-chain conformations are restricted to residues exposed on the surface; those inside the protein and close to the ligand-binding site are very similar in all structures. Although three of the models (*Acy*MBP-1, -2a and -2b) represent a construct that originally included the N-terminal extension found *in vivo* (Herrmann et al., 1996), no electron density is visible for this segment in the structures. Western blots using antibody directed to the His-tag confirmed that some or all of this region had been lost either during storage or crystallization. The construct used to obtain the *Acy*MBP-3

Table 3.2: Structural comparisons

Model	***Acy*MBP**	***Eco*MBP**	***Pfu*MBP**	***Tli*MBP**
***Acy*MBP**	374	1.60	1.74	1.81
***Eco*MBP**	321 (31.2)	370	1.65	1.83
***Pfu*MBP**	299 (32.3)	309 (29.8)	381	1.91
***Tli*MBP**	288 (25.0)	311 (23.8)	310 (26.6)	407

Numbers at top right (blue) are r.m.s. differences obtained using a cut-off of 3.5 Å in the superposition in LSQMAN (fast-force algorithm). Numbers at bottom left (green) are the number of residues matching (with % amino-acid sequence identity in parentheses), and those on the diagonal (boldface) are the number of residues in the structures being compared.

structures similarly showed no ordered electron density prior to residue 30 of the mature sequence.

Structures are currently available for MBPs from *E. coli*, *P. furiosus* and *T. litoralis*. These structures and *Acy*MBP are compared in Table 3.2. The most similar to *Acy*MBP is *Eco*MBP, in which 86 % of the residues match with an r.m.s. difference of 1.6 Å using a 3.5 Å cutoff (referring to the *Acy*MBP-1 structure). Among the residues in this core structure, 31 % of the amino acids are identical. A comparable result was obtained with *Pfu*MBP, with 80 % of its residues matching to equivalents in *Acy*MBP. Each of these three proteins is more distantly related to the trehalose/maltose receptor *Tli*MBP (which is also present as a second transport receptor in *P. furiosus*, presumably as the result of horizontal gene transfer (Diruggiero et al., 2000).

3.4.4 Ligand binding

Electron density for maltose (α-D-glucosyl-1,4-α-D-glucose) was visible in the cleft between the two domains, with full occupancy in all structures (Figure 3.2A). In addition, both Fo-Fc and 2Fo-Fc maps showed a partial occupancy of an additional glucose ring at the reducing end of the sugar in *Acy*MBP-2 and -3. The electron density for this third ring was stronger in both *Acy*MBP-2 structures than in those of *Acy*MBP-3; it was completely absent in *Acy*MBP-1. This third ring presumably represents a maltotriose impurity that varies in quantity among the different batches of maltose used for purification and co-crystallization. It is unlikely that the electron density for the third ring results from the binding of maltose in two different modes, as the density for the first two rings is in all cases as strong as that of the surrounding protein atoms. To investigate the ratio of maltose : maltotriose, refinement was carried out with different occupancies for the third ring. The best fit to both 2Fo-Fc and Fo-Fc maps was used to determine the closest match. The fractional occupancy of maltotriose in *Acy*MBP-2

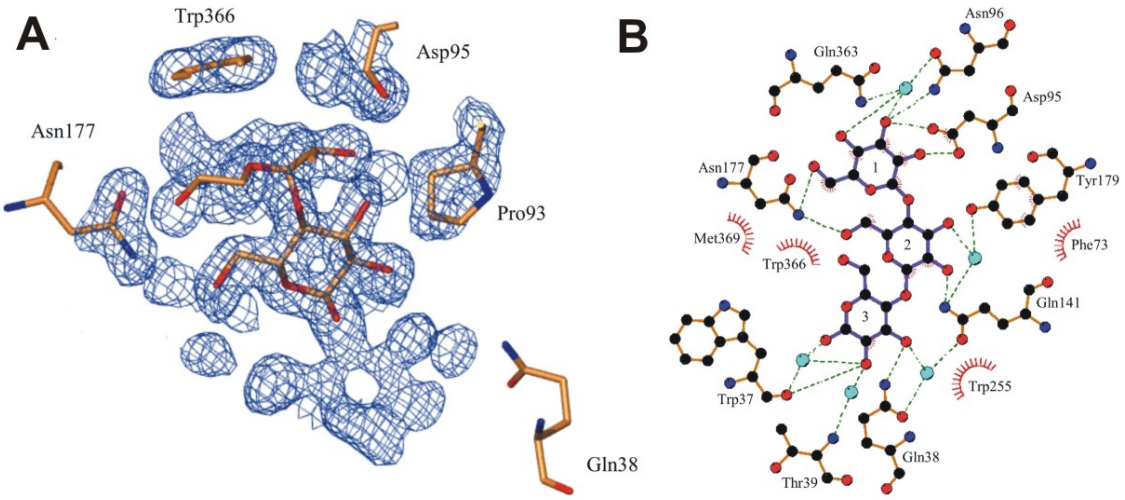

Figure 3.2: Ligand binding in *Acy*MBP A) Maltose (and maltotriose) density in the last electron density map prior to addition of ligand in the *Acy*MBP-2a model. The structure of maltose from the final refined structure is illustrated with a ball-and-stick model. B) LIGPLOT cartoon illustrating maltotriose binding in *Acy*MBP-1.

was in this way estimated to be 0.65 in both molecules, and in *Acy*MBP-3a and -3b, the corresponding values are 0.30 and 0.45, respectively.

Maltose makes direct hydrogen bonds with residues Asp95, Asn96, Gln141 and Asn177 (Figure 3.2B). Several indirect hydrogen bonds are mediated by water molecules, and link the disaccharide to residues Tyr179 and Gln363. As is typical for sugar-binding proteins, 18 non-polar interactions largely involve contacts with aromatic residues, in this case Tyr179, Trp255, Met356, Trp366 and Met369 of the C-terminal domain, and to a lesser extent Phe73 of the N-terminal domain. Binding of the β-anomer of the disaccharide would be precluded by steric clashes with Trp255. The third α-D-glycosyl ring of maltotriose forms additional hydrogen bonds to Trp37 (main-chain oxygen), Gln38, and Gln141, and through water molecules to Trp37 (main-chain oxygen), Thr39 (main-chain nitrogen) and Gln141. There are also further stacking interactions with Trp255. Assuming the same sugar conformation, extension of the trisaccharide at the reducing end would lead to clashes near residue 67 in the closed form. Proceeding from the non-reducing end, intra-molecular hydrogen bonds link O2 of the first sugar unit to O3 of the second, and O2 of the second sugar to O3 of the third; the result is a curved trisaccharide (Figure 3.3A), consistent with the form expected for the free sugar in solution.

The two fully-occupied sites in the *Acy*MBP structures also have full occupancy in the available *Pfu*MBP structure (Evdokimov et al., 2001) (Figure 3.3B). When maltose is bound to *Eco*MBP, the same two sites are preferred (Spurlino et al., 1991) (Figure 3.3C). Many of the interactions in these two sites are conserved in the three proteins.

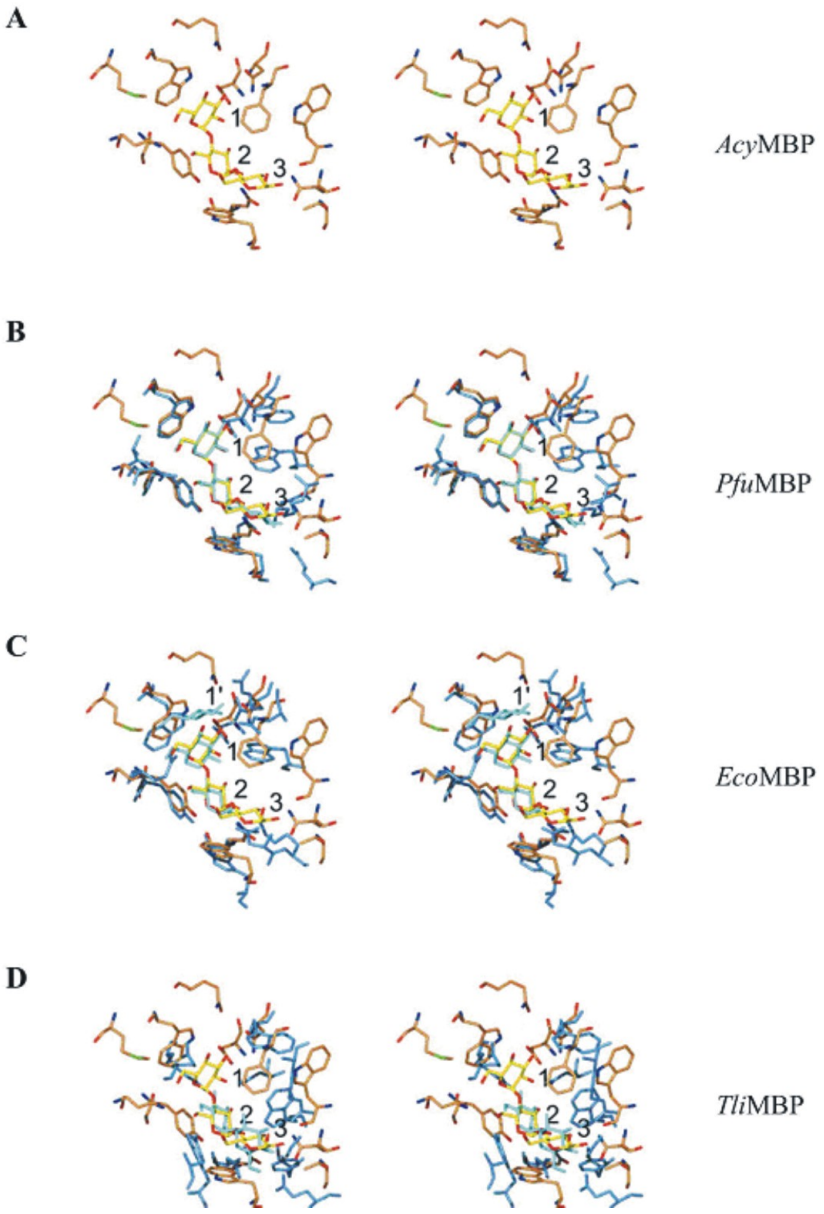

Figure 3.3: Comparison of ligand binding in available MBP structures In each case, *Acy*MBP is shown as a stick model with normal atomic coloring, and the protein to be compared is shown with cyan (darker for protein atoms). The view is similar to that in Figure 3.1. A) *Acy*MBP. The aromatic residues interacting with the ligand and the residues involved in hydrogen bonding are shown, and the three sugar sites are labeled, beginning from the non-reducing end. B) Equivalent view of *Pfu*MBP, showing maltotriose bound in the same three subsites observed for *Acy*MBP. Although the residues mediating hydrogen bonds are sometimes different, the same total number of interactions is formed. Most interacting aromatic residues are also conserved. C) Comparison with *Eco*MBP. Two of the subsites used by *Acy*MBP and *Pfu*MBP are used in binding the same sugar in *Eco*MBP. The third sugar unit is, however, placed differently in *Eco*MBP (labeled as 1'); the loop at residues 12-17 has moved and therefore it is not possible to accommodate a glucose ring at site 3. D) Comparison with *Tli*MBP. In this case, sugar is positioned in sites 2 and 3, as defined above, but with different ring orientations, and no conserved interactions.

Counterparts to *Acy*MBP's Asp95, Tyr179, Phe 180, Trp255 and Trp366 are present in both *Pfu*MBP and *Eco*MBP. A third glycosyl ring was also observed with partial occupancy in *Pfu*MBP at a position very similar to that seen in *Acy*MBP (site 3 in Figure 3.3B); this was likewise believed to represent maltotriose, a common contaminant of maltose. When maltotriose is bound to *Eco*MBP, however, the third ring is placed at the other end of the binding site, *i.e.* at the non-reducing end of the maltose (site 1' in Figure 3.3C). In *Acy*MBP, residues 37-41 form hydrogen bonds to the glycosyl unit in site 3, and corresponding interactions are found in *Pfu*MBP. This site is blocked in the *E. coli* protein; the loop at residues 12-17 is different and therefore it is not possible to accommodate a glycosyl ring at this position. At the same time, local changes in the 1'-site of the *Acy*MBP and *Pfu*MBP structures prevent them from binding a third glycosyl unit in the same way as *Eco*MBP. Binding of trehalose (α-D-glucosyl-1,1-α-D-glucose) to *Tli*MBP positions the two glycosyl units in sites 2 and 3 of *Acy*MBP and *Pfu*MBP (Figure 3.3D). The plane of the sugar ring in site 2 has a very similar orientation in *Tli*MBP, although no interactions are conserved. The plane of the sugar ring in site 3 is almost anti-parallel to that in *Acy*MBP and *Pfu*MBP, and so it is not surprising that no interactions are conserved.

3.4.5 Electrostatic properties

It had been noted during the biochemical analysis of the truncated form of *Acy*MBP that the experimentally measured isoelectric point ($\sim$ 10) was very different from that calculated based on its amino acid sequence (6.5). The *Acy*MBP structures provided a clear explanation for this unusually large discrepancy: the protein has a striking excess of basic residues on its surface. This observation prompted us to undertake a closer investigation of charged residues in all of the available MBP structures. The comparisons showed that *Acy*MBP, *Eco*MBP and *Tli*MBP have a similar number of basic residues on their surfaces, while *Pfu*MBP has slightly fewer (Figure 3.4A). Since our main interest lies in understanding the adaptation of proteins to environmental conditions, we have been consistent in considering histidine with the positively charged residues, although some of these will not be charged in the normal surroundings of the mesophile. Fewer acidic residues are exposed in *Acy*MBP than in any of the other MBPs, leading to an overall positive charge. Furthermore, the total number of both positively- and negatively-charged residues was seen to be much smaller for *Acy*MBP than for the other proteins (Figure 3.4B). While the number of acidic and basic residues in the thermoacidophile's protein is approximately equal, most of its basic residues are exposed on the surface whereas most acidic residues are buried in the structure. The differences

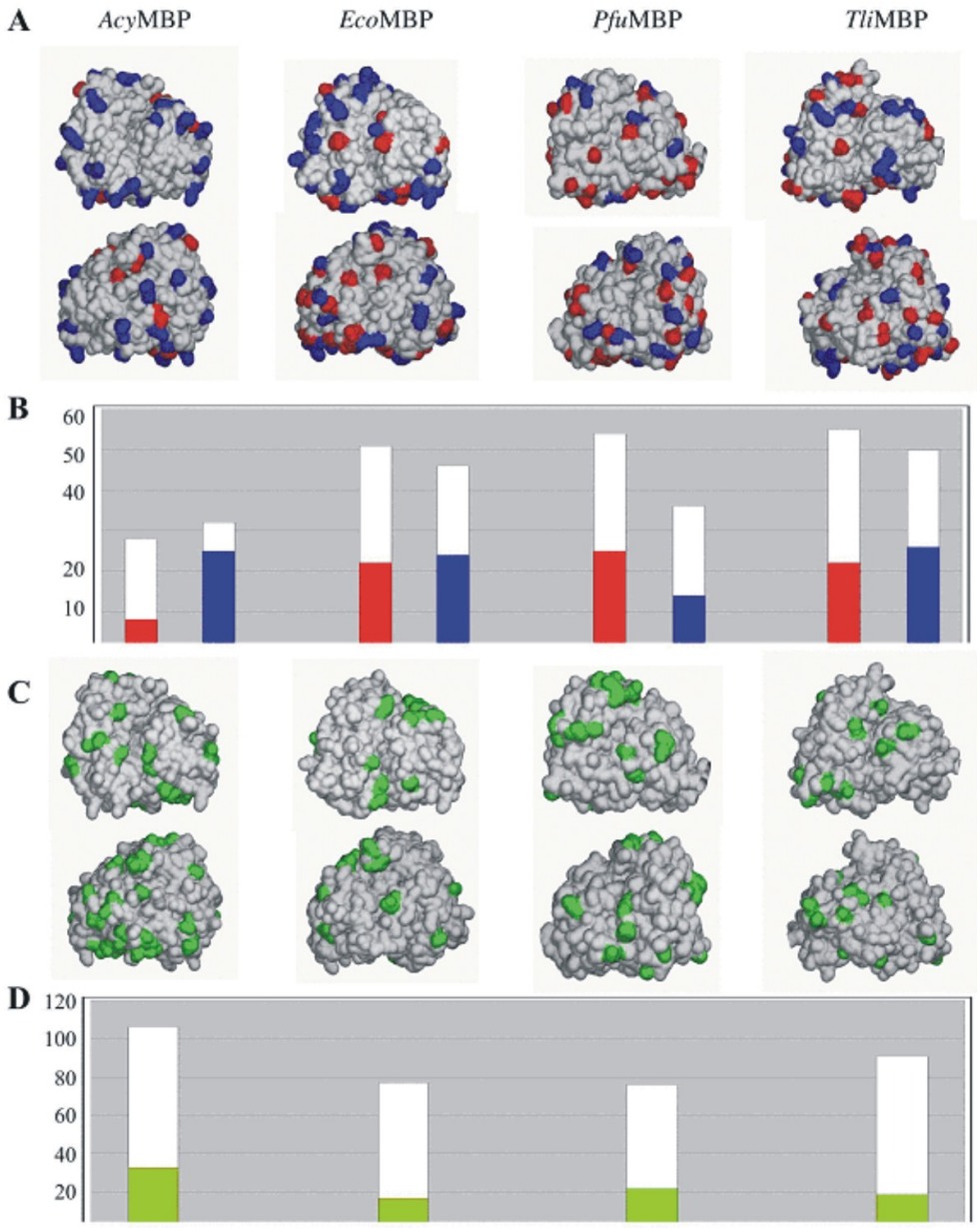

Figure 3.4: Distribution of charged and polar residues A) Acidic (red) and basic (blue) residues on the surfaces of the various proteins are shown. In the top views, the orientation is the same as that in Figure 3.1, while those at the bottom are rotated 180 ° around a vertical axis. Surfaces were prepared in the program 'O'. B) The absolute numbers of charged residues for each case in panel A are presented in a histogram, with exposed residues shown with the corresponding color, and buried residues in white. C) Polar residues (in this case, asparagine, glutamine, serine and threonine) on the surfaces are colored green in the same proteins. D) The absolute numbers of polar residues on the surface and buried in each instance in panel C are presented in a histogram, with exposed residues shown in the corresponding color.

are particularly striking on the face of the protein near the carbohydrate-binding site, where acidic residues are totally absent in *Acy*MBP. As an increase in the number of salt bridges has been correlated with thermostability of proteins, we manually searched for ion pairs in the various proteins at the graphics terminal, using a cut-off distance of 3.3 Å. This inspection identified eight ion pairs in *Eco*MBP (six buried), 13 in *Pfu*MBP (nine buried) and nine *Tli*MBP (seven buried). In each of the *Acy*MBP structures, only four salt bridges were observed, two of which were buried. Thus, fewer salt bridges are buried in *Acy*MBP than in the other MBP structures, but the number exposed on its surface does not appear to be outside the normal range.

Charged residues in the other MBPs are frequently substituted for uncharged or nonpolar residues in *Acy*MBP, both on the surface and in the interior of the protein (Figure 3.4C and D). An increased number of polar residues in *Acy*MBP is particularly evident on the side facing away from the binding cleft (Figure 3.4C). Since the surface near the binding cleft has fewer charged residues as well, it will have a more nonpolar character than the equivalent face in the other MBPs. The same pattern can be seen in more detail using structure-based sequence alignments (Figure 3.5). To further investigate these patterns, we looked at the sequences of a number of other MBPs that were identified by a BLAST20 search as being relatives of *Acy*MBP. Only a limited number of hyperthermoacidophilic MBPs appeared in this list, although large numbers of thermophilic and mesophilic sequences were located, as well as ones from the mildly acid-tolerant lactobacilli. The most similar of these sequences were selected for further analysis, since the sequence/structural relationships were expected to be easiest to interpret in this set; some more distant relatives were also included, to determine whether any trends were general within the structural family. Results are summarized in Table 3.3. It immediately becomes apparent that, like *Acy*MBP, the hyperthermoacidophilic MBPs have fewer charged residues than average. Hyperthermophiles, thermophiles and mesophiles all have higher contents of both acidic and basic residues than the (hyper)thermoacidophiles. Among the acidic residues, the content of both aspartate and glutamate is reduced; a similar trend is apparent for the mildly acid-tolerant lactic acid bacteria as well. For basic residues, lysine accounts for most of the differences, primarily because it is much more common than either arginine or histidine. In contrast, the contents of polar but uncharged residues (most consistently asparagine and glutamine, the structural homologues of aspartate and glutamate) are generally increased. Sequence alignments (not shown) confirmed that the patterns observed for the hyperthermoacidophilic proteins were the same as those for the thermoacidophilic *Acy*MBP.

Table 3.3: Amino-acid composition of MBPs

organism	growth T_{opt} [°C]	No Res.	BLAST *E*	BLAST id [%] (overlap)	Asp	Asn	Glu	Gln	Arg	Lys	His	Total Asn + Gln	Total Asp + Glu	Total Arg + Lys + His	Pro
(* hyper)thermoacidophiles															
*Acy*MBP	57	374	0.0	–	5.1	5.1	2.4	8.0	1.3	5.9	1.3	49	28	32	6.7
Thermoplasma acidophilum	55-60	437	1e-09	21 (363)	3.0	7.8	4.1	4.8	0.2	3.9	1.1	55	31	23	5.0
*Sulfolobus solfataricus**	80	450	5e-23	27 (407)	2.0	6.7	3.6	7.8	1.3	2.4	0.9	65	25	21	6.4
acid-tolerant mesophiles															
Lactobacillus plantarum	37	370	2e-32	29 (334)	5.4	7.6	2.7	4.1	0.8	13.5	0.3	43	30	54	3.9
Lactococcus lactis	25-35	369	7e-21	27 (380)	8.1	6.5	2.4	5.7	0.0	9.8	0.0	45	39	36	5.1
Lactobacillus gasseri	37	381	9e-16	25 (377)	4.7	7.6	4.7	5.2	0.3	12.9	1.0	49	36	54	5.2
(* hyper)thermophiles															
Thermoactinomyces vulgaris	40-50	397	3e-39	29 (381)	5.0	5.8	6.0	6.0	1.0	11.3	0.5	47	44	51	6.5
Thermoanaerobacter tengcongensis	75	372	1e-79	43 (369)	7.0	5.4	4.6	5.1	0.3	9.7	1.1	39	43	41	7.0
Thermotoga maritima	70-80	373	4e-40	33 (337)	5.9	5.1	7.5	3.8	2.1	7.5	0.3	33	50	37	6.2
*Tli*MBP*	85	409	2e-32	30 (334)	5.4	4.4	8.1	4.9	2.7	8.1	1.5	38	55	50	6.4
*Pfu*MBP*	100	381	1e-31	30 (350)	5.5	4.2	8.9	4.5	0.5	7.9	1.0	33	55	36	7.6
mesophiles															
Clostridium perfringens	40	379	2e-76	41 (375)	7.1	5.0	5.8	5.0	0.5	9.2	0.8	38	49	40	5.0
Deinococcus radiodurans	26-30	375	6e-48	34 (354)	4.5	5.9	2.1	4.5	2.1	8.0	0.5	39	25	40	5.9
*Eco*MBP	37	370	9e-41	31(383)	6.5	5.7	7.3	2.4	1.6	10.0	0.8	30	51	46	5.7
Enterobacter aerogenes	25-35	407	3e-41	30 (382)	6.2	5.9	5.9	2.4	1.6	10.5	0.8	31	45	48	5.9
Shigella flexneri	30-37	370	3e-40	30 (383)	6.5	5.7	7.3	2.4	1.6	10.0	0.8	30	51	46	5.7
Salmonella enterica	37	370	3e-40	32 (384)	6.8	5.7	5.9	3.2	1.6	10.3	0.5	33	47	46	5.7

Optimal growth of the organism (American Type Culture Collection, http://www.atcc.org/SearchCatalogs/Bacteria.cfm) is not necessarily the same as the temperature optimum for protein activity. Related proteins were identified using a BLAST search with the folded region of *Acy*MBP as a probe. Sequences selected as described in the text were obtained from GenBank as follows: *Acy*MBP, gi|6686562|; *Sulfolobus solfataricus*, gi|15898024|; *Thermoplasma acidophilum*, gi|16081289|; *Lactobacillus plantarum*, gi|28377115|; *Lactococcus lactis*, gi|15673665|; *Lactobacillus gasseri*, gi|23003417|; *Thermoanaerobacter tengcongensis*, gi|20808247|; *Thermotoga maritima*, gi|15643960|; *Thermoactinomyces vulgaris*, gi|13537290|; *Tli*MBP, gi|13787046|; *Pfu*MBP, gi|13096471|; *Clostridium perfringens*, gi|18311325|; *Deinococcus radiodurans*, gi|15805588|; *Eco*MBP, gi|2781044|. The mature regions corresponding to the structures compared here were calculated with the aid of SignalP, and then analyzed using ProtParam.

3.4.6 Other features

The various MBP structures were assessed using a number of criteria that have been associated with thermostability, i.e. higher secondary structural content, fewer/smaller internal cavities, clusters of hydrophobic residues, proline disposition, and so forth. Secondary structural content was evaluated using DSSP (Kabsch and Sander, 1983); the results are summarized in Table 3.4. The content of helix and sheet does not vary substantially among the various proteins, remaining relatively constant in terms of the fraction of total structure.

Using the program VOIDOO (Kleywegt and Jones, 1994b), we searched for internal cavities in the various proteins; the results are summarized in Table 3.5. Both the number of cavities, and their total volume, is greater in *Eco*MBP than in any of the other proteins. Thus all of the extremophile MBPs are more closely packed than that of the mesophile.

It has been reported that proper placement of prolines with respect to secondary structural elements can enhance protein stability (Watanabe et al., 1991; Watanabe et al., 1996; Prieto and Serrano, 1997). The total proline content is in general slightly higher in the (hyper)thermophilic MBPs than in the mesophiles (Table 3.3). Proline residues are more frequently found at the second position of β-turns in the (hyper)thermophile structures (*Acy*MBP, 4 cases; *Pfu*MBP, 5 cases; *Tli*MBP, 7 cases) than in the mesophilic one (*Eco*MBP, 1 case). The distribution of prolines at the N- and C-terminal ends of helices does not appear to differ significantly, while only *Pfu*MBP shows an appreciable decrease in the frequency of prolines within elements of secondary structure (Table 3.4). It was noted earlier that *Pfu*MBP has several dramatic clusters of isoleucine and aromatic residues, features that were suggested to increase its thermal stability relative to *Eco*MBP (Evdokimov et al., 2001). However, this phenomenon is not apparent in either *Acy*MBP or *Tli*MBP.

3.5 Discussion

In vivo, the mature form of *Acy*MBP includes an $\sim$ 30-residue N-terminal sequence that is believed to act as a linker between the binding domains and the lipid anchor (a palmitate covalently attached to Cys1) (Herrmann et al., 1996). In the present study, two of the crystals were prepared using full-length *Acy*MBP. Another crystal represented a truncated version in which much of the N-terminal sequence was removed by the action of a protease during protein preparation. No electron density corresponding to the N-terminal segment was visible in any of the maps. Tests of the supposedly full-length

Table 3.4: Comparison of residues involved in secondary structure

	α-helix	3_{10}-helix	β-strand	β-bridge	Total α	Total β	Total sec.str.	Total res.	Pro inside sec.str.	Pro outside sec.str.
*Acy*MBP	159	12	68	3	171	71	242	374	11	14
	(42.5)	(3.2)	(18.2)	(0.8)	(45.7)	(19.0)	(64.7)		(44.0)	(56.0)
*Eco*MBP	157	6	72	5	163	77	240	370	9	12
	(42.4)	(1.6)	(19.5)	(1.4)	(44.1)	(20.8)	(64.9)		(42.9)	(57.1)
*Pfu*MBP	158	17	65	8	175	73	248	381	9	19
	(42.0)	(4.5)	(17.3)	(2.1)	(46.5)	(19.4)	(66.0)		(32.1)	(67.9)
*Tli*MBP	182	10	65	9	192	74	266	407	11	15
	(44.8)	(2.5)	(16.0)	(2.2)	(47.3)	(18.2)	(65.5)		(42.3)	(57.7)

Comparison of absolute numbers (and percentages in parentheses)

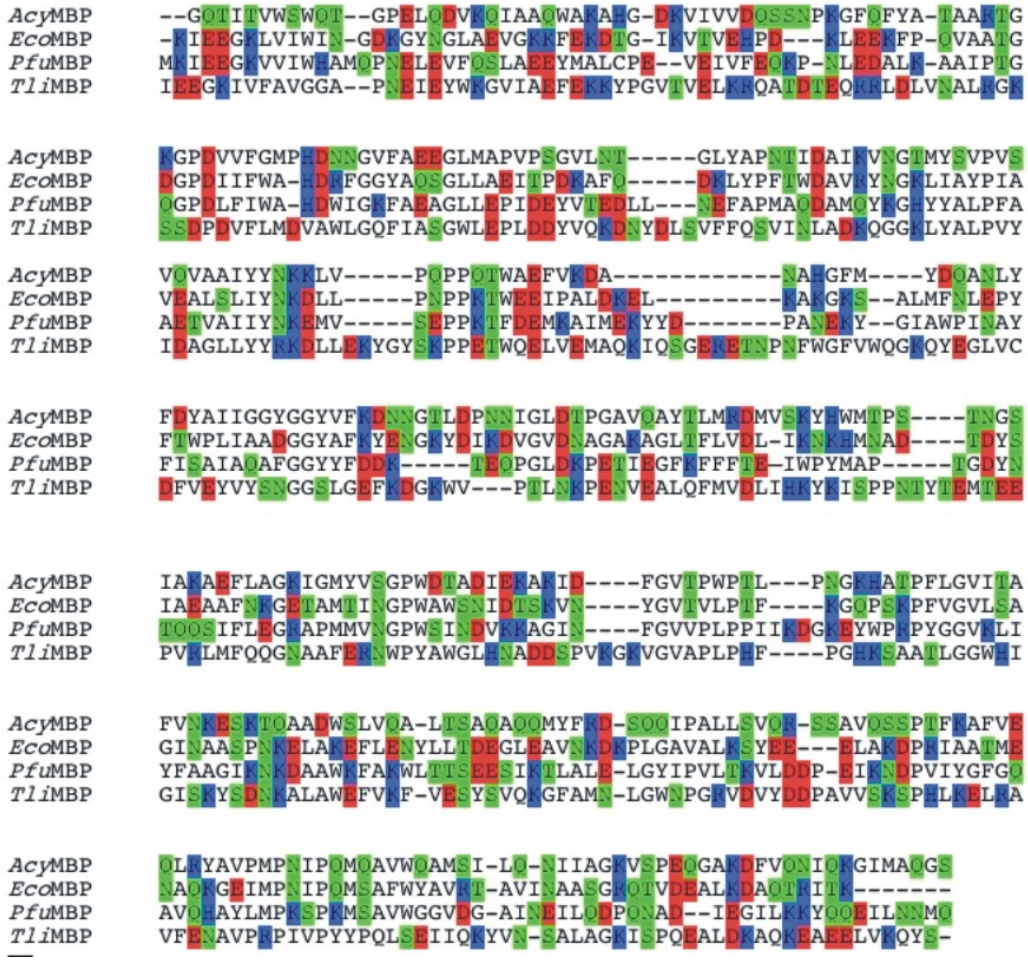

Figure 3.5: Structure-based sequence alignments Alignment of the structures of *Acy*MBP, *Eco*MBP, *Pfu*MBP and *Tli*MBP was carried out using the default parameters in LSQMAN. Residues were then colored according to whether they are acidic (Asp, Glu; red), basic (Arg, Lys, His; blue) and polar (in this case, Asn, Gln, Ser, Thr; green).

