Structural Basis for the Binding of Compatible Solutes by ProX from the Hyperthermophilic Archaeon *Archeoglobus fulgidus*

Received for publication, March 31, 2004, and in revised form, July 15, 2004 Published, JBC Papers in Press, August 11, 2004, DOI 10.1074/jbc.M403540200

André Schiefner‡§, Gudrun Holtmann¶, Kay Diederichs‡, Wolfram Welte‡, and Erhard Bremer¶

From the ‡Fachbereich Biologie, Universität Konstanz, Universitätstraße 10, D-78457 Konstanz, Germany and the ¶Fachbereich Biologie, Laboratorium für Mikrobiologie, Philipps-Universität Marburg, Karl-von-Frisch Strasse, D-35032 Marburg, Germany

Compatible solutes such as glycine betaine and proline betaine serve as protein stabilizers because of their preferential exclusion from protein surfaces. To use extracellular sources of this class of compounds as osmo-, cryo-, 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 that are able to bind the transported substance with high affinity. Therefore, binding proteins that bind compatible solutes have to avoid the exclusion of their ligands within the binding pocket. 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 ligand-binding protein ProX from the hyperthermophilic archaeon *Archeoglobus 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 trimethylammonium. 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. The comparison of the ligand binding sites of ProX from *A. fulgidus* and the recently solved structure of ProX from *Escherichia coli* revealed a very similar solution for the problem of compatible solute binding, although both proteins share only a low degree of sequence identity. The residues involved in ligand binding are functionally equivalent but not conserved in the primary sequence.

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 (1–5). Because bacteria lack systems to actively transport water across the cytoplasmic membrane, osmotic processes solely determine their intracellular water content. This requires an active adjustment of their intracellular solute pool to prevent dehydration under hypertonic growth conditions and bursting under hypotonic circumstances (6). 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 (2, 3, 5, 7–10). 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 (7). Important representatives of compatible solutes are the trimethylammonium compound glycine betaine (GB, \(N,N,N\)-trimethylglycine)\(^1\) and the dimethylammonium 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* (2, 3, 10) and *Archaea* (10–12), but also by fungal, plant, animal, and even human cells (13–18).

In addition to their well established role as osmoprotectants, compatible solutes also function as protein stabilizers both in *vitro* (19–21) and *in vivo* (22). 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 (23). This hypothesis predicts that compatible solutes are excluded from the immediate hydration shell of proteins, presumably because of their unfavorable interactions with protein surfaces (24, 25). The resulting disequilibrium provides a thermodynamic force to minimize the surface of the protein 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 (26). The ability of compatible solutes to stabilize proteins probably also explains their *in vivo* function as microbial stress protectants against heat and chill stress (5, 27–31).

Many microorganisms are able to acquire compatible solutes, in particular glycine betaine, from environmental sources through high-affinity transport systems (1, 32, 33). One example of such a glycine betaine transporter is the ProU system from *Escherichia coli* (34, 35), a member of the binding protein-dependent ATP-binding cassette (ABC) superfamily of transporters (36, 37).

In the meantime, several Gram-negative and Gram-positive microorganisms have been shown to contain homologues of the *E. coli* ProU glycine betaine transport system. Recently, we have identified such a ProU-type glycine betaine uptake sys-

---

\(^1\) The abbreviations used are: GB, glycine betaine; PB, proline betaine; MES, 4-morpholineethanesulfonic acid; TM, trimethylammonium; r.m.s., root mean square.
Binding of Compatible Solutes by ProX from *A. fulgidus* 

48271

**ProX Protein—** To overproduce the *A. fulgidus* ProX protein in *E. coli*, we used the proX plasmid pGH26. This plasmid is a derivative of the expression vector pASK-IBA6 (IBA, Göttingen, Germany) that carries an anhydrotetracycline-inducible tet promoter allowing selective proX gene expression. pGH26 carries the proX gene from *A. fulgidus* strain VC16 (40, 38) as a 386-bp-PCR fragment flanked by Bax 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. 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 residue of the mature ProX protein is replaced by a glycine to avoid a lipid modification of the soluble ProX protein. The 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. 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 cysteine. The N-terminal sequence of the *A. fulgidus* proX gene in pGH26 was verified by DNA sequence analysis (46) using a set of synthetic oligonucleotide primers appropriately spaced along the proX coding region.

