Cation-π Interactions as Determinants for Binding of the Compatible Solutes Glycine Betaine and Proline Betaine by the Periplasmic Ligand-binding Protein ProX from *Escherichia coli*

Received for publication, September 3, 2003, and in revised form, November 5, 2003
Published, JBC Papers in Press, November 11, 2003, DOI 10.1074/jbc.M309771200

Andre Schiefner‡§, Jason Breed‡§§, Linda Bösser, Susanne Kneip, Jutta Gadei,
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, Philips-Universität Marburg, Karl-von-Frisch Strasse, D-35032 Marburg, Germany*

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

The water content of bacterial cells is determined solely by osmotic processes as bacteria lack systems for active water transport into or out of the cell in response to an increase or a decrease of osmolarity in the environment (1, 2). To cope with hyperosmotic conditions, non-halophilic microorganisms generally amass large quantities of a particular group of organic osmolytes, the so-called compatible solutes (3), either by de novo synthesis or by direct uptake from the environment (1, 4–7). Compatible solutes are operationally defined as organic osmolytes that can be accumulated by the cell to exceedingly high concentrations without disturbing vital cellular functions or the correct folding of proteins (3). Their intracellular accumulation counteracts the osmotic efflux of water from the cell and thus contributes to the restoration of turgor (8) and the resumption of growth of the bacterial cells under conditions of low water activity. In addition to contributing to the maintenance of cellular water content at high external osmolarities (9), compatible solutes also counteract destabilization of native structures of proteins upon freezing, heating, desiccation, and exposure to high ionic strength both in vitro and in vivo (10–15).

The intracellular accumulation of compatible solutes as a strategy for adaptation to high osmolarity has been widely adopted not only by bacteria (5) and Archaea (16, 17) but also by fungal, plant, animal, and even human cells (18–20). Moreover a few classes of compounds are used universally throughout the kingdoms, reflecting fundamental constraints on the type of solutes that are compatible with macromolecular and cellular functions (16, 21, 22). One of the most ubiquitous compatible solutes used both by prokaryotic and eukaryotic cells, is the trimethylammonium compound glycine betaine (*N,N,N-trimethylglycine, GB*). The exact biochemical mechanism(s) through which compatible solutes act as protein stabilizers is not completely understood. Their functioning is generally explained in terms of the preferential exclusion model (23). This hypothesis predicts that compatible solutes are excluded from the immediate hydration shell of proteins, which is apparently due to unfavorable interactions with protein surfaces (24, 25). This causes a non-homogeneous distribution of compatible solutes within the cell water and a preferential hydration of protein surfaces. Experimental verification came from neutron diffraction studies (26). The nonuniform distribution of compatible solutes within the cell water results in a thermodynamic force that drives proteins to occupy a smaller volume to reduce the amount of excluded water, thereby stabilizing the native structure of proteins.

Many Gram-negative and Gram-positive bacterial species that have been found to accumulate GB are able to acquire this compatible solute from environmental sources through high affinity transport systems (1, 6, 27, 28). GB transport in *Escherichia coli* is under osmotic control (27, 29, 30) and is mediated by two transport systems: the single component H⁺-compatible

1 The abbreviations used are: GB, glycine betaine (*N,N,N-trimethylglycine*); PB, proline betaine (*N,N-dimethyl-L-proline*); X-gal, 5-bromo-4-chloro-3-indolyl-D-galactopyranoside; PIPES, 1,4-piperazine diethanesulfonic acid; NCBI, National Center for Biotechnology Information.
solute cotransporter ProP (31) and the multicomponent transport system ProU (32, 33). The ProU system is a member of the ATP-binding cassette superfamily (34, 35) consisting of two cytoplasmic, membrane-associated ATPases, ProV; two integral membrane proteins, ProW; and a periplasmic ligand-binding protein, ProX (33). Access of GB present in the environment to the periplasmic ProX protein is provided by passive diffusion across the outer membrane through the OmpC and OmpF general porins of E. coli (36). ProX binds GB avidly with a KD of ~1 μM (32, 37, 38) delivering the ligand to the substrate translocation complex that is embedded in the cytoplasmic membrane. Driven by ATP hydrolysis, the translocation complex then transports the ligand into the cytoplasm (34, 35, 39). In addition to GB, the ProU transporter also serves as a high affinity uptake system for proline betaine (N,N-dimethyl-L-proline, PB), and this compatible solute is recognized by the ProX protein with high affinity (Kd = 5 μM) as well (38). ProU also functions as an uptake system for a variety of other compatible solutes (e.g., proline and ectoine), but in contrast to GB and PB none of these compounds appear to be recognized by the ProX protein (37, 38, 40).