Table 3.5: Comparison of cavities

Structure	Number of cavities	Volume total [Å^3]
*Acy*MBP-1	5 (0.7)	104.8 (26.6)
*Eco*MBP	11	511.3
*Pfu*MBP	4	331.8
*Tli*MBP	4	124.4

Minimum volume for a cavity was set to 5 Å^3. Standard deviations for the multiple observations of the *Acy*MBP structure are shown in parentheses.

protein indicated that it, too, had lost the N-terminal segment during either storage or crystallization. The combined observations suggest that this segment is normally disordered, which agrees well with the suggested biological role: a linker should be flexible to allow MBP to interact with carbohydrates in the periplasm as well as with the membrane-bound transport proteins. Presumably the linker lacks sequences that are recognized by proteases in its natural environment.

Like other MBPs, *Acy*MBP is designed to bind small α(1,4) polymers of glucose of varying size. Both aromatic stacking and hydrogen-bonding interactions are important in binding the cognate ligand tightly (K_D for maltose is 1.5 μM (Hülsmann et al., 2000)). The presence of three clear glycosyl-binding subsites in the structure suggests that maltotriose should be the optimal ligand, and indeed, maltotriose is seen in the electron density (occupancy 0.35 - 0.65 in the various final structures), although it can have been only a minor contaminant (< 3 %) of the sugar added during purification and crystallization. Theoretically, there is sufficient space for longer sugars at the non-reducing end, but additional glycosyl units would protrude from the binding site and lack good contacts with the protein. A preference for maltose and maltotriose would make sense in biological terms, since these are the main products when starch is digested by the amylopullulanase of *A. acidocaldarius* (Schwermann et al., 1994).

Comparison of the *Acy*MBP structures with related proteins from a mesophile (*E. coli*) and two hyperthermophiles (*P. furiosus* and *T. litoralis*) gave other insights into structure and function. For example, the overall structural differences provide a good basis for understanding the various approaches to ligand binding. Two subsites (sites 1 and 2 in Figure 3.3) are shared by three of the proteins (*Acy*MBP, *Pfu*MBP and *Eco*MBP), and represent their common mode of binding to maltose. Like *Acy*MBP, *Pfu*MBP shows a strong preference for maltotriose over maltose or trehalose, placing the third sugar ring in site 3. *Eco*MBP also binds more tightly to maltotriose, but for several reasons binds the third glycosyl unit in a different subsite (site 1' in Figure 3.3). Therefore, *Acy*MBP

and *Pfu*MBP will deliver a trisaccharide to their membrane permease partners in a different way than *Eco*MBP, which further suggests that differences in the permeases will exist in these organisms.

Although the contributing residues are not always conserved, the number and position of hydrogen bonds for the two central glycosyl units (sites 1 and 2) are very similar in these three structures. Regardless of the mode of binding the third glycosyl unit, each protein utilizes a similar number of hydrogen bonds in this site. Further, *Acy*MBP, *Eco*MBP and *Pfu*MBP have similar layers of aromatic residues interacting with the relatively non-polar faces of the sugar rings, which provide much of the energy of binding. As might be expected, these interactions can be provided in a fairly non-specific fashion, although the large flat surfaces of aromatic residues are a natural complement to the shape and character of a glycosyl ring (Vyas, 1991). In sites 1-3, the largest surface (at left in Figure 3.3) is contributed by the C-terminal domain, and the residues involved are quite well conserved. By contrast, the interactions supplied by Phe45 in the N-terminal domain of the *Acy*MBP structure (at right in Figure 3.3) can be provided instead by residues from a different part of the sequence: Trp65 in *Pfu*MBP and Trp62 in *Eco*MBP. The aromatic component in *E. coli*'s site 1' (Tyr341) is missing in *Acy*MBP and *Pfu*MBP, and inspection of a number of MBP sequences suggests the presence/absence of this tyrosine will be the best predictor of binding mode for the trisaccharide; the effects of insertions/deletions near site 3 are more difficult to evaluate.

Despite the sequence/structural homology, ligand binding in *Tli*MBP is substantially different than that of the other three MBPs, and resemblance is only in the approximate position of the bound ligand. *Tli*MBP has two sites in common with *Acy*MBP and *Pfu*MBP (sites 2 and 3 in Figure 3.3), but it has a completely different approach to binding in each of those sites. In contrast to the other proteins, *Tli*MBP accepts trehalose as well as maltose, but does not bind to longer oligosaccharides. The *P. furiosus* genome contains a second binding protein corresponding very closely to *Tli*MBP that can presumably take on the role of disaccharide binding in that organism.

Since acidophilic organisms can regulate their internal pH (Bakker, 1990; Cobley and Cox, 1983; Krulwich et al., 1978), only proteins exposed to the external environment (*i.e.* either located in the periplasm or secreted, such as the binding proteins) must actually be stable under acidic conditions. *Acy*MBP thus represents a very rare opportunity to study a protein that is both thermostable and acid-stable. However, the problems of thermostability and acidostability are by no means the same. Factors that are thought to be correlated with the ability of a protein to remain stable and active at high temperatures include an increased number of hydrogen bonds, additional or improved electrostatic interactions including those mediated by salt bridges, optimized hydrophobic interac-

tions, increased compactness or packing density, increased secondary-structural content, truncation of solvent-exposed loops, greater number of prolines in loops or other key positions, and increased polar compared with non-polar surface areas (Ladenstein and Antranikian, 1998; Kumar and Nussinov, 2001; Sanchez-Ruiz and Makhatadze, 2001). Much less is known about acidostability or thermoacidostability. The availability of a range of MBP structures and sequences allows us to assess their features in the light of differing physical settings.

Matzke et al. (1997) analyzed the sequences of α-amylases using the crystal structure of the Taka enzyme from *Aspergillus oryzae* and predicted a reduced charge density on the surface of the secreted *A. acidocaldarius* enzyme. It was believed that this adaptation would prevent electrostatic repulsion of charged groups at low pH. Our analysis of *Acy*MBP showed that there are indeed fewer charged residues on its surface, and furthermore that most of these carry a positive charge. Although the numbers of basic and acidic residues are equal in *Acy*MBP (28 versus 29), most acidic residues are buried within the protein. At the pH optimal for the organism ($\sim$ 3.5), the few exposed acidic groups will in general be protonated, giving this protein a rather dramatic positive charge surplus. This is a different strategy to that observed in the acid-stable (but not thermostable) xylanase from *Aspergillus kawachii* in which many acidic residues are exposed (although even that protein is expected to be positively charged at the pH optimum of 2) (Fushinobu et al., 1998). The preponderance of basic residues would seem to have the advantage that their high pKa's are outside the pH range for optimum activity (pH 2.5 - 7), which would effectively protect the protein from destabilizing changes in surface charge if the pH in the environment varies. This is in fact the case with more "normal" proteins: the pKa's of the vast majority of their surface groups (3 - 5 for acidic residues and 9 - 11 for basic residues) are outside the usual pH range of their environment. The only exception is histidine with a pKa of 6.05; histidine, however, is relatively rare in the compared MBPs. Base-stable proteins might show a corresponding reduction in the number of basic residues. If so many surface residues carry the same charge, it is reasonable to postulate that charge repulsion will be a problem if their number is too large. Overall, the thermoacidophilic amylopullulanase was shown to contain approximately 30 % fewer charged residues than its closest relatives (Schwermann et al., 1994; Matzke et al., 1997). Analysis of the amino-acid sequences shows that (hyper)thermoacidophilic MBPs also have a smaller total number of charged residues than their (hyper)thermophilic or mesophilic counterparts. These results suggest that the trend is a general one, and linked to acidostability specifically.

The number of salt bridges buried within *Acy*MBP is smaller than for its hyperthermophilic as well as mesophilic counterparts, while the numbers exposed on the surface

are roughly the same. This is in accordance with previous suggestions that buried salt bridges can actually be destabilizing (Waldburger et al., 1995; Waldburger et al., 1996), and that an increased number of such salt bridges is tolerated in thermophilic proteins, rather than being an asset (Elcock, 1998). The results presented here suggest that reducing the number of buried salt bridges may be a factor in acidostability.
An increase in the number of polar but uncharged residues is found both on the surface and inside *Acy*MBP compared to the other proteins. Such substitutions on the surface would be expected to help maintain the polar-outside/nonpolar-inside balance that is critical for a folded protein in an aqueous environment (Schwermann et al., 1994).
The content of secondary structure does not appear to vary significantly in the various MBPs (Table 3.4). Proline content was slightly higher than average in *Acy*MBP (6.7 % compared to $\sim$ 5.7 % for mesophiles, Table 3.3), but does not appear to be correlated with a strengthening of the existing secondary structure. Further, the slightly higher proline content is also observed for the hyperthermophilic MBPs, and so may be linked to thermo- rather than acido-stability; an increased number of prolines at the second position of β-turns is a recurring feature in the (hyper)thermophilic proteins.
*Acy*MBP and the hyperthermophilic MBPs have fewer and smaller cavities compared to their mesophilic counterpart, thus suggesting that more thermostable MBPs are generally more compact (Table 3.5). Other factors, such as the number of insertions and deletions in the sequences/structures, do not appear to be correlated with differences in stability (Figure 3.5). *Acy*MBP is a monomer (confirmed here by dynamic light scattering), and so improved interactions within a multimer (Russell and Taylor, 1995) are not a viable route to its stabilization. The results presented here provide a number of testable hypotheses that can be applied in further work using existing, as well as new, structures and sequences.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SCHN 274/6-3;6-4), the Fonds der Chemischen Industrie (to E.S.), grants from the Swedish Research Foundation (VR) to S.L.M. as well as by Uppsala University and the Swedish University of Agricultural Sciences. The authors would like to thank Rebecca Fleischer for constructing pRF1, C. Schulze-Briese and T. Tomizaki for support with data collection at SLS, and Gerard Kleywegt for helpful discussions.

4 ProX from Escherichia coli

4.1 Abstract

Compatible solutes such as glycine betaine and proline betaine are accumulated to exceedingly high intracellular levels by many organisms in response to high osmolarity to offset the loss of cell water. They are excluded from the immediate hydration shell of proteins and thereby stabilize their native structure. Despite of their exclusion from protein surfaces, the periplasmic ligand-binding protein ProX from the *Escherichia coli* ABC transport system ProU binds the compatible solutes glycine betaine and proline betaine with high affinity and specificity. In order to understand the mechanism of compatible solute binding, we determined the high resolution structure of ProX in complex with its ligands glycine betaine and proline betaine. This crystallographic study revealed that cation-π interactions between the positive charge of the quaternary amine of the ligands and three tryptophan residues forming a rectangular aromatic box are the key determinants of the high affinity binding of compatible solutes by ProX. The structural analysis was combined with site-directed mutagenesis of the ligand binding pocket to estimate the contributions of the tryptophan residues involved in binding.

4.2 Introduction

The water content of bacterial cells is determined solely by osmotic processes as bacteria lack systems for active water transport into or out of the cell in response to an increase or a decrease of osmolarity in the environment (Kempf and Bremer, 1998; Booth and Louis, 1999). To cope with hyperosmotic conditions, non-halophilic microorganisms generally amass large quantities of a particular group of organic osmolytes, the so-called compatible solutes (Brown, 1976), either by *de novo* synthesis or by direct uptake from the environment (Kempf and Bremer, 1998; Csonka and Hanson, 1991; Galinski and Trüper, 1994; Bremer and Krämer, 2000; Welsh, 2000). Compatible solutes are operationally defined as organic osmolytes that can be accumulated by the cell to exceedingly high concentrations without disturbing vital cellular functions or the correct folding of

proteins (Brown, 1976). Their intracellular accumulation counteracts the osmotic efflux of water from the cell and thus contributes to the restoration of turgor (Wood, 1999) and the resumption of growth of the bacterial cells under conditions of low water activity. In addition to contributing to the maintenance of cellular water content at high external osmolarities (Record Jr. et al., 1998), compatible solutes also counteract destabilization of native structures of proteins upon freezing, heating, desiccation and exposure to high ionic strength both, *in vitro* and *in vivo* (Lippert and Galinski, 1992; Caldas et al., 1999; Canovas et al., 1999; Bourot et al., 2000; Canovas et al., 2001; Mendum and Smith, 2002).

The intracellular accumulation of compatible solutes as a strategy for adaptation to high osmolarity has been widely adopted not only by bacteria (Galinski and Trüper, 1994) and archaea (da Costa et al., 1998; Roberts, 2000) but also by fungal, plant, animal, and even human cells (Burg et al., 1997; Rhodes and Hanson, 1993; Hohmann, 2002). Moreover, a few classes of compounds are used universally throughout the kingdoms, reflecting fundamental constraints on the type of solutes that are compatible with macromolecular and cellular functions (da Costa et al., 1998; le Rudulier et al., 1984; Yancey, 1994). One of the most ubiquitous compatible solutes used both by prokaryotic and eukaryotic cells, is the trimethyl ammonium compound glycine betaine (N,N,N-trimethyl glycine, GB). The exact biochemical mechanism(s) through which compatible solutes act as protein stabilizers is not completely understood. Their functioning is generally explained in terms of the preferential exclusion model (Arakawa and Timasheff, 1985). This hypothesis predicts that compatible solutes are excluded from the immediate hydration shell of proteins, which is apparently due to unfavorable interactions with protein surfaces (Qu et al., 1998; Bolen and Baskakov, 2001). This causes a non-homogeneous distribution of compatible solutes within the cell water and a preferential hydration of protein surfaces. Experimental verification came from neutron diffraction studies (Lehmann and Zaccai, 1984). The non-uniform distribution of compatible solutes within the cell water results in a thermodynamic force that drives proteins to occupy a smaller volume in order to reduce the amount of excluded water, thereby stabilizing the native structure of proteins. Many Gram-negative and Gram-positive bacterial species that have been found to accumulate GB are able to acquire this compatible solute from environmental sources through high-affinity transport systems (Kempf and Bremer, 1998; Bremer and Krämer, 2000; Csonka and Epstein, 1996; Sleator and Hill, 2002). GB transport in *Escherichia coli* is under osmotic control (Csonka and Epstein, 1996; Perroud and le Rudulier, 1985; Lucht and Bremer, 1994) and is mediated by two transport systems: the single-component H^+-compatible solute cotransporter ProP (Culham et al., 1993) and the multi-component transport system ProU (May et al., 1986; Gowrishankar, 1989). The ProU system

is a member of the ATP-binding cassette (ABC) super family (Higgins, 1992; Boos and Lucht, 1996) consisting of two cytoplasmic, membrane-associated ATPases ProV, two integral membrane proteins ProW and a periplasmic ligand-binding protein ProX (Gowrishankar, 1989). Access of GB present in the environment to the periplasmic ProX protein is provided by passive diffusion across the outer membrane through the OmpC and OmpF general porins of *E. coli* (Faatz et al., 1988). ProX binds GB avidly with a K_D of approximately 1 μM (May et al., 1986; Barron et al., 1987; Haardt et al., 1995) delivering the ligand to the substrate translocation complex that is embedded in the cytoplasmic membrane. Driven by ATP hydrolysis, the translocation complex then transports the ligand into the cytoplasm (Higgins, 1992; Boos and Lucht, 1996; Fetsch and Davidson, 2002). In addition to GB, the ProU transporter also serves as a high-affinity uptake system for proline betaine (N,N-dimethyl-L-proline, PB), and this compatible solute is recognized by the ProX protein with high affinity (K_D = 5 μM) as well (Haardt et al., 1995). ProU also functions as an uptake system for a variety of other compatible solutes (*e.g.* proline and ectoine) but in contrast to GB and PB none of these compounds appear to be recognized by the ProX protein (Barron et al., 1987; Haardt et al., 1995; Jebbar et al., 1992).

The substrate-binding protein binds its ligand(s) selectively and with high affinity, which is thought to ensure the substrate specificity and directionality of the overall transport reaction for a given binding-protein dependent transport system (Boos and Lucht, 1996). Periplasmic binding proteins are generally composed of two domains connected by one to three polypeptide chains forming a hinge between them. In the ligand-free open conformation the two rigid domains are flexibly linked by the hinge. This has been shown by the structural analysis of various open states of the ribose-binding-protein and D-allose-binding-protein (Magnusson et al., 2002; Björkman and Mowbray, 1998). Ligand binding induces a large conformational change in the hinge region which moves both domains toward each other. After this domain movement the ligand is engulfed in a predefined cleft between the two domains which refers to the closed conformation of the binding protein.

Compatible solutes are apparently preferentially excluded from the immediate hydration shell of proteins (Arakawa and Timasheff, 1985; Qu et al., 1998; Bolen and Baskakov, 2001). Yet a ligand-binding protein such as ProX binds the compatible solutes GB and PB with high affinity and high specificity (May et al., 1986; Barron et al., 1987; Haardt et al., 1995; Higgins et al., 1987). This observation raises the question as to how such a high-affinity interaction between a compatible solute and a protein can take place. In order to understand the molecular determinants that govern substrate recognition and binding of compatible solutes by ProX, we determined the high-resolution crystal

structure of the ProX protein in complex with each of its ligands, *i.e.* ProX-GB and ProX-PB. This crystallographic analysis was combined with site-directed mutagenesis of the residues forming the ligand-binding pocket of ProX to determine their relative importance for high-affinity binding of the compatible solute GB by the ProX protein from *E. coli*. The analysis revealed cation-π interactions between the quaternary amine of GB and PB and the indole groups of three tryptophan residues.

4.3 Materials and Methods

4.3.1 Expression and purification of wild-type and mutant ProX

The wild-type *E. coli* ProX protein was overproduced and purified to homogeneity essentially as described previously (Breed et al., 2001). In brief, the *proX* gene was overexpressed under the control of the bacterial phage T7ϕ10 promoter present on the *proX*$^+$ low-copy-number plasmid pSK7 (Cm^r) in the *E. coli* strain PD141 (λDE3) (Breed et al., 2001). This strain harbors a chromosomal copy of the gene for the phage T7 RNA polymerase under the control of an IPTG-inducible *lacPO/lacI* promoter system, thereby allowing the selective expression of the *proX* gene present on plasmid pSK7 from the T7ϕ10 promoter upon the addition of IPTG to the cell culture. Cells of strain PD141 (λDE3) (pSK7) were grown in a minimal medium (Breed et al., 2001) in the presence of 30 g ml^{-1} chloramphenicol because components of rich media (*e.g.* yeast extract) are known to contain GB (Dulaney et al., 1968) that could potentially be scavenged and bound by ProX. In this way we ensured that the purified ProX protein was free of ligand since the *E. coli* strains used for overproduction are unable to synthesize GB. Overproduction of ProX was initiated by the addition of 1 mM IPTG when the cell culture had reached an OD_{578} of approximately 1.0 - 1.5. After 1 h of further growth, the cells were harvested by centrifugation and the periplasmic proteins were released from the cells by cold osmotic shock (Neu and Heppel, 1965). Insoluble material was removed from the released periplasmic proteins by ultracentifugation and the supernatant was then subjected to FPLC chromatography on a DEAE-Sepharose Fast Flow column (Amersham Bioscience, Germany). Proteins were eluted from this anion exchange column with an increasing Tris-HCl (pH 8.3) gradient (16 - 400 mM). ProX-containing fractions were combined, and ammonium sulfate was added to a final concentration of 1.5 M. This protein solution was then subjected to hydrophobic interaction chromatography on a Phenyl-Sepharose column (Amersham Bioscience, Germany). The ProX protein was eluted with a decreasing ammonium sulfate gradient (1.5 - 0 M ammonium sulfate dissolved in 10 mM Tris-HCl, pH 8.3). The purified ProX protein was then dialyzed

overnight against 5 L of a 10 mM Tris-HCl (pH 7.3) buffer. The ProX protein was free from other contaminating proteins, as judged from SDS-polyacrylamide gels stained with Coomassie brilliant blue.
ProX proteins with mutations in the substrate-binding site were purified essentially as wild-type ProX protein except that the *E. coli* strain LinE2 (PD141 Δ(*proU::spc*)608) was used for the overexpression of the mutant *proX* genes present on derivatives of plasmid pSK7. Strain LinE2 carries a chromosomal deletion of the entire *proU* operon (Haardt et al., 1995), thereby avoiding the contamination of the mutant ProX protein preparations with the wild-type ProX protein during protein purification.

4.3.2 Genetic construction of bacterial strains

To construct an *E. coli* strain that lacked the *E. coli proU* operon and would allow phage T7ϕ10-mediated overexpression of plasmid-encoded mutant *proX* genes, we prepared a P1*vir* lysate (Silhavy et al., 1984) on the *proU-lacZ* fusion strain BK16 (MC4100 ϕ(*proU-lacZ*)hyb2 λp*lac*Mu15 (Kan^r)) (Kempf and Bremer, unpublished results). Using this phage lysate, we transduced the *lacZ* gene fusion positioned next to the λp*lac*Mu15 (Kan^r) prophage (Bremer et al., 1988) into the *E. coli* strain PD141 (MC4100 (λDE3)) (Dersch and Bremer, unpublished results) by selecting for kanamycin resistant colonies (50 μg ml^{-1}) in the presence of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (300 g ml^{-1}) on Luria Bertani (LB) agar plates. One of these transductants was strain LinE1. We then replaced the ϕ(*proU-lacZ*)hyb2 λp*lac*Mu15 (Kan^r) prophage in strain LinE1 with the Δ(*proU::spc*)608 deletion from strain MKH13 (Haardt et al., 1995) by transducing strain LinE1 with a P1*vir* lysate prepared on strain MKH13 and selecting for spectinomycin-resistant colonies (100 μg ml^{-1}) in the presence of X-Gal on LB agar plates. The desired transductants carrying the Δ(*proU::spc*)608 deletion were identified as spectinomycin-resistant colonies that were kanamycin sensitive and LacZ$^-$. One of these transductants was strain LinE2. The absence of the ProX protein from strain LinE2 was verified by Western blotting analysis using a ProX antiserum and whole cell extracts prepared from cells grown in the presence of 250 mM sodium chloride in minimal medium.

4.3.3 Site-directed mutagenesis of the proX gene

To probe the contributions of residues Trp65, Trp140 and Trp188 in ProX to GB binding, we mutated the corresponding codons in the *proX* gene via site-directed mutagenesis using the Quick-change site-directed mutagenesis kit (Stratagene) and custom synthesized primers (MWG-Biotech) containing the desired mutations. We replaced the codon for

each of these Trp residues in *proX* with codons encoding the amino acids Ala, Leu, Phe, Tyr, Asp or Glu in plasmid pSK7 by following the protocol provided by Stratagene. The entire coding region of the mutant *proX* genes were then sequenced to ensure the presence of the desired mutation and the absence of unwanted alterations in the *proX* coding region. DNA sequencing was carried out by the chain terminating method (Sanger et al., 1977) with the Thermo Sequenase fluorescent-labeled primer cycle sequencing kit (Amersham Biosciences). The DNA sequencing reactions were primed with synthetic oligonucleotides labeled at their 5-prime end with the infrared dye IRD-800 (MWG-Biotech), and the products were analyzed using a LI-COR DNA Sequencer (model 4000; MWG Biotech). Each of the mutant *proX* alleles chosen for overproduction of the altered ProX proteins contained only the desired mutations in the codons encoding residues Trp65, Trp140 or Trp188 (pLB2 188Ala; pLB3 188Leu; pLB4 188Phe; pLB5 188Tyr; pLB6 140Ala; pLB7 140Leu; pLB8 140Phe; pLB9 140Tyr; pLB10 65Ala; pLB11 65Leu; pLB12 65Phe; pLB13 65Tyr; pLB14 188Asp; pLB15 188Glu; pLB16 140Asp; pLB17 140Glu; pLB18 65Asp; pLB19 65Glu). A double mutant containing replacements of the amino acids at positions Trp65 and Trp140 were constructed by re-mutating plasmid pLB10 to yield plasmid pLB20 (65Ala, 140Ala). The mutant *proX* pLB plasmids were each introduced by electroporation into the *E. coli* strain LinE2 (Δ(*proU::spc*)608), and 4 L of minimal medium with 0.4 %(w/v) glucose as the carbon source were used for the overproduction of the mutant ProX proteins. Usually, we obtained between 10 - 20 mg of purified ProX protein from 4 L of cell culture.

4.3.4 Binding of GB to ProX and its mutant derivatives

The binding affinity of the wild-type ProX protein and its mutant derivatives constructed by site-directed mutagenesis of *proX* was measured by using an ammonium sulfate precipitation assay (Richarme and Kepes, 1983). Purified ProX protein (5 μM, determined using the BCA assay) was incubated with 9 different concentrations of radio-labelled GB (1, 2, 5, 7, 10, 20, 30, 50, and 110 μM) in a 100 μl reaction volume of 10 mM Tris-HCl pH 7.3 for 5 min at room temperature. Then the ProX protein was precipitated with a 900 μl solution of ice-cold saturated ammonium sulfate solution. After an incubation of 10 min on ice, the precipitated ProX protein was collected by filtration onto a nitrocellulose filter (pore size 0.45 μm; Schleicher & Schuell), the filter was then washed with 10 ml of an ice-cold ammonium sulfate solution and the radioactivity retained by ProX on the filter was determined by scintillation counting. Each measurement was repeated three times for each concentration in order to determine the binding constant and their standard deviation. A mutant is treated as non-binding if no binding occurs

Table 4.1: Intensity statistics of the heavy atom derivatives and native data sets

	Derivative		**Native**	
Data set	$(CH_3)_3PbAc$	CH_3HgCl	glycine betaine	proline betaine
Ligand	glycine betaine	glycine betaine	glycine betaine	proline betaine
Wavelength [Å]	1.54179	1.54179	0.8439	1.54179
Total rotation range [°]	100	100	90	100
Temperature [K]	291	291	100	291
Cell (a, b, c) [Å]	48.0, 55.0, 115.8	47.9, 55.1, 115.7	47.59, 53.86, 115.43	47.71, 54.98, 115.73
Resolution limits [Å]	20.0-2.48 (2.55-2.48)	20.0-2.48 (2.60-2.48)	15-1.57 (1.67-1.57)	15-2.05 (2.25-2.05)
No. of reflections	43801	43602	142884	75377
No. of unique reflections	20031	20826	40866	19587
Completeness [%]	95.4 (95.1)	99.2 (99.5)	96.6 (95.2)	99.0 (99.4)
$I/\sigma I$	15.7 (6.3)	19.0 (10.0)	30.0 (18.6)	15.1 (5.65)
R_{sym}	4.1 (12.0)	2.8 (6.7)	2.6 (4.7)	6.0 (23.9)
R_{meas}*	5.3 (15.5)	3.6 (8.8)	3.0 (5.5)	6.6 (27.5)
$R_{mrgd\text{-}F}$*	6.4 (16.6)	3.6 (7.9)	2.2 (4.8)	7.7 (23.3)
B-factor Wilson plot [$Å^2$]	27.8	32.4	7.7	33.3
Phasing power	1.39	1.33		
Figure of merit overall	0.62			

* R_{meas} and $R_{mrgd\text{-}F}$ are better indicators for data quality than R_{sym}, R_{meas} is redundancy independent and $R_{mrgd\text{-}F}$ is a measure for the quality of the reduced amplitudes (Diederichs and Karplus, 1997; Weiss and Hilgenfeld, 1997). (values in parentheses refer to the highest resolution shell)

at concentrations of 110 μM GB, neither specific nor unspecific (Table 4.3). According to the properties of the compatible solute GB presented in the introduction, we do not expect an unspecific binding of GB to ProX at this rather low concentrations.

4.3.5 Data collection und structure refinement

Crystals of ProX-GB and ProX-PB were grown by vapor-phase equilibration at 18 °C as described (Breed et al., 2001). In brief, protein (10 mg/ml in 10 mM Tris pH 8.3) was diluted 1:2 with a reservoir solution containing 26 - 28 %(w/v) PEG4000 and 50 mM PIPES pH 6.2 - 6.4. The crystals grew within 4 - 6 weeks and belong to the space group $P2_12_12_1$ with unit cell parameters of a = 48.0, b = 55.0, and c = 115.7 Å containing one monomer per asymmetric unit. In complex with GB ProX crystals were quite resistant to damage by heavy atoms, both high concentrations and long soak times were possible. Two derivatives ProX-GB were obtained by soaking of the crystals either with 10 mM methylmercuric(II)chloride (CH_3HgCl) for 48 h or with 5 mM trimethyllead acetate ($(CH_3)_3PbAc$) for 72 h at room temperature. Data sets of the derivative crystals and a native crystal of ProX-PB were collected at room temperature using a rotating anode generator (Schneider, Offenburg, Germany) equipped with a Mar345 imaging plate detector (MarResearch, Hamburg, Germany). A native data set of ProX-GB was collected at the EMBL outstation DESY Hamburg beamline BW7B under cryogenic conditions. For this purpose the crystals were soaked in reservoir solution with stepwise increasing glycerol concentrations up to a final concentration of 25 %(v/v) glycerol and then transferred to liquid nitrogen. All data sets were processed using XDS (Kabsch, 1993) as summarized in Table 4.1. Heavy atom positions and initial phases were calculated using SOLVE (Terwilliger and Berendzen, 1999) followed by solvent flattening using DM (Cowtan, 1994). Chain tracing and model building has been done with the graphical interface 'O' (Jones et al., 1991) followed by alternating cycles of model building and refinement using REFMAC5 (Murshudov et al., 1997). The quality of the obtained models was validated with the program PROCHECK (Laskowski et al., 1993). Figures with molecule presentations were prepared with the programs MolScript (Kraulis, 1991), BobScript (Esnouf, 1997) and Raster3D (Merrit and Bacon, 1997).

Sequence analysis

DNA and protein sequences were assembled and analyzed with the Lasergene program (DNASTAR, Ltd.) on an Apple Macintosh computer. Searches for protein homologues to the *E. coli* ProX protein were performed at the National Center for Biotechnology Information (NCBI) by using the BLAST programs with standard values (Altschul et al.,

Table 4.2: Refinement statistics of the native structures

	glycine betaine	proline betaine
Resolution limits [Å]	12.26-1.59	15.0-2.05
Total No. of reflections	39378	19430
Reflections in working set	37409	18626
reflections in test set	1969	804
R [%]	15.0	17.2
R_{free} [%]	18.6	21.3
No. of amino acid residues	309	309
No. of protein atoms	2390*	2382
No. of metal ions	1	1
No. of water molecules	429	59
No. of ligand atoms	8	10
B-factor for all atoms [Å^2]	13.5	29.0
B-factor of protein atoms [Å^2]	11.7	29.0
R.m.s.d. bonds [Å]	0.013	0.010
R.m.s.d. angles [°]	1.385	1.239

* - side chains of N90 and M169 are modelled in two alternate conformations

1990). Protein sequences were aligned with the CLUSTAL algorithm provided with the Lasergene program.

Chemicals

GB and PB were purchased from Sigma Chemie and Extrasynthése, respectively. Radiolabelled GB ([1-^{14}C]-glycine betaine; 55 Ci mmol^{-1}) was obtained from ARC Inc., St Louis, MO.

PDB accession codes

The coordinate data sets of the ProX-GB and ProX-PB structures are available in the PDB with accession codes 1R9L and 1R9Q, respectively.