Rich media (e.g., yeast extract) frequently contain GB (47). 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 (45). A 10-liter flask was inoculated with 5 liters of minimal medium 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 A₅₇₈ 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 an optical density of A₅₇₈ of 5–0.8. Increased 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 2 more hours to allow ProX production and were subsequently harvested by centrifugation (10 min, 3000 × 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 F (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, 21,000 × g). Insoluble material of the supernatant was subsequently removed by ultracentrifugation (30 min, 120,000 × g). The cleared protein extract was then loaded onto a 10-ml Strep-Tactin column (IBA, Göttingen, Germany) equilibrated with buffer W (100 mM Tris-HCl, pH 8.0, 0.5 M sucrose) and washed with 5 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 M desthiobiotin). Then both the partially unreleased OmpA signal sequence and/or the Strep-Tag affinity sequence were cleaved with the protease factor Xa (1 μg of factor Xa per 200 μg of ProX protein) for 16 h at room temperature in a buffer containing 200 mM (NH₄)₂SO₄, 100 mM NaCl, and 1 mM EDTA. To remove 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 (Bio-Rad, München, Germany) equilibrated with buffer A (50 mM Tris-HCl, pH 8.0). The column was washed with 20 ml of buffer A and the proteins were eluted with a linear NaCl gradient. ProX eluted at a NaCl concentration of 250 mM from the UnoQ6 column. The protein was stored until further use at

---

**MATERIALS AND METHODS**

**Bacterial Strains—** The *E. coli* strain BL21 (Codon Plus RIL), which harbors a plasmid encoding rare trRNAs for arginine, isoleucine, and

---

2 G. Holtmann and E. Bremer, unpublished results.
The ProX protein from *A. fulgidus* contains five methionines. For the labeling of the ProX protein with L-selenomethionine, the method of metabolic inhibition was employed (48, 49). 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 harvested by centrifugation (10 min, 3,000 g) and resuspended in the same volume of M9 minimal medium (45). This entire cell suspension was then used to inoculate 4 liters of M9 medium at 37 °C. Cultures were grown to an OD of 0.5 (crystal form II). Unliganded ProX crystallized in hanging drops against 100 mM Tris-HCl, pH 8.0, 250 mM NaCl mixed with equal amounts of the reservoir solutions described below. Liganded ProX crystallized in hanging drops using a reservoir solution containing 0.2 mM CaCl2, 0.1 M sodium cacodylate, pH 6.0–6.5, 10–12% (v/v) PEG 4000. The crystals reached a final size of 250 × 200 × 150 μm after 3 days. They belong to the space group P21 (crystal form I). In a different setup, liganded ProX crystals were obtained by setting drops equilibrated against a reservoir containing 30% (v/v) PEG1500. The crystals grew to a maximum size of 50 × 50 × 50 μm within 2 months and belong to the space group P42,22 (crystal form II). Unliganded ProX crystallized in hanging drops against a reservoir solution containing 0.3 mM MgCl2, 0.1 M Tris, pH 7.0–9.0, 35% (v/v) PEG4000. The first crystals appeared after 2–3 months, reached a final size of 300 × 300 × 150 μm 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 MES, pH 6.5, 25–30% (v/v) ethylene glycol. These crystals grew within 4 weeks, reached a final size of 200 × 150 × 20 μm, and belong to space group P22,22 (crystal form IV). All data sets were collected at a synchrotron source 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 s they were frozen in liquid nitrogen. Crystals of crystal form IV could be soaked by directly transferring to a reservoir containing 0.1 M Tris, 0.1 M MES, pH 6.5, 25–30% (v/v) ethylene glycol. These crystals showed density compensation. Selenomethionine-labeled protein was crystallized in crystal form I and was treated as the native crystals prior to freezing.

**RESULTS AND DISCUSSION**

**Crystallography**—Crystals of ProX have been obtained under four different conditions yielding crystals in four different space groups. The liganded closed form crystallized in P21 and P42,22, whereas the open unliganded form crystallized in C2 and P22,22. In the following, crystallization of liganded ProX in P21 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 trimethylamine (see Table I). Trimethylamine has a pKa of 9.81 and is therefore protonated under the crystallization conditions we have used. This trimethylammonium (TM) resembles the head group of GB. However, 3,3-dimethylbutyric acid, where a carbon atom replaces the nitrogen atom compared with GB, is not able to substitute GB. 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 pocket 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 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.

