THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 46, Issue of November 12, pp. 48270–48281, 2004

Printed in U.S.A.

Structural Basis for the Binding of Compatible Solutes by ProX from the Hyperthermophilic Archaeon *Archaeoglobus 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, Universitätsstrasse 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 *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 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-

* This work was supported by Deutsche Forschungsgemeinschaft through SFB 395, the Graduiertenkolleg "Proteinfunktion auf atomarer Ebene," the TR-SFB 11 (Konstanz, Zürich), the Fonds der Chemischen Industrie (to E. B.), and the Max-Planck-Institute for terrestrial Microbiology (Marburg). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed. E-mail: andre.schiefner@uni-konstanz.de.

[1] The abbreviations used are: GB, glycine betaine; PB, proline betaine; MES, 4-morpholineethanesulfonic acid; TM, trimethylammonium; r.m.s., root mean square.

tem in *Archaeoglobus fulgidus*.[2] *A. fulgidus* is a sulfate-reducing archaeon with an optimal growth temperature of 83 °C. This hyperthermophilic microorganism was originally isolated from hot sediments of a marine hydrothermal vent system (38). Marine sediments are known to contain glycine betaine (39), 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 *A. fulgidus* strain VC16 (40), the ProU transporter of this organism consists of two copies of an ATPase (ProV), two integral membrane proteins (ProW1 and ProW2), and an extracellular ligand-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 (36). Binding proteins of binding protein-dependent ABC transporters 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 (41). This has been shown by the structural analysis of various open states of the ribose-binding protein and D-allose-binding protein (42, 43). Ligand binding induces a large conformational change in the hinge region that 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* (40) 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.[2] Substrate binding assays with radiolabeled glycine betaine and competition experiments with other compatible solutes revealed that ProX binds both GB and PB with high affinity and specificity at room temperature with an apparent K_D of 60 and 50 nM, respectively. GB and PB are well known osmoprotectants for many microbial species (10), 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*.[2]

The amino acid sequences of the ligand-binding proteins ProX from *E. coli* (34) and ProX from *A. fulgidus* (40) share only low sequence identity (<12%), and in particular the three tryptophan residues involved in substrate binding in ProX from *E. coli* (44) are not conserved in ProX from *A. fulgidus*. To broaden our understanding of the specific interactions of ligand-binding proteins with compatible solutes, 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. This allowed us to trace the structural rearrangements that occur upon ligand binding and permitted the identification of the amino acid residues that are involved in the high affinity binding of the ligands GB and PB within the binding pocket.

MATERIALS AND METHODS

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 (Heidelberg, Germany) 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 production 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 (AmpR) and the Codon Plus RIL plasmid (CmlR). The recombinant strain was propagated on Luria-Bertani (LB) medium (45).

Overproduction and Purification of the Recombinant A. fulgidus ProX Protein—To overproduce the *A. fulgidus* ProX protein in *E. coli*, we used the *proX*$^+$ plasmid pGH26.[2] 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 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.[2] 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 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.[2] 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 (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 containing 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 A_{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 (A_{578} 0.5–0.8). 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 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, 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). After washing with 5 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 of factor Xa per 200 μg of ProX protein) for 16 h at room temperature in a buffer containing 100 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM $CaCl_2$. 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

[2] G. Holtmann and E. Bremer, unpublished results.

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 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 recovered by centrifugation (5 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 in a 10-liter flask. Cells were grown to an A_{578} of 0.3, and 100 mg liter^{-1} L-lysine, 100 mg liter^{-1} L-phenylalanine, 100 mg liter^{-1} threonine, 50 mg liter^{-1} isoleucine, 50 mg liter^{-1} leucine, 50 mg liter^{-1} valine, and 50 mg liter^{-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 h of further growth, cells were harvested by centrifugation (10 min, 3,000 × *g*). The protein was then purified from these cells as described above for the unmodified ProX; the L-selenomethionine-labeled ProX showed the same behavior in each purification step as the unlabeled protein. The incorporation of L-selenomethionine into the purified ProX protein was verified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry using a sample of the unlabeled ProX protein as a reference.

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 NaCl 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 $ZnAc_2$, 0.1 M sodium cacodylate, pH 6.0–6.5, 10–12% (w/v) PEG 4000. The crystals reached a final size of 250 × 200 × 150 μm^3 after 3 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 × 50 × 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 $MgCl_2$, 0.1 M Tris, pH 7.0–9.0, 35% (w/v) PEG4000. The first crystals appeared after 2–3 months, reached a final size of 300 × 300 × 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 $ZnAc_2$, 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^3, and belong to space group $P2_12_12_1$ (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 forms II–IV could be directly transferred into liquid nitrogen. Selenomethionine-labeled protein was crystallized in crystal form I and was treated as the native crystals prior to freezing.

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 (50) as summarized in Table I. Initial phases of the selenomethionine-labeled ProX in crystal form I were determined by anomalous dispersion at the two wavelengths λ = 0.9794 Å (λI) and λ = 1.044 Å (λII). A first set of heavy atom sites was found with SHELXD (51) but the map calculated with phases obtained by SHELXE afterward 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 that 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 8 of the 11 sites found by SHELXD turned out to be selenium, the others were treated as zinc because the highest peak in a fluorescence spectrum corresponded to that of ZnKα. Refinement of this set of sites with the program SHARP (52) located five additional sites. Finally, a set of 10 selenium sites and 6 zinc sites were refined in SHARP followed by solvent flattening using RESOLVE (53). This resulted in very good phases up to 2-Å resolution leading to an electron density map that could be automatically interpreted by Arp/Warp (54). Missing residues and minor corrections of the model were built using the graphical interface "O" (55) followed by refinement with the program REFMAC5 (56).

