The Membrane-Associated Lipoprotein-9 GmpC from *Staphylococcus aureus* Binds the Dipeptide GlyMet via Side Chain Interactions[†,‡]

Wade A. Williams,[§,||,⊥] Rong-gaung Zhang,[⊥,+] Min Zhou,[+] Grazyna Joachimiak,[+] Piotr Gornicki,[@] Dominique Missiakas,[*,§,||] and Andrzej Joachimiak[*,||,+]

Committee on Microbiology, Department of Molecular Genetics and Cell Biology, and Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, and Structural Biology Center, Biosciences, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Argonne, Illinois 60439

Received June 1, 2004; Revised Manuscript Received October 12, 2004

ABSTRACT: Bacterial dipeptide ABC transporters function to import a wide range of dipeptide substrates. This ability to transport a wide variety of dipeptides is conferred by the cognate substrate binding protein (SBP) of these transporters. SBPs bind dipeptides with little regard for their amino acid content. Here, we report the 1.7 Å resolution structure of lipoprotein-9 (SA0422) of *Staphylococcus aureus* in complex with the dipeptide glycylmethionine. Experimental characterization of the subcellular location of the protein confirmed that SA0422 is an acylated, peripheral membrane protein. This is the first structure determined for an SBP of a Gram-positive dipeptide ABC transporter. Usually, binding of dipeptides occurs in a binding pocket that is largely hydrated and able to accommodate the side chains of several different amino acid residues. Unlike any other known SBP, lipoprotein-9 binds the side chains of the glycylmethionine dipeptide through very specific interactions. Lipoprotein-9 shares significant structural and sequence homology with the MetQ family of methionine SBP. Sequence comparisons between MetQ-like proteins and lipoprotein-9 suggest that the residues forming the tight interactions with the methionine side chains of the ligand are highly conserved between lipoprotein-9 and MetQ homologues, while the residues involved in coordinating the glycine residue are not. Modeling of the *Vibrio cholerae* MetQ and lipoprotein-9 binding pockets can account for lipoprotein-9 substrate specificity toward glycylmethionine. For this reason, we have designated lipoprotein-9 GmpC, for glycylmethionine binding protein.

The uptake of peptides from the environment by bacteria is an important process that not only supplies bacteria with nutrients but also allows them to sense environmental conditions and to initiate appropriate signaling cascades (*1*−*3*). These peptide uptake systems are usually members of the ATP-binding cassette (ABC)[1] family of transporters.[2] The uptake ABC transporters are multisubunit complexes composed of integral membrane proteins that function as a permease, peripheral membrane ATP binding proteins that hydrolyze ATP, and extracellular substrate binding proteins (SBPs) that act as receptors for the substrate to be transported (*1*, *4*). Although structurally conserved, these transporters function in the uptake of a very diverse range of molecules. The SBP components of these systems to a large extent determine the substrate specificity of the ABC transporters with which they are associated. In Gram-negative bacteria, these SBPs are secreted into the periplasm and retained in this compartment by the outer membrane (*1*). However, in Gram-positive bacteria, the lack of an outer membrane necessitates the tethering of these proteins to the plasma membrane by a lipid anchor or by fusion to an integral membrane component of the transporter (*1*, *5*). All SBPs studied to date are structurally similar and bind their substrate through a conserved mechanism, termed the Venus' flytrap mechanism (*6*). The unliganded SBP is usually found in an

[†] This work was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract W-31-109-Eng-38 to A.J. and U.S. Public Health Service Grants GM62414 and AI055838 to A.J. and D.M., respectively. W.A.W. acknowledges support from MCB Training Grant T32GM007183 awarded by the National Institute of General Medical Sciences to the University of Chicago.

[‡] The atomic coordinates for GmpC (SA0422) have been deposited in the Brookhaven Protein Data Bank as entry 1P99.

* To whom correspondence should be addressed. D.M.: phone, (773) 834-8161; fax, (773) 834-8150; e-mail, dmissiak@bsd.uchicago.edu. A.J.: phone, (630) 252-3926; fax, (630) 252-6126; e-mail, andrzejj@anl.gov.

[§] Committee on Microbiology, University of Chicago.

[||] Department of Molecular Genetics and Cell Biology, University of Chicago.

[⊥] These authors made equal contributions to this project.

[@] Department of Biochemistry and Molecular Biology, University of Chicago.

[+] Argonne National Laboratory.

[1] Abbreviations: ABC, ATP-binding cassette; ATP, adenosyl triphosphate; FOM, figure of merit; GlyMet, glycylmethionine dipeptide; Gly-SeMet, glycylselenomethionine dipeptide; rTEV, recombinant tobacco etch virus protease; SAD, single-wavelength anomalous diffraction; SBP, substrate binding protein; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; SeMet, selenomethionine; tRNA, transfer ribonucleic acid.

[2] The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory under contract with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the government.

10.1021/bi048877o CCC: $27.50
Published on Web 12/02/2004

open conformation with the substrate binding pocket exposed to solvent. Once the substrate binds, SPBs adopt a closed conformation in which the substrate is tightly bound and buried. The liganded SBP interacts specifically with its cognate permease, triggering ATP binding at two nonequivalent sites found within the cytoplasmic components of the transporter. The subsequent hydrolysis of the ATP supplies the energy for the long-rage conformational changes required to move the substrate from the SBP to the permease and across the plasma membrane (*6*−*9*).

Di- and oligopeptide SBPs studied to date are unique in that they can bind structurally diverse peptide substrates with high affinity. A given SBP can facilitate the import of a wide range of peptides with little regard for amino acid composition or even length (*10*−*12*). The crystal structures of *Escherichia coli* DppA and *Salmonella typhimurium* OppA, a dipeptide and oligopeptide SBP, respectively, show that the binding pockets of these proteins contain large hydrated regions that can accommodate the side chains of several different amino acids (*13*−*16*). Binding to the substrate is governed by high-affinity interactions between the SBP and the peptide backbone of the substrate. On the contrary, transport of a single amino acid by an SBP is usually very specific. For example, HisJ of *S. typhimurium* binds histidine specifically (*17*, *18*). *E. coli* LivJ binds a small group of closely related amino acids (leucine, isoleucine, and valine) (*19*, *20*). Although the overall structures of single-amino acid and peptide-binding SBPs are similar, these proteins form distinct families of SBPs that can be distinguished on the basis of amino acid sequence.