The substrate-binding protein binds its ligand(s) selectively and with high affinity, which is thought to ensure the substrate specificity and directionality of the overall transport reaction for a given binding protein-dependent transport system (35). Periplasmic binding proteins are generally composed of two domains connected by one to three polypeptide chains forming a hinge between them. In the ligand-free open conformation the two rigid domains are flexibly linked by the hinge. This has been shown by the structural analysis of various open states of the ribose-binding protein and α-allose-binding protein (41, 42). Ligand binding induces a large conformational change in the hinge region that moves both domains toward each other. After this movement domain the ligand is engulfed in a predefined cleft between the two domains that refers to the closed conformation of the binding protein.

Compatible solutes are apparently preferentially excluded from the immediate hydration shell of proteins (23–25). Yet a ligand-binding protein such as ProX binds the compatible solutes GB and PB with high affinity and high specificity (32, 37, 38, 43). This observation raises the question as to how such a high affinity interaction between a compatible solute and a protein can take place. To understand the molecular determinants that govern substrate recognition and binding of compatible solutes by ProX, we determined the high resolution crystal structure of the ProX protein in complex with each of its ligands, i.e., ProX-GB and ProX-PB. This crystallographic analysis was combined with site-directed mutagenesis of the residues forming the ligand binding pocket of ProX to determine their relative importance for high affinity binding of the compatible solute GB by the ProX protein from E. coli. The analysis revealed cation–π interactions between the quaternary amine of GB and PB and the indole groups of three tryptophan residues.

**EXPERIMENTAL PROCEDURES**

**Overproduction and Purification of Wild-type and Mutant ProX Protein**—The wild-type E. coli ProX protein was overproduced and purified to homogeneity essentially as described previously (44). In brief, the proX gene was overexpressed under the control of the bacterial phage T7Δ10 promoter present on the proX lower copy number plasmid pSK7 (Cm') in the E. coli strain PD141 (AΔE3) (44). This strain harbors a chromosomal copy of the gene for the phage T7 RNA polymerase under the control of an isopropyl-1-thio-β-D-galactopyranoside-inducible lacPOlacI promoter system, thereby allowing the selective expression of the proX gene present on the plasmid pSK7 from the T7Δ10 promoter upon the addition of isopropyl-1-thio-β-D-galactopyranoside to the cell culture. Cells of strain PD141 (AΔE3) (pSK7) were grown in a minimal medium (44) in the presence of 30 μg ml⁻¹ chloramphenicol because components of rich media (e.g. yeast extract) are known to contain GB (45) that could potentially be scavenged and bound by ProX. In this way we ensured that the purified ProX protein was free of ligand since the E. coli strains used for overproduction are unable to synthesize GB. Overproduction of ProX was initiated by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside when the cell culture had reached an A570 of ~1.0–1.5. After 1 h of further growth, the cells were harvested by centrifugation, and the periplasmic proteins were released from the cells by cold osmotic shock (46). Insoluble material was removed from the released periplasmic proteins by ultracentrifugation, and the supernatant was then subjected to fast protein liquid chromatography on a DEAE-Sepharose fast flow column (Amersham Biosciences). Proteins were eluted from this anion exchange column with an increasing Tris-

**Site-directed Mutagenesis of the proX Gene**—To probe the contributions of residues Trp65, Trp140, or Trp188 (pLB2, W188A; pLB3, W188L; pLB4, W65Y; pLB10, W65A; pLB11, W65L; pLB12, W65F; pLB13, W65Y; pLB18, W65D; and pLB19, W65E) as well (38). ProU was cloned and amino acid sulfate dissolved in 10 mM Tris-HCl buffer. The ProX protein was eluted with a decreasing ammonium sulfate gradient (1.5–0 mM ammonium sulfate dissolved in 10 mM Tris-HCl, pH 8.3). The purified ProX protein was then dialyzed overnight against 5 liters of 10 mM Tris-HCl (pH 7.3) buffer. The ProX protein was free from other contaminating proteins as judged by SDS-polyacrylamide gels stained with Coomassie Brilliant Blue.

ProX proteins with mutations in the substrate binding site were purified essentially as wild-type ProX protein except that the E. coli strain LinE2 (PD141 ΔproU::spc6088) was used for the overexpression of the mutant proX genes present on derivatives of plasmid pSK7. Strain LinE2 carries a chromosomal deletion of the entire proU operon (38), thereby avoiding the contamination of the mutant ProX protein preparations with the wild-type ProX protein during protein purification.