The liganded structure in space group $P4_32_12$ (crystal form II) was solved by molecular replacement with MOLREP (57) 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 4 copies of domain A (two for each rotation peak) could be located in the asymmetric unit that were refined using REFMAC5. To locate the B domains, a molecular replacement was carried out against the $F_o - F_c$ 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 (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 Exrasynthese (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.

Data 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-GB, ProX-TM, and ProX ligand free, respectively.

RESULTS AND DISCUSSION

Crystallization—Crystals of ProX have been obtained under four different conditions yielding crystals in four different space groups. The liganded closed form 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 trimethylamine (see Table I). Trimethylamine has a pK_a 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 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 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 $P2_1$ (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

TABLE I
Data collection statistics

Dataset	Selenomethionine derivative		Refined ProX structures			
	λI	λII	GB	PB	TM	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 (degree)	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
(α, β, γ) (degree)	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 asymetric unit	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	178,597	189,447	110,649	189,413	180,063	339,517
Unique reflections	97,236	100,862	15,622	52,348	49,535	90,922
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)
$I/\sigma(I)$	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} [a]	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-F[a]	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					

[a] R_{meas} and *R*mrgd-F are better indicators for data quality than R_{sym}, R_{meas} is redundancy independent and Rmrgd-F is a measure for the quality of the reduced amplitudes (84, 85). Values in parentheses refer to the highest resolution shell.

TABLE II
Refinement statistics

	Glycine betaine	Proline betaine	Trimethyl ammonium	Ligand-free
Resolution limits (Å)	50–2.1	50–1.9	50–1.9	50–1.8
Total number of reflections	15,622	52,348	49,535	90,922
Reflections in working set	14,849	49,688	47,096	86,386
Reflections in test set	773	2,660	2,439	4,536
R (%)	20.1	18.0	21.6	18.6
R (%)	24.9	21.9	24.5	21.4
No. of amino acid residues	270	540	540	1,080
No of protein atoms	2,161	4,322	4,322	8,644
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. deviation bonds (Å)	0.014	0.015	0.011	0.013
R.m.s. deviation angles (degree)	1.429	1.434	1.309	1.364

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, Asp^{109}–Asp^{110}-Tyr^{111}–Ala^{112} and Ala^{210}–Leu^{211}–Pro^{212}–Pro^{213}, which form with Asp^{110}-Tyr^{111} and Leu^{211}–Pro^{212}, 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 group $P2_12_12_1$. 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 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 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 Asp^{110} and Tyr^{111} and the other between Leu^{211} and Pro^{212}, 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 1.017 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.