**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 two-wavelength anomalous dispersion of ProX-PB in space group P21 (see Table I). 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 II). This is in agreement with the expectation that these five residues 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 corresponding to the two domains were found for domain A in the cross-rotation function, whereas no useful signal was detected for domain B. In the translation function 4 copies of domain A (two for each rotation peak) could be located in the asymmetric unit that were refined using REFMACS. To locate the B domains, a molecular replacement was carried out against the F ″ = F′ map of the initial A domains model, resulting in the correct position of two copies of domain B. This was repeated 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 (58). Secondary structure elements and solvent accessible surface of the ProX structure were determined using the program DSSP (59). Figures with molecule presentations were prepared using MolScript (60) and Raster3D (61).

**Chemicals**—Glycine betaine and desthiobiotin were purchased from Sigma. Proline betaine (stachydrine-HCl) was obtained from Exrasythese (Genay, France). Factor Xa was purchased from New England Biolabs (Schwalbach, Germany). L-Selenomethionine was obtained from Fisher Scientific. Trimethylamine solution was purchased from Hampton Research.

**Base Searches**—Searches for proteins that are homologous to ProX from *A. fulgidus* were conducted using the BLAST network server (62).

**Protein Data Bank Codes**—All atomic coordinates and structure factors resulting from this work have been deposited in the Protein Data Bank (63). The accession codes are 1SW1, 1SW2, 1SW4, and 1SW5 for ProX-PB, ProX-GM, ProX-TM, and ProX ligand free, respectively.
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 (41) (see Fig. 1). This can generally be subdivided into two globular domains: 1) domain A (blue) from residue 6 to 110 and from residue 212 to 275, and 2) domain B (yellow) from residue 111 to 211 connected by the two-switch segments, Asp109–Asp110–Tyr111–Ala112 and Ala110–Leu211–Pro212–Pro213, which form with Asp109, Tyr111 and Leu211–Pro212, and the two hinges between domains A and B. Each domain consists of a 4-stranded β-sheet flanked by α-helices on both sides.

Overall Conformational Changes Induced upon Ligand Binding—The ligand-free, open conformation of ProX (see Fig. 1) 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 atomic coordinates of ProX in space groups P2₁, P2₁, and C2. In C2 the asymmetric unit contains four molecules with very similar angles between 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 Å² and 44.9 Å², which are more than twice as high as those of the other B domains (19.2 Å² and 19.6 Å²) 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 ligand binding studies of ProX that were performed at room temperature.²

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 (64–66). With regard to this finding, we analyzed the difference in the arrangement of the two domains in the open and closed forms with the program NCSGROUPS. ³ The basic task of the program is to introduce hinges to optimize the least square fit of the Cα positions between two molecules. It turned out that the introduction of only two hinges, one between Asp109 and Tyr111 and the other between Leu211 and Pro212, was sufficient to superimpose the two molecules very well (domain A r.m.s. deviation of 0.521 Å and domain B r.m.s. deviation of 0.517 Å). This is equivalent to a total rotation of domain B by –58° with respect to domain A (Fig. 1). The total rotation has two components: 1) the hinge angle between the two domains of –40° with its axis going through the above determined hinges in the

³ K. Diederichs, unpublished results.
polypeptide and 2) a rotation perpendicular to the hinge axis of 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. deviation between domain B of the open form compared with 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 residue 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 more or less as rigid groups, there are a few changes of the secondary structure in two regions of ProX as indicated by the program DSSP (59). 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 Arg49, which plays an important role in ligand binding as discussed below. Furthermore, residues 222–225 (domain A), which are in turn and 310-helix conformation (in the open form), become rearranged to a short α-helix (in the closed form).