Here we report the high-resolution crystal structure and functional assignment of SA0422 (lipoprotein-9), a novel dipeptide SBP from *Staphylococcus aureus*. The crystal structure reveals the presence of the glycylmethionine dipeptide tightly bound within the substrate binding pocket of the protein. Lipoprotein-9 is related to the MetQ family of SBPs that binds methionine specifically. Lipoprotein-9 is the first example of an SBP to cocrystallize with a specific dipeptide and exhibits strong binding selectivity. The protein was renamed GmpC for glycylmethionine binding protein C.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Plasmids. S. aureus strain Newman used in this study has been described previously (*21*). The *gmpC::erm* mutant is a strain of the laboratory collection (D.M.) and will be described elsewhere. All staphylococci strains were grown at 37 °C in tryptic soy broth supplemented with erythromycin (10 μg/mL) when necessary. The *S. aureus* SA0422 open reading frame was amplified by polymerase chain reaction, with the *Nde*I and *Bam*HI sites engineered at codon 19 (removing the protein signal sequence) and immediately downstream of the translation stop codon, respectively. Vector pET15b (Novagen) was used for gene expression, fusing the mature SAV0422 coding sequence in-frame with a six-histidine tag and the rTEV cleavage site. This resulted in a 33-amino acid extension at the processed N-terminus of the polypeptide. However, this extension is not shown in the determined structure, and residues are numbered using the full-length (precursor), wild-type GmpC sequence.

Protein Purification and Crystallization. E. coli strain BL21(DE3) carrying the pMAGIC helper plasmid expressing rare tRNAs and the pET15b-*gmpC*$^+$ derivative was grown in the presence of chloramphenicol and kanamycin. Gene expression was induced with 1 mM isopropyl D-thiogalactoside for 20 h at 37 °C. Cells were harvested, suspended in 50 mM phosphate buffer (pH 8.0) containing 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, and 10% glycerol, and lysed by sonication. The histidine-tagged protein was purified by affinity chromatography using nickel-HiTrap Sepharose HP resin (Amersham Biosciences). The histidine tag was not removed. The protein was concentrated over a Centricon plus-20 instrument (Amicon) (5 kDa cutoff). An 80 mg/mL protein stock solution in 20 mM HEPES-HCl (pH 8.0), 200 mM NaCl, and 1 mM DTT was used for crystallization. Equal volumes of a GmpC protein stock solution and buffers were mixed in hanging droplets and equilibrated against 1 mL aliquots of solutions from the sparse matrix crystallization screening kits. The best crystals were obtained using vapor diffusion and hanging droplets of 50 mM Bis-Tris buffer (pH 6.5) containing 50 mM ammonium sulfate and 27% pentaerythritol ethoxylate 15/4 at 16 °C. The selenomethionine (SeMet)-labeled GmpC was prepared using the methionine biosynthesis inhibition method as described previously (*22*) but did not yield any crystals using the aforementioned conditions. However, crystals were obtained by microseeding. The crushed native crystals were used as seeds in the drops of the SeMet-labeled protein prepared and pre-equilibrated for 2 days using the aforementioned conditions. Crystals were rinsed in a cryoprotectant solution containing 25% glycerol and flash-frozen in liquid nitrogen.

Data Collection. Diffraction data were collected at 100 K at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The one-wavelength inverse-beam SAD data set [peak at 12.6603 keV (0.9794 Å)] was collected from a SeMet-labeled protein crystal at 100 K. One crystal (0.1 mm × 0.2 mm × 0.2 mm) was used to collect all SAD data to 1.70 Å resolution, with a 3 s exposure per degree per frame using a crystal to detector distance of 160 mm. The total oscillation range was 160° as predicted using the strategy module within the HKL2000 suite (*23*). The space group was $P2_12_12_1$ with the following cell dimensions: $a = 40.496$ Å, $b = 73.883$ Å, $c = 92.246$ Å, and $\alpha = \beta = \gamma = 90°$. The data were processed and scaled with HKL2000 (*23*) (Table 1) to an R_{merge} of 6.9%.

Structure Determination and Refinement. The structure of GmpC was determined by SAD phasing using CNS (*24*). SAD phases were calculated using the CNS suite and improved using the density modification method as implemented by CNS. The initial model was built using ARP/wARP (*25*), and the final model was finished manually using QUANTA (Accelrys, Inc.). The model was refined to 1.7 Å using CNS against the averaged peak data. The final values of *R*-factor and R_{free} are 20.0 and 23.6%, respectively, with excellent stereochemistry (Table 1). The stereochemistry of the structure was checked with PROCHECK (*26*) and the Ramachandran plot. The main chain torsion angles for all residues are in allowed regions. Twenty-four residues at the N-terminus are disordered because of a lack of electron densities.

Table 1: Summary of Crystal and SAD Data Collection

unit cell parameters	$a = 40.50$ Å, $b = 73.88$ Å, $c = 92.25$ Å, $\alpha = \beta = \gamma = 90°$
space group	$P2_12_12_1$
molecular mass[a] (Da)	33810
no. of molecules per asymmetric unit	1
no. of selenomethionine residues per asymmetric unit	1
wavelength (Å)	0.9791
resolution limit (Å)	1.70
no. of unique reflections	31255
overall data completeness (%)	99.4
overall data redundancy	10.45
overall R_{merge} (%)	6.9
figure of merit (FOM)	0.2967
phasing power	1.6309
Refinement Statistics	
resolution range (Å)	50.0−1.7 (1.81−1.7)[b]
no. of reflections (unique)	58916 (9343)[b]
no. of reflections (observed)	55440 (7699)[b]
percent reflections observed	94.1 (82.4)[b]
σ cutoff	0.0
overall R factor (%)	20.0 (30.5)[b]
R_{free} (%)	23.6 (31.6)[b]
rms deviations from ideal geometry	
bond lengths (Å)	0.006
angles (deg)	1.3
dihedrals (deg)	22.6
impropers (deg)	0.81
no. of protein non-hydrogen atoms	1977
no. of water molecules	292
mean B factor (Å^2)	26.5
Ramachandran plot statistics (%)	
residues in most favored regions	98.3
residues in additional allowed regions	1.7
residues in disallowed regions	0.0

[a] Molecular mass of the polypeptide with 33-residue tag fused to residues 19−280 of GmpC. [b] Values in parentheses are for the highest-resolution shell.