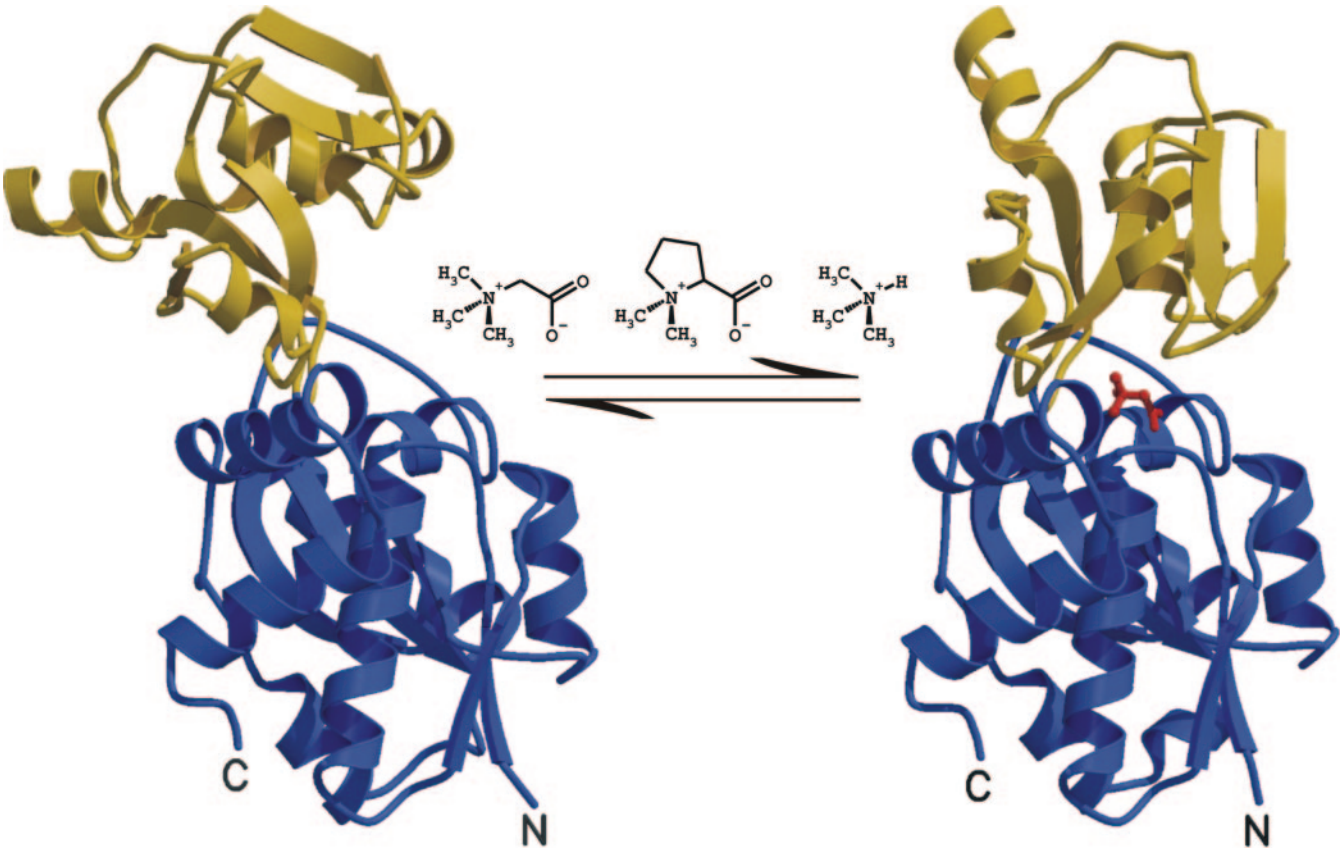

FIG. 1. **Overall structure of the open and closed conformation that ProX undergoes in a large conformational change from its unliganded open conformation (*left*) to the liganded closed form (*right*).** In between (*above* the *arrow*) the three structures, ligands are shown that have been found to induce this conformational change, GB, PB, and TM, from left to right. Domain A (*blue*) is shown in the same orientation for both conformations, whereas 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*.

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 Arg^{149}, 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 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 $Å^2$ between the open and closed forms, resulting in a buried surface increase of 476 A^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 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. 2*A* 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} (domain A), Arg^{149} (domain B), and Thr^{66} (domain A), respectively. However, the betaine ligand is not completely covered by the ligand binding site. A water molecule, which is not in direct contact with the ligand and held in place by residues Tyr^{111}, Glu^{145}, 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 (70), cation-dipole, van der Waals interaction, and non-classical hydrogen bonds (71–73). All distances between the carbon atoms bonded by the quaternary nitrogen and the tyrosine phenyl groups of Tyr^{63}, Tyr^{111}, Tyr^{190}, and Tyr^{214} as well as main chain carbonyl oxygen of Asp^{109} have been determined. These distances were compared with a compiled list of van der Waals radii published by Li and Nussinov (74). 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, 1.82, and 1.52 Å, respectively, with distance distributions having standard deviations of ~0.5 Å (74). 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 and 4 Å or between 3.2 and 3.7 Å, respectively. According to these criteria, GB forms 5, 6, 6, 3, and 3 contacts with Tyr^{63}, Tyr^{111}, Tyr^{190}, Tyr^{214}, and Asp^{109}-CO, respectively. From the arrangement of the tyrosine residues in the binding site, Tyr^{63} and Tyr^{214} from domain A, and Tyr^{111} and Tyr^{190} 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, toward the main chain carbonyl oxygen of Asp^{109}. However, the position of PB is tilted with respect to the GB position because of the proline ring (Fig. 2*B*). Because the binding site is too small to accommodate a proline ring in the direction of Tyr^{190}, PB is pushed into the direction of Tyr^{214}, resulting in a tilting of the Cα–N axis of PB with respect to that of GB. Although this escape of PB is not sufficient, the side chains of Tyr^{190} and Tyr^{214} adapt to the bulkier ligand by slightly rotating their phenyl rings of the binding site. Nevertheless, the mean distances between Tyr^{190} and Tyr^{214} and the PB liganding side chains are slightly shorter than expected from the distance distribution mentioned above (minimal distance found is 3.1 Å). PB forms 6, 8, 3, and 3 contacts with Tyr^{63}, Tyr^{111}, Tyr^{214}, and Asp^{109}-CO, respectively, with C1, C2, Cδ, and Cα. There is no direct contact of the quaternary amine to Tyr^{190}, 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 (see Fig. 2*C*). This seems to be the favored conformation for an isolated trimethylammonium or a rotable trimethylammonium group as it is the case for GB. Additionally, a chloride ion binds in the position where the carboxyl tail is bound in the ProX-GB structure. This chloride ion interacts with all residues 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 rotable 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 carbonyls. 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 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···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 Cα–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.