4.4 Results

4.4.1 Overall structure

The structure of ProX from *E. coli* has been solved in the closed conformation in complex with each of the two ligands glycine betaine (GB) and proline betaine (PB). Initial phases for ProX-GB were obtained by multiple isomorphous replacement using two heavy atom derivatives CH_3HgCl and $(CH_3)_3PbAc$ (see Table 4.1). After solvent flattening the maps were sufficiently clear to trace the polypeptide. The initial model was refined against a native ProX-GB data set to 1.6 Å resolution. Clear difference electron density could

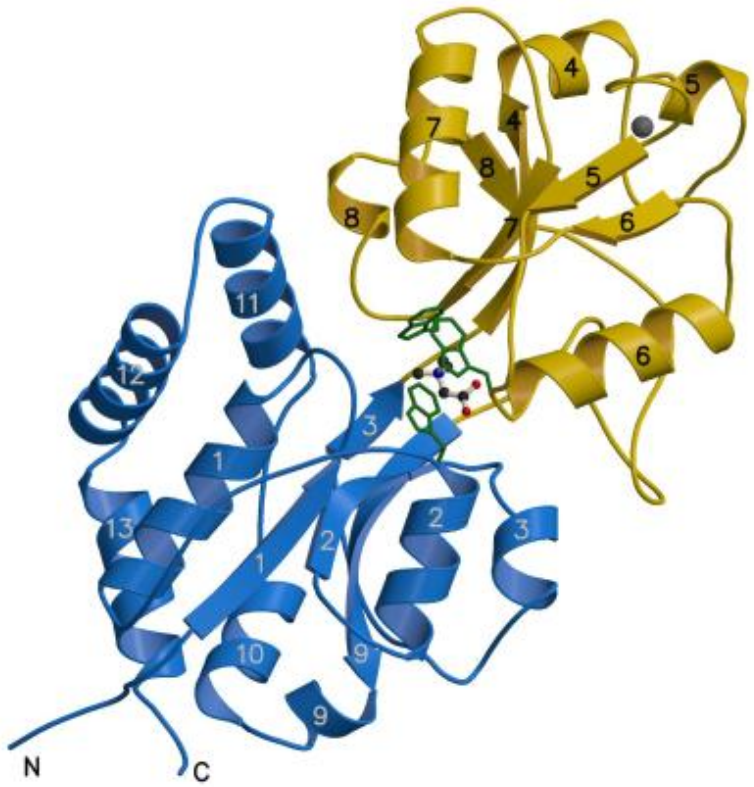

Figure 4.1: Ribbon diagram of ProX Domain A (residues 1 - 92, 234 - 309) and domain B (residues 93 - 233) are colored blue and yellow, respectively. The two switch segments 90 - 94 and 231 - 235 connect both domains. Helices and strands are consecutively numbered and the N- and C-termini are labelled at the bottom. The ligand GB is drawn as ball-and-stick and is enclosed by the three Trp residues (green) involved in ligand binding.

be seen for the bound ligand. The final model contains all 309 residues of the mature protein, a bound GB and a metal ion (see Table 4.2).

The corresponding structure of ProX-PB was solved by molecular replacement with the former structure (see Table 4.2). ProX-GB and ProX-PB are ellipsoidal with approximate dimensions of 75 x 38 x 20 $Å^3$. The structure can be subdivided into two globular domains: domain A (blue) from 1-92 + 234-309 and B (yellow) from 93-233 connected by the two switch segments 90-94 and 231-235 forming the hinge region (Figure 4.1). Each domain consists of a four-stranded β-sheet flanked by α-helices on both sides as usually found among the members of the binding protein family.

In general binding proteins are subdivided into two structural groups with the different domain folds I and II which correlate with the number of switches of the polypeptide between the two domains (Quiocho and Ledvina, 1996). Group one shows three switches and group II two switches. An additional group is represented by the structures of

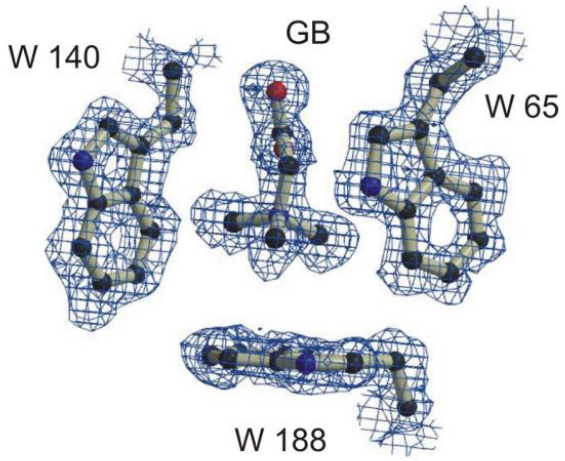

Figure 4.2: Electron density of the tryptophan box Binding pocket of ProX-GB with 2Fo-Fc map contoured at 1 σ after refinement to 1.6 Å resolution. The tryptophan residues form a rectangular box with Trp188 as bottom and Trp65 and Trp140 as sides.

FhuD and BtuF from *E.coli*, periplasmic binding proteins specific for ferrichrome and vitamin B_{12}, respectively (Clarke et al., 2000; Borths et al., 2002). These proteins have only one switch between their two domains. According to this classification ProX has a periplasmic binding-protein type II fold.

4.4.2 The ligand binding site

As expected from other binding protein structures, the ligand binding site is located in the cleft between the two globular domains (Figure 4.1). The two adjacent switch segments in the hinge region allow the formation of a deep cleft which entirely buries the betaine ligand and thus prevents it from contacting the solvent.
Around the quaternary amine of the betaine ligand, the binding pocket is formed by the indole groups of the three tryptophans Trp65, Trp140 and Trp188. Figure 4.2 shows the electron density for the GB ligand surrounded by the tryptophan residues refined at 1.6 Å resolution. These planar indole groups are arranged like three faces of a rectangular box with the planes of Trp65 and Trp140 being almost parallel to each other, while that of Trp188 is perpendicular to these. Trp65 at the C-terminus of strand 2 is the only one of the tryptophans provided by domain A and is closest to the switch segments (Figure 4.3). Domain B provides the other two of the tryptophan residues Trp140 and Trp188. Trp188 is placed at the C-terminal end of strand 7, and Trp140 is furthest from the switch segments and located in the loop between strand 5 and helix 6 which apparently is stabilized by the disulfide bond between Cys136 and Cys142. This disulfide bond between Cys136

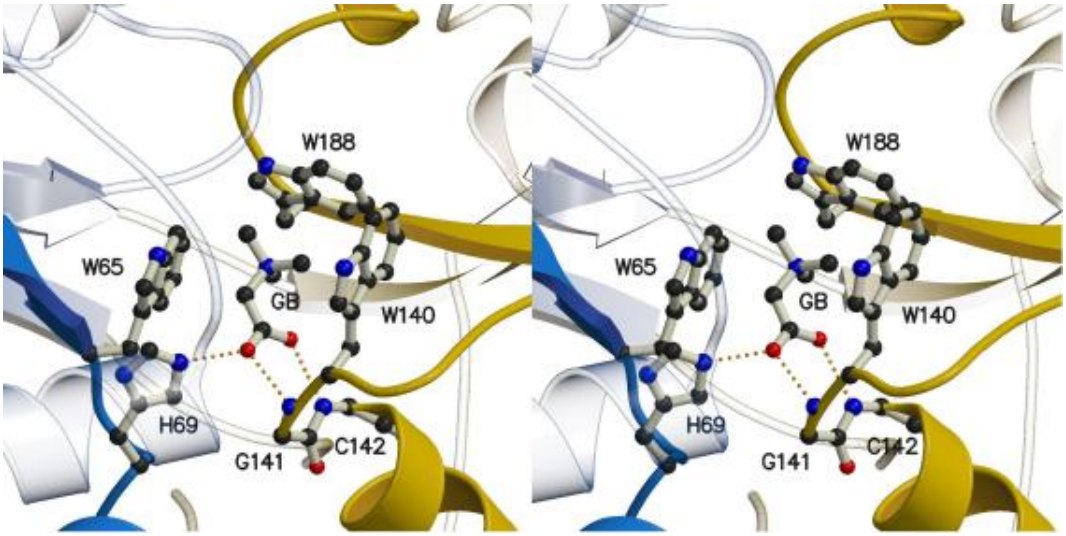

Figure 4.3: Stereo picture of the binding pocket Domains are colored blue and yellow as in Figure 1. The ligand GB and residues involved in binding are drawn as ball-and-stick and labelled with residue name and number. Hydrogen bonds are represented as dashed lines whereas the cation-π interaction is not explicitly depicted. The side chain of Cys142 is not drawn to improve clarity.

and Cys142 is close to the ligand binding site. Whereas other binding protein structures such as putrescine-binding-protein, dipeptide-binding-protein, leucine-binding-protein, and leucine/isoleucine/valine-binding-protein with PDB accession codes 1A99, 1DPE, 2LPB, and 2LIV, respectively (Berman et al., 2000), also contain disulfide bonds, all of these are placed distantly from the ligand binding site. The carboxylic group of the GB ligand protrudes out of the indole box forming hydrogen bonds with the backbone amide hydrogens of Gly141 and Cys142 of domain B. These two residues are part of the loop between the disulfide-linked residues Cys136 and Cys142 which also contains the indole ring of Trp140. Additionally, the carboxylic group forms a hydrogen bond with the imidazole ring of His69 from domain A (Figure 4.3). Residues involved in the binding of PB are essentially the same as can be seen in Figure 4.4 showing a superposition of the ProX-PB structure (ligand yellow) onto the ProX-GB structure (ligand blue). The two methyl groups C1 and C2 as well as the Cδ (from the proline ring) of the quaternary amine form similar contacts with the three indole rings and the carboxylic group forms identical hydrogen bonds as in ProX-GB. Also the PB ring atoms Cβ and Cγ contribute additional van der Waals interactions with the indole ring of Trp65 and Cγ with Leu68 side chain as well.

Mainly, there are two contributions involved in the binding of a quaternary ammonium group in an aromatic environment, cation-π and van der Waals interactions. To understand the nature of the interaction between the tryptophans and the ligand, all distances

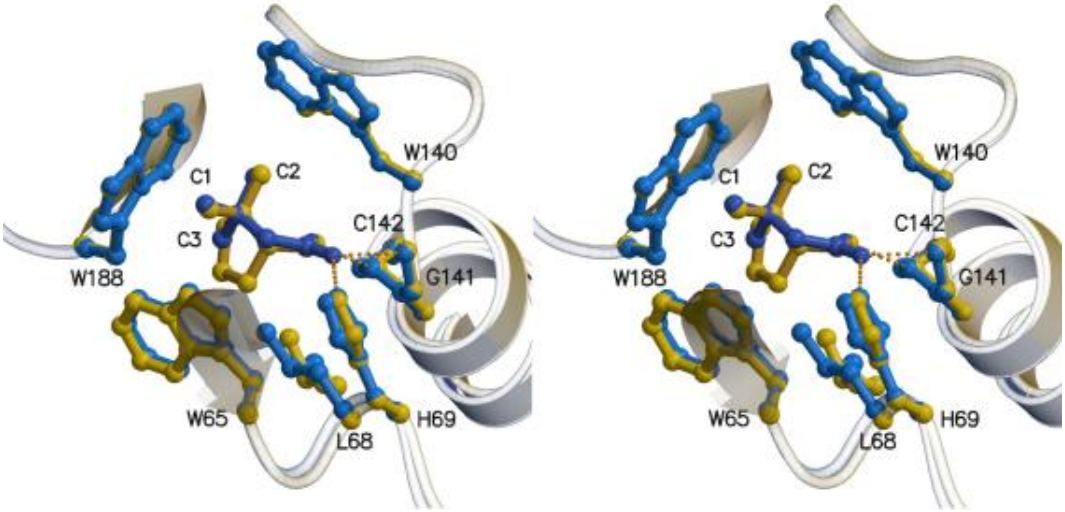

Figure 4.4: Superposition of ProX-PB onto ProX-GB Stereo picture of the superposition of the ProX-PB structure (yellow) onto the ProX-GB structure(blue). The Cα positions of the Trp65, Trp140 and Trp188 were used to superimpose the ligand binding sites. Secondary structure elements are shown for side chains involved in ligand binding, moreover the point of view has been changed to improve clarity. Additionally, Leu68 is shown which is suitably spaced to form van der Waals contacts with the Cγ atom of the ring structure of proline. The methyl groups of the ligand are labelled according to the numbers of the ProX-GB structure in the pdb file (C3 $\equiv$ Cδ in ProX-PB).

between the carbon atoms bonded to the quaternary nitrogen and the ring atoms of the three indole groups were determined (Figure 4.5). These distances were compared with a compiled list of van der Waals radii published by Li and Nussinov (1998). If the influence of the quaternary ammonium charge on the contact distances is neglected, the methyl- or methylene groups and the aromatic ring atoms possess van der Waals radii of 1.92 Å and 1.82 Å, respectively. This results in a mean distance of 3.74 Å between those carbon atoms having a standard deviation of approximately 0.5 Å (Li and Nussinov, 1998). We therefore consider an indole ring atom to be in contact with a methyl or a methylene group if their mutual distance is between 3.5 Å and 4 Å. The indole ring atoms which fullfill this criterion are color coded in Figure 4.5. According to this criterion, GB forms nine, five, and four contacts with Trp188, Trp65, and Trp140 indole groups, respectively. The number of contacts of the indole rings suggests that the three indole groups contribute differently to the enthalpy of GB binding. Trp188 is contacted by each of the three methyl groups because the Cα-N bond is oriented perpendicular to the aromatic bottom of the box (Figure 4.3). Trp65 is contacted by one methyl group C1 and the Cα and Trp140 is contacted only by the methyl group C2.
In case of PB, seven, two, and five contacts are found of the quaternary amine with the

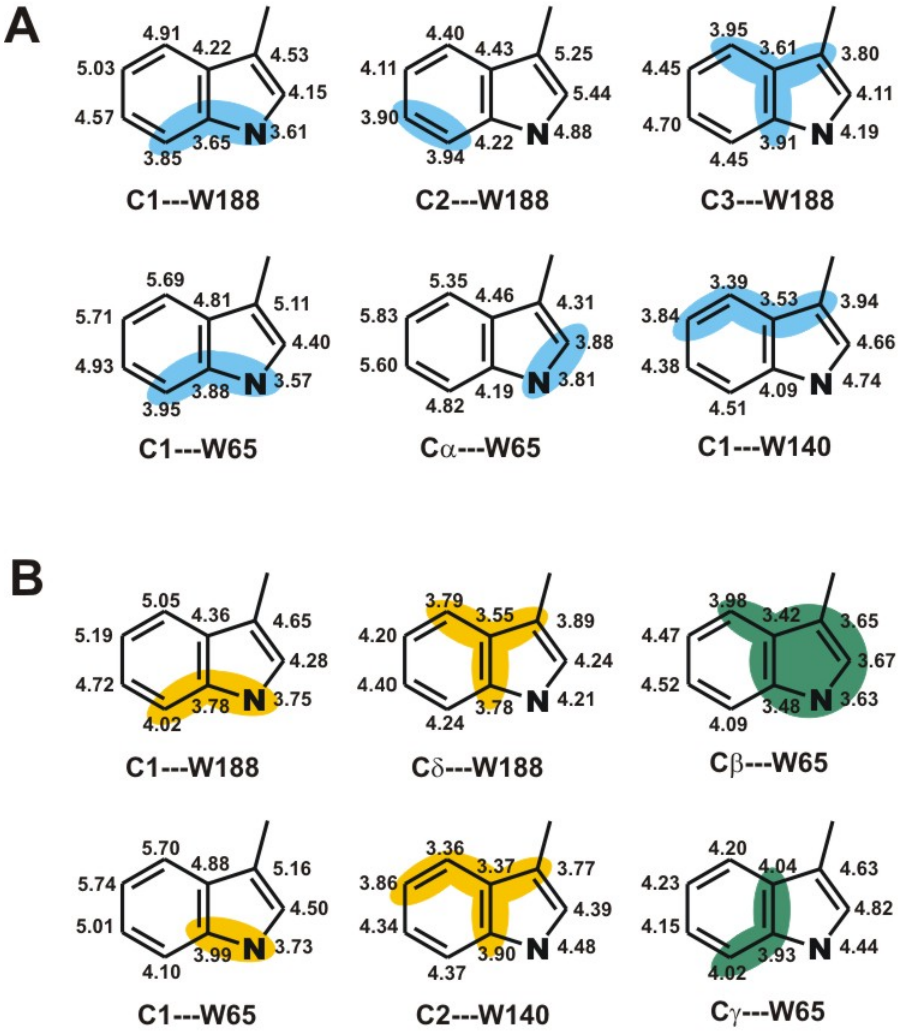

Figure 4.5: Distances between the ligands GB (A) and PB (B) and the indole groups All carbon atoms of the indole rings involved in cation-π interaction (matching the criterion to be within 3.5-4.0 Å distance) are in cyan for GB and yellow for PB (color coding for the ligand is the same as in Figure 4.4). The nomenclature for PB is similar as for GB only C3 in GB is as Cδ part of the proline ring in PB. The quaternary amine of GB forms nine, five, and four contacts whereas the one PB forms seven, two, five contacts with the indole groups of Trp188, Trp65 and Trp140, respectively. PB also forms nine van der Waals contacts between Cβ and Cγ with the indole ring of Trp65 shown in green.

Table 4.3: Binding affinities of GB to wildtype and mutated ProX

	Trp 65		Trp 140		Trp 188	
	K_D [μM]	SD	K_D [μM]	SD	K_D [μM]	SD
wildtype	4	0.1	4	0.1	4	0.1
Trp → Phe	4	0.1	3	0.3	4	0.4
Trp → Tyr	13	0.6	4	0.5	5	0.2
Trp → Ala	50	3.3	56	10	–	–
Trp → Leu	–	–	14	1.4	–	–
Trp → Asp	–	–	8	0.8	–	–
Trp → Glu	–	–	21	2.5	–	–
Trp65 & Trp140 → Ala			–			

Non-binding ProX mutants are indicated by –, these proteins show no GB binding at the maximum concentration of 110 μM used in the experiments.

indole rings of Trp188, Trp65 and Trp140, respectively. One difference compared to the ProX-GB structure is that the Cα is no longer in contact distance with the indole ring of Trp65, having been pushed aside by the PB ring atoms Cβ and Cγ. These ring atoms form nine new van der Waals contacts with Trp65, colored green in Figure 4.5 because they are not part of the quaternary amine.

4.4.3 Metal binding site

A metal binding site distant from the ligand binding site has been identified in the ProX structure. The metal ion is shown in Figure 4.1 as a gray sphere in domain B between helix 5 and strand 5. Although the protein was never exposed to a buffer containing any metal cations during purification and crystallization an ion has been identified in the loop from residues 125 to 130. An anomalous difference electron density shows a strong peak of 14.5 σ at $\lambda = 0.8439$ Å. The octahedral coordination of the ion is provided by five ligands of the surface-exposed loop in domain B between helix 5 and strand 5 (Asp124, Asn126, Asp128, Lys130-CO, and Asp132) and completed by a water molecule as the sixth ligand. This arrangement of the binding site resembles an EF-hand motif. A similar situation has been found in the alginate binding protein where the metal ion has been identified to be calcium (Momma et al., 2002). Whereas the coordination of the metal is rather indicative for a calcium ion the anomalous signal at the measured wavelength is pointing to a transition metal. It was not possible to decide which metal is actually bound by ProX. The loop from residues 125 to 130 that is apparently stablized by the metal ion seems to have no function.

4.4.4 Differentiation of the indole groups by mutational studies

To estimate the contributions of each of the indole groups of the tryptophan box, site specific mutants were designed and their relative binding affinity for GB was compared with that of the wild type. The binding affinity for the wild type ProX measured by the ammonium sulfate precipitation is in very good agreement with a K_D value of 1.4 μM determined by a release assay (May et al., 1986). All three tryptophan residues were consecutively substituted by tyrosine, phenylalanine, alanine, leucine, aspartate or glutamate, as summarized in Table 4.3.
The data demonstrate that a phenyl ring can provide a similar contribution to the binding affinity as an indole ring (K_D = 3-5 μM). In contrast, non-aromatic replacements of these residues impair the binding affinity considerably. The strongest influence of non-aromatic mutations is at position 188 where no binding affinity is detectable in the Ala, Leu, Asp and Glu mutants. Position 65 is to a certain degree tolerant against Ala (K_D = 50 mM) but does not tolerate leucine and acidic residues whereas position 140 tolerates Ala, Leu, Asp, and Glu substitutions with dissociation constants being 2-10 times larger than for the wild type (Table 4.3). Thus, the binding affinity is most sensitive against substitutions at position 188. The weakest sensitivity is at position 140 confirming the relative importance of the single tryptophans derived from the number of contacts with the GB (Figure 4.5). Nevertheless, the two residues Trp65 and Trp140 are necessary for the binding as seen from the double mutant Trp65Ala and Trp140Ala (see Table 4.3).

4.4.5 A conserved sequence motif

4.6). A Blast search in the NCBI database identified many proteins which are homolgous to ProX. For clarity we selected a few representatives of the taxonomic classes. The selection is shown in Figure 4.6, the *E. coli* sequence is followed by sequences of three Enterobacteriacea (same family), one Vibrinonales (same subdivision), three Pseudomonales (same subdivision), one Rhodospirillales (same phyllum), three Rhizobiales (same phyllum) and one Cyanobacterium (same kingdom).
It can be seen that the tryptophan residues are well conserved in this group of binding proteins and in some cases they are replaced by tyrosine residues (see Figure 4.6). A restraint on the sequence is the proper orientation of the three tryptophan side chains achieved by the surrounding residues which can also be inferred from the alignment. In ProX the second residue after each of the tryptophans 65 and 188 is a conserved proline in cis-conformation which allows for a sharply bent backbone. Pro67 is placed between the two residues Trp65 and His69 which both interact with the ligand (see Figure 4.6). In the loop containing Trp140, two conserved glycine and cysteine residues are

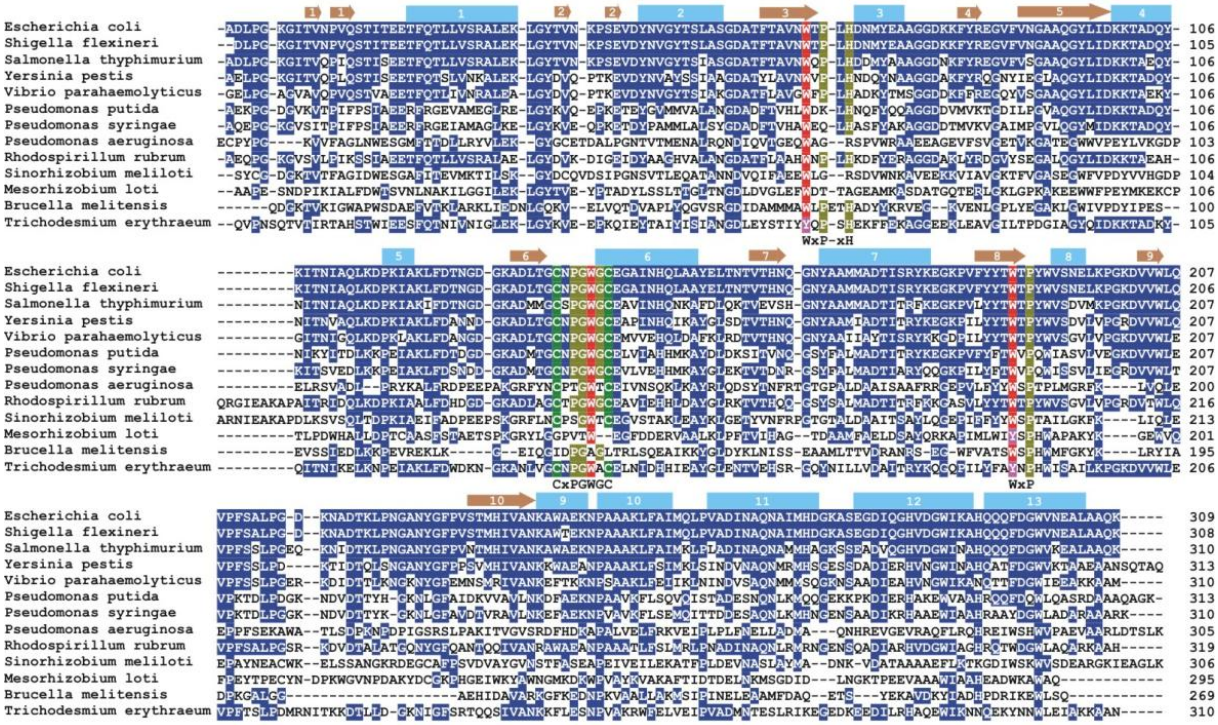

Figure 4.6: Alignment of ProX with selected sequences from a homology search Conserved residues are shown in blue, residues involved in binding of the quaternary amine head group are red (equivalent residues pink), conserved cysteine residues are green and structurally important residues are marked olive. Secondary structure elements for the ProX structure were determined using DSSP (Kabsch and Sander, 1983). Helices (light-blue) and sheets (brown) are numbered according to Figure 4.1. *E. coli* K12 (NCBI no. 16130593), *Shigella flexneri* 2a str. 301 (NCBI no. 24113983, 99% identity), *Salmonella typhimurium* LT2 (NCBI no. 16766122, 83% identity), *Yersinia pestis* (NCBI no. 22125121, 71% identity), *Vibrio parahaemolyticus* (NCBI no. 28898502, 66% identity), *Pseudomonas putida* (NCBI no. 7542433 51% identity) *Pseudomonas syringae pv tomato* (NCBI no. 28870233, 48% identity), *Pseudomonas aeruginosa* PA01 (NCBI no. 15600296 24% identity), *Rhodospirillum rubrum* (NCBI no. 22968752, 63% identity), *Sinorhizobium meliloti* (NCBI no. 8650111, 26% identity), *Mesorhizobium loti* (NCBI no. 13476104, 25% identity), *Brucella melitensis* (NCBI no. 17988895, 25% identity), *Trichodesmium erythraeum* (NCBI no. 23041742, 42% identity).

stabilized by a disulfide-link between the two cysteines Cys136 and Cys142. Again, the sharp bend of the loop may require Pro138 which is also conserved. This entire motif C_{136}xPGWGC_{142} is strictly conserved among the five closest homologues of ProX (see Figure 5.6.

4.4.6 Other cases of ligands bound by cation-π interaction

Several kinds of ligands with quaternary amines have been observed to form van der Waals contacts with proteins. In the phosphocholine binding antibody McPC603 (PDB accession code 2MCP) two methyl groups of the quaternary amine form contacts with the indole group of Trp107 and one methyl forms contacts with the phenyl ring of Tyr93. In the acetylcholine esterase complex with the decamethonium bromide inhibitor (Harel et al., 1993) and with another inhibitor BW284C51 (Felder et al., 2002), two quaternary ammonium groups at the ends of the ligands form van der Waals contacts with indole groups. Two methyl groups contact the indole of Trp84 and one methyl group contacts the indole of Trp279, at the bottom and the top of the 'gorge' (Ripoli et al., 1993), respectively. The former of these contacts is believed to resemble that of the quaternary ammonium group of acetylcholine upon catalysis.
In order to understand if there is a general architecture of these quaternary amine binding sites in proteins we superimposed ProX-GB with two other structures having completely different folds and substrates. For the comparison we used the structures of the phosphatidylcholine binding protein (Roderick et al., 2002) and the HP1 chromodomain (Jacobs and Khorasanizadeh, 2002) with PDB accession codes 1LN1 and 1KNE, respectively. Because of their different folds only the quaternary amine head group of their ligands was superimposed using LSQMAN (Kleywegt and Jones, 1994a). Figure 4.7 shows those amino acid side chains which contain at least one atom closer than 4.0 Å to the bound quaternary amine. As can be seen aromatic residues equivalent to Trp65 and Trp188 are also present in the two other proteins whereas the third residue of the aromatic box is less conserved.

4.5 Discussion

4.5.1 Binding a compatible solute to a protein

Compatible solutes are of practical interest as they can stabilize labile proteins in a functionally active form *in vitro* over extended periods or act as cryoprotectants. The protective value has been traced to their exclusion from direct contact with the protein surface. As the surface increases in the initial conformational transitions associated

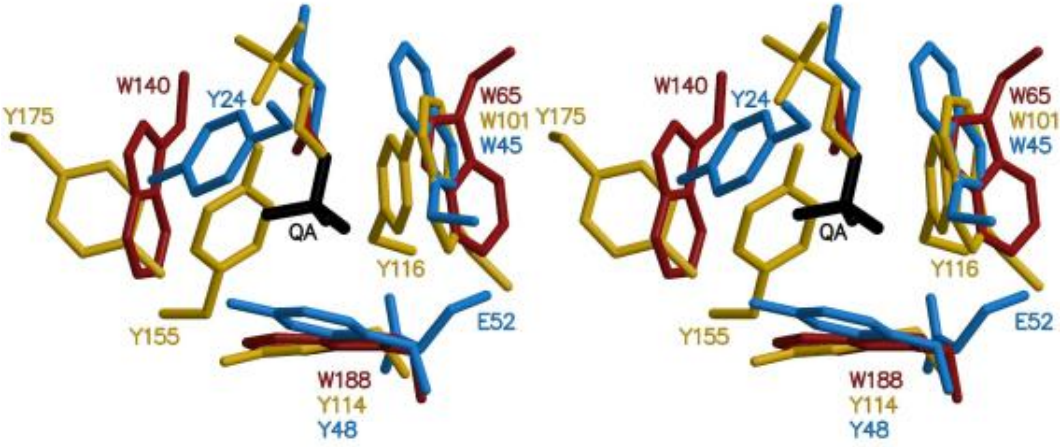

Figure 4.7: Superposition of three quaternary amine binding proteins The quaternary amine in the centre of the binding site (black) was superimposed using LSQMAN (Kleywegt and Jones, 1994a). Each of the protein residue side chains and the quaternary amine elongations (or parts of them, in parentheses) are color-coded: red ProX (GB), yellow phosphatidylcholine transfer protein (phosphocholine), blue chromodomain HP1 (three-methylated lysine). All amino acid side chains of the proteins which have at least one atom closer than 4 Å to quaternary amine are shown in the figure.

with denaturation, denaturation is energetically disfavored in presence of a compatible solute (Arakawa and Timasheff, 1985; Lee and Timasheff, 1981; Timasheff, 1993; Kendrick et al., 1997). This work addresses the question as to how such a substance known to avoid interaction with protein surfaces can nevertheless be complexed with high affinity by a protein and what are the critical structural features of such a binding site. Compared to other small ions or dipolar molecules the quaternary amine group is exceptionally bulky. The positive charge is virtually distributed over a larger volume compared to a metal ion, resulting in a smaller surface potential. The underlying reason for the unfavorable interaction must be a successful competition of water with quaternary ammonium compounds for binding sites on the protein surface. For common protein surfaces the preferred interaction partner may be water because of its small size and consequently stronger surface potential so that it can form an electrically and sterically complementary interface with the protein surface better than the quaternary ammonium ion. This is in accord with explanations given for other substances classified as compatible solutes, which are found to be preferentially excluded from the protein surface. In case of ribonuclease and the compatible solute glycerol, the preferential hydration has been directly proven by neutron small angle scattering (Lehmann and Zaccai, 1984). Furthermore, this exclusion of compatible solutes also resembles the phenomenon

of excluded volume, which generally exists in concentrated solutions of large molecules (Minton, 2001).
However, if proteins need to bind such quaternary ammonium cations, a site can be tailored, as it is found in ProX. This special site should be a cavity just large enough to accommodate the bulky cation and has to possess an evenly negative surface potential. Indole and phenyl groups from tryptophans, tyrosines and phenylalanines have been shown to possess such an evenly negative surface potential (Mecozzi et al., 1996). In ProX, the rectangular indole box is well designed to accommodate quaternary amine ligands and thus provides electrostatic and van der Waals complementarity. An important structural feature are two conserved *cis*-prolines which follow the residues Trp65 and Trp188 and might be also in *cis*-conformation in the related structures (see Figure 4.6). These seem to play a crucial role in positioning Trp65 and Trp188 by allowing a sharp bend in the polypeptide backbone to form this almost rectangular box of tryptophan residues as found in ProX. The interaction energy has thus been well optimized under evolutionary pressure.