Shilton et al. (67) found that decrease of the accessible surface area on the opposite side of the binding cleft may stabilize the open conformation in the absence of the ligand. To check whether there is a similar effect in the ProX structure we calculated the accessible surface area for every residue of both the open and closed conformations with the program DSSP. The differences for all residues between the open and closed forms 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 that were between the two domains on the opposite side of the binding cleft. This approach identified 11 residues that show decreases of the accessible surface of more than 25 Å² between the open and closed forms, resulting in a buried surface increase of 476 Å² 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 closed forms 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 (68). According to this criterion, a total of 19 salt bridges is formed in ProX in both the open and closed forms, 13 of them are hydrogen bonded with a distance of less than 3.5 Å (69). ProX has no salt bridges with histidines involved, because it does not contain any histidine residues. During the conformational change 5 salt bridges are broken and 5 new ones are formed. Interestingly, the ratio of observed ion pairs changes. Whereas in the open form a similar number of Glu–Lys, Glu–Arg, Asp–Lys, and Asp–Arg pairs was observed (6, 3, 4, and 6, respectively), in the closed form 8 Glu–Lys and 8 Asp–Arg pairs but only 2 Glu–Arg and 1 Asp–Lys pairs were found. The open conformation is apparently not stabilized by salt bridges compared with the closed form. We find only one interdomain salt bridge in the open conformation compared with 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.

**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 Fig. 1. In the case of 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. Fig. 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 Asp\(^{109}\) as the bottom part of the box, and the four tyrosine residues Tyr\(^{63}\), Tyr\(^{111}\), Tyr\(^{190}\), and Tyr\(^{214}\) 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 Asp\(^{109}\) the tyrosine side chains provide a negative surface potential that 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 Lys\(^{13}\), Thr\(^{66}\), and Arg\(^{149}\) 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 Lys\(^{13}\) and Arg\(^{149}\).

Basically, all ligands have the same orientation within the binding site. The trimethylammonium group of GB and 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 rotatable trimethylammonium because of the bulky proline ring. Nevertheless, all three ligands interact in the same way with the quaternary amine binding part of the binding site. Their quaternary amine head groups all point toward the main chain carbonyl oxygen of Asp\(^{109}\) and are surrounded by the four tyrosine faces of Tyr\(^{63}\), Tyr\(^{111}\), Tyr\(^{190}\), and Tyr\(^{214}\).

In a recent publication from our laboratory (44) 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 carboxyls. The nitrogen attached methyl or methylene 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 (75) these C–H groups are soft donors that are able to form hydrogen bonds with soft acceptors like π-systems and carbonyl oxygens. Scheiner et al. (76) found by quantum mechanical calculations that C–H–O hydrogen bonds between CO–H and H\(_2\)O 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–O). The distance between C and O in a C–H–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 Asp\(^{109}\) are 3.3, 3.2, and 3.5 Å, which is very close to the ideal distance. Therefore we assume that GB mainly interacts via non-classical hydrogen bonds with the main chain carbonyl of Asp\(^{109}\). This is very similar for TM with mean distances of 3.4, 3.3, and 3.4 Å, but slightly distorted for PB with mean distances of 3.9, 3.3, and 3.5 Å because of the tilting of the Co–N axis as explained above. For the same reason the interaction between the nitrogen attached methyl or methylene groups of the ligand and the tyrosine π-systems has two components, a cation–π and non-classical hydrogen bond contribution, by C–H⋯π hydrogen bonds (77, 78).

The above mentioned explanations also fit the observation that 3,3-dimethylbutyric acid is not bound by ProX. Without a quaternary nitrogen carrying a positive charge that polarizes the bound methyl or methylene groups, it is unable to interact with the ProX ligand binding site.
Fig. 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 described in the legend to Fig. 1. All residues involved in binding are presented as a ball-and-stick model and labeled with their residue names. The ligands are highlighted with brown sticks. To compare the ligand positions, the binding sites were superimposed.
Conformational Changes of the Ligand Binding Site upon Ligand Binding—Fig. 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 the ligand binding show virtually the same orientation in the open and closed forms (right part of Fig. 3). Residues Tyr<sup>63</sup>, Tyr<sup>214</sup>, Lys<sup>13</sup>, and Thr<sup>66</sup> superimpose very well. Only the main chain carbonyl of Asp<sup>109</sup> from domain A is slightly out of place compared with the closed form because of the enormous main chain rearrangement between Asp<sup>110</sup> and Tyr<sup>111</sup> upon closing. The residues contributed by domain B behave quite differently. Tyr<sup>111</sup> and Tyr<sup>180</sup> 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 Arg<sup>149</sup> shows only small changes between the open and closed conformations although it undergoes a large movement as part of domain B.