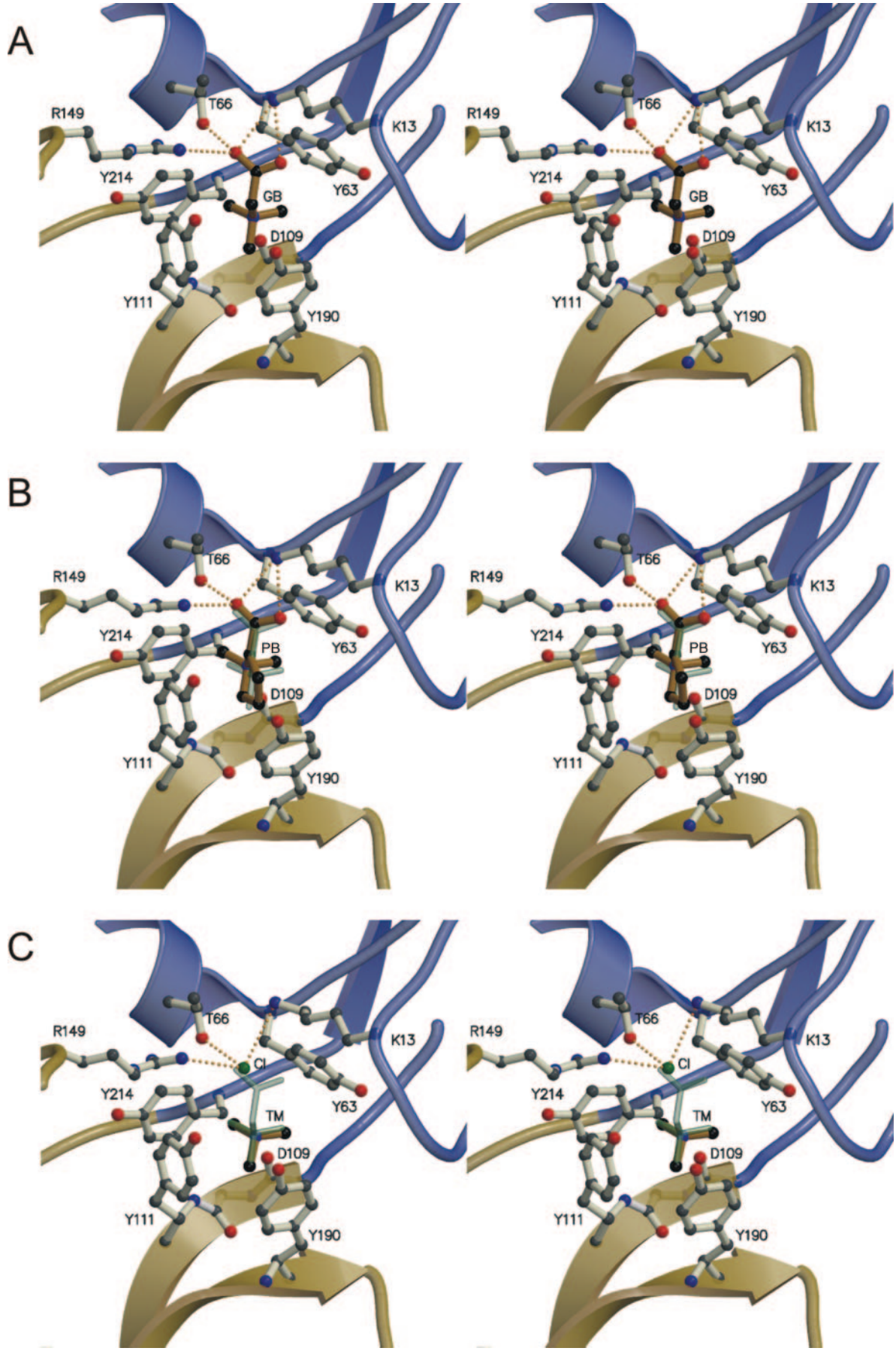

FIG. 2. **Binding site of ProX with three different ligands.** Stereo picture of the ProX binding sites with the three different ligands. The domains are colored as described in the legend to Fig. 1. All residues involved in binding are presented as a *ball-and-stick* model and labeled with their residue names. The ligands are highlighted with *brown sticks*. To compare the ligand positions, the binding sites were superimposed

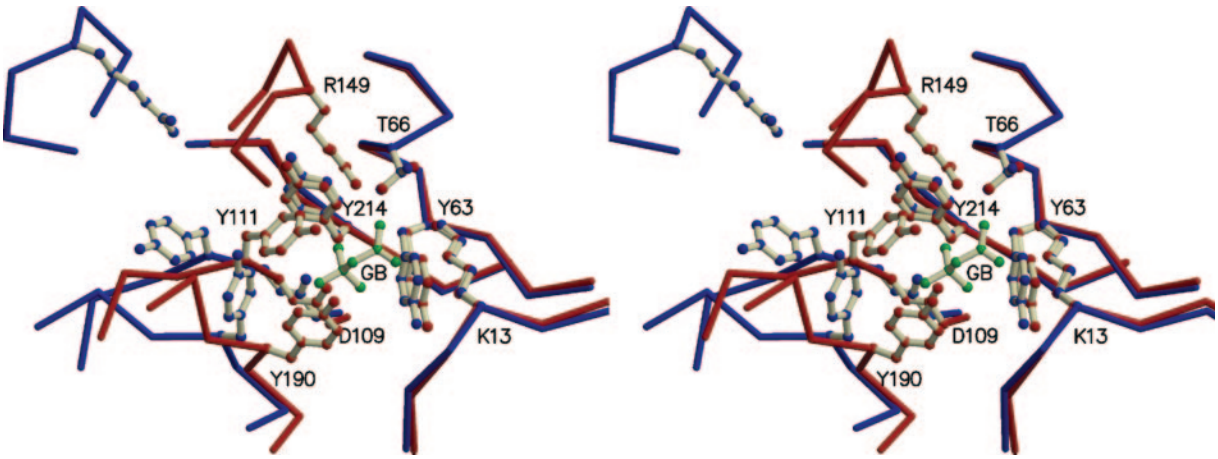

FIG. 3. **Superposition of the ProX binding site in the open and closed conformations.** Stereo picture of the superposition of ProX in the open conformation (*blue*) onto the closed conformation (*red*). GB (*green*) is shown as the ligand in the closed form. Residues involved in ligand binding and the ligand are drawn as *ball-and-stick* model and *labeled* with their residue names.