4.5.2 Mutational studies

Our mutational data show that at position 188 the phenyl ring of tyrosine or phenylalanine is able to substitute for the indole group whereas binding is abolished in the non-aromatic mutants. In ProX-GB residue Trp188 forms more van der Waals contacts (nine) than the other two tryptophans of the box (five and four) and it contacts all methyl groups of the trimethyl betaine ligand (Figure 4.5). Trp188 is the only residue of the box which cannot be replaced by a non-aromatic apolar or acidic residue without a complete loss of binding (Table 4.3). This indicates an essential contribution of the cation-π interaction between Trp188 and the betaine ligand. As mutational and structural data both hint at a critical influence of this residue, it seems to have a special function in ligand binding.
At the first look positions 65 and 140 appear to be of approximately equal importance for binding of GB, whereas for PB binding position 65 appears to be more important (Figure 4.5). In ProX-GB the indole rings of residues Trp65 and Trp140 form similar numbers of van der Waals contacts with the ligand (Figure 4.5). Both positions tolerate to a certain degree substitution against alanine and can not be distinguished by that. However replacement of the indole group by a bulky aliphatic or a carboxylic group impairs the binding much stronger in position 65 than in 140 (Table 4.3). This finding can be interpreted in two ways either as stronger steric or electrostatic restrictions in position 65 or as a more important role of Trp65 in ligand binding. The binding site in the

closed, ligand bound conformation appears to offer more space for a large non-aromatic side chain in position 140. An aspartate residue seems to be able to substitute for the tryptophan in position 140 providing electrostatic interactions with a K_D of 8 μM (Table 4.3). But the larger glutamate residue is worse suitable than the aliphatic side chain of leucine which is most likely for steric reasons. The complete loss of binding activity in the double mutant Trp65Ala, Trp140Ala demonstrates that at least two aromatic residues are necessary for GB binding. The comparison between ProX and two other quaternary amine binding proteins being structurally not related shows a convergent evolution of the binding site (Figure 4.7). Two aromatic rings almost perpendicular to each other seem to be required, whereas further aromatic residues and their orientation with respect to the ligand determine the strength of the interaction between ligand and protein. In the superposition in Figure 4.7 the ProX residues Trp65 and Trp188 are in the positions of the two conserved aromatic residues found also in the two other proteins. This is in accord with the findings of our mutational data which show that Trp65 has a more important role in ligand binding than Trp140. In summary, it could be shown that Trp188 in pair with either Trp65 or Trp140 is absolutely necessary for a proper ligand binding whereas the third tryptophan residue enforces the binding affinity in ProX. The difference in the binding studies between Trp65 and Trp140 reflects their importance in ligand binding.

4.5.3 Quaternary amine derivatives as neurotransmitters

Because of the identical quaternary amine head groups of GB and acetylcholine one tends to compare the interactions between them and their receptors. According to the finding of Dougherty and colleagues (Zhong et al., 1998), the cation-π energy of the acetylcholine quaternary amine bound to Trp149 in the α-subunit of the nAChR from *Torpedo californica* is critical for channel response. The structure of the acetylcholine binding protein (AChBP) from snail (Brejc et al., 2001) that can be taken as model for the extracelluar part of the nAChR suggests an analogous role for Trp143 in AChBP. However, it is not known yet how the neurotransmitter acetylcholine is exactly bound in the expected aromatic environment. From our data we suggest that Trp188 in ProX has a similar function as Trp149 in nAChR and Trp143 in AChBP. Additionally, a second aromatic residue in a similar position as Trp65 in ProX has to be present in the nAChR in order to bind the acetylcholine head group properly.
Furthermore, from the identity of the head groups we assume a similar exclusion of this part of the acetylcholine molecule from protein surfaces as found for the compatible solute glycine betaine. Only a protein with a highly specific binding site based on cation-

π interaction in form of an aromatic box is able to bind acetylcholine. We thus speculate that quaternary amine derivatives are especially suited to act as neurotransmitters in an environment crowded with a huge number of different proteins, like it is in the synaptic cleft. Where unspecific binding would lead to an inefficient use of the signaling molecule.

Acknowledgements

Financial support for this study was provided by the Deutsche Forschungsgemeinschaft through SFB-395, the Graduiertenkolleg "Proteinfunktion auf atomarer Ebene", the Max-Planck-Institute for terrestrial Microbiology (Marburg) and the Fonds der Chemischen Industrie (to E.B.).

5 ProX from Archaeoglobus fulgidus

5.1 Abstract

Compatible solutes such as glycine betaine and proline betaine serve as protein stabilizers due to their preferential exclusion from protein surfaces. To use extracellular sources of this class of compounds as osmo- or thermoprotectants, bacteria and archaea have developed high affinity uptake systems of the ATP-binding cassette type. These transport systems require periplasmic or extracellular binding-proteins which are able to bind the transported substance with high affinity. Therefore, binding-proteins which bind compatible solutes have to avoid the exclusion of their ligands within the binding site. In the present study we addressed the question to how compatible solutes can be effectively bound by a protein at temperatures around 83 °C as this is done by the binding protein ProX from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. We solved the structures of ProX without ligand and in complex with both of its natural ligands glycine betaine and proline betaine as well as in complex with the artificial ligand trimethyl ammonium. The comparison of the ligand binding sites of ProX from *A. fulgidus* and ProX from *E. coli* revealed an alternative solution for the problem of compatible solute binding. Cation-π interactions and non-classical hydrogen bonds between four tyrosine residues, a main chain carbonyl oxygen and the ligand have been identified to be the key determinants in binding the quaternary amines of the three investigated ligands. Furthermore, the ligand-free structure of ProX allowed us to trace the structural rearrangements that occur upon ligand binding.

5.2 Introduction

Variations in the supply of water and the concomitant changes in salinity and osmolarity are among the most important environmental parameters affecting the growth of microorganisms (Csonka and Epstein, 1996; Galinski and Trüper, 1994; Kempf and Bremer, 1998; Ventosa et al., 1998; Welsh, 2000). Because bacteria lack systems to actively transport water across the cytoplasmic membrane their intracellular water con-

tent is solely determined by osmotic processes. This requires an active adjustment of their intracellular solute pool to prevent dehydration under hypertonic growth conditions and bursting under hypotonic circumstances (Booth and Louis, 1999). To retain a suitable level of cellular water and to maintain turgor within a physiologically acceptable range under high osmolarity growth conditions, many bacterial species accumulate large amounts of a particular class of organic osmolytes, the so called compatible solutes. This can be accomplished either through synthesis or uptake from the environment (Record Jr. et al., 1998; Wood, 1999; Brown, 1976; da Costa et al., 1998; Galinski and Trüper, 1994; Kempf and Bremer, 1998; Welsh, 2000). Compatible solutes are non-interfering with cellular functions and can be amassed up to molar concentrations in the cytoplasm without disturbing essential cellular processes and the functioning of cell components (Record Jr. et al., 1998). Important representatives of compatible solutes are the trimethylammonium compound glycine betaine (GB, N,N,N-trimethyl glycine) and the dimethyl ammonium compound proline betaine (PB, N,N-dimethyl-L-proline). The intracellular accumulation of compatible solutes as a strategy for adaptation to high osmolarity has been widely adopted not only by bacteria and archaea (Galinski and Trüper, 1994; Kempf and Bremer, 1998; da Costa et al., 1998; Martin et al., 1999; Roessler and Muller, 2001), but also by fungal, plant, animal and even human cells (Burg et al., 1997; Hohmann, 2002; le Rudulier et al., 1984; McNeil et al., 1999; Rhodes and Hanson, 1993; Yancey, 1994).

In addition to their well-established role as osmoprotectants, compatible solutes also function as protein stabilizers both *in vitro* (Canovas et al., 1999; Courtenay et al., 2000; Lippert and Galinski, 1992) and *in vivo* (Bourot et al., 2000). The exact biological mechanism(s) by which these compounds affect protein stability is not completely understood, but their functioning is generally explained in terms of the preferential exclusion model (Arakawa and Timasheff, 1985). This hypothesis predicts that compatible solutes are excluded from the immediate hydration shell of proteins, presumably due to their unfavorable interactions with protein surfaces (Bolen and Baskakov, 2001; Qu et al., 1998). The resulting disequilibrium provides a thermodynamic force to minimize the surface of the protein in order to reduce the amount of hydration water and thereby stabilizes the native structure of proteins and favors the formation of protein assemblies. An experimental confirmation of this behavior came from neutron diffraction studies (Lehmann and Zaccai, 1984). The ability of compatible solutes to stabilize proteins probably also explains their *in vivo* function as microbial stress protectants against heat and chill stress (Brigulla et al., 2003; Caldas et al., 1999; Canovas et al., 2001; Mendum and Smith, 2002; Santos and da Costa, 2002; Welsh, 2000).

Many microorganisms are able to acquire compatible solutes, in particular glycine be-

taine, from environmental sources through high-affinity transport systems (Bremer and Krämer, 2000; Csonka and Epstein, 1996; Sleator and Hill, 2002). One example of such a glycine betaine transporter is the ProU system from *Escherichia coli* (Gowrishankar, 1989; Lucht and Bremer, 1994) a member of the binding-protein-dependent ATP-binding cassette (ABC) super family of transporters (Boos and Lucht, 1996; Lucht and Bremer, 1994). In the meantime, several Gram-negative and Gram-positive microorganisms have been shown to contain a homologue of the *E. coli* ProU glycine betaine transport system. Recently, we have identified such a ProU-type glycine betaine uptake system in *Archaeoglobus fulgidus* (Holtmann and Bremer, unpublished results). *A. fulgidus* is a sulfate-reducing archaeon with an optimal growth temperature of 83 °C. This hyperthermophile microorganism was originally isolated from hot sediments of a marine hydrothermal vent system (Stetter et al., 1987). Marine sediments are known to contain glycine betaine (Ghoul et al., 1990), but the concentration of this compatible solute in natural settings is likely to be very low and variable. In this habitat, microorganisms need an effective uptake route to scavenge the compatible solute from very dilute solutions at high temperatures. As deduced from the genome sequence of the *A. fulgidus* strain VC16 (Klenk et al., 1997), the ProU transporter of this organism consists of an ATPase (ProV), two integral membrane proteins (ProW1 and ProW2) and an extracellular substrate-binding protein (ProX). In general substrate-binding proteins bind their ligand(s) selectively and with high affinity, which is thought to ensure the substrate specificity and directionality of the overall transport reaction for a given binding protein-dependent transport system (Boos and Lucht, 1996). Periplasmic binding proteins are composed of two domains connected by one to three polypeptide chains forming a hinge between them. In the ligand-free open conformation the two rigid domains are flexibly linked by the hinge (Quiocho and Ledvina, 1996). This has been shown by the structural analysis of various open states of the ribose-binding-protein and D-allose-binding-protein (Magnusson et al., 2002; Björkman and Mowbray, 1998). Ligand binding induces a large conformational change in the hinge region which moves both domains toward each other. After this domain rearrangement, the ligand is engulfed in a predefined cleft between the two domains which refers to the closed conformation of the binding protein.
Inspection of the *proX* sequence from *A. fulgidus* (Klenk et al., 1997) suggests that ProX is an extracellular lipoprotein that is tethered to the cytoplasmic membrane via a lipid modification at the N-terminal cysteine residue. Heterologous overproduction of the ProX protein in a soluble form in *E. coli* and its subsequent purification allowed us to assess the biochemical properties of this ligand-binding protein (Holtmann and Bremer, unpublished results). Substrate binding assays with radio-labelled glycine betaine and competition experiments with other compatible solutes revealed that ProX binds

both GB and PB with high affinity and specificity with an apparent K_D of 60 nM and 50 nM, respectively at room temperature. GB and PB are well known osmoprotectants for many microbial species (da Costa et al., 1998), but both do not have this function in *A. fulgidus*. *A. fulgidus* is not able to grow in a mineral-salt-based medium lacking yeast extract at 90 °C; however both GB and PB could rescue its growth to a large extent at this elevated temperature. Consequently, both compatible solutes serve as effective thermoprotectants for the hyperthermophilic archaeon *A. fulgidus* (Holtmann and Bremer, unpublished results).
The amino acid sequences of the ProX proteins from *E. coli* (Gowrishankar, 1989) and *A. fulgidus* (Klenk et al., 1997) are only very distantly related (sequence identity < 12 %), and in particular the three tryptophan residues involved in substrate binding in ProX from *E. coli* ProX (Schiefner et al., 2004) are not conserved in ProX from *A. fulgidus*. In order to broaden our understanding of the specific interactions of compatible solutes with ligand-binding proteins, we solved the structure of ProX from *A. fulgidus* without ligand and in complex with its natural ligands GB and PB as well as the artificial ligand trimethylammonium. These structures revealed cation-π interactions and non-classical hydrogen bonds to be essential for ligand binding by ProX from *A. fulgidus*. The configuration of the ligand binding site as well as the residues involved in ligand binding differ from that of ProX from *E. coli*. Furthermore we also solved the structure of the ligand-free open conformation of ProX from *A. fulgidus*, which allowed us to trace the structural rearrangements that occur upon substrate binding.

5.3 Materials and Methods

5.3.1 Bacterial strains

The *E. coli* strain BL21 (Codon Plus RIL), which harbors a plasmid encoding rare tRNAs for arginine, isoleucine and leucine, was purchased from Stratagene and used for the heterologous overproduction of the *A. fulgidus* ProX protein. Without the presence of the Codon Plus RIL plasmid in strain BL21 there was no substantial overproduction of the ProX protein. The *proX*$^+$ overexpression plasmid pGH26 was introduced by electrotransformation into strain BL21 (Codon Plus RIL). The resulting transformants were propagated in the presence of both ampicillin (150 μg ml^{-1}) and chloramphenicol (30 μg ml^{-1}) to simultaneously select for pGH26 (Amp^r) and the Codon Plus RIL plasmid (Cm^r). The recombinant strain was propagated on Luria-Bertani (LB) medium (Miller, 1992).

5.3.2 Expression and purification of recombinant ProX

To overproduce the *A. fulgidus* ProX protein in *E. coli*, we used the *proX*$^+$ plasmid pGH26 (Holtmann and Bremer, unpublished results). This plasmid is a derivative of the expression vector pASK-IBA6 (IBA Göttingen) that carries an anhydrotetracycline-inducible *tet* promoter allowing selective *proX* gene expression. pGH26 carries the *proX* gene from the *A. fulgidus* strain VC16 (Klenk et al., 1997; Stetter et al., 1987) as a 869-bp PCR fragment flanked by BsaI restriction sites. In *A. fulgidus*, the authentic ProX protein is likely to be a lipoprotein that is acylated at its N-terminal cysteine residue allowing its anchoring into the cytoplasmic membrane (Holtmann and Bremer, unpublished results). To overproduce the ProX protein in a soluble form in *E. coli*, the *proX* gene present in pGH26 is engineered in such a way that it lacks its own signal sequence and the N-terminal cysteine is replaced by a glycine to avoid lipid modification of the soluble protein. The *proX* coding sequence for the mature *A. fulgidus* ProX was inserted into the expression vector pASK-IBA6 in frame with an upstream located OmpA signal sequence and a Strep-Tag affinity peptide. This construction allows the secretion of the ProX protein into the periplasm of *E. coli* and its recovery in a soluble form via affinity chromatography. We observed that a substantial portion of the ProX protein heterologously produced in *E. coli* still carried an unprocessed OmpA signal sequence (Holtmann and Bremer, unpublished results). To remove both the unprocessed OmpA signal sequence and the Strep-Tag affinity tag from the recombinant *A. fulgidus* ProX, the protein preparation was treated with the protease factor Xa using the cleavage site of the pASK-IBA6 vector between the Strep-Tag and ProX. This method yielded soluble ProX protein without any N-terminal extensions and the amino acid sequence starting with a glycine residue instead of a cysteine residue. The nucleotide sequence of the *A. fulgidus proX* gene in pGH26 was verified by DNA sequence analysis (Sanger et al., 1977) using a set of synthetic oligonucleotide primers appropriately spaced along the *proX* coding region.

Rich media (*e.g.* yeast extract) frequently contain GB (Dulaney et al., 1968). To obtain *A. fulgidus* ProX preparations that were free of its ligands GB and PB, the overproducing strain BL21 (Codon Plus RIL) (pGH26) was grown in a chemically defined minimal medium (MMA) (Miller, 1992). A ten-liter flask containing five liters of MMA supplemented with 150 μg ml^{-1} ampicillin, 30 μg ml^{-1} chloramphenicol, 0.5 %(w/v) glucose as the carbon source and 0.2 %(w/v) casamino acids was inoculated to an OD_{578} of 0.1 from an overnight culture prepared in the same growth medium. The cells were grown at 37 °C with vigorous stirring until the culture had reached mid-exponential phase (OD_{578} 0.5 - 0.8). Transcription of the *proX* gene from the plasmid-encoded *tet* promoter in

pGH26 was then induced by the addition of anhydrotetracycline (final concentration 0.2 μg ml^{-1}). Cells were grown for two more hours to allow ProX production and were subsequently harvested by centrifugation (10 min, 3000 x g).
To release the periplasmic proteins from the ProX overproducing *E. coli* cells, the cell paste was resuspended in 50 ml of ice cold buffer P (100 mM Tris-HCl pH 8.0, 500 mM sucrose and 1 mM EDTA) to allow the formation of spheroplasts. After 30 min incubation on ice, the spheroplasts were separated from the periplasmic protein extract by centrifugation (15 min, 21000 x g). Insoluble material of the supernatant was subsequently removed by ultracentrifugation (30 min, 120000 x g). The cleared protein extract was then loaded onto a 10 ml Strep-Tactin column (IBA, Göttingen) equilibrated with buffer W (100 mM Tris-HCl pH 8.0). After washing with five bed volumes of buffer W, the bound proteins were eluted from the affinity resin with buffer E (100 mM Tris-HCl pH 8.0, 2.5 mM desthiobiotin). Then both the partially unprocessed OmpA signal sequence and or the Strep-Tag affinity sequence were cleaved with the protease factor Xa (1 μg factor Xa per 200 μg ProX protein) for 16 hours at room temperature in a buffer containing 100 mM Tris-HCl pH 8.0, 100 mM sodium chloride and 1 mM calcium chloride. To remove the factor Xa and other minor contaminating *E. coli* proteins from the *A. fulgidus* ProX protein solution, the solution was diluted 1:2 with deionized water and loaded onto a UnoQ6 column (BioRad) equilibrated with buffer A (50 mM Tris-HCl pH 8.0). The column was washed with 20 ml buffer A and the proteins were eluted with a linear sodium chloride gradient. ProX eluted at a sodium chloride concentration of 250 mM from the UnoQ6 column. The protein was stored until further use at 4 °C. In general, 2 mg of pure ProX protein were obtained per 1 liter culture.
The ProX protein from *A. fulgidus* contains five methionines. For the labelling of the ProX protein with L-selenomethionine, the method of metabolic inhibition was employed (Doublie, 1997; van Duyne et al., 1993). A 100 ml preculture of strain BL21 (Codon Plus RIL) (pGH26) was grown for 20 h in LB medium containing the appropriate antibiotics at 37 °C. Cells from 16 ml of this culture were recovered by centrifugation (5 min, 3000 x g) and resuspended in the same volume of M9 minimal medium (Miller, 1992). This entire cell suspension was then used to inoculate 4 liters of M9 medium in a ten-liter flask. Cells were grown to an OD_{578} of 0.3, and 100 mg l^{-1} L-lysine, 100 mg l^{-1} L-phenylalanine, 100 mg l^{-1} threonine, 50 mg l^{-1} isoleucine, 50 mg l^{-1} leucine, 50 mg l^{-1} valine and 50 mg l^{-1} L-selenomethionine were added as solid powder to the cell culture. Fifteen minutes after addition of the amino acids, the expression of the *A. fulgidus proX* gene was induced by adding anhydrotetracycline (0.2 μg ml^{-1}) to the culture. After 6 hours of further growth, cells were harvested by centrifugation (10 min, 3000 x g). The protein was then purified from these cells as described above for the unmodified

ProX; the L-selenomethionine-labelled ProX showed the same behavior in each purification step as the unlabelled protein. The incorporation of L-selenomethionine into the purified ProX protein was verified by MALDI-TOF mass-spectrometry using a sample of the unlabelled ProX protein as a reference.

5.3.3 Crystallization

Crystals were grown using the vapor diffusion method at 18 °C. The protein solution containing 10 mg/ml protein in 50 mM Tris-HCl pH 8.0, 250 mM sodium chloride was mixed with equal amounts of the reservoir solutions described below. Liganded ProX crystallized in hanging drops using a reservoir solution containing 0.2 M zinc acetate, 0.1 M sodium cacodylate pH 6.0 - 6.5, 10 - 12 %(w/v) PEG4000. The crystals reached a final size of 250 x 200 x 150 μm^3 after three days. They belong to the space group $P2_1$ (crystal form I). In a different setup, liganded ProX crystallized in sitting drops equilibrated against a reservoir containing 30 %(w/v) PEG1500. The crystals grew to a maximum size of 50 x 50 x 50 μm^3 within 2 months and belong to the space group $P4_32_12$ (crystal form II). Unliganded ProX crystallized in hanging drops against a reservoir solution containing 0.3 M magnesium chloride, 0.1 M Tris-HCl pH 7.0 - 9.0, 35 %(w/v) PEG4000. First crystals appeared after 2 - 3 months, reached a final size of 300 x 300 x 150 μm^3 and belong to space group C2 (crystal form III). Again using a different setup, unliganded ProX crystallized in hanging drops equilibrated against a reservoir containing 0.1 M zinc acetate, 0.1 M NaMES pH 6.5, 25 - 30 %(v/v) ethylene glycol. These crystals grew within 4 weeks, reached a final size of 200 x 150 x 20 μm^3 and belong to space group $P2_12_12_1$ (crystal form IV). All data sets were collected at synchrotron sources from frozen crystals. Crystals of crystal form I were transferred into the reservoir solution saturated with solid glucose and after a soaking time of 20 seconds they were frozen in liquid nitrogen. Crystals of the crystal forms II - IV could be directly transferred into liquid nitrogen. Selenomethionine labelled protein was crystallized in the crystal form I and was treated as the native crystals prior to freezing.

5.3.4 Data collection and Refinement

Data collection was carried out at the Swiss Light Source SLS Villigen (CH) beamline X06SA. All data sets were processed using XDS (Kabsch, 1993) as summarized in Table 5.1. Initial phases of the selenomethionine labelled ProX in crystal form I were determined by anomalous dispersion at the two wavelengths $\lambda = 0.9794$ Å (λI) and $\lambda =$ 1.044 Å (λII). A first set of heavy atom sites was found with SHELXD (Schneider and Sheldrick, 2002) but the map calculated with phases obtained by SHELXE afterwards

showed poor connectivity and was not sufficient to be interpreted with a protein model. These preliminary phases were used to calculate anomalous difference maps to locate the selenium sites at the two wavelengths λI and λII. From the comparison of the site occupancies which are proportional to the f″-values of the element at the wavelengths λI and λII it could be concluded that there were at least two types of anomalous scatterers present. Only eight of the eleven sites found by SHELXD turned out to be selenium, the others were treated as zinc because the highest peak in the fluorescence spectrum corresponded to that of ZnKα. Refinement of this set of sites with the program SHARP (de la Fortelle and Bricogne, 1997) located five additional sites. Finally, a set of ten selenium sites and six zinc sites were refined in SHARP followed by solvent flattening using RESOLVE (Terwilliger, 2001). This resulted in very good phases up to 2 Å resolution leading to an electron density map which could be automatically interpreted by Arp/Warp (Perrakis et al., 1999). Missing residues and minor corrections of the model were built using the graphical interface 'O' (Jones et al., 1991) followed by refinement with the program REFMAC5 (Murshudov et al., 1997).
The liganded structure in space group $P4_32_12$ (crystal form II) was solved by molecular replacement with MOLREP (Vagin and Teplyakov, 1997) using chain I of the model in space group $P2_1$. Also the unliganded structures of ProX were solved by molecular replacement with MOLREP. First the structure in space group C2 (crystal form III) was solved with the two domains of the closed structure as separate search models. Two peaks of approximately the same height were found for domain A in the cross rotation function whereas no useful signal was detected for domain B. In the translation function four copies of domain A (two for each rotation peak) could be located in the asymmetric unit which were refined using REFMAC5. To locate the B domains, a molecular replacement was carried out against the Fo-Fc map of the initial A domains model, resulting in the correct position of two copies of domain B. This was repeated twice to complete the model. The structure of the unliganded ProX in crystal form IV was solved by molecular replacement with chain I of the model in space group C2.
The quality of the obtained models was validated with the program PROCHECK (Laskowski et al., 1993). Secondary structure elements and solvent accessible surface of the ProX structure were determined using DSSP (Kabsch and Sander, 1983). Figures with molecule presentations were prepared using the programs MolScript (Kraulis, 1991), and Raster3D (Merrit and Bacon, 1997).

Table 5.1: Data collection statistics

	SeMet derivatives		**Refined ProX structures**			
Dataset	λI	λII	ProX-GB	ProX-PB	ProX-TM	ProX ligand-free
Ligand	Proline betaine	Proline betaine	Glycine betaine	Proline betaine	Trimethyl ammonium	–
Wavelength [Å]	0.9794	1.0440	0.9774	1.0440	0.9778	0.9778
Total rotation range [°]	141.5	180	90	180	180	180
Temperature [K]	100	100	100	100	100	100
Space group	$P2_1$	$P2_1$	$P4_32_12$	$P2_1$	$P2_1$	C2
Cell (a, b, c) [Å]	64.0, 77.9, 67.6	64.4, 78.4, 67.8	75.9, 75.9, 88.4	64.4, 77.9, 67.6	63.7, 75.7, 67.0	161.5, 56.5, 116.1
(a, b, g) [°]	90, 91.7, 90	90, 91.7, 90	90, 90, 90	90, 91.7, 90	90, 91.4, 90	90, 110.6, 90
Molecules per asu	2	2	1	2	2	4
Resolution limits [Å]	50-1.8 (1.9-1.8)	50-1.9 (2.0-1.9)	50-2.1 (2.2-2.1)	50-1.9 (2.0-1.9)	50-1.9 (2.0-1.9)	50-1.8 (1.9-1.8)
Number of reflections	178597	189447	110649	189413	180063	339517
Unique reflections	97236	100862	15622	52348	49535	90922
Completeness [%]	80.4 (83.4)	96.5 (90.9)	99.8 (100.0)	98.4 (99.6)	98.6 (98.9)	99.6 (99.8)
I/σI	9.3 (2.5)	10.3 (4.1)	10.8 (3.8)	13.8 (5.4)	9.7 (4.1)	10.7 (3.9)
R_{sym}	4.7 (25.5)	4.5 (14.3)	10.8 (54.7)	5.5 (17.2)	8.6 (32.2)	7.7 (33.2)
R_{meas}*	6.2 (33.6)	6.2 (19.3)	11.7 (59.1)	6.5 (20.4)	10.1 (38.1)	9.1 (38.8)
$R_{mrgd\text{-}F}$*	8.3 (44.6)	7.2 (24.8)	9.6 (38.3)	6.1 (22.6)	10.4 (34.4)	9.3 (34.7)
B-factor (Wilson plot) [$Å^2$]	31.2	29.4	40.1	29.4	28.4	26.8
Figure of merit overall	0.49					

R_{meas} and $R_{mrgd\text{-}F}$ are better indicators for quality than R_{sym}. R_{meas} is redundancy independent and $R_{mrgd\text{-}F}$ is a measure for the quality of the reduced amplitudes (Diederichs and Karplus, 1997; Weiss and Hilgenfeld, 1997). Values in parentheses refer to the highest resolution data shell.

Chemicals

Glycine betaine and desthiobiotin were purchased from Sigma. Proline betaine (stachydrine-HCl) was obtained from Exrasynthése. Factor Xa was purchased from New England Biolabs. L-Seleno-methionine was obtained from Fisher Scientfic. Trimethyl amine solution was purchased from Hampton Research.

Database searches

Searches for proteins that are homologous to ProX from *A. fulgidus* were conducted using the BLAST network server (Altschul and Gish, 1996).

PDB accession codes

All atomic coordinates and structure factors resulting from this work have been deposited in the Protein Data Bank (Berman et al., 2000). The accession codes are 1SW1, 1SW2, 1SW4, and 1SW5 for ProX-PB, ProX-GB, ProX-TM, and ProX ligand-free, respectively.

5.4 Results and Discussion

5.4.1 Crystallization

Crystals of ProX have been obtained under four different conditions yielding crystals in four different space groups. The liganded closed from crystallized in $P2_1$ and $P4_32_12$ whereas the open unliganded form crystallized in C2 and $P2_12_12_1$. In the following, crystallization of liganded ProX in $P2_1$ has been taken as test for ligand binding, because the ProX protein does not crystallize without any ligand under these conditions. Various compounds have been added to test if they are potential ligands of ProX. Well diffracting crystals were obtained with the natural ligands GB and PB. Crystals of the same quality and approximately of the same size were obtained by the addition of trimethyl amine (see Table 5.1). Trimethyl amine has a pKa of 9.81 and is therefore protonated under the crystallization conditions we have used. This trimethyl ammonium (TM) resembles the head group of GB. However, 3,3-dimethyl butyric acid, where a carbon atom replaces the nitrogen atom compared to GB, is not able to substitute GB. In order to test whether an ordinary alkaline metal ion is a good mimic for the positive charge of the natural ligand GB, the chlorides of Li^+, Na^+, K^+, Rb^+ and Cs^+ were added to the drop with final concentrations of 10 mM. None of these ions was able to close the binding protein in the same way as the natural ligands GB and PB did. This might have two reasons, either the interaction between the ligand and the ProX requires more than a positive charge or

Table 5.2: Refinement statistics

	ProX-GB	ProX-PB	ProX-TM	Ligand-free
Resolution limits [Å]	50-2.1	50-1.9	50-1.9	50-1.8
Total number of reflections	15622	52348	49535	90922
Reflections in working set	14849	49688	47096	86386
Reflections in test set	773	2660	2439	4536
R [%]	20.1	18.0	21.6	18.6
R_{free} [%]	24.9	21.9	24.5	21.4
No. of amino acid residues	270	540	540	1080
No. of protein atoms	2161	4322	4322	8644
No. of ions	–	6	9	9
No. of water molecules	46	266	245	327
No. of ligand atoms	8	20	8	–
B-factor for all atoms [Å^2]	38.4	32.2	26.0	21.4
B-factor of protein atoms [Å^2]	38.4	31.9	25.8	21.2
R.m.s.d. bonds [Å]	0.014	0.015	0.011	0.013
R.m.s.d. angles [°]	1.429	1.434	1.309	1.364

the metal ions cannot get rid of their hydration shell which would make them too bulky for the closed binding pocket. This question cannot be answered by our experiment.

5.4.2 Overall structure

The structure of ProX from *A.fulgidus* has been solved in the open conformation as well as in the closed conformation in complex with each of its natural ligands GB, PB, and the artificial ligand TM. Initial phases were obtained by a two wavelength MAD approach of ProX-PB in space group $P2_1$ (see Table 5.1). All other structures were determined by molecular replacement. The final models contain 270 residues, the five N-terminal residues of the expressed protein are disordered in all crystal forms and are not part of the models (see Table 5.2). This is in agreement with the expectation that these five residues have to be especially flexible because they provide the linker between ProX and its lipidic membrane anchor.