**Genetic Construction of Bacterial Strains for Expression of Mutant ProX**—To construct an E. coli strain that lacked the E. coli proU operon we would allow phage λD10-mediated transduction of plasmid-encoded mutant proU genes, we prepared a P1vir lysate (47) on the proU lacZ fusion strain BK16 (MC4100 proU::lacZ2h2::λp lacMu15 (Kan')). Using this phage lysate, we transduced the lacZ gene fusion positioned next to the λp lacMu15 (Kan') prophage (49) into the E. coli strain PD141 (MC4100 (ΔDE3)) by selecting for kanamycin-resistant colonies (50 μg ml⁻¹) in the presence of X-gal on LB agar plates. One of these transductants was strain LinE1. We then replaced the proU::lacZ2h2::λp lacMu15 (Kan') prophage in strain LinE1 with the ΔproU::spc6088 deletion from strain MKH13 (38) by transducing strain LinE1 with a P1vir lysate prepared on strain MKH13 and selecting for spectinomycin-resistant colonies (100 μg ml⁻¹) in the presence of X-gal on LB agar plates. One of these transductants carrying the ΔproU::spc6088 deletion were identified as spectino-

1 B. Kempf and E. Bremer, unpublished results.
2 P. Dersch and E. Bremer, unpublished results.
ments of the amino acids at positions Trp65 and Trp140 were constructed by remutating plasmid pLB10 to yield plasmid pLB20 (W65A, W140A). The mutant proX pLB plasmids were each introduced by electroporation into the E. coli strain LinE2 (ΔproU::E::spc608), and 4 liters of minimal medium with 0.4% (w/v) glucose as the carbon source were used for the overproduction of the mutant ProX proteins. Usually we obtained between 10 and 20 mg of purified ProX protein from 4 liters of cell culture.

Binding of Radiolabeled GB to ProX and Its Mutant Derivatives—

The binding affinity of the wild-type ProX protein and its mutant derivatives constructed by site-directed mutagenesis of proX was measured by using an ammonium sulfate precipitation assay (52). Purified ProX protein (5 μM, determined using the BCA assay) was incubated with nine different concentrations of radiolabeled GB (1, 2, 5, 7, 10, 20, 30, 50, and 110 μM) in a 100-μl reaction volume of 10 mM Tris-HCl (pH 7.3) for 5 min at room temperature. Then the ProX protein was precipitated with 900 μl of ice-cold saturated ammonium sulfate solution. After an incubation of 10 min on ice, the precipitated ProX protein was collected by filtration onto a nitrocellulose filter (pore size, 0.45 μm; Schleicher & Schuell), the filter was then washed with 10 ml of an ice-cold saturated ammonium sulfate solution, and the radioactivity retained by ProX on the filter was determined by scintillation counting. Each measurement was repeated three times for each concentration to determine the binding constant and the S.D. A mutant is treated as non-binding if no binding, either specific or unspecific, occurs at a concentration of 110 μM.

Binding of Compatible Solutes by Cation-π Interactions

TABLE I

Intensity statistics of the heavy atom derivatives and native data sets

| Data set | Derivative | Native |
|----------|------------|--------|
| Ligand   | Glycine betaine (CH$_3$)$_2$PbAc | Glycine betaine | Glycine betaine Proline betaine |
| Wavelength (Å) | 1.54179 | 1.54179 | 0.8439 |
| Total rotation range (°) | 100 | 100 | 100 |
| Temperature (K) | 291 | 291 | 291 |
| Cell (a, b, c) (Å) | 48.0, 55.0, 115.8 | 47.9, 55.1, 115.7 | 47.59, 53.86, 115.43 |
| Resolution limits (Å) | 20.0–2.48 (2.55–2.48) | 20.0–2.48 (2.60–2.48) | 15–1.57 (1.67–1.57) |
| Number of reflections | 43,801 | 43,602 | 142,884 |
| Unique reflections | 20,031 | 20,826 | 40,867 |
| Completeness (%) | 95.4 (95.1) | 99.2 (99.5) | 98.6 (95.2) |
| I/σ(I) | 15.7 (6.3) | 19.0 (10.0) | 30.0 (16.8) |
| R$_{c}$sym | 4.1 (12.0) | 2.6 (6.7) | 2.6 (4.7) |
| R$_{merge}$ | 5.3 (15.5) | 3.6 (8.8) | 3.0 (5.5) |
| R$_{meas}$ | 6.4 (16.6) | 3.6 (7.9) | 2.2 (4.8) |
| B-factor Wilson plot (Å$^2$) | 27.8 | 32.4 | 7.7 |
| Phasing power overall | 1.39 | 1.33 | 30.3 |
| Figure of merit overall | 0.62 | |

$^a$ R$_{meas}$ and R$_{merge}$ are better indicators for data quality than R$_{sym}$. R$_{meas}$ is redundancy-independent, and R$_{merge}$ is a measure for the quality of the reduced amplitudes (48, 50).
periplasmic binding proteins specific for ferrichrome and vitamin B₁₂, respectively (64, 65). These proteins have only one switch between their two domains. According to this classification, ProX has a periplasmic binding protein type II fold.