Cell Fractionation. S. aureus strains grown in tryptic soy broth were suspended in minimal medium, and subjected to labeling with 100 μCi of ^{35}S-labeled Promix (Amersham Biosciences) for 5 min at 37 °C. Samples were spun, and supernatants were precipitated with trichloroacetic acid and boiled in 4% SDS (medium fraction). Cells were suspended in a hypertonic solution [500 mM sucrose, 10 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5)] and digested with lysostaphin at 37 °C for 10 min. The samples were spun and separated into the supernatant and pellet. Supernatants were precipitated with trichloroacetic acid and boiled in 4% SDS (cell wall fraction). The protoplasts contained in the pellets were lysed in a hypotonic solution [0.1 M NaCl, 0.1 M Tris-HCl (pH 7.5), and 0.01 M $MgCl_2$]. The fraction containing the membrane proteins (membrane fraction) was separated from the soluble proteins (cytoplasmic fraction) by ultracentrifugation for 30 min at 100000*g*, and samples were precipitated with trichloroacetic acid and boiled in 4% SDS. All the samples were immunoprecipitated with α-GmpC or α-PrsA rabbit polyclonal antisera in radio-immunoprecipitation assay buffer with protein A−Sepharose (Sigma). Immunoprecipitates were washed in the radio-immunoprecipitation buffer, boiled in sample buffer, and analyzed by 15% SDS−PAGE and a PhosphorImager.

Labeling of GmpC with Tritiated Palmitic Acid. S. aureus strains grown in tryptic soy broth were suspended in minimal medium, and subjected to labeling with 50 μCi of [9,10-^{3}H-(N)]palmitic acid (American Radiolabeled Chemicals) for 60 min at 37 °C. The incorporation of tritiated palmitic acid was quenched by precipitation with trichloroacetic acid followed by an acetone wash. Precipitates were suspended in 0.5 M Tris-HCl buffer (pH 7.5), and the cell wall was digested with lysostaphin for 60 min at 37 °C. Total proteins were precipitated with trichloroacetic acid and boiled in 4% SDS. All samples were immunoprecipitated with α-GmpC rabbit polyclonal antisera and analyzed as described above.

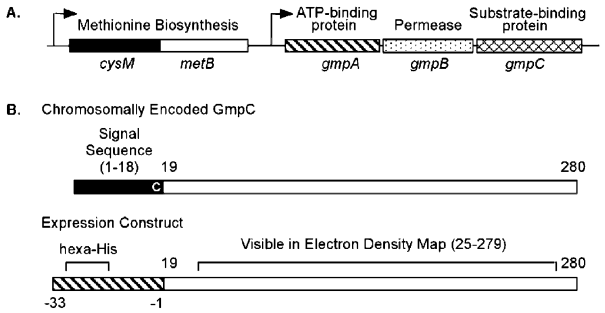

FIGURE 1: (A) Genomic organization of *S. aureus gmp*. GmpC is located in an operon with the putative ABC transporter ATP-binding protein and permease, GmpA and GmpB, respectively. Upstream of the *gmp* operon is a two-gene operon encoding the methionine biosynthetic enzymes cysteine synthase (CysM) and cystathionine γ-synthase (MetB). Putative transcriptional start sites are depicted with arrows. (B) Schematic representation of the expression construct used in this study. The chromosomally encoded GmpC is synthesized with an 18-amino acid signal sequence, including a predicted lipid modification site (Cys18, denoted with a C). For expression and purification purposes, the region encoding residues 19−280 of the full-length GmpC was fused to a 33-amino acid extension (numbered −33 to −1) containing a hexahistidine epitope tag. This fusion protein was subjected to crystallographic analysis which produced an electron density map in which the residues corresponding to residues 25−279 of the chromosomally encoded GmpC were visible while the remainder of the construct was unstructured.

Membrane Extractions. S. aureus strains grown in tryptic soy broth were suspended in a hypertonic solution and cell walls digested with lysostaphin at 37 °C for 10 min, as described above. Protoplasts were harvested by centrifugation and lysed using French pressure, and the membrane fraction was separated by ultracentrifugation for 30 min at 100000*g*. The membrane fraction was suspended in either 50 mM Tris-HCl buffer (pH 7), 100 mM NaCl, and 10 mM $MgCl_2$ (total membrane) or the same buffer containing 8 M urea (urea extraction) or 0.05 M NaOH (NaOH extraction) and incubated at 23 °C for 15 min. Proteins in the total membrane preparations were precipitate with trichloroacetic acid and boiled in 4% SDS. For the membrane extraction preparations, the insoluble material was removed by ultracentrifugation for 30 min at 100000*g*, and proteins in the supernatants were precipitated with trichloroacetic acid and boiled in 4% SDS. All samples were analyzed by SDS−PAGE and Western blot analysis using polyclonal α-GmpC primary antibodies.

Biotinylation of Surface Proteins. Overnight cultures of *S. aureus* were suspended at an optical density of 0.01 (660 nm) in tryptic soy broth containing the iron chelator 2′,2′-Dipyridyl (1 mM). Growth under iron-depleted conditions is necessary to induce the expression of *isdG*, encoding a cytoplasmic protein used as a control for this experiment. Such growth conditions did not affect the production and localization of GmpC (data not shown). Cultures were