The overall structure shows a typical binding protein type II fold (Quiocho and Ledvina, 1996) (see Figure 5.1). This can generally be subdivided into two globular domains: 1) domain A (blue) from residue 6 to residue 110 and from residue 212 to residue 275 and 2) domain B (yellow) from residue 111 to residue 211 connected by the two switch segments Asp109-Asp110-Tyr111-Ala112 and Ala210-Leu211-Pro212-Pro213 that form with Asp110-Tyr111 and Leu211-Pro212 the two hinges between domain A and B. Each domain consists of a four-stranded β-sheet flanked by α-helices on both sides.

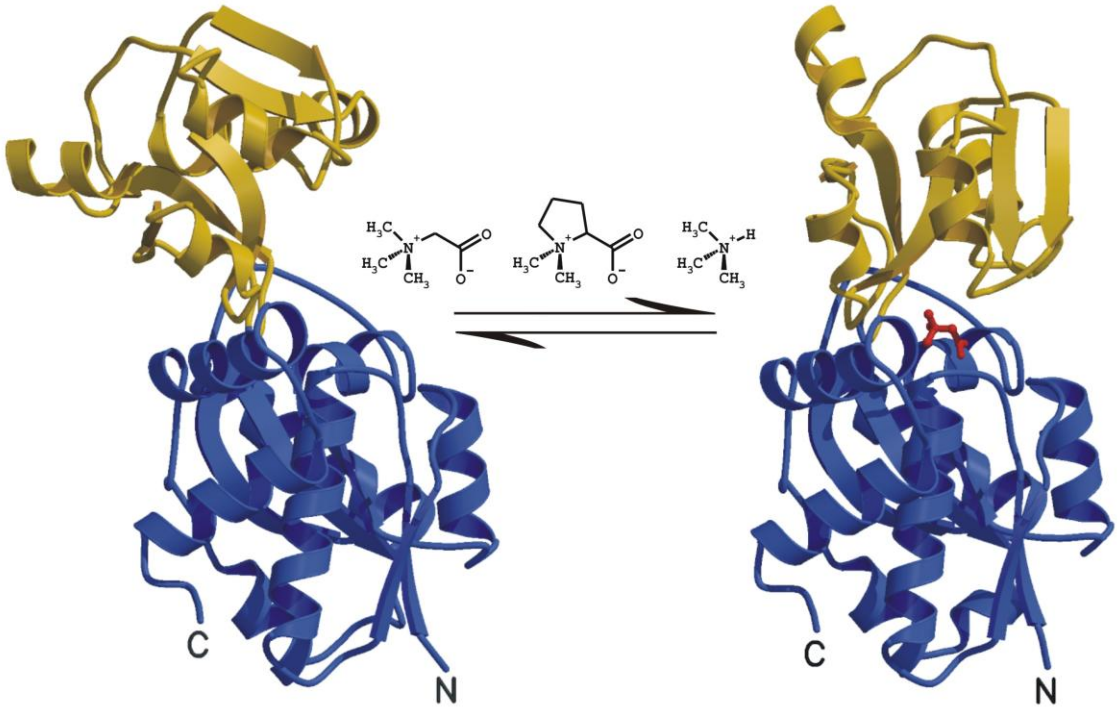

Figure 5.1: Overall structure open and closed conformation ProX undergoes a large conformational change from its unliganded open conformation (left) to the liganded closed form (right). In between (above the arrow) the three ligands are shown which have been found to induce this conformational change, GB, PB and TM, from left to right. Domain A (blue) is shown in the same orientation for both conformations, while the comparison shows that domain B (yellow) is relocated with respect to domain A. In the closed conformation the ligand GB is highlighted in red.

5.4.3 Overall conformational changes induced upon ligand binding

The ligand free, open conformation of ProX has been crystallized in two different space groups. In this paper only the open conformation in space group C2 is discussed because of the better resolution obtained from C2 crystals and the very similar coordinates of ProX in space group $P2_12_12_1$. In C2 the asymmetric unit contains four molecules with very similar angles between the two domains A and B. The only major difference is that domain B of the molecules 3 and 4 is less ordered than the B domains of molecules 1 and 2. This is reflected by their high mean B-factors 47.1 Å^2 and 44.9 Å^2 which are more than twice as high as those of the other B domains (19.2 Å^2 and 19.6 Å^2) resulting in a poorly defined electron density in some regions of domain B in molecules 3 and 4. This can be interpreted as an intrinsic motion of the two B domains with respect to their appropriate A domains, which is rather surprising because at low temperatures a thermophilic protein is expected to show a rigid conformation. Apparently, the hinge provides enough flexibility at low temperatures as well, being in agreement with binding studies of ProX that were performed at room temperature (Holtmann and Bremer, unpublished results).

In the following discussion of the open conformation only molecules 1 and 2 are used. The comparison of the ligand free and liganded conformation of other binding proteins showed an approximate rigid body motion of the two domains (Sharff et al., 1992; Oh et al., 1993; Gerstein et al., 1993). With regard to this finding, we analyzed the difference in the arrangement of the two domains in the open and in the closed form with the program NCSGROUPS (Diederichs, unpublished results). The basic task of the program is to introduce hinges, which give a better least square fit of the Cα positions between two molecules. It turned out that the introduction of only two hinges, one between Asp110 and Tyr111 and the other between Leu211 and Pro212, was sufficient to superimpose the two molecules very well (domain A r.m.s.d. of 0.521 Å and domain B r.m.s.d. of 1.017 Å). This is equivalent to a total rotation of domain B by $\sim$ 58 ° with respect to domain A (Figure 5.1). The total rotation has two components: 1) the hinge angle between the two domains of $\sim$ 40 ° with its axis going through the above determined hinges in the polypeptide and 2) a rotation perpendicular to the hinge axis of $\sim$ 42 °. If two more hinges are introduced separating domain B in the three segments 111 - 188, 189 - 193, and 194 - 211, the r.m.s.d. between domain B of the open form compared to that of the closed form is reduced to 0.720 Å. Segment 194 - 211 is crossing the hinge region almost perpendicular to that on the opposite side of the ligand binding site and forms with its terminal residues contacts with domain A. Interestingly, segment 111 - 188 that is farthest from the hinge has the highest B-factor in all four polypeptides in

the asymmetric unit of space group C2. For that reason we expect different degrees of mobility of domain B in the open conformation, the hinge associated segments 189 - 193 and 194 - 211 are less mobile and stabilized by contacts to domain A.
Although the domains behave as rigid groups, there are a few changes of the secondary structure in two regions of ProX as indicated by DSSP (Kabsch and Sander, 1983). The α-helical conformation (in the open form) of residues 144 - 148 (domain B) change either to isolated-β-bridge- or to turn-conformation (in the closed form). This conformational change may be caused by the proximity to Arg149 which plays an important role in ligand binding as discussed below. Furthermore, residues 222 - 225 (domain A) which are in turn- and 3_{10}-helix-conformation (in the open form) become rearranged to a short α-helix (in the open form).
Shilton et al. (1996) found that decrease of the accessible surface area on the opposite side of the binding cleft may stabilize the open conformation in absence of the ligand. To check wether there is a similar effect in the ProX structure we calculated the accessible surface area for every residue of both the open and the closed conformation with DSSP. The differences for all residues between the open and the closed form were plotted. Only changes in the accessible surface of all residues with more than the standard deviation of this function were further analyzed. Among those all residues were selected which were between the two domains on the opposite side of the binding cleft. This approach identified eleven residues that show decreases of the accessible surface of more than 25 Å^2 between the open and the closed form, resulting in a buried surface increase of 476 Å^2 in this area. Five of these residues are part of the above mentioned segments 189 - 193 and 194 - 211 which may have a special function during closure of the binding site.
Furthermore, all salt bridges formed in the open and the closed form have been analyzed. Salt bridge formation is inferred for a pair of oppositely charged residues Asp, Glu and Arg, Lys or His, if the participating oxygen and nitrogen atoms of the involved residues lie within a distance of less than 4 Å (Barlow and Thornton, 1983). ProX has no salt bridges with histidines involved, since it does not contain any histidine residues. According to this criterion, a total of 19 salt bridges is formed in ProX in both the open and in the closed form, 13 of them are hydrogen bonded with a distance of less than 3.5 Å (Kumar and Nussinov, 1999). During the conformational change five salt bridges are broken and five new ones are formed. Interestingly, the ratio of observed ion pairs changes. While in the open form a similar number of Glu-Lys, Glu-Arg, Asp-Lys, and Asp-Arg pairs was observed (six, three, four, and six, respectively), in the closed form eight Glu-Lys and eight Asp-Arg pairs but only two Glu-Arg and one Asp-Lys pairs were found. The open conformation is apparently not stabilized by salt bridges compared to the closed form. We find only one interdomain salt bridge in the open conformation

compared to three in the closed liganded form of ProX.
Altogether our data support the observation from other binding proteins that a larger buried surface area on the opposite side of the ligand binding site is found in the open conformation. This may delay the domain movement in the open conformation to shift the equilibrium between the open and the closed conformation without ligand toward the open conformation.

5.4.4 Ligand binding site in the closed conformation

The binding site is located in the cleft between the two domains A and B as shown in Figure 5.1. In the case of the natural substrates GB and PB, the ligand binding site can be subdivided into two parts, one binding the quaternary ammonium head group and the other binding the carboxylic tail of these compounds. Figure 5.2A shows all residues involved in GB binding. The quaternary ammonium head group is caged in a box formed by the main chain carbonyl of Asp109 as the bottom part of the box, and the four tyrosine residues Tyr63, Tyr111, Tyr190, and Tyr214 as the four faces of the box, being almost perpendicular to each other. This arrangement gives the impression of an aromatic girdle of tyrosines around the quaternary amine. Together with the main chain oxygen atom of Asp109 the tyrosine side chains provide a negative surface potential which is complementary to the cationic quaternary ammonium head group of GB. The carboxylic tail of GB is pointing outward of this partially negatively charged environment forming two salt bridges and one hydrogen bond with Lys13 (domain A), Arg149 (domain B) and Thr66 (domain A), respectively. However, the betaine ligand is not completely covered by the ligand binding site. A water molecule, that is not in direct contact with the ligand and held in place by residues Tyr111, Glu145 and a further water molecule, separates the betaine ligand from the bulk solvent.
The binding of the quaternary ammonium head group of GB by ProX is mediated by different forces, cation-π (Dougherty, 1996), cation-dipole, van der Waals interaction and non-classical hydrogen bonds (Taylor and Kennard, 1982; Derewenda et al., 1995; Weiss et al., 2001). All distances between the carbon atoms bonded by the quaternary nitrogen and the tyrosine phenyl groups of Tyr63, Tyr111, Tyr190, Tyr214 as well as main chain carbonyl oxygen of Asp109 have been determined. These distances were compared to a compiled list of van der Waals radii published by Li and Nussinov (1998). For this analysis the influence of the quaternary ammonium charge on the contact distances was neglected. The methyl- or methylene groups, the aromatic ring atoms, and the carbonyl oxygens possess van der Waals radii of 1.92 Å, 1.82 Å, and 1.52 Å, respectively, with distance distributions having standard deviations of approximately 0.5 Å (Li

and Nussinov, 1998). Therefore we considered a methyl- or a methylene group to be in contact with a phenyl ring atom or a carbonyl oxygen atom if their mutual distance was between 3.5 Å and 4 Å or between 3.2 Å and 3.7 Å, respectively. According to these criteria, GB forms five, six, six, three, and three contacts with Tyr63, Tyr111, Tyr190, Tyr214 and Asp109-CO, respectively. From the arrangement of the tyrosine residues in the binding site, Tyr63 and Tyr214 from domain A, and Tyr111 and Tyr190 from domain B, we expect an almost equal importance of all tyrosine residues in binding of the betaine ligand.

PB binds basically in the same way into the binding pocket of ProX as GB does. The trimethyl tripod (C1, C2 and Cδ of the proline ring) points into the same direction as the trimethyl head group of GB does, towards the main chain carbonyl oxygen of Asp109. However, the position of PB is tilted with respect to the GB position due to the proline ring (Figure 5.2B). Since the binding site is too small to accommodate a proline ring in the direction of Tyr190, PB is pushed into the direction of Tyr214, resulting in a tilting of the N-Cα-axis of PB with respect to that of GB. Though this escape of PB is not sufficient, the side chains of Tyr190 and Tyr214 adapt to the bulkier ligand by slightly rotating their phenyl rings out of the binding site. Nevertheless, the mean distances between Tyr190 and Tyr214 and the PB liganding side chains are slightly shorter than expected from the distance distribution mentioned above (minimal distance found is 3.12 Å). PB forms six, eight, three, and three contacts with Tyr63, Tyr111, Tyr214 and Asp109-CO, respectively with C1, C2, Cδ, and Cα. There is no direct contact of the quaternary amine to Tyr190, this residue is only in contact with the Cβ and Cγ of the PB proline ring. The carboxylic tail of PB is in a very similar position as that of GB, it forms the same contacts with comparable distances.

TM binds in the same way as the quaternary amine head group of GB does (see Figure 5.2C). This seems to be the favored conformation for an isolated trimethyl group or a rotable trimethyl ammonium group as it is the case for GB. Additionally, a chloride ion binds in the position where the carboxyl tail is bound in the GB structure. This chloride ion interacts with all residues Lys13, Thr66, and Arg149 that are involved in binding of the carboxylic tail of the natural ligands GB and PB. It apparently compensates for the gap caused by the missing carboxylic group and the positive charges of Lys13 and Arg149.

Basically, all ligands are oriented in the same way in the binding site. The trimethylammonium group of GB and of TM are located in the same position. Only the dimethylammonium group of PB is slightly shifted from the optimal position of an isolated or freely rotable trimethyl ammonium due to the bulky proline ring. Nevertheless, all three ligands interact in the same way with the quaternary amine binding part of the binding

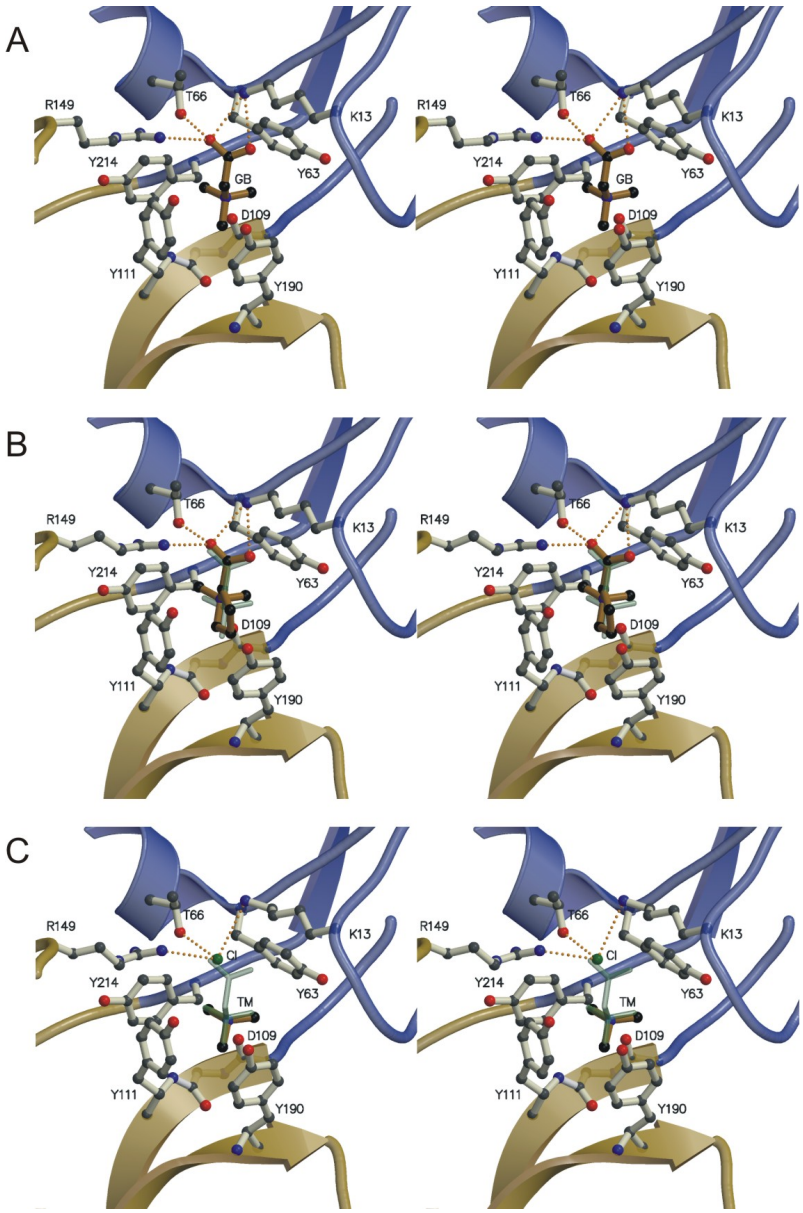

Figure 5.2: Binding site of ProX with three different ligands Stereo picture of the ProX binding sites with the three different ligands. The domains are colored as in Figure 5.1. All residues involved in binding are presented as a ball-and-stick model and labelled with their residue names. The ligands are highlighted with brown sticks. To compare the ligand positions, the binding sites were superimposed with LSQKAB (CCP4, 1994). A) Shows the binding site with GB: the quaternary amine is bound by the four tyrosine side chains of Tyr63, Tyr111, Tyr190, Tyr214, and the main chain carbonyl oxygen of Asp109 (the interaction itself is not depicted). Salt bridge formation and hydrogen bonding between the ligand carboxylate and Lys13, Arg149, and Thr66 are drawn as dashed lines. B) PB binds essentially in the same way as GB does. The trimethyl head group of PB consisting of C1, C2 and Cδ (of the proline ring) is bound in the same orientation as GB but is slightly tilted with respect to the GB. To visualize the difference to the GB structure the GB ligand has been superimposed as grey transparent stick-model. C) The trimethyl group of trimethylammonium occupies almost the same position in the binding site as the trimethyl head group of GB does. Additionally, a chloride ion (green) is bound in the position where the carboxylic tails of the natural ligands GB and PB are bound. For comparison with the GB structure, the GB ligand has been superimposed as light-green transparent stick-model.

site. Their quaternary amine head groups point all towards the main chain carbonyl oxygen of Asp109.
In a recent publication from our lab (Schiefner et al., 2004) we stated that quaternary amines are bound by a certain arrangement of aromatic residues forming an aromatic box. The present work demonstrates that quaternary ammonium groups can also be bound by partially charged oxygens, like main chain carbonyls. The methyl groups of GB, PB and TM are polarized by the quaternary nitrogen making their hydrogen atoms more acidic and thereby enabling them to act as hydrogen bond donors. According to the HSAB concept of Pearson (1963) these C-H groups are soft donors which are able to form hydrogen bonds with soft acceptors like π-systems and carbonyl oxygens. Scheiner et al. (2001) found by quantum mechanical calulations that C-H$\cdots$O hydrogen bonds between Cα-H and H_2O are about half as strong as a classical hydrogen bond and are less sensitive to geometrical distortions than classical hydrogen bonds (e.g. O-H$\cdots$O). The distance between C and O in a C-H$\cdots$O hydrogen bond is in the range from 3.31 to 3.35 Å. In the ProX-GB structure the distances between the methyl carbons of GB and the main chain carbonyl of Asp109 are with 3.33, 3.21 and 3.51 Å very close to the ideal distance. Therefore we assume that GB mainly interacts via non-classical hydrogen bonds with the main chain carbonyl of Asp109. This is very similar for TM with mean distances of 3.38, 3.34 and 3.35 Å, but slightly different for PB with mean distances of 3.86, 3.34 and 3.45 Å due to the tilting of the N-Cα axis as explained above. Furthermore, we speculate that the interaction between the ligands and the tyrosine residues has two components, a cation-π and non-classical hydrogen bond contributions by C-H$\cdots\pi$ hydrogen bonds (Brandl et al., 2001).
The above mentioned explanations also fit the observation that 3,3-dimethyl butyric acid is not bound by ProX. Without a quaternary nitrogen carrying a positive charge that polarizes the bound methyl groups, it is unable to interact with the ProX ligand binding site.

5.4.5 Conformational changes of the ligand binding site

Figure 5.3 shows the superposition of the open unliganded form onto the closed liganded form of ProX. The residues of domain A that are involved in ligand binding show virtually the same orientation in the open and in the closed form (right part of Figure 5.3). Residues Tyr63, Tyr214, Lys13 and Thr66 superimpose very well. Only the main chain carbonyl of Asp109 from domain A is slightly out of place compared to the closed form due to the enormous main chain rearrangement between Asp110 and Tyr111 upon closing. The residues contributed by domain B behave quite differently.

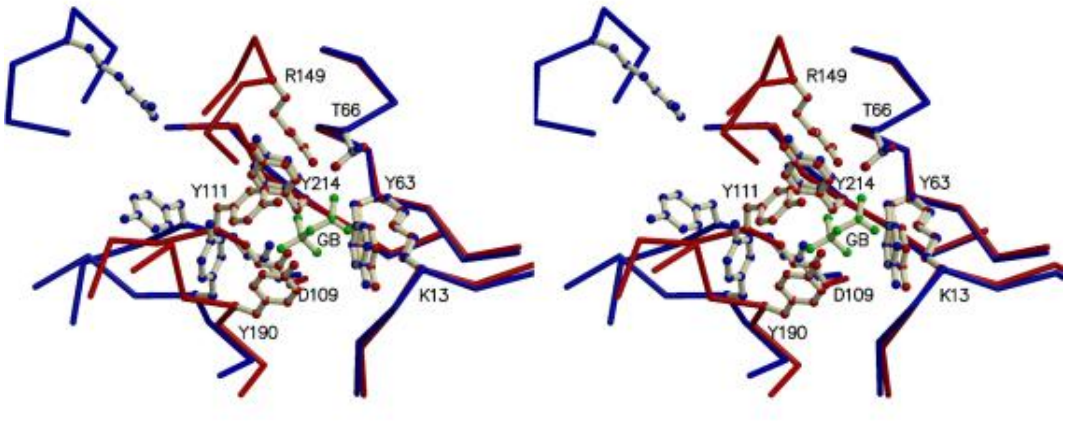

Figure 5.3: ProX binding site in the open and in the closed conformation Stereo picture of the superposition of ProX in the open conformation (blue) onto the closed conformation (red). GB (green) is shown as the ligand in the closed form. Residues involved in ligand binding and the ligand are drawn as ball-and-stick model and labelled with their residue names.

Tyr111 and Tyr190 are not only moved as parts of domain B but they undergo a major conformational change to adopt the conformation of the closed ligand binding site. The side chain conformation of Arg149 shows only small changes between the open and the closed conformation although it undergoes a large movement as part of domain B.
Not shown in Figure 5.3 are the contacts between the ligand binding residues and their neighboring residues. There is no change for the residues of domain A between the open and the closed conformation, residues Tyr63, Tyr214, Lys13 are stabilized either by hydrogen bond formation (Tyr63-Glu17, Tyr63-Glu18, Tyr214-Asp151, Lys13-Thr45) or salt bridges (Lys13-Glu62). The side chain of Thr66 is not involved in any contacts in the open form. Many changes happen in domain B between the open and the closed conformation. In the open form C-H$\cdots\pi$ interactions (Brandl et al., 2001; Steiner and Koellner, 2001) (Tyr190-Phe146, Tyr111-Tyr190) and hydrophobic contacts (Tyr190-Pro188) are formed. This changes upon closing, in closed conformation the orientation of the phenyl rings is stablized by C-H$\cdots\pi$ interactions (Tyr111-Phe146 donor and acceptor being exchanged, Tyr190-Phe15) and by van der Waals contacts (Tyr190-Met175, Tyr214-Pro212). Furthermore, in the closed liganded conformation tyrosine residues Tyr111 and Tyr190 form hydrogen bonds as well (Tyr111-Asp143; Tyr190-Asp143). Additionally, two interdomain hydrogen bonds (Tyr214-Asp151, Thr66-Arg149) and two salt bridges (Arg149-Glu145, Arg149-Asp151) are formed in the closed conformation.

Overall, the conformational change during closing of ProX appears rather like a closing hand than a venus-flytrap as proposed by Mao et al. (1982). All residues provided by domain A, including Tyr63 and Tyr214, Asp109-CO as well as Lys13 and Thr66, remain at their positions which can be referred to as the palm of a hand. The residues of domain B, however, undergo a large conformational shift, like the fingers of a grasping hand (with segment 111 - 188 as the finger tips). From the number of contacts and the conformation of the residues from domain A involved in ligand binding, we expect that the incoming ligand first binds to domain A, where it almost adopts its destined position. The binding of the betaine ligand reduces the negative surface potential in the binding site making domain A complementary to domain B of the closed form. Thereby the equilibrium between the open and the closed form is shifted towards the closed form. In domain B only residues Tyr111 and Tyr190 have to be rearranged to close the binding site and to engulf the ligand. This point of view agrees with the observations from kinetic experiments (Miller III. et al., 1983) that the binding of the ligand occurs in a series of reactions of first order.

5.4.6 Comparison of related sequences

Many sequences related to ProX from *A. fulgidus* have been identified in a Blast search against the NCBI database. The fact that among these sequences ProX from *E. coli* did not appear will be discussed below. Six of the identified sequences have been selected for an alignment with ProX from *A. fulgidus* (see Figure 5.4). A protein from the archaeon *Methanosarcina mazei* shows with 55 % the highest identity to the *A. fulgidus* ProX protein found in the Blast search. All other selected sequences were from bacteria: *Bacillus subtilis*, *Clostridium acetobutylicum* (Gram-positive with low GC content), *Streptococcus mutans* (Gram-positive with high GC content), and *Rhodopseudomonas palustris* (α-Proteobacteria). The sequences of the compatible solute binding proteins OpuBC and OpuCC from *B. subtilis* with about 30 % identity were chosen because *B. subtilis* used to be a model organism for the investigation of the physiological effects of compatible solutes on cell growth under hyperosmotic conditions (Kempf and Bremer, 1998). Whereas the binding proteins of *A. fulgidus*, *M. mazei* and *B. subtilis* are supposed to be membrane anchored via lipidic modification at their N-termini, other binding proteins like those from *C. acetobutylicum*, *S. mutans*, and *R. palustris* are fused to their ABC transporter components located in the cytoplasmic membrane.
In the alignment shown (Figure 5.4), the positions of the tyrosine residues involved in ligand binding are clearly conserved with only two exceptions: 1) position equivalent to Y63 in *S. mutans* and 2) *C. acetobutylicum* position equivalent to Y190. In both cases

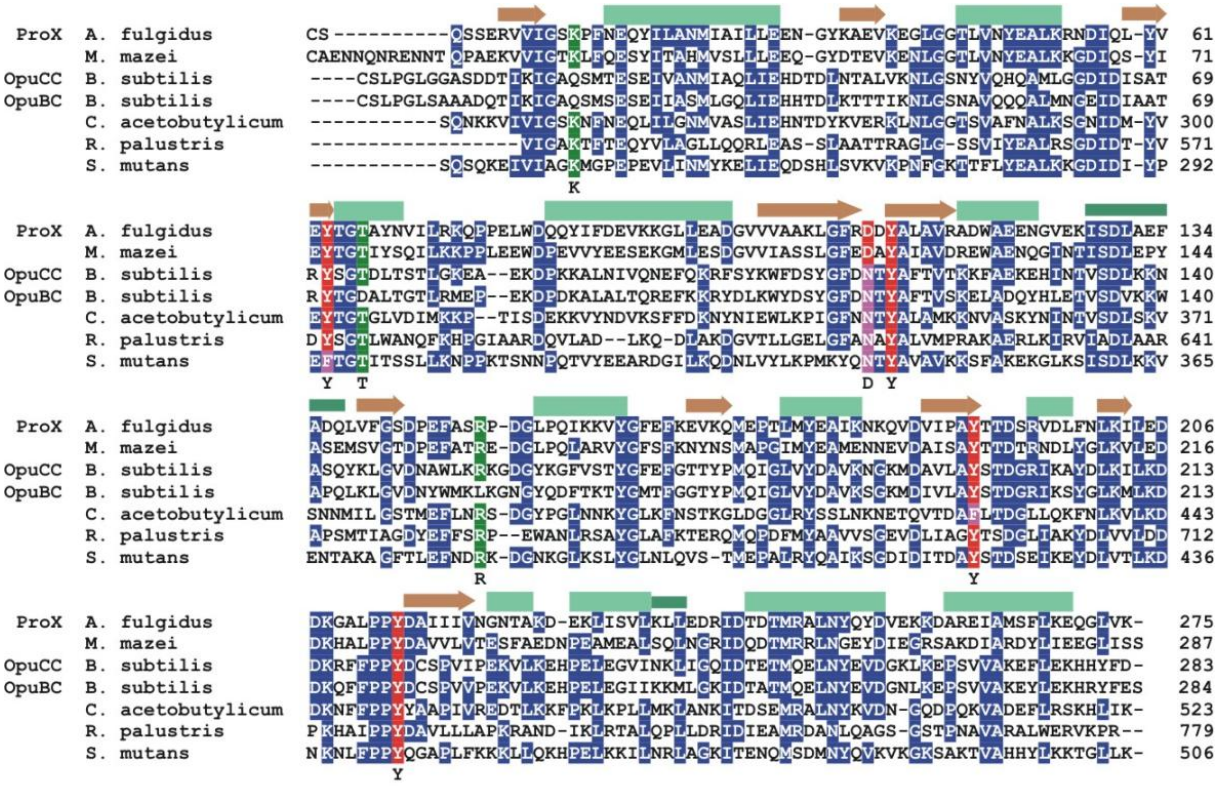

Figure 5.4: Sequence alignment between ProX and six related sequences *Archaeoglobus fulgidus*, *Methanosarcina mazei*, *Bacillus subtilis* OpuCC and OpuBC, *Clostridium acetobutylicum*, *Streptococcus mutans*, and *Rhodopseudomonas palustris*. Conserved residues are marked by blue background, conserved residues that are involved in binding of the quaternary amine head group are labelled red, non-conserved residues that can contribute in the same way are labelled violet. Residues which are involved in salt bridge or hydrogen bond formation with the ligand carboxylate of GB or PB are labelled green. Secondary structures determined for ProX are highlighted on top of the alignment. Helices are drawn as light-green boxes, 3_{10}-helices as dark-green slim boxes and sheets as brown arrows. All residues of ProX involved in ligand binding are marked at the bottom of the alignment.

the tyrosine residue is replaced by phenylalanine. A phenylalanine is supposed to be able to bind a quaternary amine in the same way as a tyrosine does. That has been shown by mutational studies of the binding pocket of ProX from *E. coli* (Schiefner et al., 2004). Therefore we expect the same mode of binding for those proteins. The fifth residue involved in binding of the ligand quaternary amine in ProX is Asp109. This residue is only conserved in *M. mazei*, in all other sequences it is replaced by an asparagine. Since this interaction is mediated via the main chain carbonyl oxygen, aspartic acid and asparagine are equivalent in this position. The residues of ProX interacting with the carboxylic tail of the ligand are conserved for all proteins except OpuBC and OpuCC. None of the three residues Lys13, Thr66 and Arg149 are conserved in OpuBC, the reason for this might be that OpuBC is optimized to bind choline instead of GB where the carboxylic group is reduced to an -CH_2OH group. On the other hand, in OpuCC, that binds choline and GB, only Lys13 is replaced by a glutamine. This single mutation seems to allow OpuCC to recognize both the carboxylic group of GB and the -CH_2OH group of choline.