The Ligand Binding Site—As expected from other binding protein structures such as putrescine-binding protein, dipeptide-binding protein, leucine-binding protein, and leucine/isoleucine/valine-binding protein with Protein Data Bank accession codes 1A99, 1DPE, 2LPB, and 2LIV, respectively (66), also contain disulfide bonds, all of these are placed distantly from the ligand binding site. The carboxylic group of the GB ligand protrudes out of the indole box forming hydrogen bonds with the backbone amide hydrogens of Gly141 and Cys142 of domain B. These two residues are part of the loop between the disulfide-linked residues Cys136 and Cys142, which also contains the indole ring of Trp140. Additionally, the carboxylic group forms a hydrogen bond with the imidazole ring of His89 from domain A (Fig. 3).

Residues involved in the binding of PB are essentially the same as can be seen in Fig. 4 showing a superposition of the ProX-PB structure (ligand, yellow) onto the ProX-GB structure (ligand, blue). The two methyl groups, C-1 and C-2, as well as the Cβ (from the proline ring) of the quaternary amine form similar contacts with the three indole rings, and the carboxylic group forms hydrogen bonds identical to those in ProX-GB. Also the PB ring atoms Cβ and Cy contribute additional van der Waals interactions with the indole ring of Trp65 and Cy with Leu68 side chain as well.

Mainly there are two contributions involved in the binding of a quaternary ammonium group in an aromatic environment, cation–π and van der Waals interactions. To understand the nature of the interaction between the tryptophans and the ligand, all distances between the carbon atoms bonded to the quaternary nitrogen and the ring atoms of the three indole groups were determined (Fig. 5). These distances were compared with a compiled list of van der Waals radii published by Li and Nussinov (67). If the influence of the quaternary amonium charge on the contact distances is neglected, the methyl or methylene groups and the aromatic ring atoms possess van der Waals radii of 1.92 and 1.82 Å, respectively. This results in a mean distance of 3.74 Å between those carbon atoms having a S.D. of −0.5 Å (67). We therefore consider an indole ring atom to be in contact with a methyl or a methylene group if their mutual distance is between 3.5 and 4 Å. The indole ring atoms that fulfill this criterion are color-coded in Fig. 5. According to this criterion, GB forms nine, five, and four contacts with Trp188, Trp65, and Trp140 indole groups, respectively. The number of contacts of the indole rings suggests that the three indole groups contribute differently to the enthalpy of GB binding.

---

**TABLE II**

| Glycine betaine | Proline betaine |
|-----------------|-----------------|
| Resolution limits (Å) | 12.26–1.59 | 15.0–2.05 |
| Total number of reflections | 39,378 | 18,430 |
| Reflections in working set | 37,409 | 18,626 |
| Reflections in test set | 1,969 | 804 |
| R (%) | 15.0 | 17.2 |
| R<sub>free</sub> (%) | 18.6 | 21.3 |
| No. of amino acid residues | 309 | 309 |
| No. of protein atoms | 2,390<sup>a</sup> | 2,382 |
| No. of metal ions | 1 | 1 |
| No. of water molecules | 429 | 59 |
| B-factor for all atoms (Å<sup>2</sup>) | 13.5 | 29.0 |
| B-factor for all atoms (Å<sup>2</sup>) | 11.7 | 29.0 |
| r.m.s.d.<sup>b</sup> bonds (Å) | 0.013 | 0.010 |
| r.m.s.d. angles (°) | 1.385 | 1.239 |

<sup>a</sup> Side chains of Asn<sup>50</sup> and Met<sup>100</sup> are modeled in two alternate conformations.

<sup>b</sup> r.m.s.d., root mean square deviation.
Trp$^{188}$ is contacted by each of the three methyl groups because the Ca–N bond is oriented perpendicular to the aromatic bottom of the box (Fig. 3). Trp$^{65}$ is contacted by the methyl group C-1 and the Ca, and Trp$^{140}$ is contacted only by the methyl group C-2.

In the case of PB, seven, two, and five contacts are found of the quaternary amine with the indole rings of Trp$^{188}$, Trp$^{65}$, and Trp$^{140}$, respectively. One difference compared with the ProX-GB structure is that the Ca is no longer in contact distance with the indole ring of Trp$^{65}$, having been pushed aside by the PB ring atoms Cβ and Cγ. These ring atoms form nine new van der Waals contacts with Trp$^{65}$, colored green in Fig. 5 because they are not part of the quaternary amine.