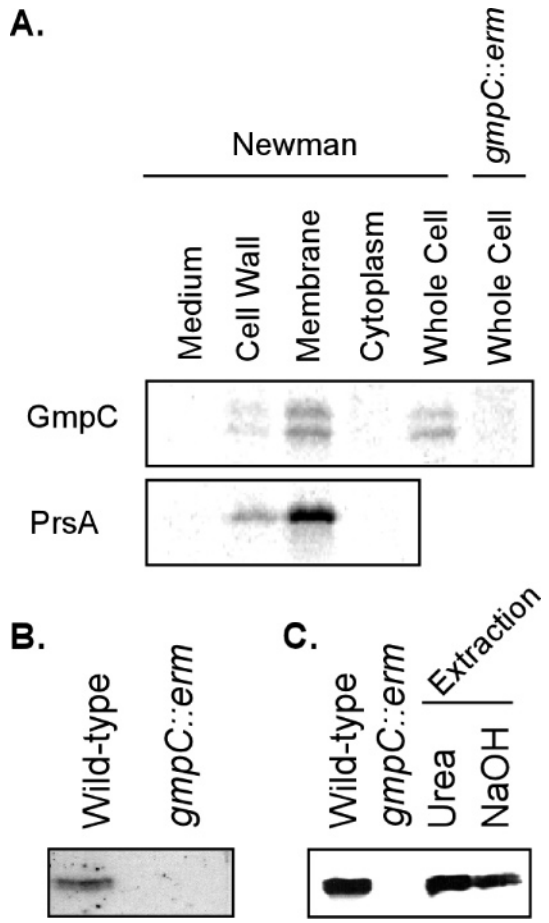

FIGURE 2: Cellular localization of GmpC. (A) GmpC is found in the membrane of staphylococci. Newman cultures of wild-type and *gmpC* mutant cells were radiolabeled with [^{35}S]methionine and cysteine, and fractionated into medium (MD), cell wall (CW), membrane (M), and cytosolic (C) compartments. Only the whole cell sample is shown for the mutant. Samples were subjected to immunoprecipitation with α-GmpC or α-PrsA antibodies, and analyzed via SDS−PAGE (15%) and on a PhosphorImager. (B) GmpC is acylated *in vivo*. Newman cultures of wild-type and *gmpC* mutant cells were radiolabeled with [9,10-^{3}H(N)]palmitic acid. Total cell extracts were obtained and subjected to immunoprecipitation with α-GmpC rabbit polyclonal antisera. Samples were analyzed by SDS−PAGE (12%) and autoradiography. (C) GmpC is peripherally associated with the membrane. Membranes prepared from wild-type staphylococci (left lane) were treated with 8 M urea, 100 mM NaCl, and 10 mM $MgCl_2$ (urea extraction) or 0.05 M NaOH, 100 mM NaCl, and 10 mM $MgCl_2$ (NaOH extraction) and incubated at 23 °C for 15 min. Proteins solubilized during these extraction procedures were separated by SDS−PAGE (12%), transferred to a PVDF membrane, and revealed using polyclonal α-GmpC primary antibodies. A sample containing the membrane fraction of a *gmpC* mutant strain is shown as a control.

allowed to reach an optical density of 0.8 at 660 nm. Cells were washed three times with phosphate buffer (PBS), suspended in 1 mL of PBS, and incubated with the membrane impermeable compound EZ-link Sulfo-NHS-SS-Biotin (Pierce) at 22 °C for 30 min. The reactions were quenched with 0.5 M Tris-HCl (pH 7.5). Cells were incubated with lysostaphin at 37 °C for 10 min to digest the cell wall, and the samples were precipitated with trichloroacetic acid and boiled in 4% SDS. The samples were further incubated with ImmunoPure immobilized streptavidin cross-linked to 6% agarose beads (Pierce) in precipitation assay buffer as recommended by the manufacturer. Biotin−streptavidin precipitates were washed in precipitation assay buffer, boiled in sample buffer,

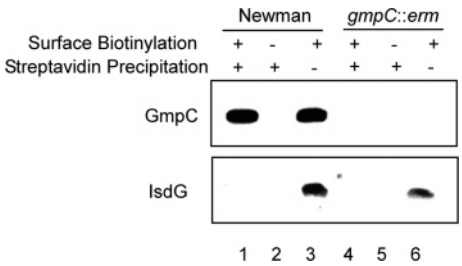

FIGURE 3: Display of GmpC on the staphylococcal surface. Newman cultures of wild-type and *gmpC* mutant cells were incubated with (+) or without (−) biotin using the membrane impermeable compound Sulfo-HNS-SS-Biotin. Samples were further incubated with or without streptavidin immobilized on agarose beads [Streptavidin Precipitation (+) or (−), respetively] and analyzed by SDS−PAGE followed by Western blotting with α-GmpC or α-IsdG antibodies.

and analyzed by SDS−PAGE (12%) and Western blotting using polyclonal α-GmpC or α-IsdG primary antibodies.

RESULTS AND DISCUSSION

Genomic Organization of the gmp Locus. Sequence analysis of the annotated genomic sequence of the clinical isolate *S. aureus* strain N315 (*27*) revealed that *gmpC* is located downstream of a predicted ABC transporter ATP-binding protein (SA0420 *gmpA*) and a predicted ABC-type permease protein (SA0421 *gmpB*) (Figure 1). A single transcriptional start site can be identified upstream of *gmpA*, suggesting that the three genes are part of an operon. A two-gene operon encoding enzymes for methionine biosynthesis, cysteine synthase (CysM, SA0418), and cystathionine β-lyase/γ-synthase (MetB, SA0419) is located upstream of the *gmp* operon (Figure 1). Pfam classification (Pfam03180.7) of GmpC (SA0422) puts the protein in the lipoprotein-9 family. This family of lipoproteins is composed of more than 200 members, and they are found exclusively in bacteria (185 species) (COG1464) (*28*). The function of these lipoproteins (including SA0422) is unknown. Members of this group include: ABC transporter substrate-binding and SBPs, outer membrane signal peptide proteins, ABC-type metal ion transport systems, D-Met-binding lipoprotein (MetQ family), and surface antigen. Members of the COG1464 group do not share significant sequence similarity with SBPs of well-characterized ABC-type transport systems. GmpC appears to be mostly related to NlpA (COG1464.1), classified as an ABC-type metal ion transport system, periplasmic component, and surface antigen. It is presumed that members of this family may be involved in inorganic ion transport and metabolism. However, NlpA is not encoded within a canonical ABC-type operon and consequently may not function as a transporter.