5.4.7 Metal binding sites

In the anomalous difference density maps of the refined $P2_1$ crystal structures a total of 17 anomalous density peaks above 5 σ have been found in the asymmetric unit at λ = 1.0440 Å. Due to the crystallization conditions and the results of a fluorescence scan, the crystal lattice was likely to contain zinc or arsenic. To decide whether the anomalous scatterers bound by ProX are zinc or cacodylate ions, two datasets of the same crystal were collected at λ = 1.0440 Å and λ = 1.0507 Å. The six anomalous peaks above 7.5 σ were assigned to be zinc ions and have been modelled in the final structures in $P2_1$. In all cases the zinc ions are bound by carboxylic groups of glutamic acid or aspartic acid side chains. The rather high content of acidic residues (17.8 %) might be the reason for the unexpected metal binding affinity of ProX.
All these six zinc ions are involved in crystal contacts, especially three of them are tightly bound in the interface between the two molecules of the asymmetric unit, mediating the only crystal contact in this lattice dimension. One is located on the non-crystallographic symmetry axis bound by Glu238 and Asp239 of both molecules, the other two are coordinated by chain I Asp242, Asp244 and chain II Asp81 and vice versa.

5.4.8 Comparison of ProX from A.fulgidus and from E.coli

In the beginning of the structural analysis of ProX from *A. fulgidus* we noticed that an alignment based only on the sequences of ProX from *A. fulgidus* (*Afu*ProX) and ProX

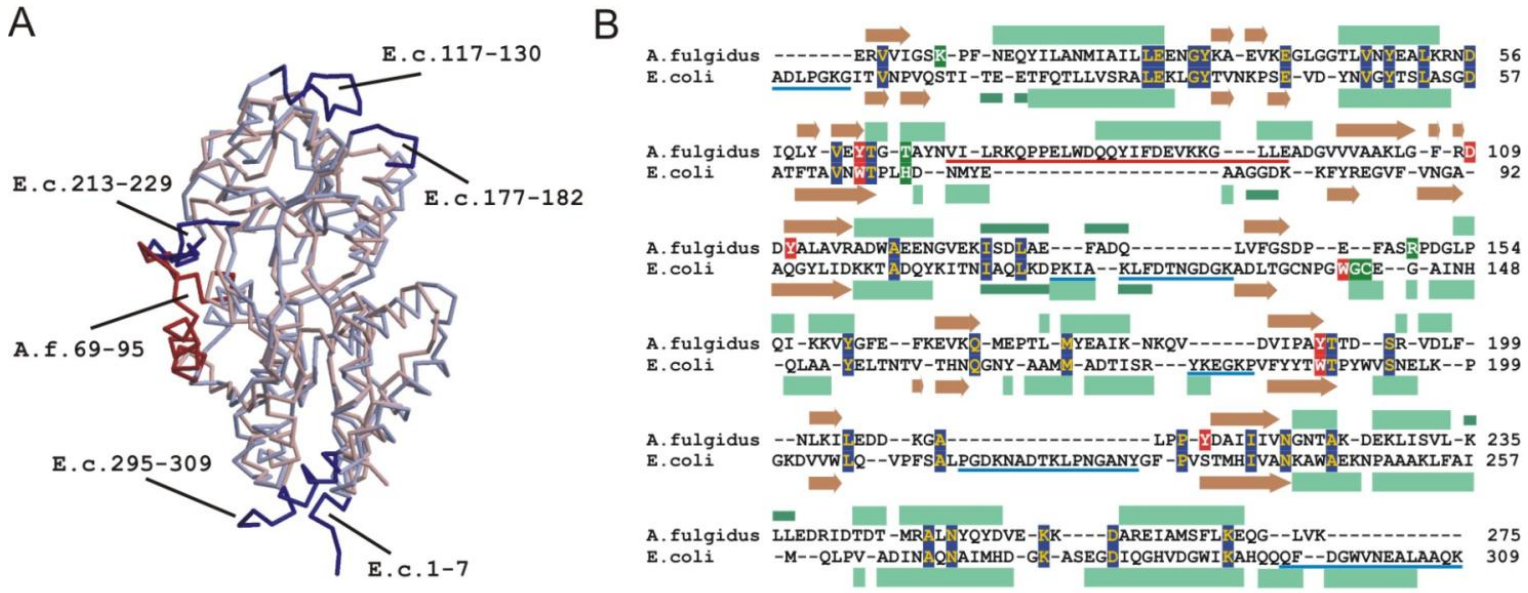

Figure 5.5: Differences in the overall ProX structures from *A. fulgidus* and *E. coli* in the closed conformation A) The structure of ProX from *E. coli* (in light-blue) superimposed onto the structure of ProX from *A. fulgidus* (in light-red). Larger differences with a series of five or more residues differing more than 3.5 Å in their Cα positions between the two structures are highlighted: ProX from *E. coli* shown in dark blue and ProX from *A. fulgidus* shown in dark red. Smaller variations are treated as local displacements of the polypeptide chain and are not separately marked. b) Structure based sequence alignment of ProX from *A. fulgidus* and from *E. coli*. Residues conserved in both sequences are labelled blue, residues involved in the binding of the quaternary amine are shown in red, and residues involved in hydrogen bond or salt bridge formation with the ligand carboxylate are labelled green. Secondary structure elements of the structures are drawn as light-green boxes for helices (dark-green slim boxes for 3_{10}-helices) and brown arrows for sheets. Large differences as defined in part A of the figure are underlined with the same color: *E. coli* in dark-blue and *A. fulgidus* in dark-red.

from *E. coli* (*Eco*ProX) could hardly be done properly. After the structural analysis it became clear that this is not only due to the low sequence identity but also to shifts of the primary sequence with respect to the domains.
The structures of *Afu*ProX and *Eco*ProX (pdb codes 1R9L and 1R9Q) were superimposed using a global-superposition-distance-based Needleman-Wunsch alignment as implemented in the program LSQMAN (Kleywegt and Jones, 1994a). All Cα postions within a cutoff distance of 3.5 Å were treated as matched. This alignment gave a total of 214 residues which are structurally equivalent, although only 31 residues are identical, resulting in an amino acid sequence identity of only 11.5 % for the 270 structurally defined residues of *Afu*ProX. Obviously, the fold of both molecules is the same but some parts of the polypeptide differ. Figure 5.5A shows the Cα trace presentation of *Afu*ProX in red and of *Eco*ProX in blue. The structure based sequence alignment is given in Figure 5.5B. At the first sight, the structure of the *Afu*ProX appears to be more compact than that of *Eco*ProX. The N-terminus of the *Eco*ProX structure is seven residues longer due to the fact that in the *Afu*ProX structure the first five residues are disordered and could not be modelled as discussed above. On the other hand, residues 295 to 309 of *Eco*ProX are missing in *Afu*ProX although they apparently have a stable α-helical conformation in *Eco*ProX. Particularly interesting is the difference in the hinge region shown on the left of Figure 5.5A. There are large loop regions in both molecules that do not match at all. These are mainly responsible for the difficulty to obtain a sequence alignment. In the case of *Afu*ProX this region including 27 residues is part of domain A and is stablized by α-helical- and turn-conformation. In contrast, in *Eco*ProX the loop composed of 14 residues is part of domain B and has mainly turn conformation as structural element. Additionally, two loops are elongated in *Eco*ProX, the first formed by the residues 124 - 130 is involved in the coordination of a metal ion of unknown function (Schiefner et al., 2004), the second loop, on the other hand, seems to be only a minor elongation. The 27 residues spanning extension of domain A in *Afu*ProX may have a function in stabilizing the open conformation, since it forms additional contacts with domain B in the open conformation. These contacts contribute to more than 2/5 to the buried surface in the open conformation as described above.

5.4.9 Comparison of the ligand binding sites

Although the general folds of *Afu*ProX and *Eco*ProX are the same, the problem of compatible solute binding has been solved in two different ways. Figure 5.6 shows the superposition of the two binding sites (*A. fulgidus* in red and *E. coli* in blue) with the protein side chains involved in the binding of the quaternary amine head groups of the

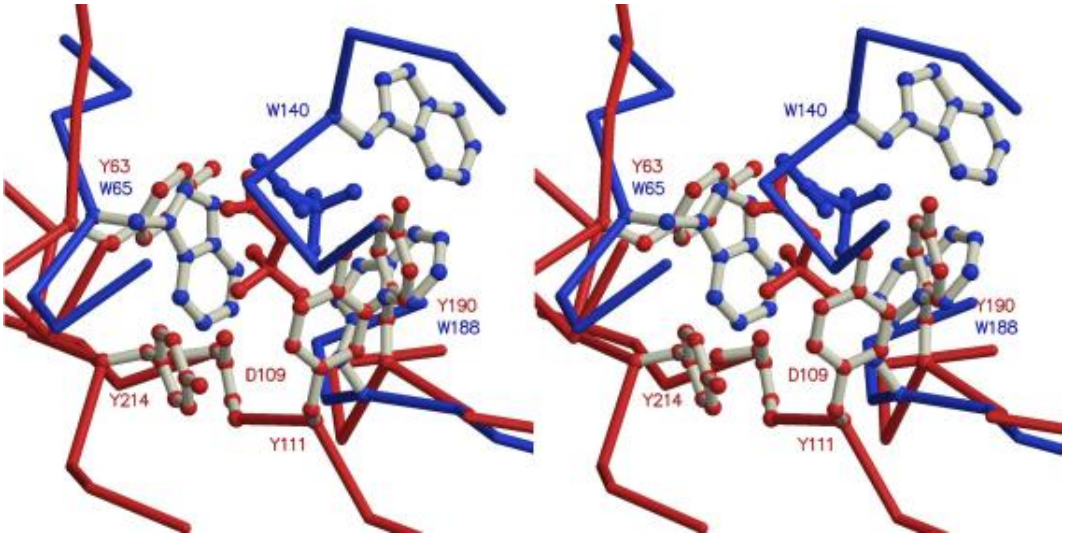

Figure 5.6: Differences in ligand binding between ProX from *A.fulgidus* and from *E. coli* Superposition of the ligand binding sites of ProX-GB from *E. coli* onto ProX-GB from *A. fulgidus*. All residues involved in binding of the quaternary amine are drawn as a ball-and-stick model with gray sticks. Ligands and the corresponding Cα-trace are colored blue for ProX from *E. coli* and red for ProX from *A. fulgidus*.

ligands. The two ligands are in a different orientation with respect to the polypeptide backbone, and the residue types involved in ligand binding are different as well. Whereas the quaternary amine is bound by four tyrosines and a main chain carbonyl oxygen in *Afu*ProX, this is achieved by three tryptophan residues in *Eco*ProX. Interestingly, two of the aromatic positions seem to be conserved (*Afu*Tyr63-*Eco*Trp65 and *Afu*Tyr190-*Eco*Trp188). However, they do not interact in the same way with the ligand. The similar spacing in the primary sequence of these two residues seems to be accidental because the large loop of 69-95 in *Afu*ProX is compensated by the two small loops 117-130 + 177-182 in *Eco*ProX (see Figure 5.5). The binding site is also organized differently. If the ligand is divided into quaternary amine head group and carboxylic tail, the differences become even more obvious. In *Afu*ProX the head group is bound by the four side chains of Tyr63, Tyr111, Tyr190, and Tyr214, and the main chain the carbonyl of Asp109. The tail is held in place only by side chain interactions with Lys13, Thr66, and Arg149. In contrast *Eco*ProX binds the head group only by side chain contacts of the three tryptophan residues Trp65, Trp140, and Trp188, whereas in binding of the tail both the main chain contacts (nitrogen atoms of Gly141 and Cys142) as well as side chain contacts (His69) are involved (Schiefner et al., 2004). From the sequence alignment of *Afu*ProX and the structural comparison between *Afu*ProX and *Eco*ProX we conclude that at least two classes of binding-proteins exist which bind compatible solutes of the

quaternary ammonium type. These are not distinct by their three-dimensional structure but in their way of binding the ligand.

5.4.10 Analysis of the thermophilic properties

There is no shift to higher secondary structure content in the thermophilic protein from *A. fulgidus* compared to the mesophilic one of *E. coli*. The secondary structure content is very similar or even slightly lower in *Afu*ProX with 56.7 % (18.2 % β-sheet and 38.5 % α-helix) compared to *Eco*ProX with 57.6 % (17.8 % β-sheet and 39.8 % α-helix).
The comparison of the amino acid compositions of *Afu*ProX and *Eco*ProX correlate quite well with the known differences of amino acid distributions between thermophiles and mesophiles. Kumar et al. (2000) identified in a statistical study the general trend that the number of amino acid residues that are capable to form both short and long range interactions are increased in thermophilic proteins. Therefore Arg and Tyr residues occur more frequently, while Ser and Cys residues are less frequent. A further trend toward thermostability is the increased occurrence of long side chains that are able to form salt bridges. Both of these trends are clearly found in the structure of *Afu*ProX, the percentage of Arg is increased from 1 to 3.6 % and that of Glu is increased from 4.2 to 9.5 %. Overall, the content of charged residues is increased by 10.8 %. Interestingly, there are no His residues in *Afu*ProX, although His could be a potential partner in salt bridges as well. The reason might be the fact that His is able to form salt bridges only in a relatively narrow pH range. A total of 19 salt bridges (13 H-bonded) have been found in *Afu*ProX compared to only eight salt bridges (five H-bonded) in *Eco*ProX. We observed an unexpected decrease in the number of tryptophan residues, although these residues should be favored by thermophilic proteins. This decrease might be due to a high tryptophan abundance in *Eco*ProX where three tryptophan residues are used to bind the quaternary amine of the ligand.

Acknowledgements

Financial support for this study was provided by the Deutsche Forschungsgemeinschaft through SFB 395, the Graduiertenkolleg "Proteinfunktion auf atomarer Ebene", the TR-SFB 11 (Konstanz-Zürich), and the Fonds der Chemischen Industrie (to E.B.). We also thank Clemens Schulze-Briese and Takashi Tomizaki for their SLS support and Kinga Gerber for careful reading of the manuscript.

6 AcrB from Escherichia coli

6.1 Abstract

A His-tagged derivative of the multidrug efflux pump AcrB could be crystallized in three different space groups (R3, R32 and P321). Experimental MAD-phasing maps from R32 $AcrB_{His}$ crystals were obtained to a resolution of 3.5 Å. Datasets of native and substrate soaked $AcrB_{His}$ crystals were collected at the Swiss Light Source X06SA beamline up to a resolution of 2.7 Å and refinement of these data provided good quality electron density maps, which allowed us to complement the published AcrB structure (PDB code 1IWG). Introduction of amino acids 860 to 865 and 868 lacking in the 1IWG structure and deletion of a highly disordered region (amino acids 669-678) improved R_{free} and average B-factors in the 2.7 Å model. We could not identify significant densities indicating specific antibiotic binding sites in the AcrB R32 space group datasets under the soaking conditions tested.

6.2 Introduction

Bacteria have evolved 3 general mechanisms of resistance towards antibiotics: (i) resistance by target modification; (ii) resistance by antibiotic modification and (iii) resistance through the action of antibiotic efflux pumps. Membrane protein pumps responsible for the latter type of resistance can be classified as (i) ABC-type antibiotic resistance pumps, which utilize the free energy of ATP hydrolysis to energize the efflux of toxic compounds over the membrane or (ii) secondary antibiotic efflux pumps which use the electrochemical gradient of ions across the membrane in an antiport modus to energize efflux processes.

Secondary antibiotic/multiple drug resistance transporters can be divided into four families (Busch and Saier Jr., 2002; Busch and Saier Jr., 2003): (i) Major Facilitator Superfamily (MFS); (ii) Small Multidrug Resistance family (SMR); Multidrug And Toxic compound Extrusion family (MATE) and (iv) Resistance Nodulation cell Division superfamily (RND, TC# 2.A.6). Members of the RND superfamily involved in the

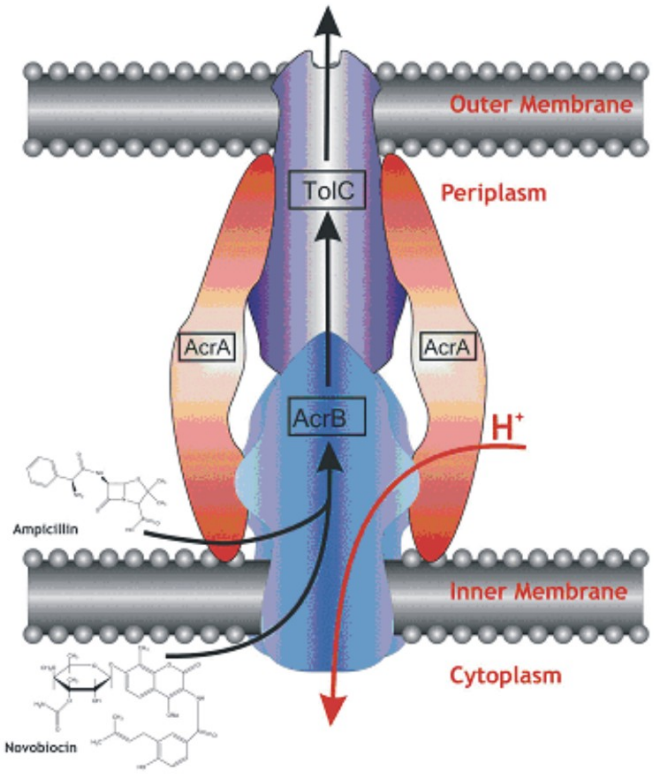

Figure 6.1: Hypothetical structure of tri-partite efflux pumps found in Gram-negative bacteria *e.g.* the AcrAB-TolC efflux pump from *E. coli* (adapted from Nikaido and Zgurskaya, 2001).

transport of antibiotics belong to the family of the (largely Gram-negative bacterial) Hydrophobe/Amphiphile Efflux-1 (HAE1) Family (TC# 2.A.6.2). The proton motive force-driven RND/HAE1 drug/H^+ antiporters usually function within a tripartite system including an outer membrane channel (outer membrane factor family (OMF, TC# 1.B.17)) and a member of the membrane fusion protein family (MFP, TC# 8.A.1) (Figure 6.1). The drug efflux is believed to occur via the initial binding of the drug to the RND component and subsequent transport through the OMF. The role of the MFP is still puzzling, although it has been shown that it is absolutely essential for the function of the whole pump (Ma et al., 1993; Zgurskaya and Nikaido, 1999a; Zgurskaya and Nikaido, 1999b). Besides the well-studied RND-MFP-OMF systems in *Pseudomonas aeruginosa* (e.g. MexAB-OprM and MexCD-OprJ)(Poole, 2001), the AcrAB-TolC tripartite system is currently the most studied due to the structural information on the AcrB and TolC components (Yu et al., 2003; Murakami et al., 2002; Koronakis et al., 2000). The substrate specificity of the AcrAB-TolC system is rather broad: anionic, cationic and zwitterionic as well as neutral compounds are transported, including fluoroquinolones, macrolides and phenylpropanoids. Most of the substrates are hydrophobic and can tra-

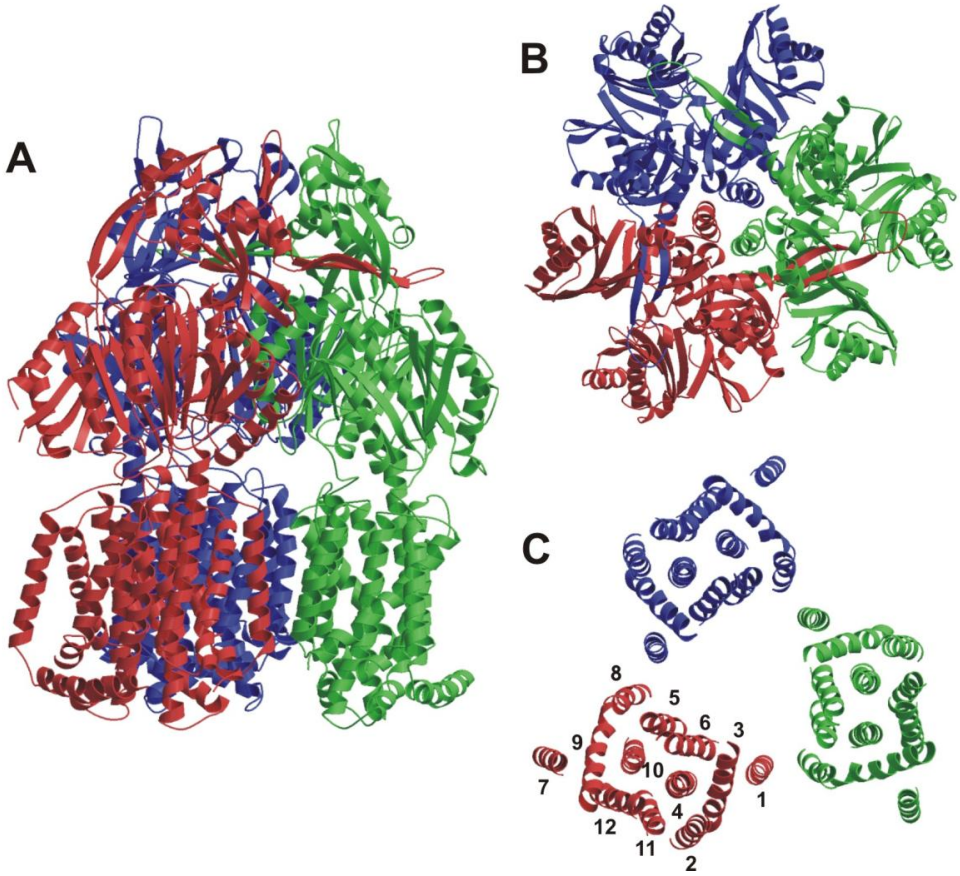

Figure 6.2: Structural model of AcrB (A) Backbone ribbon diagram side view of AcrB trimer; (B) Topology of the transmembrane helices top view along the crystallographic three-fold axis. The three monomers enclose a 30 - 35 Å wide cavity (C) Backbone periplasmic top view along the crystallographic three-fold axis. The long loops protruding from one monomer into the next provide the main interaction within the AcrB trimer. The protomers are individually colored. Figures are MolScript drawings based on the coordinates deposited in the pdb databank entry 1IWG (Murakami et al., 2002).

verse the inner membrane. A notable exception are the β-lactams transported by the system (Nikaido and Zgurskaya, 2001). TolC has been structurally characterized to a resolution of 2.1 Å by crystallographic means (Koronakis et al., 2000). Apart from the membrane-embedded regions (about 4 nm thickness), a TolC trimer has a 10 nm periplasmic barrel domain composed of twelve α-helices, which form a long conduit through which the drugs might be transported. The crystallization and structural analysis of AcrB has been described recently (Murakami et al., 2002; Pos and Diederichs, 2002). Like TolC, AcrB forms a homotrimer, which is believed to be its functional unit. Each AcrB monomer contains twelve membrane spanning α-helices and a large periplasmic domain (Figure 6.2A). The transmembrane α-helices in the AcrB trimer are organized in a way that these encircle a 35 Å wide cavity, which is believed to be

filled with phospholipids *in vivo* (Figure 6.2B). Within each monomer, transmembrane helices 4 and 10, encircled by the other transmembrane helices of the monomer, harbor the residues Lys 940 (helix 10) and Asp 407 and 408 (helix 4), which are postulated to be important for the H^+ translocation (Figure 6.2B and Murakami et al., 2002). The periplasmic domain can be divided into two major parts: The TolC docking domain and the pore domain. The TolC docking domain exhibits a funnel-like structure narrowing to the central pore located in the pore domain. The internal diameter on the top of the funnel is about the same as the diameter of the TolC periplasmic conduit. The central pore structure consists of three α-helices, one donated by each AcrB monomer (Figure 6.2C). Another remarkable feature is a long loop protruding from one monomer through the neighboring monomer. These loops appear to make the main and almost only interaction between the monomers. At the proximal end of the central pore, the structure opens up to a central cavity, leading to the cavity in the transmembrane part. Moreover, three entrances or vestibules reaching towards the central cavity are located just above the membrane plane. From the AcrB native structure a drug transport model has been postulated (Murakami et al., 2002; Murakami and Yamaguchi, 2003) which involves the transport of membrane permeable drugs from the inner leaflet of the cytoplasmic membrane through a groove formed by transmembrane helices 8 and 9 towards the central cavity. β-lactam antibiotics are believed to enter the central cavity via the vestibules. Once in the central cavity, the substrates are believed to be transported into the external medium through the central pore (which has to open first), through the funnel part of the AcrB periplasmic domain and finally through TolC . Currently, the elucidation of the molecular mechanism(s) of transport of the wide variety of substrates is a major challenge. The substrate binding and specificity has already been addressed by structural and functional studies. It has been shown that the periplasmic part of the RND component is responsible for the substrate specificity of the whole tripartite system. This has been demonstrated for AcrB, MexB, MexD, and MexY (Tikhonova et al., 2002; Eda et al., 2003; Elkins and Nikaido, 2002). Additionally, structural analysis of binding of antibiotics to AcrB (based on 3.5 - 3.8 Å data) has been reported recently (Yu et al., 2003). Binding of four AcrB substrates, rhodamine G6, ethidium, dequalinium and ciprofloxacin is observed in the central cavity. For all substrates except ciprofloxacin, Phe 386 (helix 3) was one of the main hydrophobic contacts, in fact in the case of ethidium it seemed to be the only one. For dequalinium, only the top quinolinium moiety was close to the charged residues Asp 99 and Asp101 but the bottom quinolinium moiety was close to only Phe386. For ciprofloxacin, Phe458 and 459 (helix 5) were the main ligands. Binding of three substrate molecules per trimer was observed as enforced by the three-fold symmetry of the R32 space group. Another observation was an 1 °

outward tilting of the periplasmic domain, postulated to be induced by the substrate binding (Yu et al., 2003). We report here on the crystallographic analysis of diffraction data obtained (up to 2.7 Å) with crystals from native $AcrB_{His}$, a selenomethionine substituted derivative and substrate soaked derivatives of $AcrB_{His}$.

6.3 Material and Methods

6.3.1 Crystal optimization and SeMet substitution

Expression of $acrB_{His}$, preparation of membranes, purification and initial crystallization of $AcrB_{His}$ are described in Pos and Diederichs (2002). To obtain SeMet substituted AcrB, *E. coli* C43(DE3)/ pET24$acrB_{His}$ was grown in methionine deprived minimal media including selenomethionine. Yields were 6 mg of SeMet substituted $AcrB_{His}$ from 1L of minimal media culture. SeMet crystals (200 x 200 x 100 μm^3) were grown using 0.1 M NaHEPES pH 7.5, 10 - 60 mM sodium chloride and 8 %(w/v) PEG4000 as precipitant and protein at 39 mg ml^{-1} in a hanging drop (1:2 dilution) experimental setup over a 0.5 ml precipitant solution in the reservoir. Crystals which belong to the space group R32 were frozen in liquid ethane using glycerol as cryoprotectant. $AcrB_{His}$ R3 crystals were obtained using 17 mg ml^{-1} protein diluted 1:2 in 0.1 M NaHEPES pH 7.5, 50 mM sodium chloride, 5 %(v/v) glycerol and 5 %(w/v) PEG4000 as precipitant in a hanging drop experimental setup over a 0.5 ml precipitant solution in the reservoir. P321 crystals (about 100 x 100 x 10 μm^3) were obtained using 70 mM sodium citrate, pH 4.6, 8 %(v/v) glycerol and 16 %(v/v) PEG400 as precipitant (1:2 dilution, hanging drop). For freezing these crystals, solid glucose was added to the droplet containing the crystal and allowed to saturate the drop in a stepwise manner. Large single native $AcrB_{His}$ (R32) crystals with dimensions up to 1000 x 800 x 400 μm^3 were obtained by adding isopropyl-β- small D-thiogalactopyranoside, n-decanoylsucrose, n-nonyl-β-D-thiomaltoside or n-octyl-β-D-thiomaltoside (at 1 %(w/v), 2.5 mM, 3.2 mM and 9 mM final concentration, respectively). $AcrB_{His}$ was used at 16.8 mg ml^{-1}. Crystals were transferred in 11 steps to 0.1 M NaHEPES pH 7.5, 5 %(w/v) PEG4000, 0.1 M sodium chloride and 0.05 %(w/v) CHM solutions containing 2-30 %(v/v) glycerol in 5 to 10 minutes intervals. After the final soaking step at 30 %(v/v) glycerol, the crystals were directly frozen into liquid propane or ethane and stored in liquid nitrogen. In soaking experiments, crystals were treated using the same procedure in the presence of 0.5 to 5 mM bile salt or antibiotic.

6.3.2 X-ray diffraction data set analysis and refinement procedure

Datasets from native crystals and heavy atom derivatives were collected at the DESY, Hamburg, EMBL beamlines BW7A, BW7B and X11. Native and substrate co-crystals as well as the MAD data were measured either at beamline X06SA of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland) or at beamline ID29 of the ESRF (Grenoble, France). Data reduction was done with the XDS Software package (Kabsch, 1993). The selenomethionine substructure of the SeMet derivative crystals was solved with SHELXD (Schneider and Sheldrick, 2002). A total of 35 selenium sites out of 42 theoretically possible sites could be detected. Further refinement of the heavy atom positions was carried out with the program SHARP (Bricogne et al., 2003). Solvent flattening with RESOLVE (Terwilliger, 2001) resulted in interpretable maps at a resolution of 3.5 Å. The structures in the other space groups were solved by molecular replacement using MOLREP (Vagin and Teplyakov, 1997). As search models 1IWG and a modified version of 1IWG (model2, see results) were used. All structures were refined with the program REFMAC5 (Murshudov et al., 1997) starting with 20 cycles of rigid body refinement with four rigid domains consisting of residues (i) 1-30 and 334-530, (ii) 31-333, (iii) 561-859 and (iv) 531-560 and 860-1031, followed by 20 cycles of restrained refinement. Model building was done using the program 'O' (Jones et al., 1991). Figure 6.2 was created using MOLSCRIPT (Kraulis, 1991) and Figures 6.3 and 6.4 were created using DINO (www.dino3d.org).

6.4 Results and Discussion

6.4.1 Diffraction quality of $AcrB_{His}$ crystals

The first crystals of $AcrB_{His}$ in spacegroup R32 were obtained in a screening procedure using the MembFac screen of Hampton (solution 30) and were rather sensitive to radiation damage (Pos and Diederichs, 2002). The initial cryoprotection procedures led to cracking of larger crystals while freezing them in liquid nitrogen. These problems could be overcome by the optimization of the crystallization conditions and freezing protocol. Crystals grown in the presence of n-nonyl-β-D-thiomaltoside and n-octyl-β-D-thiomaltoside were transferred into cryoprotection solutions containing 1 mM erythromycin or novobiocin, respectively in the case of substrate binding studies. After the final soaking step, the crystals were directly frozen in liquid propane or ethane and stored in liquid nitrogen. This procedure improved the diffraction of the crystals to a resolution of 2.7 and 2.8 Å, for the native and substrate soaked crystals, respectively. AcrB could be crystallized in two other space groups, R3 and P321. R3 crystals were

Table 6.1: Data collection and crystallographic analysis

	SeMet					
	peak	**inflection**	**remote**	**native**	**native**	**native**
Space group		R32		R32	R3	P321
Cell axes [Å]		a = 146.2		a = 143.7	a = 145.4	a = 133.1
		c = 520.0		c = 513.9	c = 519.2	c = 192.2
Wavelength [Å]	0.9792	0.9791	0.9756	0.9774	0.9796	0.9790
Resolution [Å]	3.5	3.5	3.5	2.65	3.2	3.5
No. unique reflections	52275	52273	50259	59390	66801	25414
Completeness [%]	99.9	99.9	96.1	99.2	99.0	99.9
I/σI	7.0	7.4	9.4	15.5	5.4	10.3
R_{sym} [%]	30.1	32.4	23.3	8.2	9.9	19.8
$R_{mrgd\text{-}F}$* [%]	24.1	23	19.8	7.3	21.5	16.8
Phasing power						
(anomalous, acentric)	1.6	0.9	1.5			
Figure of merit		0.47				

* $R_{mrgd\text{-}F}$ is a measure for the quality of the reduced amplitudes (Diederichs and Karplus, 1997; Weiss and Hilgenfeld, 1997)

obtained using the same crystallization conditions as described for R32 crystals, but with addition of 5 %(v/v) glycerol in the mother liquor and 2.5 %(v/v) glycerol in the initial protein droplet condition. These crystals (350 x 350 x 300 μm^3) diffracted to 3.0 Å resolution. P321 crystals (about 100 x 100 x 10 μm^3) were obtained using 70 mM sodium citrate pH 4.6, 8 %(v/v) glycerol and 16 %(v/v) PEG400 as precipitant. These crystals were frozen using glucose as cryoprotectant. One dataset has been collected to a resolution of 3.5 Å.