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 ligand binding show virtually the same orientation in the open and closed forms (*right part* of Fig. 3). Residues Tyr^{63}, Tyr^{214}, Lys^{13}, and Thr^{66} superimpose very well. Only the main chain carbonyl of Asp^{109} from domain A is slightly out of place compared with the closed form because of the enormous main chain rearrangement between Asp^{110} and Tyr^{111} upon closing. The residues contributed by domain B behave quite differently. Tyr^{111} and Tyr^{190} 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^{149} 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^{63}, Tyr^{214}, and Lys^{13} are stabilized either by hydrogen bond formation (Tyr^{63}–Glu^{17}, Tyr^{63}–Glu^{18}, Tyr^{214}–Asp^{151}, Lys^{13}–Thr^{45}) or salt bridges (Lys^{13}–Glu^{62}). The side chain of Thr^{66} 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$\cdots\pi$ interactions (77, 78) (Tyr^{190}–Phe^{146}, Tyr^{111}–Tyr^{190}) and hydrophobic contacts (Tyr^{190}–Pro^{188}) are formed. This changes upon closing: in the closed conformation the orientation of the phenyl rings is stabilized by C–H$\cdots\pi$ interactions (Tyr^{111}–Phe^{146} donor and acceptor being exchanged, Tyr^{190}–Phe^{15}), by van der Waals contacts (Tyr^{190}–Met^{175}, Tyr^{214}–Pro^{212}), and hydrogen bonds (Tyr^{111}–Asp^{143}; Tyr^{190}–Asp^{143}). Additionally, two interdomain hydrogen bonds (Tyr^{214}–Asp^{151}, Thr^{66}–Arg^{149}) and two salt bridges (Arg^{149}–Glu^{145}, Arg^{149}–Asp^{151}) 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^{63} and Tyr^{214}, Asp^{109}–CO as well as Lys^{13} and Thr^{66}, 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^{111} and Tyr^{190} 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-

with LSQKAB (83). *A*, shows the binding site with GB: the quaternary amine is bound by the four tyrosine side chains of Tyr^{63}, Tyr^{111}, Tyr^{190}, Tyr^{214}, and the main chain carbonyl oxygen of Asp^{109} (the interaction itself is not depicted). Salt bridge formation and hydrogen bonding between the ligand carboxylate and Lys^{13}, Arg^{149}, and Thr^{66} 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 a *gray 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 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 a *gray transparent stick* model.

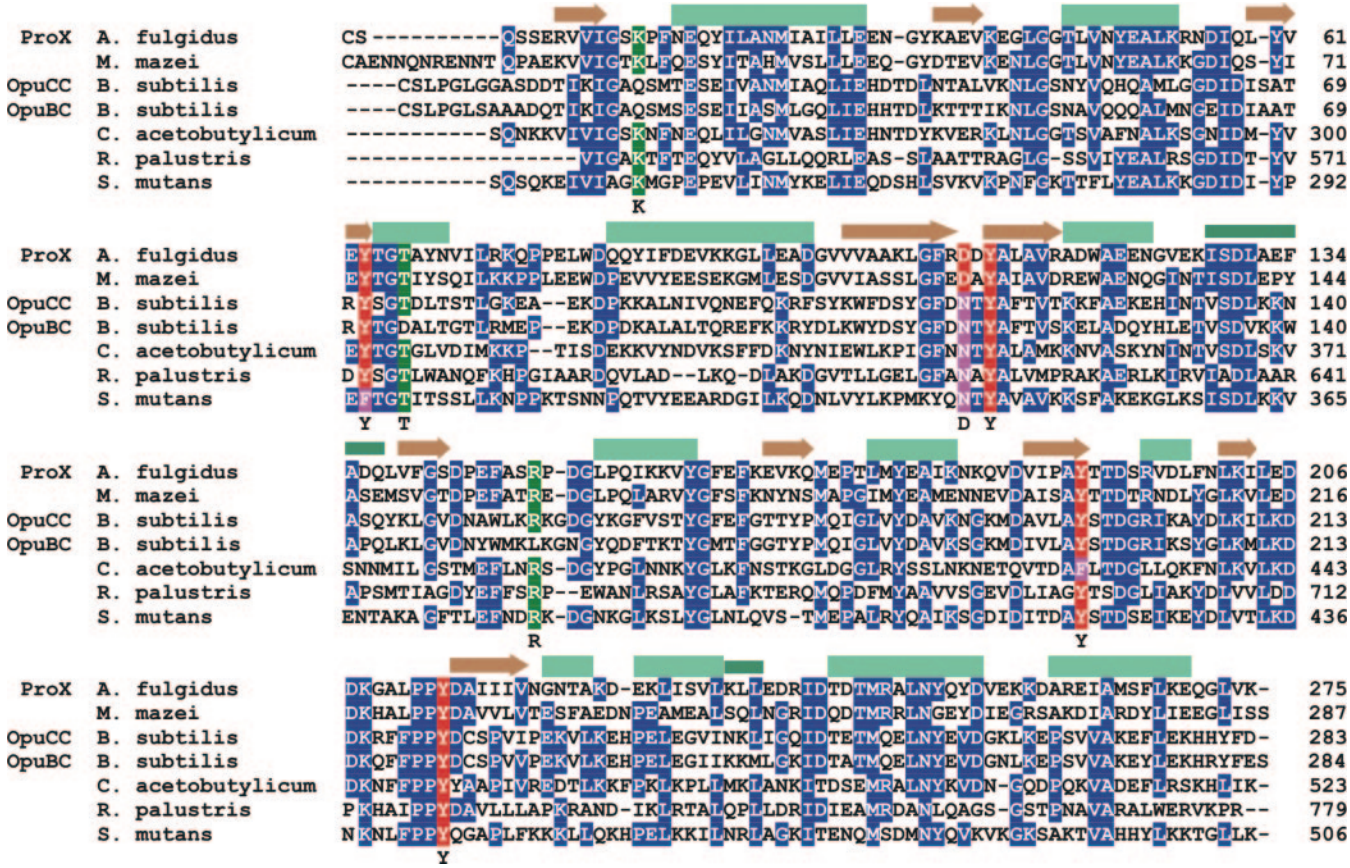

FIG. 4. **Sequence alignment between ProX from *A. fulgidus* and six related sequences from *M. mazei*, *B. subtilis* (OpuCC and OpuBC), *C. acetobutylicum*, *S. mutans*, and *R. palustris*.** Conserved residues are marked by *blue background*, conserved residues that are involved in binding of the quaternary amine head group are labeled *red*, non-conserved residues that can contribute in the same way are labeled *pink*. Residues that are involved in salt bridge or hydrogen bond formation with the ligand carboxylate of GB or PB are labeled *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.