Not shown in Fig. 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 conformations, residues Tyr<sup>63</sup>, Tyr<sup>214</sup>, and Lys<sup>13</sup> are stabilized either by hydrogen bond formation (Tyr<sup>63</sup>–Glu<sup>146</sup>, Tyr<sup>63</sup>–Glu<sup>118</sup>, Tyr<sup>214</sup>–Asp<sup>151</sup>, Lys<sup>13</sup>–Thr<sup>66</sup>) or salt bridges (Lys<sup>13</sup>–Glu<sup>62</sup>). The side chain of Thr<sup>66</sup> is not involved in any contacts in the open form. Many changes happen in domain B between the open and closed conformations. In the open form C–H–π interactions (77, 78) (Tyr<sup>190</sup>–Phe<sup>146</sup>, Tyr<sup>111</sup>–Tyr<sup>180</sup>) and hydrophobic contacts (Tyr<sup>180</sup>–Pro<sup>188</sup>) are formed. This changes upon closing: in the closed conformation the orientation of the phenyl rings is stabilized by C–H–π interactions (Tyr<sup>111</sup>–Phe<sup>146</sup> donor and acceptor being exchanged, Tyr<sup>190</sup>–Phe<sup>151</sup>), by van der Waals contacts (Tyr<sup>190</sup>–Met<sup>177</sup>, Tyr<sup>214</sup>–Pro<sup>212</sup>), and hydrogen bonds (Tyr<sup>111</sup>–Asp<sup>143</sup>, Tyr<sup>180</sup>–Asp<sup>143</sup>). Additionally, two interdomain hydrogen bonds (Tyr<sup>214</sup>–Asp<sup>151</sup>, Thr<sup>66</sup>–Arg<sup>149</sup>) and two salt bridges (Arg<sup>149</sup>–Glu<sup>145</sup>, Arg<sup>149</sup>–Asp<sup>151</sup>) 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. (79). All residues provided by domain A, including Tyr<sup>63</sup> and Tyr<sup>114</sup>, Asp<sup>109</sup>–CO as well as Lys<sup>13</sup> and Thr<sup>66</sup>, remain at their positions that 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 that has the highest mobility). 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 closed forms is shifted toward the closed form. In domain B only residues Tyr<sup>111</sup> and Tyr<sup>180</sup> 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 (80) that the binding of the ligand occurs in a series of reactions of first order.

Comparison of Related Sequences—Many sequences related to ProX from A. fulgidus have been identified in a Blast search against the NCBI data base. 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 Fig. 4). A protein from the archaeon Methanosarcina mazei shows with 55% the highest identity to ProX from A. fulgidus 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* is used as a model organism for the investigation of the physiological effects of compatible solutes on cell growth under hyperosmotic conditions (3). Whereas the binding proteins of *A. fulgidus*, *M. mazei*, and *B. subtilis* are supposed to be membrane an-
chored 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 in Fig. 4, the positions of the tyrosine residues involved in ligand binding are clearly conserved with only two exceptions: 1) the position equivalent to Tyr63 in *S. mutans* and 2) the *C. acetobutylicum* position equivalent to Tyr190. In both cases 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* (44). 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. Because 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 for 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 -CH₂OH group. On the other hand, in OpuCC, which 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 -CH₂OH group of choline.

**Metal Binding Sites**—In the anomalous difference density maps of the refined P2₁ crystal structures a total of 17 anomalous density peaks above 5 σ have been found in the asymmetric unit at λ = 1.0440 A. Because of 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 modeled in the final structures in P2₁. 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 of these 6 zinc ions are involved in crystal contacts, especially three of them that 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 and Asp244, and chain II Asp241, and vice versa.

**Comparison of the ProX Structure from A. fulgidus and ProX 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* could hardly be done properly. After the structural analysis it became clear that this is not only because of the low sequence identity but also because of shifts in the primary sequence with respect to the domains.