6.4.2 Structure solution and refinement

In order to obtain phase information, R32 crystals were soaked with various heavy atom compounds. Despite good datasets (to 3.0 Å resolution), phases could not be determined due to severe non-isomorphism. Especially the c-axis of the crystal unit cell varied drastically (between 490 and 520 Å), which made direct comparison of the datasets impossible. We therefore produced crystals of SeMet substituted $AcrB_{His}$ and collected MAD data (Table 6.1). After determining phases using the programs SHELXD, SHARP and RESOLVE, interpretable electron density maps were obtained at 3.5 Å resolution. During the process of model-building, Murakami et al. (2002) published a structure of AcrB based on 3.5 Å diffraction data. Their model (PDB database entry 1IWG) was used in the initial refinement procedures with our diffraction data. The refinement statistics listed in Table 6.2 clearly demonstrate that our experimental data from native and substrate soaked crystals produced models of AcrB with high R_{free} values after rigid

Table 6.2: Refinement statistics after rigid body and restrained refinement

Crystal name	Resolution range [Å]	R-factor [%]	R_{free} [%]	r.m.s.d. bonds [Å]	r.m.s.d. angles [°]	Average B-factor [Å^2]
M1 (R3 native)						
1IWG	19.7-3.0	36.2	41.8	0.013	1.48	61.6
model2	19.7-3.0	36.0	41.1	0.013	1.42	61.6
M209 (native)						
1IWG	19.5-2.65	34.1	37.2	0.009	1.19	100.7
model2	182.5-2.65	34.5	36.4	0.010	1.27	67.5
M210 (native)						
1IWG	19.8-3.0	37.0	39.0	0.011	1.32	110.0
model2	19.8-3.0	35.6	37.0	0.013	1.40	74.7
M237 (native)						
1IWG	46.6-3.0	33.5	36.2	0.011	1.32	102.8
model2	46.6-3.0	32.9	35.0	0.011	1.32	76.3
M238 (native+IPTG)						
1IWG	46.6-2.9	37.3	39.8	0.011	1.32	118.7
model2	46.6-2.9	35.6	38.7	0.012	1.37	91.5
M198 (Hoechst33342)						
1IWG	47.7-2.9	37.6	39.7	0.010	1.28	109.0
model2	47.7-2.9	34.7	38.4	0.011	1.29	85.4
M215 (taurocholate)						
1IWG	47.1-3.1	40.1	42.6	0.011	1.32	112.2
model2	47.1-3.1	37.6	39.4	0.013	1.41	90.9
M216 (taurocholate)						
1IWG	46.6-3.0	35.5	39.3	0.011	1.32	114.7
model2	46.6-3.0	33.5	37.7	0.011	1.34	88.4
M222 (lomefloxacin)						
1IWG	47.7-3.1	40.0	41.4	0.013	1.41	112.6
model2	47.7-3.1	38.8	40.3	0.015	1.50	95.1
M227 (erythromycin)						
1IWG	46.6-2.9	34.0	39.9	0.011	1.28	98.1
model2	46.6-2.9	34.3	40.1	0.011	1.26	73.4
M228 (novobiocin)						
1IWG	47.7-2.8	33.1	35.8	0.010	1.27	105.6
model2	47.7-2.8	33.9	36.8	0.010	1.26	81.6
M240 (erythromycin)						
1IWG	46.1-3.1	36.1	40.2	0.011	1.36	124.7
model2	46.1-3.1	33.4	37.8	0.013	1.42	104.1
M241 (erythromycin)						
1IWG	47.0-3.0	33.8	37.7	0.012	1.42	109.3
model2	47.1-3.0	34.8	37.1	0.013	1.43	92.9

body and restrained refinement. The highest R_{free} values were observed with data from a R32 crystal soaked with taurocholate (M215, 3.1 Å, R_{free} = 42.6 %) and with the native R3 crystal dataset (M1, 3.0 Å, R_{free} = 41.8 %). The best R_{free} values were obtained with the native crystal dataset M237 (3.0 Å, R_{free} = 36.2 %) and the data from the novobiocin soaked crystal M228 (2.8 Å, R_{free} = 35.8 %). In comparison, the 1IWG structure published by Murakami et al. (2002) yielded an R_{free} value of 35.5 % (at 3.5 Å) and the unliganded structure 1OY6 published by Yu et al. (2003) showed an R_{free} of 33.0 % at 3.7 Å. The average B-factor from the refined structures based on 1IWG are in almost all cases above 100 $Å^2$, with the notable exception of the M1 3.0 Å dataset in R3 (61.6 $Å^2$). High average B-factors were also found for the structures 1IWG (Murakami et al., 2002) and 1OYE (Yu et al., 2003). Due to disorder, the 1IWG AcrB model lacks regions 1-6, 499-512, 860-868 and 1037-1049 (Murakami et al., 2002). Inspection of our Fo-Fc and 2Fo-Fc maps obtained with datasets M1, M209, M227 and M228 (Table 6.2) revealed the presence of interpretable density in the region of residues 860-868 (Figure 6.3A and 6.3B). We amended the model in this region by implementing amino acid residues 860 to 865 and 868. Furthermore, we corrected the side chain location of Glu112 (pore region) and His596 and deleted amino acids 669-678 due to poor electron density. The modified model (designated model2) resulted in slightly better R_{free} factors in the refinement cycle compared to the same refinement done with 1IWG (Table 6.2). The best R_{free}-factor was obtained with the datasets from native R32 crystals using this modified model. Refinement against data from crystal M237 yielded an R-factor of 32.9 % and R_{free} of 35.0 %. The electron density in the omitted region 669-678 after refinement with model2 is difficult to interpret and requires further analysis before an alternate chain progression can be determined. Remarkably, the B-factors dropped approximately 10 to 30 % compared to the refined 1IWG structure (Table 6.2).

6.4.3 Substrate binding studies

We have tested several compounds (erythromycin, novobiocin, taurocholate, taurodeoxycholate, lomefloxacin, ciprofloxacin, deoxycholate, chloramphenicol and Hoechst 33342) for binding to $AcrB_{His}$ by crystal soaking experiments and occasionally by co-crystallization. All data obtained from compound soaked crystals resulted in higher R-factors, with the exception of M228 (novobiocin soak, R_{free} = 35.8 % with 1IWG and 36.8 % with model2). Very high R-factors were obtained with the lomefloxacin and erythromycin soaked crystal datasets M222 and M227 (= 40%). In comparison, Yu et al. (2003) reported almost equal R_{free} factors for the unliganded and liganded structures (32 - 34 %) at 3.5 to 3.8 Å. Our experimental setup differs from that described in Yu et al. (2003)

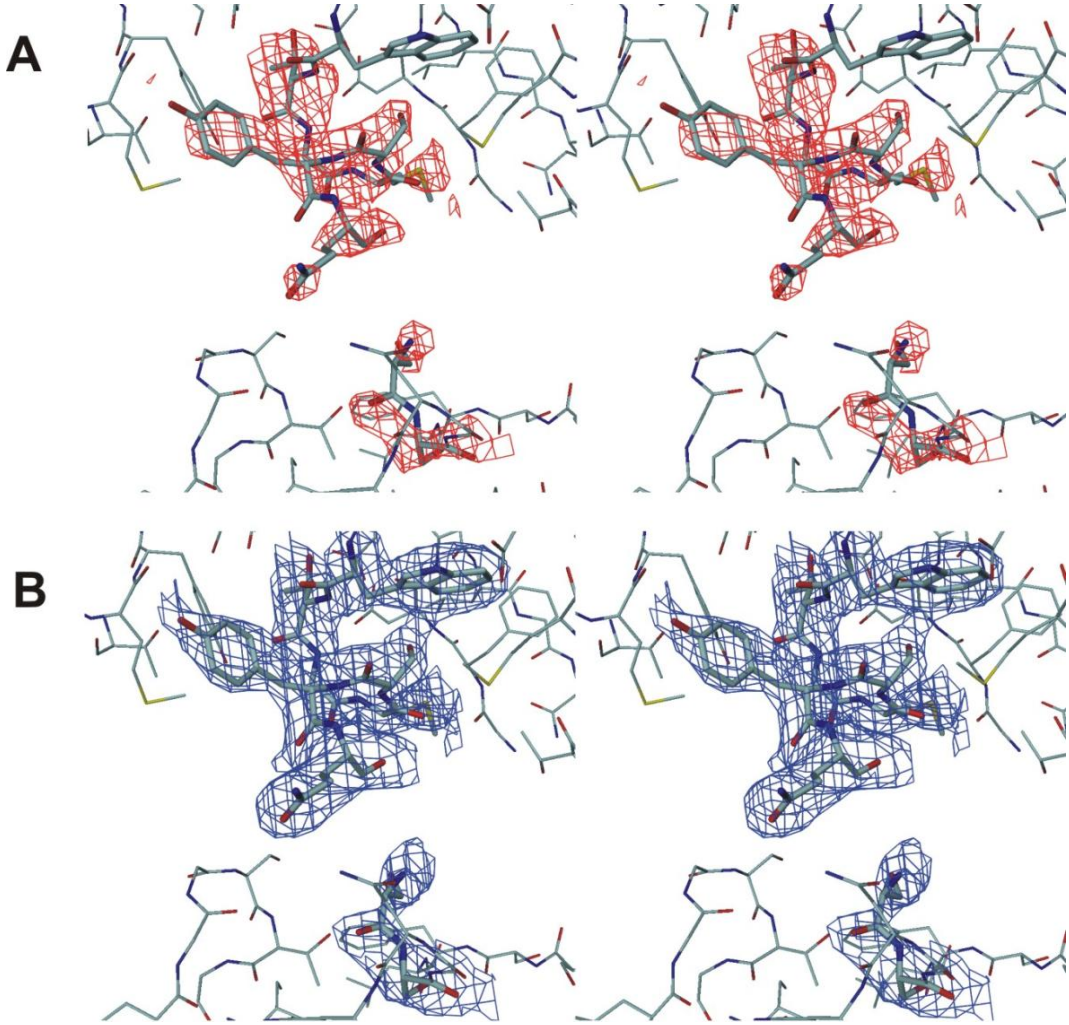

Figure 6.3: Stereo view displaying the region 860-868 A) the Fo-Fc map at 2.7 Å contoured at 3 σ in red and B) 2Fo-Fc electron density map at 2.7 Å contoured at 1 σ in blue. Shown in bold sticks from top to bottom are the residues Trp859 and the amended residues Thr860, Gly861, Met862, Ser863, Tyr864, Gln865, Leu868 and Ser869.

by (i) the presence of a His-tag at the C-terminus of the protein, (ii) a higher pH of the soaking buffer (7.5 versus 5.6 and 6.6) and (iii) the use of CHM (0.05 %) rather than DDM (0.1 %) as detergent. Furthermore, in the determination of drug binding sites, Yu et al. (2003) used electron density omit maps directly after data scaling and molecular replacement, before doing any refinement. We calculated the Fo-Fc and 2Fo-Fc maps (Figure 6.4) after molecular replacement and refinement (rigid body and restrained refinement with REFMAC5). Inspection of Fo-Fc difference maps using our substrate soaked crystal datasets did not reveal any significant positive density features different from the ones observed with the native datasets (Figure 6.4). The 2Fo-Fc maps of M209 (native), M227 (erythromycin soaked crystal) and M228 (novobiocin soaked crystal) at

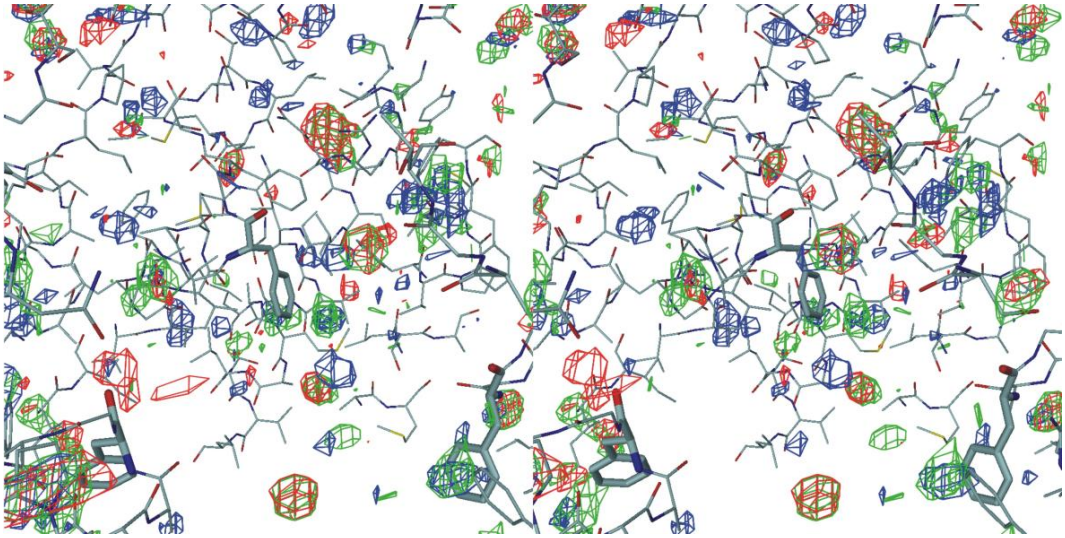

Figure 6.4: Trimer central cavity Region near Phe386 (bold stick): Stereo view of Fo-Fc electron density maps at 2.9 Å contoured at 3 σ: blue: M209 (native), green: M227 (erythromycin soaked), red: M228 (novobiocin soaked).

2.7 to 2.8 Å all showed similar density features (not shown). Despite the good resolution of the density maps (2.8 Å), we were not able to identify difference density features that could be attributed to bound ligands. A possible reason for this is the low quality of the phases derived from a high R-factor model. We expect that a future improvement of the AcrB model (with lower R_{free}) might lead to the unambiguous identification of substrate binding sites, allowing a deeper understanding of the functional aspects of AcrB.

6.4.4 Acknowledgement

We like to thank Clemens Schulze-Briese, Takashi Tomizaki and all other members of the X06SA beamline at the SLS (Paul Scherrer Institute, Villigen, Switzerland) for their excellent support.

7 Summary

This thesis reports several crystal structures of proteins involved in active transport across bacterial membranes. The majority of the proteins are parts of binding protein-dependent ABC importers but one them is an exporter, which is involved in resistance formation of bacteria against antibiotics.

ABC transporters form the largest family of homologous transport proteins. Representatives of this family have been found in all investigated organisms. Therefore, it is of general interest, how those transport systems work. Although good progress has been made in the field within the last years, the molecular mechanism of transport is not completely understood.

The focus of the first part of the presented work was directed on the structural determination of an intact binding protein-dependent ABC transporter. For this purpose the trehalose/maltose transporter $MalFGK_2$ from the hyperthermophilic archaeon *Thermococcus litoralis* has been chosen. This is distinguished from its well studied homologue $MalFGK_2$ from *E. coli* by its optimum of activity at 85 °C. Thermophilic proteins or complexes are expected to be less flexible at room temperature, making them more suitable for crystallization. However, the best crystals obtained during this doctoral work diffracted only to a resolution of 5 Å, which is not sufficient to determine the atomic structure.

Furthermore, in the context of the work, the atomic structures of three binding proteins from three different organisms have been determined. Thereby new insights have been gained as to how binding proteins are adapted to specific tasks and conditions.

MalE from *Alicyclobacillus acidocaldarius* was studied in order to understand its acido- and thermostability. Since the Gram-positive bacterium *A. acidocaldarius* usually grows optimally under fairly extreme conditions as pH 3.6 and 57 °C, all of its non-cytoplasmic proteins, which are exposed to high temperatures as well as low pH, have to be adapted to work optimally without being damaged over a longer period of time. Compared to MalE homologues from non-acidophilic bacteria and archaea it turned out that above all the number of charged residues is reduced on the protein surface. Although the number of positively and negatively charged residues is almost the same, substantially more

positively charged residues are exposed. The resulting positive surface charge likely contributes to the acidostability.

In the context of this work special attention was payed on the high affinity binding of compatible solutes by binding proteins. Compatible solutes, like the quaternary ammonium compounds glycine betaine and proline betaine, are characterized by their property to be excluded from protein surfaces. This property enables cells to accumulate high concentrations of these compounds without affecting structure and function of their proteins at the same time. Furthermore, the resulting non-uniform distribution of compatible solutes within the cell has stabilizing effects on the structure of proteins. To understand, what enables binding proteins to bind compatible solutes with high affinity, the structures of ProX from *Escherichia coli* and of ProX from *Archaeoglobus fulgidus* have been determined. It turned out that aromatic side chains like those of tryptophan and tyrosine in a defined sterical arrangement can perform this task. ProX from *E. coli* interacts with the positive charge of the quaternary amine by three tryptophan side chains which are approximately perpendicularly oriented to each other. In ProX from *A. fulgidus* a similar architecture is found but with different sterical features. There are no tryptophan side chains but tyrosines and a main chain oxygen involved in the binding of the quaternary amine. In both cases the binding is mediated by a combination of cation-π interactions and non-classical hydrogen bonds.

The last part of the presented work concentrated on structural investigations of the multi-drug transporter AcrB from *E. coli* that is instrumentally involved in the formation of resistances in pathogenic Gram-negative bacteria. The major goal of this work was to improve the quality of the existing AcrB model and the localization of substrate binding sites by the structural determination of AcrB-substrate complexes. Although the AcrB model has been improved its quality still seems to be not sufficient to localize any substrate binding sites.

8 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurden die Strukturen von verschiedenen Proteinen bestimmt, die am aktiven Transport über bakterielle Membranen beteiligt sind. Bei der Mehrzahl dieser Proteine handelt es sich um Teile von Bindeprotein-abhängigen ABC importern. Ein weiteres ist ein Exporter, der an der Resistenzbildung von Bakterien gegen Antibiotika beteiligt ist.
ABC transporter stellen die größte Familie homologer Transportproteine dar. Bislang wurden Vertreter dieser Familie in allen näher untersuchten Organismen gefunden. Es ist daher von allgemeinem Interesse, wie diese Transportsysteme funktionieren. Trotz großer Fortschritte auf diesem Gebiet, innerhalb der letzten Jahre, ist der genaue Transportmechanismus noch nicht aufgeklärt.
Der erste Teil der Arbeit war auf die Bestimmung der Struktur eines intakten Bindeprotein-abhängigen ABC transporters gerichtet. Hierfür wurde der Trehalose/Maltose transporter $MalFGK_2$ aus dem hyperthermophilen Archaeon *Thermococcus litoralis* ausgewählt. Dieser zeichnet sich gegenüber seinem gut studierten Homologen $MalFGK_2$ aus *Escherichia coli* dadurch aus, dass er bei 85 °C sein Aktivitätsoptimum hat. Erwartungsgemäß sind solche Proteine oder Komplexe bei Raumtemperatur weniger flexibel, wodurch sie sich eher für Kristallisationsversuche eignen. Die besten Kristalle, die innerhalb der Doktorarbeit erhalten wurden, beugten nur bis 5 Å, was nicht ausreicht um eine atomare Struktur zu berechnen.
Desweiteren wurden im Rahmen dieser Arbeit die atomaren Strukturen von drei Bindeproteinen aus drei verschiedenen Organismen aufgeklärt. In diesem Zusammenhang konnten neue Erkenntnisse darüber gewonnen werden, wie Bindeproteine an spezifische Aufgaben und Gegebenheiten angepasst sind.
MalE aus *Alicyclobacillus acidocaldarius* wurde untersucht, um seine Säure- und Hitzestabilität zu verstehen. Da das Gram-positive Bakterium *A. acidocaldarius* gewöhnlich unter recht extremen Bedingungen, wie pH 3.6 and 57 °C optimal wächst, müssen all seine nicht-cytoplasmatischen Proteine, die sowohl hohen Temperaturen als auch dem niedrigen pH des extrazellulären Milieus ausgesetzt sind, so angepasst sein, dass sie optimal arbeiten können und auf längere Sicht keinen Schaden nehmen. Im Vergleich

mit MalE Homologen aus Bakterien und Archaea stellte sich heraus, dass vor allem die Zahl der geladenen Reste auf der Proteinoberfläche reduziert ist. Und obwohl die Zahl der positiv und negativ geladenen Reste ungefähr gleich ist, so sind doch erheblich mehr positiv geladene exponiert. Daraus resultiert eine positive Oberflächenladung, die wahrscheinlich zur Säurestabilität beiträgt.
Ein besonderes Interesse, im Rahmen dieser Arbeit, wurde der hochaffinen Bindung von compatiblen Soluten durch Bindeproteine gewidmet. Compatible Solute, wie die quaternären Ammoniumverbindungen Glycin Betain und Prolin Betain, zeichnen sich dadurch aus, dass sie nicht direkt mit Proteinoberflächen wechselwirken können. Das ermöglicht es Zellen diese Verbindungen zu sehr hohen Konzentrationen anzureichern, ohne damit gleichzeitig Struktur und Funktion ihrer Proteine zu beeinträchtigen. Vielmehr wirkt sich die resultierende nicht-gleichförmige Verteilung compatibler Solute innerhalb der Zelle sogar stabilisierend auf die Struktur von Proteinen aus. Um zu verstehen, wie es Bindeproteinen dennoch möglich ist compatible Solute mit hoher Affinität zu binden, wurden die Strukturen von ProX aus *Escherichia coli* und ProX aus *Archaeoglobus fulgidus* aufgeklärt. Es stellte sich heraus, dass aromatische Aminosäureseitenketten, wie die von Tryptophan und Tyrosin, in einer definierten räumlichen Anordnung diese Aufgabe erfüllen können. ProX aus *E. coli* wechselwirkt mit der positive Ladung des quaternären Amins durch drei Tryptophanseitenketten, die annäherd rechtwinkligig zueinander orientiert sind. Sehr ähnlich, aber in räumlich veränderter Anordnung, erfolgt die Bindung compatibler Solute durch ProX aus *A. fulgidus*. Diesmal sind es nicht Tryptophanseitenketten, sondern Tyrosine und ein Hauptkettensauerstoff, die an der Bindung des quaternären amins beteiligt sind. In beiden Fällen wird die Bindung durch eine Kombination aus Kation-π Wechselwirkung und nicht-klassischen Wasserstoffbrücken vermittelt.
Der letzte Teil der vorliegenden Arbeit konzentrierte sich auf strukturelle Untersuchungen, des Multi-Medikamenten Transporters AcrB aus *E. coli*, der maßgeblich an der Resistenzbildung pathogener Gram-negativer Bakterien beteiligt ist. Das Hauptziel dieser Arbeit war die Verbesserung der Qualität des existierenden AcrB Modells und die Lokalisierung von Substratenbindestellen durch die Bestimmung von AcrB-Substratkomplexen. Obwohl das AcrB Modell verbessert werden konnte, erscheint seine Qualität immer noch nicht ausreichend, um irgendeine Substratbindestelle lokalisieren zu können.

9 List of Publications

Schiefner, A., Diederichs, K., Hashimoto, K., Boos, W., and Welte, W. (2002) Crystallization and preliminary X-ray analysis of the trehalose/maltose ABC transporter $MalFGK_2$ from *Thermococcus litoralis*. *Acta Cryst* D58:2147-2149

Schiefner, A., Breed, J., Bösser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W., and Bremer, E. (2004) Cation-π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *J Biol Chem* 279:5588-5596

Schäfer, K., Magnusson, U., Scheffel, F., **Schiefner, A.**, Sandgren, M. O. J., Diederichs, K., Welte, W., Hülsmann, A., Schneider, E., and Mowbray, S. L. (2004) X-ray structures of the maltose-maltodextrin-binding protein of the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* provide insight into acid stability of proteins. *J Mol Biol* 335:261-274

Gerber, K., **Schiefner, A.**, Seige, P., Diederichs, K., Boos, W., and Welte, W. (2004) Crystallization and preliminary X-ray analysis of Aes, an acetyl-esterase from *Escherichia coli*. *Acta Cryst* D60:531-533

Pos, K. M., **Schiefner, A.**, Seeger, M. A., and Diederichs, K. (2004) Crystallographic analysis of AcrB. *FEBS Lett* 564:333-339

Schiefner, A., Holtmann, G., Diederichs, K., Welte, W., Bremer, E. Cation-π interactions and non-classical hydrogen bonds determine the binding of compatible solutes by ProX from the hyperthermophilic archaeon *Archeoglobus fulgidus*. *Manuscript submitted*

References

Albers, S. V., Elferink, M. G., Charlebois, R. L., Sensen, C. W., Driessen, A. J., and Konings, W. N. (1999). Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. *J Bacteriol*, 181:4285–4291. 3.2

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*. Garland publishing, N. Y., 4 edition.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J Mol Biol*, 215:403–410. 3.3.4, 4.3.5

Altschul, S. F. and Gish, W. (1996). Local alignment statistics. *Methods Enzymol*, 266:460–480. 5.3.4

Arakawa, T. and Timasheff, S. (1985). The stabilization of proteins by osmolytes. *Biochem J*, 47:411–414. 4.2, 4.5.1, 5.2

Bakker, E. P. (1990). The role of alkali-cation transport in energy coupling of neutrophilic and acidophilic bacteria: an assessment of methods and concepts. *FEMS Microbiol Rev*, 75:319–334. 3.2, 3.5

Barlow, D. and Thornton, J. (1983). Ion-pairs in proteins. *J Mol Biol*, 168:867–885. 5.4.3

Barron, A., Jung, J., and M., V. (1987). Purification and characterization of a glycine betaine binding protein from *Escherichia coli*. *J Biol Chem*, 262:11841–11846. 4.2

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2003). GenBank. *Nucleic Acids Res*, 31:23–27. 3.3.4

Berman, H., Westbrock, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., and Bourne, P. (2000). The Protein Data Bank. *Nucleic Acid Research*, 28:235–242. 3.3.3, 3.3.4, 4.4.2, 5.3.4

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. T., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112:535–542. 3.3.4

Björkman, A. and Mowbray, S. (1998). Multiple open forms of ribose-binding protein trace the path of its conformational change. *J Mol Biol*, 279:651–664. 1.3.3, 4.2, 5.2

Böhm, A., Diez, J., Diederichs, K., Welte, W., and Boos, W. (2002). Structural model of MalK, the ABC subunit of the maltose transporter of *Escherichia coli*. *J Biol Chem*, 277:3708–3717. 1.3.3

Bolen, D. and Baskakov, I. (2001). The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol*, 310:955–963. 4.2, 5.2

Boos, W. and Lucht, J. (1996). Periplasmic binding protein-dependent ABC transporters. In Neidhard, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M. Schaechter, M., and Umbarger, H. E., editors, *Escherichia coli and Salmonella Cellular and molecular biology*, volume 1, pages 1175–1209. American Society for Microbiology Press, Washington, D. C. 1.3.3, 4.2, 5.2

Boos, W. and Shuman, H. (1998). The maltose/maltodextrin system of *E. coli*; transport, metabolism, and regulation. *Microbiol Mol Biol Rev*, 62:204–229. 1.3.3, 2.2

Booth, I. and Louis, P. (1999). Managing hypoosmotic stress: aquaporins and mechanosensitive channels in *Escherichia coli*. *Curr Opin Microbiol*, 2:166–169. 4.2, 5.2

Borths, E., Locher, K., Lee, A., and Rees, D. (2002). The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proc Natl Acad Sci USA*, 99:16642–16647. 4.4.1

Bourot, S., Sire, O., Trautwetter, A., Touze, T., Wu, L., Blanco, C., and Bernard, T. (2000). Glycine betaine-assisted protein folding in a lysA mutant of *Escherichia coli*. *J Biol Chem*, 275:1050–1056. 4.2, 5.2

Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J., and Hilgenfeld, R. (2001). CH$\cdots\pi$ interactions in proteins. *J Mol Biol*, 307:357–377. 5.4.4, 5.4.5

Breed, J., Kneip, S., Gade, J., Welte, W., and Bremer, E. (2001). Purification, crystallization and preliminary crystallographic analysis of the periplasmic binding protein ProX from *Escherichia coli*. *Acta Cryst*, D57:448–450. 4.3.1, 4.3.5

Brejc, K., van Dijk, W. J., Klaassen, R. V., Schuurmans, M., van der Oost, J., Smit, A. B., and Sixma, T. K. (2001). Crystal structure of an Ach-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, 411:269–276. 4.5.3

Bremer, E. and Krämer, R. (2000). Coping with osmotic challenges: osmoregulation through accumulation and release of compatible solutes in bacteria. In Storz, G. and Hengge-Aronis, R., editors, *Bacterial stress responses*, pages 79–97. American Society for Microbiology Press, Washington, D. C. 4.2, 5.2

Bremer, E., Silhavy, T. J., and Weinstock, G. M. (1988). Transposition of λp*lac*Mu is mediated by the A protein altered at its carboxy-terminal end. *Gene*, 71:177–186. 4.3.2

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Cryst*, D59:2023–2030. 6.3.2

Brigulla, M., Hoffmann, T., Krisp, A., Volker, A., Bremer, E., and Volker, U. (2003). Chill induction of the SigB-dependent general stress response in *Bacillus subtilis* and its contribution to low-temperature adaptation. *J Bacteriol*, 185:4305–4314. 5.2

Brown, A. D. (1976). Microbial water stress. *Bacteriol Rev*, 40:803–846. 4.2, 5.2

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Cryst*, D54:905–921. 3.3.3

Burg, M., Kwon, E., and Kültz, D. (1997). Regulation of gene expression by hypertonicity. *Ann Rev Physiol*, 59:437–455. 4.2, 5.2

Busch, W. and Saier Jr., M. H. (2002). The transporter classification (TC) system, 2002. *Crit Rev Biochem Mol Biol*, 37:287–337. 6.2

Busch, W. and Saier Jr., M. H. (2003). The IUBMB-endorsed transporter classification system. *Methods Mol Biol*, 227:21–36. 6.2

Caldas, T., Demont-Caulet, N., Ghazi, A., and Richarme, G. (1999). Thermoprotection by glycine betaine and choline. *Microbiology*, 145:2543–2548. 4.2, 5.2

Canovas, D., Borges, N., Vargas, C., Ventosa, A., Nieto, J. J., and Santos, H. (1999). Role of N-γ-acetyldiaminobutyrate as an enzyme stabilizer and an intermediate in the biosynthesis of hydroxyectoine. *Appl Environ Microbiol*, 65:3774–3779. 4.2, 5.2

Canovas, D., Fletcher, S. A., Hayashi, M., and Csonka, L. N. (2001). Role of trehalose in growth at high temperature of *Salmonella enterica* serovar *typhimurium*. *J Bacteriol*, 183:3365–3371. 4.2, 5.2

CCP4 (1994). The CCP4 suite: Programs for protein crystallography. *Acta Cryst*, D50:760–763. 3.3.2, 3.3.3, 5.2

Chang, G. (2003). Structure of MsbA from *Vibrio cholera*: a multidrug resistance ABC transporter homolog in a closed conformation. *J Mol Biol*, 330:419–430. 1.3.2

Chang, G. and Roth, C. B. (2001). Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP cassette (ABC) transporters. *Science*, 293:1793–1800. 1.3.2

Chen, H. L., Gabrilovich, D., Tampé, R., Girgis, K. R., Nadaf, S., and Carbone, D. P. (1996). A functionally defective allele of TAP1 results in loss of MHC class I antigen presentation in a human lung cancer. *Nature Genet*, 13:210–213. 1.3.1