**Metal Binding Site**—A metal binding site distant from the ligand binding site has been identified in the ProX structure. The metal ion is shown in Fig. 1 as a gray sphere in domain B between helix 5 and strand 5. Although the protein was never exposed to a buffer containing any metal cations during purification and crystallization an ion has been identified to be calcium (68). The octahedral coordination of the ion is provided by five ligands of the surface-exposed loop in domain B between helix 5 and strand 5 (Asp$^{128}$, Asn$^{129}$, Asp$^{128}$, Lys$^{130}$, and Asp$^{132}$) and completed by a water molecule as the sixth ligand. This arrangement of the ligand binding site resembles an EF-hand motif. A similar situation has been found in the aldolase-binding protein where the metal ion has been identified to be calcium (68).

Whereas the coordination of the metal is rather indicative for a calcium ion the anomalous signal at the measured wavelength is pointing to a transition metal. It was not possible to decide which metal is actually bound by ProX. The loop from residues 125 to 130 that is apparently stabilized by the metal ion seems to have no function.

**Differentiation of the Three Indole Groups as Assessed by Mutations and Structural Data**—To estimate the contributions of each of the indole groups of the tryptophan box, site-specific mutants were designed, and their relative binding affinity for GB was compared with that of the wild type. The binding affinity for the wild-type ProX measured by the ammonium sulfate precipitation is in very good agreement with a $K_D$ value of 1.4 μM determined by a release assay (32). All three tryptophan residues were consecutively substituted by tyrosine, phenylalanine, alanine, aspartate, or glutamate as summarized in Table III. The data demonstrate that a phenyl ring can provide a contribution to the binding affinity similar to that of an indole ring ($K_D$ = 3–5 μM). In contrast, non-aromatic replacements of these residues impair the binding affinity considerably. The strongest influence of non-aromatic mutations is at position 188 where no binding affinity is detectable in the Ala, Leu, Asp, and Glu mutants. Position 65 is to a certain degree tolerant against Ala ($K_D$ = 50 μM) but does not tolerate leucine and acidic residues, whereas position 140 tolerates Ala, Leu, Asp, and Glu substitutions with dissociation constants being 2–10 times larger than for the wild type (Table III). Thus, the binding affinity is most sensitive against substitutions at position 188. The weakest sensitivity is at position 140 confirming the relative importance of the single tryptophans derived from the number of contacts with the GB (Fig. 5). Nevertheless the two residues Trp$^{65}$ and Trp$^{140}$ are necessary for the binding as seen from the double mutant W65A,W140A (see Table III).

**A Conserved Sequence Motif**—A BLAST search in the NCBI data base identified many proteins that are homologous to ProX. For clarity we selected a few representatives of the taxonomic classes. The selection is shown in Fig. 6; the *E. coli* sequence is followed by sequences of three Enterobacteriaceae (same family), one Vibrionales (same subdivision), three Pseudomonales (same subdivision), one Rhodospirillales (same
phyllum), three Rhizobiales (same phylum), and one Cyanobacteria (same kingdom). It can be seen that the tryptophan residues are well conserved in this group of binding proteins, and in some cases they are replaced by tyrosine residues (see Fig. 6). A restraint on the sequence is the proper orientation of the three tryptophan side chains achieved by the surrounding residues, which can also be inferred from the alignment. In ProX the second residue after each of the tryptophans 65 and 188 is a conserved proline in cis-conformation, which allows for a sharply bent backbone. Pro67 is placed between the two residues Trp65 and His69, which both interact with the ligand (see Fig. 6). In the loop containing Trp140, two conserved glycine and cysteine residues are stabilized by a disulfide link between the two cysteines Cys136 and Cys142. Again the sharp bend of the loop may require Pro138, which is also conserved. This entire motif 136CXPGWGC142 is strictly conserved among the five closest homologues of ProX (see Fig. 6).

Other Cases of Ligands Bound by Cation-π Interaction—Several kinds of ligands with quaternary amines have been observed to form van der Waals contacts with proteins. In a phosphocholine-binding antibody McPC603 (Protein Data Bank accession code 2MCP) two methyl groups of the quaternary amine form contacts with the indole group of Trp, and one methyl forms contacts with the phenyl ring of Tyr.

To understand whether there is a general architecture of these quaternary amine binding sites in proteins we superimposed ProX-GB with two other structures having completely different folds and substrates. For the comparison we used the structures of the phosphatidylcholine-binding protein (72) and the HP1 chromodomain (73) with Protein Data Bank accession codes 1LN1 and 1KNE, respectively. Because of their different folds only the quaternary amine head group of their ligands was superimposed using LSQMAN. Fig. 7 shows those amino acid side chains, which contain at least one atom closer than 4.0 Å to the bound quaternary amine. As can be seen aromatic residues equivalent to Trp65 and Trp138 are also present in the two other proteins, whereas the third residue of the aromatic box is less conserved.