Subcellular Localization of GmpC. GmpC is predicted to be a lipoprotein (N-terminal diacylglycerol) and bears a predicted type II signal sequence with the proposed lipoprotein modification box C_{18}GSN. Upon signal peptide removal, the cysteine at position 18 of full-length GmpC becomes the first amino acid of the mature protein and is modified by acylation, a process that tethers lipoproteins to the outer leaflet of the cytoplasmic membrane. To confirm this prediction, total proteins of wild-type cells (*S. aureus* strain Newman) or a mutant unable to synthesize GmpC (*gmpC*::*erm*) was labeled with a radioactive mix containing

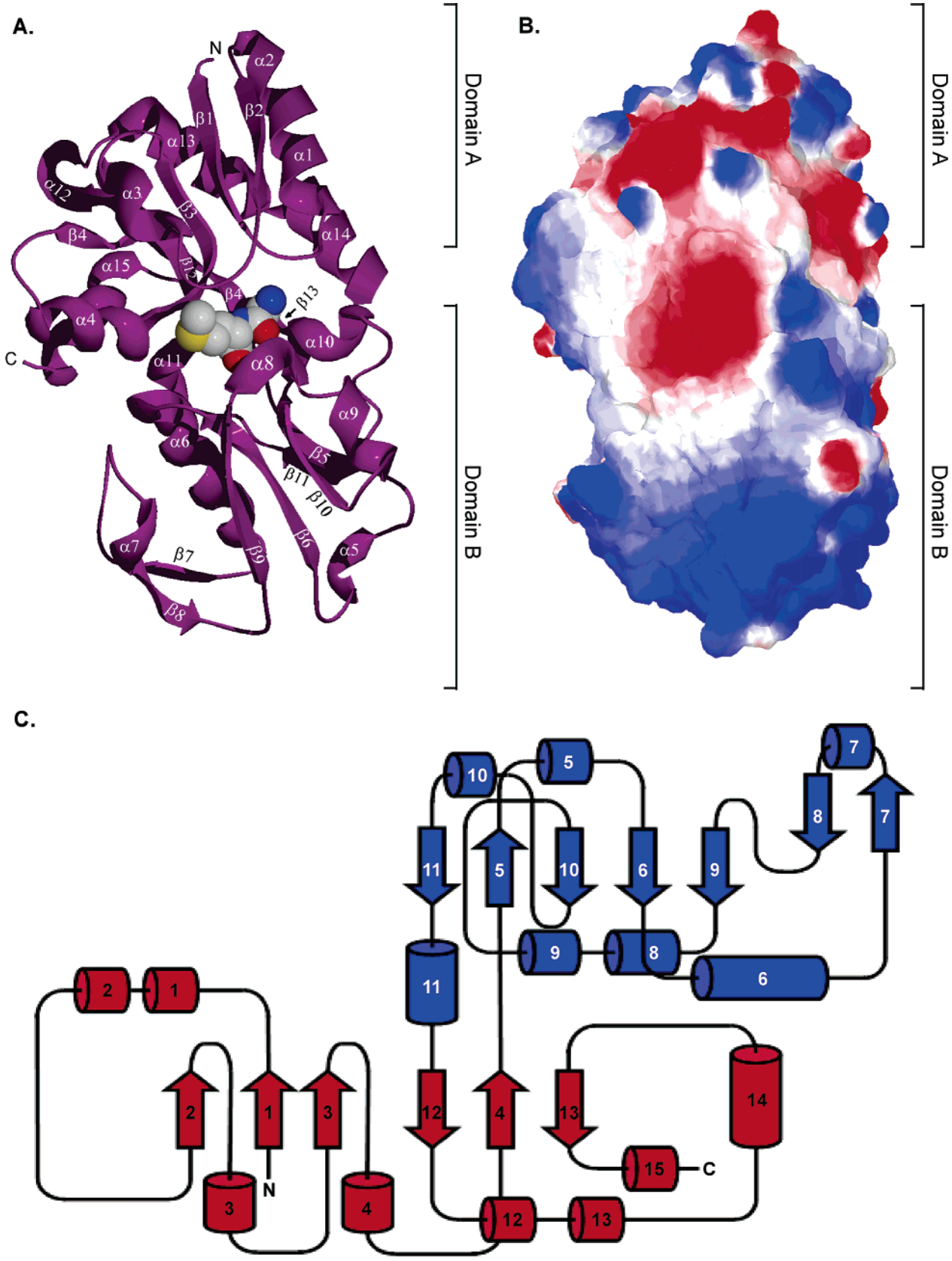

FIGURE 4: Structure of GmpC. (A) Ribbon diagram of the GmpC structure with the ligand. The oxygen, nitrogen, carbon, and sulfur atoms of the GlyMet ligand are colored red, blue, white, and yellow, respectively. (B) Solvent-accessible surface and electrostatic potential of GmpC. The ligand is not exposed to the solvent and is obstructed from view. The methionine residue of the ligand is located directly underneath the acidic patch located in the center of the molecule. Panels A and B were generated using DeepView PDB viewer (*34*) in conjunction with POV-Ray rendering software. (C) Secondary structure diagram of GmpC. The domain A secondary structure is colored in red, while the secondary structure of domain B is colored blue. Panel C was prepared using data provided by the ProCheck algorithm at EBI (*26*).

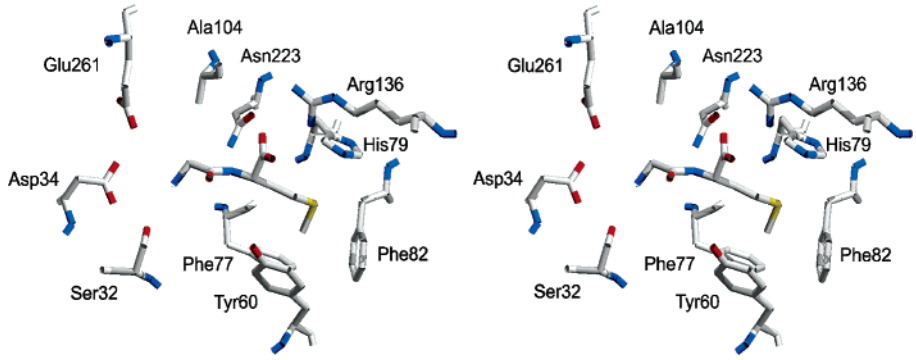

FIGURE 5: Stereoview of the ligand-binding pocket of GmpC bound to GlyMet. For clarity, only the residues of the binding pocket are shown. These figures were generated using DeepView PDB viewer (*34*) in conjunction with POV-Ray rendering software.