The structures of Af-ProX and Ec-ProX (Protein Data Bank code 1R9L) were superimposed using a global superposition-distance-based Needleman-Wunsch alignment as implemented.
in the program LSQMAN (81). All Cα positions within a cutoff distance of 3.5 Å were treated as matched. This alignment gave a total of 214 residues that 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 Af-ProX. Obviously, the fold of both molecules is the same but some smaller parts of the polypeptide chain are different. Fig. 5A shows the Cα trace presentation of Af-ProX in red and Ec-ProX in blue. The structure-based sequence alignment is given in Fig. 5B. At first sight, the structure of Af-ProX appears to be more compact than that of Ec-ProX. The N terminus of the Ec-ProX structure is 7 residues longer because of the fact that in the Af-ProX structure the first 5 residues are disordered and could not be modeled as discussed above. On the other hand, residues 295 to 309 of Ec-ProX are missing in Af-ProX although they apparently have a stable α-helical conformation in Ec-ProX. Particularly interesting is the difference in the hinge region shown on the left of Fig. 5A. There are large loop regions in both molecules that do not match at all. These are mainly responsible for the difficulty in obtaining a sequence alignment. In the case of Af-ProX this region including 27 residues is part of domain A and is stabilized by α-helical and turn conformations. In contrast, in Ec-ProX the loop composed of 14 residues is part of domain B and has mainly turn conformations as structural elements. Additionally, two loops are elongated in Ec-ProX, the first formed by residues 124–130 is involved in the coordination of a metal ion of unknown function (44), the second loop, on the other hand, seems to be only a minor elongation. The 27 residues spanning extension of domain A in Af-ProX may have a function in stabilizing the open conformation, because it forms additional contacts with domain B in the open conformation. These contacts contribute to more than two-fifths of the buried surface in the open conformation as described above.

**Comparison of the Ligand Binding Sites**—A closer look to the binding pockets reveals that the interaction of Af-ProX and...
Ec-ProX with their ligands is very similar. In both proteins a set of sterically well arranged aromatic residues is essential to form a binding site that is capable of interacting effectively with the quaternary amine of the natural ligands GB and PB. This interaction is mainly determined by the two factors, cation–π interaction and non-classical hydrogen bonding. On the other hand, the carboxylic tail is bound by conventional hydrogen bonding and/or salt bridge formation.

Fig. 6 shows a superposition of the Af-ProX and Ec-ProX ligand binding sites with bound GB. All residues being involved in the interaction with the quaternary amine are highlighted.

The GB ligand is oriented very similar within both binding sites: its quaternary amine is pointed deeply into the binding pocket and its carboxylic tail points toward the opening side. Noteworthily is that the Ca–N axis of GB is tilted by almost 90° with respect to the polypeptide backbone between the two structures. This results from the fact that Af-ProX and Ec-ProX do not use the same set of residues to bind the quaternary amine: whereas in Af-ProX four tyrosine side chains and one main chain carbonyl oxygen are involved, in Ec-ProX the same task is performed by three tryptophan side chains. Despite the tilted GB Ca–N axis two positions of the involved aromatic residues seem to be conserved. These side chains are almost perpendicular oriented to each other in both proteins. Interestingly, the residue numbers of the aromatic residues coincide (Af-ProX Tyr62 ≡ Ec-ProX Trp62 and Af-ProX Tyr190 ≡ Ec-ProX Trp188) although their spacing is given by different loops, e.g. the larger loop of 69–95 in Ec-ProX is compensated by the two smaller loops 117–130 + 177–182 in Ec-ProX (see Fig. 5).

Number and manner of the contacts between the GB quaternary amine head as well as its carboxylic tail with Af-ProX and Ec-ProX is apparently correlated with the strength of the protein–ligand interaction. According to the above mentioned criteria, Af-ProX forms 23 contacts to the GB quaternary amine, two salt bridges, and one hydrogen bond with the carboxylic tail. On the other hand, Ec-ProX forms 18 contacts with the quaternary amine head and, depending on the pH, one salt bridge and two hydrogen bonds or only three hydrogen bonds with the carboxylic tail of the ligand. From this structural data one might expect a significantly lower $K_d$ of Af-ProX for GB compared with Ec-ProX, which is clearly supported by experimental data from binding assays where Af-ProX with a $K_d = 60$ nM compared with Ec-ProX with a $K_d = 4 \mu$M (44) binds GB more than 60 times better at room temperature.

The distribution of the few identical residues in both structures and the conservation of two positions of aromatic residues indicates that both molecules might have derived from a common ancestor. Possibly, a common ancestor of both proteins might have contained two aromatic residues (Trp, Tyr, or Phe) which are equivalent as known from mutational studies of Ec-ProX (Neidhard, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Remizov, W. S., Riley, M., Schaechter, M., and Unharger, H. E., eds /Vol. 1, pp. 1210–1223, American Society for Microbiology Press, Washington, D. C.).

Acknowledgments—We thank Clemens Schulze-Briese and Takashi Tomizaki for their SLS support and Kinga Gerber for careful reading of the manuscript.

REFERENCES