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 Tyr^{63} in *S. mutans* and 2) the *C. acetobutylicum* position equivalent to Tyr^{190}. 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 Asp^{109}. 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, Lys^{13}, Thr^{66}, and Arg^{149}, 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, which binds choline and GB, only Lys^{13} 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.

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 $\lambda = 1.0440$ Å. 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 $\lambda = 1.0440$ Å and $\lambda = 1.0507$ Å. The six anomalous peaks above 7.5 σ were assigned to be zinc ions and have been modeled in the final structures in $P2_1$. 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 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 Glu^{238} and Asp^{239} of both molecules, the other two are coordinated by chain I Asp^{242} and Asp^{244}, and chain II Asp^{81}, 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* (Af-ProX) and ProX from *E. coli* (Ec-ProX) 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

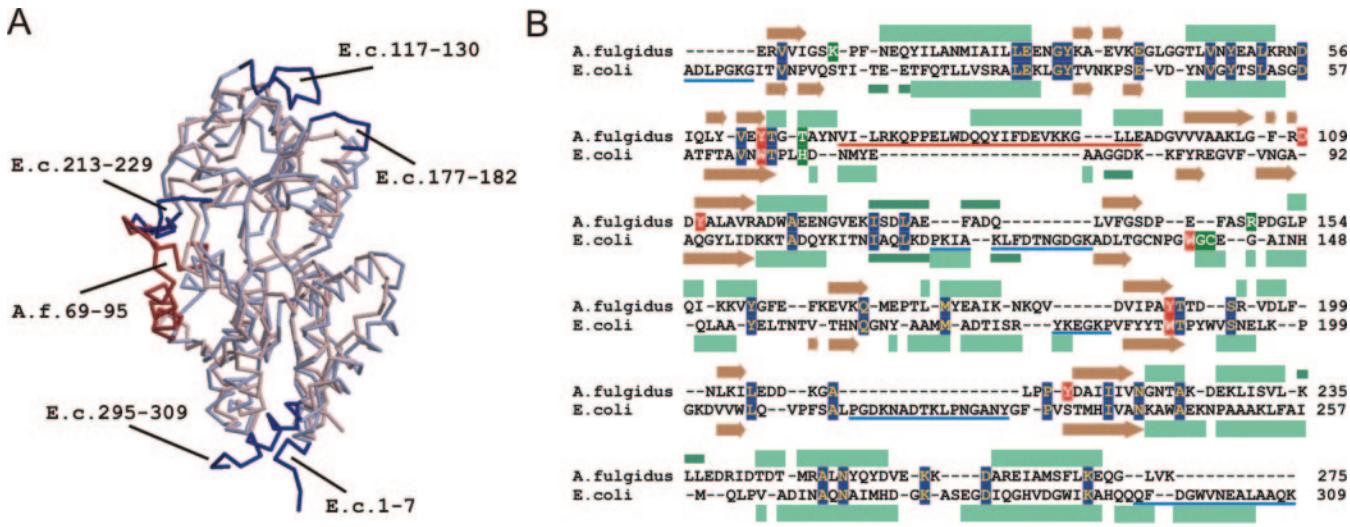

FIG. 5. **Differences in the overall ProX structures from *A. fulgidus* and *E. coli* in the closed conformation.** *A*, the structure of ProX from *E. coli* (*E.c.*) (in *light blue*) superimposed onto the structure of ProX from *A. fulgidus* (*A.f.*) (in *light red*). Larger differences with a series 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* is 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 *E. coli*. Residues conserved in both sequences are labeled *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 labeled *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 *A* of the figure are *underlined* with the same color: *E. coli* in *dark blue* and *A. fulgidus* in *dark red*.

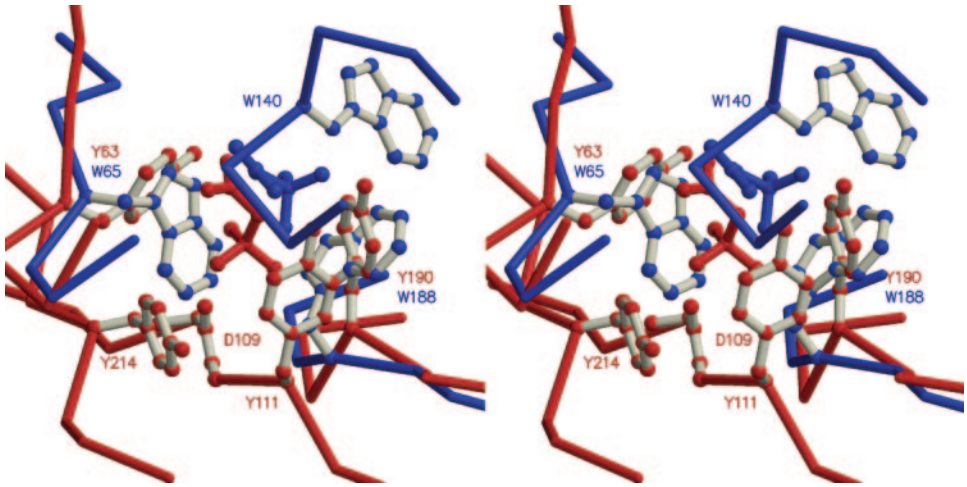

FIG. 6. **Differences in ligand binding between ProX from *A. fulgidus* and *E. coli*.** Superposition of the ligand binding sites of ProX-GB from *E. coli* onto ProX-GB from *A. fulgidus*. All residues involved in binding of the quaternary amine are drawn as a *ball-and-stick* model with *gray sticks*. Ligands and the corresponding Cα-trace are colored *blue* for ProX from *E. coli* and *red* for ProX from *A. fulgidus*.