^{35}S-labeled methionine and cysteine for 5 min. The cells were fractionated into cytoplamic, membrane, cell wall, and medium fractions. The presence of GmpC in each fraction was assessed by immunoprecipitation with specific antiserum and separation of proteins on SDS−PAGE. The gels were dried, and radioactive proteins were revealed with a PhosphorImager. A similar experiment was conducted using antibodies against the well-characterized lipoprotein PrsA. As shown in Figure 2A, the majority of GmpC is found in the membrane fraction, a localization mirrored by PrsA. Furthermore, radiolabeled GmpC species were immunoprecipitated from wild-type staphylococci, but not *gmpC* mutant cells, grown in the presence of [9,10-^{3}H(N)]palmitic acid (Figure 2B). Palmitate is incorporated into lipids and lipoproteins. Also, GmpC could be extracted from membrane preparations upon treatment with urea or sodium hydroxide (Figure 2C). Such treatments fail to extract integral membrane proteins but not lipid-anchored lipoproteins, underlying the peripheral association of such proteins with the lipid bilayer (Figure 2C). Taken together, these data strongly support the notion that *gmpC* encodes a lipoprotein. It should be noted that when GmpC was separated on an SDS−PAGE gel containing 15% acrylamide, two species with different mobilities were detected (Figure 2A). It is unclear whether the presence of two immunoreactive species reflects the presence or absence of bound ligand or the mere degradation of GmpC.

The cell surface display of GmpC was assessed using a fourth assay. Staphylococcal cells were incubated with the membrane impermeable compound EZ-link Sulfo-NHS-SS-Biotin (Pierce). Biotinylated proteins were precipitated using streptavidin beads, and specific proteins were detected using immunoblot analysis. GmpC was recovered on streptavidin beads following treatment of cells with Sulfo-NHS-SS-Biotin, while the known cytoplasmic protein IsdG was not (Figure 3, lanes 1). IsdG was only visualized upon disruption of the cell envelope (Figure 3, lane 3). Additionally, precipitation of GmpC on streptavidin beads did not occur when cells were not treated with Sulfo-NHS-SS-Biotin (Figure 3, lane 2). Finally, no GmpC immunoreactive species were detected when an isogenic *gmpC*::*erm* mutant was used (Figure 3, lanes 4−6). Together, these data indicate that GmpC is localized to the outer leaflet of the staphylococcal membrane *in vivo*.

Crystal Structure of GmpC. The crystal structure of GmpC residues 19−280 was determined at 1.7 Å resolution using the single-wavelength anomalous diffraction (SAD) method. Residues 19−24 and residue 280 were the only disordered segments for which a structure could not be assigned. In general, the overall structure of GmpC is reminiscent of structures of other SBPs and is composed of two domains separated by a groove region (Figure 4A). The GmpC monomer contains 13 β-strands and 15 α-helices. The β-strands are arranged into two sheets defining two domains, A and B. Domain A contains a twisted β-sheet comprising six β-strands ($\beta2\downarrow\beta1\uparrow\beta3\uparrow\beta12\downarrow\beta4\uparrow\beta13\downarrow$) and nine α-helices covering the concave and convex site of the sheet (Figure 4A,C). Domain B is composed of a strongly twisted β-sheet ($\beta11\downarrow\beta5\uparrow\beta10\downarrow\beta6\downarrow\beta9\downarrow$) and six α-helices. Two additional β-strands, $\beta7$ and $\beta8$, form a hairpin (Figure 4C).

A search for structurally related proteins in the Brookhaven Protein Data Bank, performed with the DALI server (*29*), identified seven proteins with a Z score of >10 (data not shown). The closest structural homologues of GmpC were found to be the molybdate SBP ModA [rmsd of 2.98 Å over 46 Cα atoms, PDB entry 1WOD (*30*)] and the lysine, arginine, ornithine SBP [rmsd of 1.55 Å over 60 Cα atoms, PDB entry 1LST (*31*)] with Z scores of 13.5 and 11.8, respectively. Also, GmpC shares structural similarities with the spermine-putrescine SBP (PDB entry 1POT), the maltose-dextrin SBP (PDB entry 4MBP), a fragment of the glutamate receptor (PDB entry 1MY0), and the phosphate and sulfate SBPs. All these proteins share very little sequence identity with GmpC (14−8%), and all structural similarities are confined to domain A.

The cavity formed at the interface between domains A and B of GmpC contains a ligand in the determined structure. Ligands bound to ModA and other structural homologues occupy very similar space, suggesting that this fold has evolved to accommodate a variety of ligands (see below).

Glycylmethionine Peptide Binding Site. After completion of the protein model, a well-defined electron density was evident in the deep cavity formed between the two domains of the protein. The excellent quality of the electron density map in this region permitted the unambiguous identification of this ligand as a glycylmethionine dipeptide (GlyMet). The methionine residue is in fact a highly ordered selenomethionine (SeMet), and the selenium atom was used for phasing purposes. The atoms of the peptide refined to a low B factor (15.37) and are very well-ordered. The GmpC protein was overproduced in the presence of SeMet and obviously copurified with the Gly-SeMet peptide as a result of *in vivo* processing of SeMet, thereby displaying the only SeMet in the complex.