Table III

| Binding affinities of GB to wild-type ProX and ProX derivatives |
|-------------|----------------|----------------|----------------|
|             | $K_D$ S.D.       | $K_D$ S.D.       | $K_D$ S.D.       |
|             | $\mu M$ | $\mu M$ | $\mu M$ | $\mu M$ | $\mu M$ | $\mu M$ |
| Wild type   | 4 0.1  4 0.1  4 0.1 | 4 0.1  3 0.3  4 0.4 | 4 0.1  3 0.3  4 0.4 |
| Trp ↔ Phe   | 13 0.6  4 0.5  5 0.2 | 13 0.6  4 0.5  5 0.2 | 13 0.6  4 0.5  5 0.2 |
| Trp ↔ Ala   | 50 3.3  56 10 —  —  | 50 3.3  56 10 —  —  | 50 3.3  56 10 —  —  |
| Trp ↔ Leu   | —    —   —   14 1.4 —  —  | —    —   —   14 1.4 —  —  | —    —   —   14 1.4 —  —  |
| Trp ↔ Asp   | —    —   —   8 0.8 —  —  | —    —   —   8 0.8 —  —  | —    —   —   8 0.8 —  —  |
| Trp ↔ Glu   | —    —   —   21 2.5 —  —  | —    —   —   21 2.5 —  —  | —    —   —   21 2.5 —  —  |
| Trp65 ↔ Ala, Trp140 | —    —   —   —  —  | —    —   —   —  —  | —    —   —   —  —  |
| — ↔ Ala     | —    —   —   —  —  | —    —   —   —  —  | —    —   —   —  —  |

---

$\mu M$ — indicates non-binding ProX mutants. These proteins show no GB binding at the maximum concentration of 110 $\mu M$ used in the experiments.

Fig. 5. Carbon-carbon distances in Å between the ligands GB (a) and PB (b) and the carbon atoms of the indole groups of the three tryptophan residues. All atoms of the indole rings involved in cation-π interaction (matching the criterion to be within 3.5–4.0Å distance) are in blue for GB and yellow for PB (color-coding for the ligand is the same as in Fig. 4). The nomenclature for PB is similar to that for GB except that 3 in GB is, as C6, part of the proline ring in PB. The quaternary amine of GB forms nine, five, and four contacts, whereas the quaternary amine of PB forms seven, two, and five contacts with the indole groups of Trp, Trp, and Trp, respectively. PB also forms van der Waals contacts between Cβ and Cγ with the indole ring of Trp shown in green.
DISCUSSION

Binding a Compatible Solute to a Protein—Compatible solutes are of practical interest as they can stabilize labile proteins in a functionally active form in vitro over extended periods or act as cryoprotectants. The protective value has been traced to their exclusion from direct contact with the protein surface. As the surface increases in the initial conformational transitions associated with denaturation, denaturation is energetically disfavored in presence of a compatible solute (23, 75–77).

This work addresses the questions as to how such a substance known to avoid interaction with protein surfaces can nevertheless be complexed with high affinity by a protein and what are the critical structural features of such a binding site.

Compared with other small ions or dipolar molecules the quaternary amine group is exceptionally bulky. The positive charge is virtually distributed over a larger volume compared with a metal ion, resulting in a smaller surface potential. The underlying reason for the unfavorable interaction must be a successful competition of water with quaternary ammonium compounds for binding sites on the protein surface. For common protein surfaces the preferred interaction partner may be water because of its small size and consequently stronger surface potential so that it can form an electrically and sterically complementary interface with the protein surface that is better than the quaternary ammonium ion. This is in accord with explanations given for other substances classified as compatible solutes, which are found to be preferentially excluded from

FIG. 6. Alignment of ProX with selected sequences from a homology search. Conserved residues are shown in blue, residues involved in binding of the quaternary amine head group are red (equivalent residues are pink), conserved cysteine residues are green, and structurally important residues are marked olive. Secondary structure elements for the ProX structure were determined using DSSP (82). Helices (light blue) and sheets (brown) are numbered according to Fig. 1. Shown are E. coli K12 (NCBI accession no. 16130593), Shigella flexneri 2a str. 301 (NCBI accession no. 24113983, 99% identity), Salmonella typhimurium LT2 (NCBI accession no. 16766122, 83% identity), Yersinia pestis (NCBI accession no. 22125121, 71% identity), Vibrio paraeaeolyticus (NCBI accession no. 28885052, 66% identity), Pseudomonas putida (NCBI accession no. 7842433, 51% identity) Pseudomonas syringae pv. tomato (NCBI accession no. 28870253, 48% identity), Pseudomonas aeruginosa PA01 (NCBI accession no. 15600586, 24% identity), Rhodopseudomonas rubrum (NCBI accession no. 22968752, 63% identity), Sinorhizobium meliloti (NCBI accession no. 22125121, 26% identity), Mesorhizobium loti (NCBI accession no. 13476104, 25% identity), Brucella melitensis (NCBI accession no. 17988895, 25% identity), and Trichodesmium erythraeum (NCBI accession no. 23041742, 42% identity).