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. 5*A* shows the Cα trace presentation of Af-ProX in *red* and Ec-ProX in *blue*. The structure-based sequence alignment is given in Fig. 5*B*. 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. 5*A*. 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 forces, 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. Noteworthy is that the Cα–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 Cα-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 $Tyr^{63} \equiv$ Ec-ProX Trp^{65} and Af-ProX $Tyr^{190} \equiv$ Ec-ProX Trp^{188}) although their spacing is given by different loops, the larger loop of 69–95 in Af-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[2] compared with Ec-ProX with a $K_D = 4$ μM (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 are equivalent as known from mutational studies of Ec-ProX (44)) being already capable of binding compatible solutes of the quaternary amine type. This ancient binding site later on might have been optimized toward higher affinity. For this purpose nature came up with two slightly different configurations based on the same molecular interactions to effectively bind compatible solutes of the quaternary amine type.

Analysis of the Thermophilic Properties—There is no shift to higher secondary structure content in the thermophilic protein from *A. fulgidus* compared with the mesophilic one of *E. coli*. The secondary structure content is very similar or even slightly lower in Af-ProX with 56.7% (18.2% β-sheet and 38.5% α-helix) compared with Ec-ProX with 57.6% (17.8% β-sheet and 39.8% α-helix).

The comparison of the amino acid compositions of Af-ProX and Ec-ProX correlate quite well with the known differences of amino acid distributions between thermophiles and mesophiles. Kumar *et al.* (82) 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, whereas 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 Af-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 Af-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 Af-ProX compared with only 8 salt bridges (5 H-bonded) in Ec-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 because of a high tryptophan abundance in Ec-ProX where three tryptophan residues are used to bind the quaternary amine of the ligand.

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

REFERENCES

1. Csonka, L. N., and Epstein, W. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhard, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) Vol. 1, pp. 1210–1223, American Society for Microbiology Press, Washington, D. C.
2. Galinski, E. A., and Trüper, H. G. (1994) *FEMS Microbiol. Rev.* **15,** 95–108
3. Kempf, B., and Bremer, E. (1998) *Arch. Microbiol.* **170,** 319–330
4. Ventosa, A., Nieto, J. J., and Oren, A. (1998) *Microbiol. Mol. Biol. Rev.* **62,** 504–544
5. Welsh, D. T. (2000) *FEMS Microbiol. Rev.* **24,** 263–290
6. Booth, I. R., and Louis, P. (1999) *Curr. Opin. Microbiol.* **2,** 166–169
7. Record, M. T., Jr., Courtenay, E. S., Cayley, D. S., and Guttman, H. J. (1998) *Trends Biochem. Sci.* **23,** 143–148
8. Wood, J. M. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 230–262
9. Brown, A. D. (1976) *Bacteriol. Rev.* **40,** 803–846
10. da Costa, M. S., Santos, H., and Galinski, E. A. (1998) *Biochem. Eng. Biotechnol.* **61,** 117–153
11. Martin, D. D., Ciulla, R. A., and Roberts, M. F. (1999) *Appl. Environ. Microbiol.* **65,** 1815–1825
12. Roessler, M., and Muller, V. (2001) *Environ. Microbiol.* **3,** 742–754
13. Burg, M., Kwon, E., and Kültz, D. (1997) *Annu. Rev. Physiol.* **59,** 437–455
14. Hohmann, S. (2002) *Int. Rev. Cytol.* **215,** 149–187
15. Le Rudulier, D., Strøm, A. R., Dandekar, A. M., Smith, L. T., and Valentine, R. C. (1984) *Science* **224,** 1064–1068
16. McNeil, S. D., Nuccio, M. L., and Hanson, A. D. (1999) *Plant Physiol.* **120,** 945–950
17. Rhodes, D., and Hanson, A. D. (1993) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44,** 357–384
18. Yancey, P. H. (1994) in *Cellular and Molecular Physiology of Cell Volume Regulation* (Strange, K., ed) pp. 81–109, CRC Press, Inc., Boca Raton, FL
19. Canovas, D., Borges, N., Vargas, C., Ventosa, A., Nieto, J. J., and Santos, H. (1999) *Appl. Environ. Microbiol.* **65,** 3774–3779
20. Courtenay, E. S., Capp, M. W., Anderson, C. F., and Record, M. T., Jr. (2000) *Biochemistry* **39,** 4455–4471
21. Lippert, K., and Galinski, A. A. (1992) *Appl. Microbiol. Biotechnol.* **37,** 61–65
22. Bourot, S., Sire, O., Trautwetter, A., Touze, T., Wu, L. F., Blanco, C., and Bernard, T. (2000) *J. Biol. Chem.* **275,** 1050–1056
23. Arakawa, T., and Timasheff, S. N. (1985) *Biochem. J.* **47,** 411–414
24. Bolen, D. W., and Baskakov, I. V. (2001) *J. Mol. Biol.* **310,** 955–963
25. Qu, Y., Bolen, C. L., and Bolen, D. W. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9268–9273
26. Lehmann, M. S., and Zaccai, G. (1984) *Biochemistry* **23,** 1939–1942
27. Brigulla, M., Hoffmann, T., Krisp, A., Volker, A., Bremer, E., and Volker, U. (2003) *J. Bacteriol.* **185,** 4305–4314
28. Caldas, T., Demont-Caulet, N., Ghazi, A., and Richarme, G. (1999) *Microbiology* **145,** 2543–2548
29. Canovas, D., Fletcher, S. A., Hayashi, M., and Csonka, L. N. (2001) *J. Bacteriol.* **183,** 3365–3371
30. Mendum, M. L., and Smith, L. T. (2002) *Appl. Environ. Microbiol.* **68,** 813–819
31. Santos, H., and da Costa, M. S. (2002) *Environ. Microbiol.* **4,** 501–509
32. Bremer, E., and Krämer, R. (2000) in *Bacterial Stress Responses* (Storz, G., and Hengge-Aronis, R., eds) pp. 79–97, American Society for Microbiology Press, Washington, D. C.
33. Sleator, R. D., and Hill, C. (2002) *FEMS Microbiol. Rev.* **26,** 49–71
34. Gowrishankar, J. (1989) *J. Bacteriol.* **171,** 1923–1931
35. Lucht, J. M., and Bremer, E. (1994) *FEMS Microbiol. Lett.* **14,** 3–20