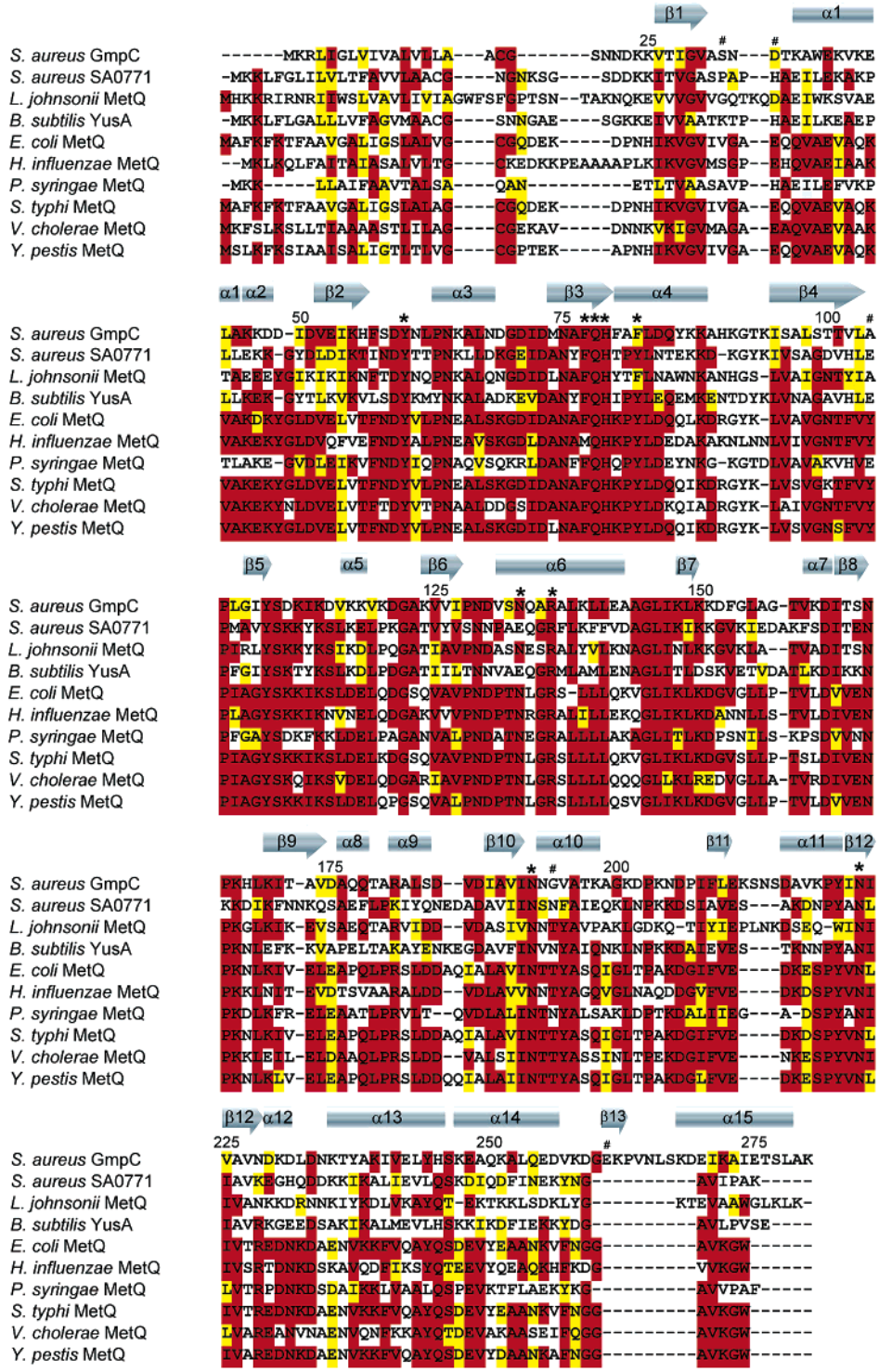

FIGURE 6: Sequence alignment of GmpC and various MetQ homologues. The sequence of GmpC was aligned to *E. coli* MetQ as well as MetQ homologues from various bacteria, including *S. aureus* SA0771. Identical and conserved residues are highlighted in red and yellow, respecitvely. Residues interacting with the methionine residue of the GlyMet substrate are denoted with an asterisk, while those interacting with the glycyl residue are denoted with a number symbol. The secondary structure of GmpC is shown above the sequence alignment. This alignment was performed using the Clustal W algorithm (*35*).

SBPs often copurify with ligands, and most SBP structures are determined with ligands. However, for dipeptide SBPs, such ligands are composed of a heterogeneous mix of peptides, rendering the visualization of the interaction between the protein and ligands impossible. To obtain a discernible ligand structure, these heterogeneous complexes are partially denatured prior to crystallization with specific peptides. Our finding suggests that unlike other dipeptide SBPs, the binding pocket of GmpC is very specific for the glycylmethionine dipeptide.

The dipeptide is trapped within the hinge region of GmpC that is articulated around three loops and two helices. One of these loops between β-strands β4 and β5 along with helix α11 join domains A and B of the protein. The remaining two loops, between β-strand β2 and helix α3 within domain A and the loop connecting β-strand β9 and helix α8 in