FIG. 7. Superposition of three quaternary amine binding proteins. The quaternary amine in the center of the binding site (black) was superimposed using LSQMAN (74). Each of the protein residue side chains and the quaternary amine elongations (or parts of them, in brackets) are color-coded: red, ProX (GB); yellow, phosphatidylcholine transfer protein (phosphocholine); blue, chromodomain HP1 (the lysine side chain has three methyl groups attached to the terminal amino group). All amino acid side chains of the proteins that have at least one atom closer than 4 Å to the quaternary amine (QA) are shown in the figure.
the protein surface. In case of ribonuclease and the compatible solute glycerol, the preferential hydration has been directly proven by neutron small angle scattering (26). Furthermore this exclusion of compatible solutes also resembles the phenomenon of excluded volume, which generally exists in concentrated solutions of large molecules (78).

However, if proteins need to bind such quaternary ammonium cations, a site can be tailored as it is found in ProX. This special site should be a cavity just large enough to accommodate the bulky cation and has to possess an evenly negative surface potential. Indole and phenyl groups from tryptophans, tyrosines, and phenylalanines have been shown to possess such an evenly negative surface potential (79). In ProX, the rectangular indole box is well designed to accommodate quaternary amine ligands and thus provides electrostatic and van der Waals complementarity. An important structural feature are two conserved cis-prolines, which follow the residues Trp65 and Trp188 and might be also in cis-conformation in the related structures (see Fig. 6). These seem to play a crucial role in positioning Trp65 and Trp188 by allowing a sharp bend in the polypeptide backbone to form this almost rectangular box of tryptophan residues as found in ProX. The interaction energy has thus been well optimized under evolutionary pressure.

Mutational Studies—Our mutational data show that at position 188 the phenyl ring of tyrosine or phenylalanine is able to substitute for the indole group, whereas binding is abolished in the non-aromatic mutants. In ProX-GB residue Trp188 forms more van der Waals contacts (nine) than the other two tryptophans of the box (five and four), and it contacts all methyl groups of the trimethyl betaine ligand (Fig. 5). Trp188 is the only residue of the box that cannot be replaced by a non-aromatic apolar or acidic residue without a complete loss of binding (Table III). This indicates an essential contribution of the cation-π interaction between Trp188 and the betaine ligand. As mutational and structural data both hint at a critical influence of this residue, it seems to have a special function in ligand binding.

At the first look positions 65 and 140 appear to be of approximately equal importance for binding of GB, whereas for PB binding position 65 appears to be more important (Fig. 5). In ProX-GB the indole rings of residues Trp65 and Trp140 form similar numbers of van der Waals contacts with the ligand (Fig. 5). Both positions tolerate to a certain degree substitution against alanine and cannot be distinguished by that. However, replacement of the indole group by a bulky aliphatic or a carboxylic group impairs the binding much more strongly in position 65 than in 140 (Table III). This finding can be interpreted in two ways, either as stronger steric or electrostatic restrictions in position 65 or as a more important role of Trp65 in ligand binding. The binding site in the closed, ligand-bound conformation appears to offer more space for a large non-aromatic side chain in position 140. An aspartate residue seems to be able to substitute for the tryptophan in position 140 providing electrostatic interactions with a $K_D$ of 8 $\mu$M (Table III). But the larger glutamate residue is less suitable than the aliphatic side chain of leucine, which is most likely for steric reasons. The complete loss of binding activity in the double mutant W65A,W140A demonstrates that at least two aromatic residues are necessary for GB binding. The comparison between ProX and two other quaternary amine-binding proteins being structurally not related shows a convergent evolution of the binding site (Fig. 7). Two aromatic rings almost perpendicular to each other seem to be required, whereas further aromatic residues and their orientation with respect to the ligand determine the strength of the interaction between ligand and protein. In the superposition in Fig. 7 the ProX residues Trp65 and Trp188 are in the positions of the two conserved aromatic residues found also in the two other proteins. This is in accord with the findings of our mutational data, which show that Trp65 has a more important role in ligand binding than Trp140. In summary, it could be shown that Trp188 in ProX in pair with either Trp65 or Trp140 is absolutely necessary for a proper ligand binding, whereas the third tryptophan residue enforces the binding affinity in ProX. The difference in the binding studies between Trp65 and Trp140 reflects their importance in ligand binding.