36. Boos, W., and Lucht, J. M. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhard, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) Vol. 1, pp. 1175–1209, American Society for Microbiology Press, Washington, D. C.
37. Higgins, C. F. (1992) *Annu. Rev. Cell Biol.* **8,** 67–113
38. Stetter, K. O., Lauerer, G., Thomm, M., and Neuner, A. (1987) *Science* **236,** 822–824
39. Ghoul, M., Bernard, T., and Cormier, M. (1990) *Appl. Environ. Microbiol.* **56,** 551–554
40. 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., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Springs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1997) *Nature* **390,** 364–370
41. Quiocho, F. A., and Ledvina, P. S. (1996) *Mol. Microbiol.* **20,** 17–25
42. Magnusson, U., Chaudhuri, B. N., Ko, J., Park, C., Jones, A., and Mowbray, S. L. (2002) *J. Biol. Chem.* **277,** 14077–14084
43. Björkman, A. J., and Mowbray, S. L. (1998) *J. Mol. Biol.* **279,** 651–664
44. Schiefner, A., Breed, J., Bösser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W., and Bremer, E. (2004) *J. Biol. Chem.* **279,** 5588–5596
45. Miller, J. H. (1992) *A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
46. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
47. Dulaney, E. L., Dulaney, D. D., and Rickes, E. L. (1968) *Dev. Ind. Microbiol.* **9,** 260–269
48. Doublie, S. (1997) *Methods Enzymol.* **276,** 523–530
49. van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124
50. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26,** 795–800
51. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D* **58**, 1772–1779
52. de la Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494
53. Terwilliger, T. C. (2001) *Acta Crystallogr. Sect. D* **57**, 1755–1762
54. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
55. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
56. Murshudov, G. N. Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D* **53**, 240–255
57. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
58. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
59. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
60. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
61. Merrit, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
62. Altschul, S. F., and Gish, W. (1996) *Methods Enzymol.* **266,** 460–480
63. Berman, H. M., Westbrock, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acid Res.* **28,** 235–242
64. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) *Biochemistry* **31,** 10657–10663
65. Oh, B.-H., Pandit, J., Kang, C.-H., Nikaido, K., Gokcen, S., Ames, G. F.-L., and Kim, S.-H. (1993) *J. Biol. Chem.* **268,** 11348–11355
66. Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M., and Chothia, C. (1993) *J. Mol. Biol.* **234,** 357–372
67. Shilton, B. H., Flocco, M. M., Nilsson, M., and Mowbray, S. L. (1996) *J. Mol. Biol.* **264,** 350–363
68. Barlow, D. J., and Thornton, J. M. (1983) *J. Mol. Biol.* **168,** 867–885
69. Kumar, S., and Nussinov, R. (1999) *J. Mol. Biol.* **293,** 1241–1255
70. Dougherty, D. A. (1996) *Science* **271,** 163–168
71. Taylor, R., and Kennard, O. (1982) *J. Am. Chem. Soc.* **104,** 5063–5070
72. Derewenda, Z. S., Lee, L., and Derewenda, U. (1995) *J. Mol. Biol.* **252,** 248–262
73. Weiss, M. S., Brandl, M., Sühnel, J., Pal, D., and Hilgenfeld, R. (2001) *Trends Biochem. Sci.* **26,** 521–521
74. Li, A.-J., and Nussinov, R. (1998) *Proteins* **32,** 111–127
75. Pearson, R. G. (1963) *J. Am. Chem. Soc.* **85,** 3533–3539
76. Scheiner, S., Kar, T., and Gu, Y. (2001) *J. Biol. Chem.* **276,** 9832–9837
77. Steiner, T., and Koellner, G. (2001) *J. Mol. Biol.* **305,** 535–557
78. Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J., and Hilgenfeld, R. (2001) *J. Mol. Biol.* **307,** 357–377
79. Mao, B., Pear, R., and McCammon, J. A. (1982) *J. Biol. Chem.* **257,** 1131–1133
80. Miller, D. M., Olson, J. S., Pflugrath, J. W., and Quiocho, F. A. (1983) *J. Biol. Chem.* **258,** 13665–13672
81. Kleywegt, G. J., and Jones, T. A. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 9–14
82. Kumar, S., Tsai, C.-J., and Nussinov, R. (2000) *Protein Eng.* **13,** 179–191
83. The CCP4 Suite (1994) *Acta Crystallogr. D* **50,** 760–763
84. Diederichs, K., and Karplus, P. A. (1997) *Nat. Struct. Biol.* **4,** 269–275
85. Weiss, M. S., and Hilgenfeld, R. (1997) *J. Appl. Crystallogr.* **30,** 203–205