Chen, J., Lu, G., Lin, J., Davidson, A. L., and Quiochio, F. A. (2003). A Tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol Cell*, 12:651–661. 1.3.2, 1.4, 1.3.3

Chen, J., Sharma, S., Quiocho, F. A., and Davidson, A. L. (2001). Trapping the transition state of an ATP-binding-cassette transporter: evidence for a concerted mechanism of maltose transport. *Proc Natl Acad Sci USA*, 98:1525–1530. 1.3.3

Clarke, T. E., Ku, S.-Y., Dougan, D. R., Vogel, H. J., and Tari, L. W. (2000). The structure of the ferric siderophore binding protein FhuD complexed with gallichrome. *Nat Struct Biol*, 7:287–291. 4.4.1

Cobley, J. G. and Cox, J. C. (1983). Energy conservation in acidophilic bacteria. *Microbiol Rev*, 47:579–595. 3.5

Cordon-Cardo, C., O'Brien, J. P., Casals, D., Rittman-Grauer, L. Biedler, J. L., Melamed, M. R., and Bertino, J. R. (1989). Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites. *PNAS*, 86:695–698. 1.3.1

Courtenay, E. S., Capp, M. W., Anderson, C. F., and Record Jr., M. T. (2000). Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants *in vivo* and for the interpretation of "osmotic stress" experiments *in vitro*. *Biochemistry*, 39:4455–4471. 5.2

Cowtan, K. (1994). DM. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, 31:34–38. 4.3.5

Csonka, L. N. and Epstein, W. (1996). Osmoregulation. In Neidhard, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors, *Escherichia coli and Salmonella Cellular and molecular biology*, volume 1, pages 1210–1223. American Society for Microbiology Press, Washington, D. C. 4.2, 5.2

Csonka, L. N. and Hanson, A. D. (1991). Prokaryotic osmoregulation: genetics and physiology. *Ann Rev Microbiol*, 45:569–606. 4.2

Culham, D., Lasby, B., Marangoni, A., Milner, J., Steer, B., van Nues R.W., and Wood, J. (1993). Isolation and sequencing of *Escherichia coli* gene *proP* reveals unusual structural features of the osmoregulatory proline/betaine transporter, ProP. *J Mol Biol*, 229:268–276. 4.2

da Costa, M. S., Santos, H., and Galinski, E. A. (1998). An overview of the role and diversity of compatible solutes in *Bacteria* and *Archaea*. *Adv Biochem Eng Biotechnol*, 61:117–153. 4.2, 5.2

Darland, G. and Brock, T. D. (1971). *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore-forming bacterium. *J Gen Microbiol*, 67:9–15. 3.2

de la Fortelle, E. and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth Enzymol*, 276:472–494. 5.3.4

Derewenda, Z., Lee, L., and Derewenda, U. (1995). The occurence of C-H$\cdots$O hydrogen bonds in proteins. *J Mol Biol*, 252:248–262. 5.4.4

Diederichs, K., Diez, J., Greller, G., Müller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W., and Welte, W. (2000). Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. *EMBO J*, 19:5951–5961. 1.3.3, 2.2

Diederichs, K. and Karplus, P. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, 4:269–275. 4.1, 5.1, 6.1

Diez, J., Diederichs, K., Greller, G., Horlacher, R., Boos, W., and Welte, W. (2001). The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis* at 1.85 Å. *J Mol Biol*, 305:905–915. 1.3.3, 3.2, 3.3.4

Diruggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W., and Weiss, R. B. (2000). Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol Microbiol*, 38:684–693. 3.4.3

Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol*, 276:523–530. 5.3.2

Dougherty, D. A. (1996). Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science*, 271:163–168. 5.4.4

Dulaney, E. L., Dulaney, D. D., and Rickes, E. L. (1968). Factors in yeast extract which relieve growth inhibition of bacteria in defined medium of high osmolarity. *Dev Ind Microbiol*, 9:260–269. 4.3.1, 5.3.2

Eda, S., Maseda, H., and Nakae, T. (2003). An elegant means of self-protection in gram-negative bacteria by recognizing and extruding xenobiotics from the periplasmic space. *J Biol Chem*, 278:2085–2088. 6.2

Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol*, 284:489–502. 3.5

Elkins, C. A. and Nikaido, H. (2002). Substrate specificity of the RND-type multidrug efflux pumps AcrB and AcrD of *Escherichia coli* is determined predominantly by two large periplasmic loops. *J Bacteriol*, 184:6490–6498. 6.2

Elkins, C. A. and Nikaido, H. (2003). 3D structure of AcrB: the archetypal multidrug efflux transporter of *Escherichia coli* likely captures substrates from periplasm. *Drug Resis Updat*, 6:9–13. 1.4.2

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graphics Mod*, 15:132–134. 4.3.5

Evdokimov, A. G., Anderson, D. E., Routzahn, K. M., and Waugh, D. S. (2001). Structural basis for oligosaccharide recognition by *Pyrococcus furiosus* maltodextrin-binding protein. *J Mol Biol*, 305:891–904. 1.3.3, 3.2, 3.3.4, 3.4.4, 3.4.6

Faatz, E., Middendorf, A., and Bremer, E. (1988). Cloned structural genes for the osmotically regulated binding-protein-dependent glycine betaine transport system (ProU) of *Escherichia coli K-12*. *Mol Microbiol*, 2:265–279. 4.2

Felder, C. E., Harel, M., Silman, I., and Sussman, J. L. (2002). Structure of a complex of the potent and specific inhibitor BW284C51 with *Torpedo californica* acetylcholineesterase. *Acta Cryst*, D58:1765–1771. 4.4.6

Fetsch, E. E. and Davidson, A. L. (2002). Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. *Proc Natl Acad Sci USA*, 99:9685–9690. 4.2

Fushinobu, S., Ito, K., Konno, M., Wakagi, T., and Matsuzawa, H. (1998). Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: biased distribution of acidic residues and importance of Asp37 for catalysis at low pH. *Protein Eng*, 11:1121–1128. 3.5

Galinski, E. A. and Trüper, H. G. (1994). Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol Rev*, 15:95–108. 4.2, 5.2

Gaudet, R. and Wiley, D. C. (2001). Structure of the ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J*, 20:4964–4972. 2.2

Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M., and Chothia, C. (1993). Domain closure in lactoferrin: two hinges produce a see-saw motion between alternative close-packed interfaces. *J Mol Biol*, 234:357–372. 5.4.3

Ghoul, M., Bernard, T., and Cormier, M. (1990). Evidence that *Escherichia coli* accumulates glycine betaine from marine sediments. *Appl Environ Microbiol*, 56:551–554. 5.2

Gouesbet, G., Jebbar, M., Talibart, R., Bernard, T., and Blanco, C. (1994). Pipecolic acid is an osmoprotectant for *Escherichia coli* taken up by the general osmoporters ProU and ProP. *Microbiol*, 140:2415–2422. 1.3.3

Gowrishankar, J. (1989). Nucleotide sequence of the osmoregulatory proU operon of *Escherichia coli*. *J Bacteriol*, 171:1923–1931. 4.2, 5.2

Greller, G., Riek, R., and Boos, W. (2001). Purification and characterization of the heterologously expressed trehalose/maltose ABC transporter complex of the hyperthermophilic archaeon *Thermococcus litoralis*. *Eur J Biochem*, 268:4011–4018.

Guex, N. and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18:2714–2723. 3.3.4

Haardt, M., Kempf, B., Faatz, E., and Bremer, E. (1995). The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Mol Gen Genet*, 246:783–786. 1.3.3, 4.2, 4.3.1, 4.3.2

Harel, M., Schalk, I., Ehret-Sabatier, L., Bouet, F., Goeldner, M., Hirth, C., Axelsen, P. H., Silman, I., and Sussman, J. L. (1993). Quaternary ligand binding to aromatic residues in the active site of acetylcholinesterase. *Proc Natl Acad Sci USA*, 90:9031–9035. 4.4.6

Harris, M. and Jones, T. A. (2001). Molray - a web interface between O and the POV-Ray ray tracer. *Acta Cryst*, D57:1201–1203. 3.3.4

Hart, P., Warth, J. D., Levesque, P. C., Collier, M. L., Geary, Y., Horowitz, B., and Hume, J. R. (1996). Cystic fibrosis gene encodes a cAMP-dependent chloride channel in heart. *PNAS*, 93:6343–6348. 1.3.1

Hernando, N., Forster, I. C., Biber, J., and Murer, H. (2000). Molecular characteristics of phosphate transporters and their regulation. *Exp Nephrol*, 8:366–375. 1.4.1

Herrmann, A., Schlösser, A., Schmid, R., and Schneider, E. (1996). Biochemical identification of a lipoprotein with maltose-binding activity in the thermoacidophilic Gram-positive bacterium *Alicyclobacillus acidocaldarius*. *Res Microbiol*, 147:733–737. 3.3.1, 3.4.1, 3.4.3, 3.5

Higgins, C. F. (1992). ABC transporters: from microorganisms to man. *Ann Rev Cell Biol*, 8:67–113. 4.2

Higgins, C. F., Sutherland, L., Cairney, J., and Booth, I. R. (1987). The osmotically regulated proU locus of *Salmonella typhimurium* encodes a periplasmic betaine-binding protein. *J General Microbiol*, 133:305–310. 4.2

Hohmann, S. (2002). Osmotic adaptation in yeast-control of the yeast osmolyte system. *Int Rev Cytol*, 215:149–187. 4.2, 5.2

Holland, I. B. and Blight, M. A. (1999). ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol*, 293:381–399. 2.2

Hopfner, K.-P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*, 101:789–800. 1.3.2

Hopfner, K.-P. and Tainer, J. A. (2003). Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr Opin Struct Biol*, 13:249–255. 1.3.2

Hülsmann, A., Lurz, R., Scheffel, F., and Schneider, E. (2000). Maltose and maltodextrin transport in the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius* is mediated by a high-affinity transport system that includes a maltose binding protein tolerant to low pH. *J Bacteriol*, 182:6292–6301. 3.2, 3.4.1, 3.5

Hung, L. W., Wang, I. X., Nikaido, K., Liu, P. Q., Ames, G. F. L., and Kim, S.-H. (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*, 396:703–707. 1.3.3, 2.2

Jacobs, S. A. and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, 295:2080–2083. 4.4.6

Jancarik, J. and Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Cryst*, 24:409–411. 2.3.2

Jebbar, M., Talibart, R., Gloux, K., Bernard, T., and Blanco, C. (1992). Osmoprotection of *Escherichia coli* by ectoine: uptake and accumulation characteristics. *J Bacteriol*, 174:5027–5035. 4.2

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst*, A47:110–119. 3.3.3, 4.3.5, 5.3.4, 6.3.2

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst*, 26:795–800. 2.3.2, 3.3.2, 4.3.5, 5.3.4, 6.3.2

Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637. 3.3.4, 3.4.6, 4.6, 5.3.4, 5.4.3

Karpowich, N., Martsinkevich, O., Millen, L., Yuan, Y., Dai, P. L., Macvey, K., Thomas, P. J., and Hunt, J. F. (2001). Crystal structures of the MJ1267 ATP-binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure Fold Des*, 9:571–586. 2.2

Kempf, B. and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high osmolality environments. *Arch Microbiol*, 170:319–330. 4.2, 5.2, 5.4.6

Kendrick, B. S., Chang, B. S., Arakawa, T., Peterson, B., Randolph, T. W., Manning, M. C., and Carpenter, J. F. (1997). Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: Role in restricted conformational mobility and compaction of native state. *Proc Natl Acad Sci USA*, 94:11917–11922. 4.5.1

Klein, I., Sarkadi, B., and Váradi (1999). An inventory of the human ABC proteins. *Biochim Biophys Acta*, 1461:237–262. 1.3.1

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Venter, J. C., et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, 390:364–370. 5.2, 5.3.2

Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst*, D52:842–857. 3.3.4

Kleywegt, G. J. and Jones, T. A. (1994a). A super position. *Joint CCP4 and ESF-EACBM newsletter on protein crystallography*, 31:9–14. 4.4.6, 4.7, 5.4.8

Kleywegt, G. J. and Jones, T. A. (1994b). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst*, D50:178–185. 3.3.4, 3.4.6

Kleywegt, G. J. and Jones, T. A. (1996). Phi/Psi-cology: Ramachandran revisited. *Structure*, 4:1395–1400. 3.1

Kleywegt, G. J. and Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol*, 277:525–545. 3.3.4

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*, 405:914–919. 6.2, 6.2

Kraulis, P. J. (1991). MolScript: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst*, 24:946–950. 3.3.4, 4.3.5, 5.3.4, 6.3.2

Krulwich, T. A., Davidson, L. F., Filip Jr., S. J., Zuckerman, R. S., and Guffanti, A. A. (1978). The protonmotive force and beta-galactoside transport in *Bacillus acidocaldarius*. *J Biol Chem*, 253:4599–4603. 3.5

Kumar, S. and Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J Mol Biol*, 293:1241–1255. 5.4.3

Kumar, S. and Nussinov, R. (2001). How do thermophilic proteins deal with heat? *Cell Mol Life Sci*, 58:1216–1233. 3.5

Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors enhancing protein thermostability. *Prot Engineering*, 13:179–191.

Ladenstein, R. and Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv Biochem Eng Biotechnol*, 61:37–85. 3.5

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst*, 26:283–291. 4.3.5, 5.3.4

le Rudulier, D., Strøm, A. R., Dandekar, A. M., Smith, L. T., and Valentine, R. C. (1984). Molecular biology of osmoregulation. *Science*, 224:1064–1068. 4.2, 5.2

Lee, J. C. and Timasheff, S. N. (1981). The stabilization of proteins by sucrose. *J Biol Chem*, 256:7193–7201. 4.5.1

Lehmann, M. S. and Zaccai, G. (1984). Neutron small-angle scattering studies of Ribonuclease in mixed aqueous solutions and determination of the preferentially bound water. *Biochemistry*, 23:1939–1942. 4.2, 4.5.1, 5.2

Li, A.-J. and Nussinov, R. (1998). A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins*, 32:111–127. 4.4.2, 5.4.4

Linton, K. J. and Higgins, C. F. (1998). The *Escherichia coli* ATP binding cassette (ABC) proteins. *Mol Microbiol*, 28:5–13. 1.3.1

Lippert, K. and Galinski, A. A. (1992). Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl Microbiol Biotechnol*, 37:61–65. 4.2, 5.2

Locher, K. P., Lee, A. T., and Rees, D. C. (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, 296:1091–1098. 1.3.2, 1.3.3

Lu, G. (1998). Fit. http://bioinfo1.mbfys.lu.se/ guoguang/fit.html. 3.3.4, 3.4.3

Lucht, J. M. and Bremer, E. (1994). Adaptation of *Escherichia coli* to high osmolarity environments: osmoregulation of the high-affinity glycine betaine transport system ProU. *FEMS Microbiol Lett*, 14:3–20. 4.2, 5.2

Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., and Hearst, J. E. (1993). Molecular cloning and characterization of acrA and acrE genes of *Escherichia coli*. *J Bacteriol*, 175:6299–6313. 6.2

Magnusson, U., Chaudhuri, B. N., Ko, J., Park, C., Jones, A., and Mowbray, S. L. (2002). Hinge-bending motion of D-allose-binding protein from *Escherichia coli*. *J Biol Chem*, 277:14077–14084. 1.3.3, 4.2, 5.2

Mao, B., Pear, R., McCammon, J. A., and Quiocho, F. A. (1982). Hinge-bending in L-arabinose-binding protein. The "venus's-flytrap" model. *J Biol Chem*, 257:1131–1133.

Martin, D. D., Ciulla, R. A., and Roberts, M. F. (1999). Osmoadaptation in archaea. *Appl Environ Microbiol*, 65:1815–1825. 5.2

Matzke, J., Herrmann, A., Schneider, E., and Bakker, E. P. (2000). Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes. *FEMS Microbiol Lett*, 183:55–61. 3.2

Matzke, J., Schwermann, B., and Bakker, E. P. (1997). Acidostable and acidophilic proteins: the example of the α-amylase from *Alicyclobacillus acidocaldarius*. *Comp Biochem Physiol*, 118:475–479. 3.5

May, G., Faatz, E., Villarejo, M., and Bremer, E. (1986). Binding protein dependent transport of glycine betaine and its osmotic regulation in *Escherichia coli* K12. *Mol Gen Genet*, 205:225–233. 4.2, 4.4.4

McNeil, S. D., Nuccio, M. L., and Hanson, A. D. (1999). Betaines and related osmoprotectants. Targets for metabolic engineering of stress resistance. *Plant Physiol*, 120:945–950. 5.2

Mecozzi, S., West Jr., A. P., and Dougherty, D. (1996). Cation-π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc Nat Acad Sci USA*, 93:10566–10571. 4.5.1

Mendum, M. L. and Smith, L. T. (2002). Characterization of glycine betaine porter I from *Listeria monocytogenes* and its roles in salt and chill tolerance. *Appl Environ Microbiol*, 68:813–819. 4.2, 5.2

Merrit, E. A. and Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. *Meth Enzymol*, 277:505–524. 4.3.5, 5.3.4

Metzler, D. E. (2001). *Biochemistry: The chemical reactions of living cells*, volume 1. Academic Press, CA, USA, 2 edition.

Miller, J. H. (1992). *A short course in bacterial genetics. A laboratory manual and handbook for Escherichia coli and related bacteria*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. 5.3.1, 5.3.2

Miller III., D. M., Olson, J. S., Pflugrath, J. W., and Quiocho, F. A. (1983). Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J Biol Chem*, 258:13665–13672. 5.4.5

Minton, A. (2001). The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem*, 276:10577–10580. 4.5.1

Momma, K., Mikami, B., Mishima, Y., Hashimoto, W., and Murata, K. (2002). Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of *Sphingomonas* sp. A1 at 2.0 Å resolution. *J Mol Biol*, 316:1051–1059. 4.4.3

Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, 419:587–593. 6.2, 6.2, 6.4.2

Murakami, S. and Yamaguchi, A. (2003). Multidrug-exporting secondary transporters. *Curr Opin Struct Biol*, 13:443–452. 6.2

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst*, D53:240–255. 3.3.3, 4.3.5, 5.3.4, 6.3.2

Navaza, J. and Saludjian, P. (1997). AMoRe: an automated molecular replacement program package. *Methods Enzymol*, 276:581–594. 3.3.3

Neu, H. C. and Heppel, L. A. (1965). The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *J Biol Chem*, 240:3685–3692. 4.3.1

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*, 10:1–6. 3.3.4

Nikaido, H., Basina, M., Nguyen, V., and Rosenberg, E. Y. (1998). Multidrug efflux pump AcrAB of *Salmonella typhimurium* excretes only those β-lactam antibiotics containing lipophilic side chains. *J Bacteriol*, 180:4686–4692. 1.4.2

Nikaido, H. and Zgurskaya, H. I. (2001). AcrAB and related multidrug efflux pumps of *Escherichia coli*. *J Mol Microbiol Biotechnol*, 3:215–218. 6.2

Oh, B.-H., Kang, C.-H., de Bondt, H., Kim, S.-H., Nikaido, K., Joshi, A. K., and Ames, G. F.-L. (1994). The bacterial periplasmic histidine-binding protein. *J Biol Chem*, 269:4135–4134. 1.3.3

Oh, B.-H., Pandit, J., Kang, C.-H., Nikaido, K., Gokcen, S., Ames, G. F.-L., and Kim, S.-H. (1993). Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J Biol Chem*, 268:11348–11355. 5.4.3

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, 276:307–326. 3.3.2

Paulsen, I. T. (2003). Multidrug efflux pumps and resistance: regulation and evolution. *Curr Opin Microbiol*, 6:446–451. 1.4.2

Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R., and Saier Jr., M. H. (2000). Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J Mol Biol*, 301:75–100. 1.4.1

Pearson, R. G. (1963). Hard and Soft Acids and Bases. *J Am Chem Soc*, 85:3533–3539.

Perrakis, A., Morris, R., and Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct Biol*, 6:458–463. 5.3.4

Perroud, B. and le Rudulier, D. (1985). Glycine betaine transport in *Escherichia coli*: osmotic modulation. *J Bacteriol*, 161:393–401. 4.2

Poole, K. (2001). Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *J Mol Microbiol Biotechnol*, 3:255–264. 6.2

Pos, K. M. and Diederichs, K. (2002). Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Cryst*, D58:1865–1867. 6.2, 6.4.1

Prieto, J. and Serrano, L. (1997). C-capping and helix stability: the Pro C-capping motif. *J Mol Biol*, 274:276–288. 3.4.6

Qu, Y., Bolen, C. L., and Bolen, D. W. (1998). Osmolyte-driven contraction of a random coil protein. *Proc Natl Acad Sci USA*, 95:9268–9273. 4.2, 5.2

Quiocho, F. A. and Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol*, 20:17–25. 4.4.1, 5.2, 5.4.2

Quiocho, F. A., Spurlino, J. C., and Rodseth, L. E. (1997). Extensive features of tight oligosaccararide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure*, 5:997–1015. 3.3.4, 3.4.2

Record Jr., M. T., Courtenay, E. S., Cayley, D. S., and Guttman, H. J. (1998). Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *TIBS*, 23:143–148. 4.2, 5.2

Rhodes, D. and Hanson, A. D. (1993). Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Ann Rev Plant Physiol Plant Mol Biol*, 44:357–384. 4.2, 5.2

Richarme, G. and Kepes, A. (1983). Study of binding protein-ligand interaction by ammonium sulfate-assisted adsorption on cellulose esters filters. *Biochim Biophys Acta*, 742:16–24. 4.3.4

Ripoli, D. R., Faerman, C. H., Axelsen, P. H., Silman, I., and Sussman, J. L. (1993). An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholineesterase. *Proc Natl Acad Sci USA*, 90:5128–5132. 4.4.6

Roberts, M. F. (2000). Osmoadaptation and osmoregulation in archaea. *Front Biosci*, 5:796–812. 4.2

Roderick, S. L., Chan, W. W., Agate, D. S., Olsen, L. R., Vetting, M. W., Rajashankar, K. R., and Cohen, D. E. (2002). Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat Struct Biol*, 9:507–511. 4.4.6

Roessler, M. and Muller, V. (2001). Osmoadaptation in bacteria and archaea: common principles and differences. *Environ Microbiol*, 3:742–754. 5.2

Russell, R. J. and Taylor, G. L. (1995). Engineering thermostability: lessons from thermophilic proteins. *Curr Opin Biotechnol*, 6:370–374. 3.5

Saier Jr., M. H. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev*, 64:354–411. 1.4.1

Sanchez-Ruiz, J. M. and Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol*, 19:132–135. 3.5

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, 74:5463–5467. 4.3.3, 5.3.2

Santos, H. and da Costa, M. S. (2002). Compatible solutes of organisms that live in hot saline environments. *Environ Microbiol*, 4:501–509. 5.2

Saurin, W., Hofnung, M., and Dassa, E. (1999). Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol*, 48:22–41. 1.3.1, 2.2

Scheiner, S., Kar, T., and Gu, Y. (2001). Strength of CαH$\cdots$O hydrogen bond of amino acid residues. *J Biol Chem*, 276:9832–9837.

Schiefner, A., Breed, J., Bösser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W., and Bremer, E. (2004). Cation-π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *J Biol Chem*, 279:5588–5596. 5.2, 5.4.4, 5.4.6, 5.4.8, 5.4.9

Schneider, E. (2003). Import of solutes by ABC transporters: the maltose and other systems. In Cole, S., Higgins, C., Holland, E. B., and Kuchler, K., editors, *ABC proteins: from bacteria to man*, pages 157–185. Academic Press. 3.2

Schneider, T. and Sheldrick, G. (2002). Substructure solution with SHELXD. *Acta Cryst*, D58:1772–1779. 5.3.4, 6.3.2

Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T., and Bakker, E. P. (1994). Purification, properties and structural aspects of a thermoacidophilic α-amylase from *Alicyclobacillus acidocaldarius* atcc 27009. Insight into acidostability of proteins. *Eur J Biochem*, 226:981–991. 3.2, 3.4.1, 3.5

Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport. *Biochemistry*, 31:10657–10663. 1.3.3, 3.2, 3.4.3, 5.4.3

Shilton, B. H., Flocco, M. M., Nilsson, M., and Mowbray, S. L. (1996). Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins. *J Mol Biol*, 264:350–363. 3.4.3

Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984). *Experiments with gene fusions*. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y. 4.3.2

Singer, S. J. and Nicolson, G. L. (1972). The fluid mosaic model of the structure of cell membranes. *Science*, 175:720–731. 1.1

Sleator, R. D. and Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev*, 26:49–71. 4.2, 5.2

Spurlino, J. C., Lu, G.-Y., and Quiocho, F. A. (1991). The 2.3 Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J Biol Chem*, 266:5202–5219. 1.3.3, 3.4.4

Steiner, T. and Koellner, G. (2001). Hydrogen bonds with π-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J Mol Biol*, 305:535–557. 5.4.5

Stetter, K. O., Lauerer, G., Thomm, M., and Neuner, A. (1987). Isolation of extremely thermophilic sulfate reducers evidence for a novel branch of archaebacteria. *Science*, 236:822–824. 5.2, 5.3.2

Sutcliffe, I. C. and Russell, R. R. (1995). Lipoproteins of Gram-positive bacteria. *J Bacteriol*, 177:1123–1128. 3.2

Taylor, R. and Kennard, O. (1982). Crystallographic evidence for the existence of C-H$\cdots$O, C-H$\cdots$N, and C-H$\cdots$Cl hydrogen bonds. *J Am Chem Soc*, 104:5063–5070. 5.4.4

Terwilliger, T. C. (2001). Maximum-likelihood density modification with pattern recognition of structural motifs. *Acta Cryst*, D57:1755–1762. 5.3.4, 6.3.2

Terwilliger, T. C. and Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Cryst*, D55:849–861. 4.3.5

Thiebaut, F., Tsuruo, T., Hamada, H., Gottesman, M. M., Pastan, I., and Willingham, M. C. (1987). Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. *PNAS*, 84:7735–7738. 1.3.1

Tikhonova, E. B., Q., W., and Zgurskaya, H. I. (2002). Chimeric analysis of the multicomponent multidrug efflux transporters from gram-negative bacteria. *J Bacteriol*, 184:6499–6507. 6.2

Timasheff, S. N. (1993). The control of protein stability and association by weak interactions with water: How do solvents affect these processes. *Ann Rev Biophys Biol Struct*, 22:67–97. 4.5.1

Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J Appl Cryst*, 30:1022–1025. 5.3.4, 6.3.2

van Bambeke, F., Balzi, E., and Tulkens, P. M. (2000). Antibiotic efflux pumps. *Biochem Pharmacol*, 60:457–470. 1.4.1

van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*, 229:105–124. 5.3.2

Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. *Microbiol Mol Biol Rev*, 62:504–544. 5.2

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol*, 1:732–740. 3.5

Waldburger, C. D., Jonsson, T., and Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc Natl Acad Sci USA*, 93:2629–2634. 3.5

Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity ? *Nat Struct Biol*, 2:122–128. 3.5

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot Eng*, 8:127–134. 3.3.4

Watanabe, K., Chishiro, K., Kitamura, K., and Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* KP1006. *J Biol Chem*, 266:24287–24294. 3.4.6

Watanabe, K., Kitamura, K., and Suzuki, Y. (1996). Analysis of the critical sites for protein thermostabilization by proline substitution in oligo-1,6-glucosidase from *Bacillus coagulans* ATCC 7050 and the evolutionary consideration of proline residues. *Appl Environ Microbiol*, 62:2066–2073. 3.4.6

Weiss, M. S., Brandl, M., Sühnel, J., Pal, D., and Hilgenfeld, R. (2001). More hydrogen bonds for the (structural) biologist. *Trends Biochem Sci*, 26:521–521. 5.4.4

Weiss, M. S. and Hilgenfeld, R. (1997). On the use of the merging R factor as a quality indicator for X-ray data. *J Appl Cryst*, 30:203–205. 4.1, 5.1, 6.1

Welsh, D. T. (2000). Ecological significance of compatible solute accumulation by microorganisms: from single cells to global climate. *FEMS Microbiol Rev*, 24:263–290. 4.2, 5.2

Wisotzkey, J. D., Jurtshuk Jr., P., Fox, G. E., Deinhard, G., and Poralla, K. (1992). Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. *Int J Syst Bacteriol*, 42:263–269. 3.2

Wood, J. M. (1999). Osmosensing by bacteria: signals and membrane-based sensors. *Microbiol Mol Biol Rev*, 63:230–262. 4.2, 5.2

Xavier, K. B., Martins, L. O., Peist, R., Kossmann, M., Boos, W., and Santos, H. (1996). High-affinity maltose/trehalose transport system in the hyperthermophilic archaeon *Thermococcus litoralis*. *J Bacteriol*, 178:4773–4777. 1.3.3, 2.2

Yancey, P. H. (1994). Compatible and counteracting solutes. In Strange, K., editor, *Cellular and molecular physiology of cell volume regulation*, pages 81–109. CRC Press, Inc., Boca Raton, USA. 4.2, 5.2

Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H., and Koshland Jr., D. E. (2003). Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science*, 300:976–980. 6.2, 6.2, 6.4.2

Yuan, Y.-R., Blecker, S., Martsinkevich, O., Millen, L., Thomas, P. J., and Hunt, J. F. (2001). The crystal structure of the MJ0796 ATP-binding cassette. *J Biol Chem*, 276:32313–32321. 2.2

Zgurskaya, H. I. and Nikaido, H. (1999a). AcrA is a highly asymmetric protein capable of spanning the periplasm. *J Mol Biol*, 285:409–420. 6.2

Zgurskaya, H. I. and Nikaido, H. (1999b). Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc Natl Acad Sci USA*, 96:7190–7195. 6.2

Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A., and Dougherty, D. A. (1998). From ab initio quantum mechanics to molecular neurobiology: A cation-π binding site in the nicotinic receptor. *Proc Natl Acad Sci USA*, 95:12088–12093. 4.5.3

Zhou, Z., White, K. A., Polissi, A., Georgopoulos, C., and Raetz, C. R. H. (1998). Function of *Escherichia coli* MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J Biol Chem*, 273:12466–12475. 1.3.2

Dank

Vor allen möchte ich Prof. Dr. Wolfram Welte danken, für die Möglichkeit meine Arbeit in seiner Arbeitsgruppe anfertigen zu können. Er war stehts bemüht meine Interessen zu fördern und bot bei Problemen gern seine Unterstützung an.
Auch möchte ich mich bei Dr. Kay Diederichs bedanken, der Ansprechpartner war, wenn kristallographische Probleme auftraten. Ausserdem schätze ich seine Bemühungen sehr, immer die neuste Software zu organisieren und bereitzustellen, um eine optimale Auswertung der Daten sicher zu stellen.
Prof. Dr. Erhard Bremer und Dr. Gudrun Holtmann danke ich für ein sehr fruchtbare Kooperation. Sie waren stehts am Fortgang unserer gemeinsamen Projekte interessiert und haben durch ihren Einsatz die erzielten Ergebnisse erst möglich gemacht.
Desweiteren bedanke ich mich bei Prof. Dr. Winfried Boos, Dr. Alex Böhm und Dr. Reinhold Horlacher für ihre Unterstützung in mikrobiologischen und molekularbiologischen Fragen.
Vergessen möchte ich auch nicht alle aktuellen und ehemaligen Mitglieder der AG Welte, die mich unterstützen, und die auch dafür sorgten, dass die meiste Zeit ein gutes Klima herrschte. Besonders bedanken möchte ich mich an dieser Stelle bei Kyoko Hashimoto, die mich lange Zeit erheblich im Labor unterstützte, aber auch bei Kinga Gerber, die stehts bereit war meine Manuskripte Korrektur zu lesen.