Quaternary Amine Derivatives as Neurotransmitters—Because of the identical quaternary amine head groups of GB and acetylcholine one tends to compare the interactions between them and their receptors. According to the finding of Dougherty and colleagues (80), the cation-π energy of the acetylcholine quaternary amine bound to Trp149 in the α-subunit of the nicotinic acetylcholine receptor from Torpedo californica is critical for channel response. The structure of the acetylcholine-binding protein from snail (81) that can be taken as a model for the extracellular part of the nicotinic acetylcholine receptor suggests an analogous role for Trp143 in acetylcholine-binding protein. However, it is not known yet how the neurotransmitter acetylcholine is exactly bound in the expected aromatic environment. From our data we suggest that Trp188 in ProX has a function similar to that of Trp149 in nicotinic acetylcholine receptor and Trp143 in acetylcholine-binding protein. Additionally a second aromatic residue in a position similar to that of Trp65 in ProX has to be present in the nicotinic acetylcholine receptor to bind the acetylcholine head group properly.

Furthermore from the identity of the head groups we assume a similar exclusion of this part of the acetylcholine molecule from protein surfaces as found for the compatible solute glycine betaine. Only a protein with a highly specific binding site based on cation-π interaction in the form of an aromatic box is able to bind acetylcholine. We thus speculate that quaternary amine derivatives are especially suited to act as neurotransmitters in an environment crowded with a huge number of different proteins like it is in the synaptic cleft where unspecific binding would lead to an inefficient use of the signaling molecule.

REFERENCES
1. Kempf, B., and Bremer, E. (1998) Arch. Microbiol. 170, 319–330
2. Booth, I. R., and Louis, P. (1999) Curr. Opin. Microbiol. 2, 166–169
3. Galinski, E. A., and Trupler, H. G. (1994) FEMS Microbiol. Rev. 15, 95–108
4. Bremer, E., and Kramer, R. (2000) in Bacterial Stress Responses (Sterz, G., and Hengge-Aronis, R., eds) pp. 79–97, American Society for Microbiology Press, Washington, D.C.
5. Walsh, D. T. (2000) FEMS Microbiol. Rev. 24, 263–290
6. Wood, J. M. (1999) Microbiol. Mol. Biol. Rev. 63, 230–262
7. Record, M. T., Jr., Coutenay, E. S., Cayley, D. S., and Gutman, H. J. (1998) Trends Biochem. Sci. 23, 143–148
8. Lippert, K., and Galinski, A. A. (1992) Appl. Microbiol. Biotechnol. 37, 61–65
9. Caldwell, T., Devont-Casle, N., Ghazi, A., and Richarme, G. (1999) Microbiol. 145, 2543–2548
10. Canovas, D., Fletcher, S. A., Hayashi, M., and Csonka, L. N. (2001) J. Bacteriol. 183, 3363–3371
11. Canovas, D., and Galinski, A. A. (1992) Appl. Microbiol. Biotechnol. 37, 61–65
12. Bourot, S., Sire, O., Trautwetter, A., Touze, T., Wu, L. F., Blanco, C., and Bernard, T. (2000) J. Biol. Chem. 275, 1050–1056
13. Da Costa, M. S., Santos, H., and Galinski, E. A. (1998) Appl. Environ. Microbiol. 64, 3774–3779
14. Daneshfar, A. R., Dandekar, A. M., Smith, L. T., and Valentine, R. C. (1984) Science 224, 1064–1068
15. Wood, J. M. (1999) Microbiol. Mol. Biol. Rev. 63, 813–819
16. Da Costa, M. S., Santos, H., and Galinski, E. A. (1998) Adv. Biochem. Eng. Biotechnol. 61, 117–153
17. Mandel, D., and Smith, L. T. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 2543–2548
18. Mandel, D., and Smith, L. T. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 2543–2548
19. Mandel, D., and Smith, L. T. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 2543–2548
20. Hohmann, S. (2002) Int. Rev. Cytol. 215, 149–187
21. Loeblvi, D., Strom, A. R., Danekar, A. M., Smith, L. T., and Valentine, R. C. (1984) Science 224, 1064–1068
22. 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
23. Arakawa, T., and Timasheff, S. N. (1985) Biochem. J. 247, 411–414
24. Qu, Y., Bolen, C. L., and Bolen, D. W. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 9268–9273