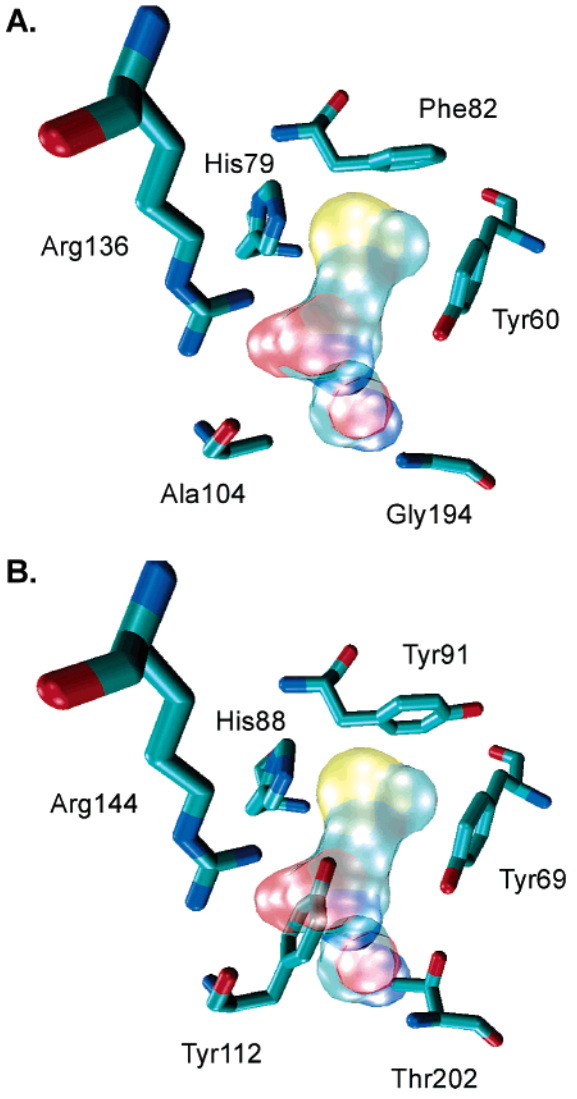

FIGURE 7: Comparison of the substrate binding pockets of GmpC and MetQ. (A) Graphical representation of the binding pocket of GmpC-like proteins. Y60, F77 (not shown for clarity), H79, and F82 form a hydrophobic pocket surrounding the methionine side chain of the substrate. R136 forms a H-bond with the methionine carboxylate (see the text for details). The GlyMet substrate is shown as a transparent space-filled model. Residues are numbered according to the GmpC amino acid sequence. (B) Model of the *V. cholerae* MetQ binding site as compared to the GmpC binding site using SWISS-MODEL. Y69, F86, H88, and Y91 form the highly conserved, hydrophobic binding pocket for methionine. R144 appears to be positioned correctly to form a salt bridge with the carboxylate of methionine. Y112 and T202, however, are positioned in a way that would prevent interaction with the dipeptide. The steric hindrance imposed by Y112 and T202 is largely responsible for MetQ specificity toward free methionine. Residues are numbered according to the *V. cholerae* MetQ amino acid sequence. This figure was generated using VMD (*36*, *37*) and rendered with POV-Ray.

conjunction with helix α8 in domain B, appear to form the clamps of the so-called "Venus' flytrap" (Figure 4A). Each clamp appears to be centered around an aspartate residue (D59 β2−α3 loop and D175 β9−α8 loop). The negatively charged area surrounding these aspartate residues is directly above the methionine residue of the ligand and constitutes a negatively charged area in the center of the molecule clearly visible on Figure 4B. Also evident on Figure 4B is a large positively charged region in domain B that is composed of two lysine-rich regions between β-strands β8 and β9 and β-strands β5 and β67. Additionally, β-strand β7 is very rich in positively charged residues. This large region of positive charges may represent a protein−protein interaction domain.

The interaction between the protein and peptide is very extensive (72 Å^2) and involves all the accessible surface of the peptide ligand which is completely buried between the domains and has no direct contact with the bulk solvent (Figure 4A,B). Fourteen residues from both domains make contacts with the peptide through numerous H-bonds and van der Waals interactions. Of the residues that establish an interaction with GlyMet, several could be inferred from structures of other SBPs. These residues include Y60, F77, H79, N133, R136, N192, N193, and N223 (two interactions). Four water molecules that are buried in this cavity mediate additional interactions between the protein and the peptide.

The methionine residue of the GlyMet substrate is interacting with Y60 (the closest residue to the methyl group of Met), H79 (the closest residue to the Se atom), F77, F82, N133, and carbon α of Q78 (Figures 4 and 6). One oxygen atom of the GlyMet carboxylate forms strong H-bonds with N192 (2.72 Å), R136 (2.76 Å), and N223 (2.95 Å). This oxygen also interacts with a water molecule that is stabilized through H-bonding with the carbonyl groups of A104 and P105. The second oxygen atom of the GlyMet carboxylate forms H-bonds with N223 and a water molecule that is stabilized through interactions with the carbonyl groups of Y221 and A104. An interaction with the N-terminal dipole of helix 10 helps in neutralizing the buried negative charge of the GlyMet carboxylate.

Strong H-bonds are observed between the N-terminus of GlyMet and the hydroxyl group of S32, and between the amide group of GlyMet and the conserved N223 residue (2.82 Å) (Figure 4A). N223 coordinates both the carboxylate and imino groups of the methionine residue of GlyMet. The N-terminus of GlyMet also forms a water-mediated H-bond with D34 and two weak direct H-bonds with D34 and E261. The glycyl-carbonyl group of GlyMet is engaged in two H-bonds, one with the imino group of G194 and another with a water molecule interacting with the hydroxyl group of Y60. Altogether, the data suggest that glycine is the only residue that can be accommodated along with methionine in this binding pocket. Indeed, the Cα atom of Gly is nearly within van der Waals distance of A104 (3.65 Å) and E261 (3.70 Å). In conclusion, GlyMet binding is highly specific, consistent with the finding that attempts to exchange the GlyMet dipeptide with GlyLys, GlyLeu, or AlaMet dipeptides failed (even at peptide concentrations reaching 10 mM).

Comparison of GmpC and MetQ Proteins. The discovery of GlyMet in the GmpC binding pocket prompted us to re-examine the related MetQ family of SBP that binds methionine specifically (*32*, *33*). BLAST searches using the genome of *S. aureus* revealed the presence of a second MetQ-related protein encoded by the gene SA0771 (Figure 6). Sequence comparison of GmpC and the MetQ proteins from several species shows that the residues involved in coordinating the methionine of GlyMet (Y60, F77, Q78, H79, and F82 in GmpC) and its carboxylate group (R136 and N223) are almost universally conserved (Figure 6). The only discrepancies are at positions 77 and 82. All MetQ homologues carry a phenylalanine at position 77 with the exception of *Haemophilus influenzae* MetQ, which carries a methionine. A phenylalanine occupies position 82 of GmpC and

MetQ of *Lactobacillus johnsonii*, whereas all other MetQ proteins bear a tyrosine at the same position. Despite these discrepancies, all of these residues should still establish a hydrophobic interaction with the side chain of the methionine substrate (Figures 4C and 7). In contrast, none of the residues involved in specifically binding the glycyl residue of GlyMet (S32, D34, A104, N193, G194, and E261 in GmpC) are conserved between GmpC and the MetQ proteins (Figure 6). Instead, in the MetQ homologues, either a threonine or an asparagine replaces G194 and a tyrosine or a glutamate replaces Ala104 with the exception of the *L. johnsonii* protein. These residues with larger amino acid side chains preclude the presence of the glycyl residue within the binding pocket (Figure 7). Indeed, modeling of the *Vibrio cholerae* MetQ three-dimensional structure using the structure of GmpC as a scaffold (performed by and deposited in the SWISS-MODEL Repository) revealed that due to steric hindrance GlyMet cannot fit into the predicted substrate binding pocket of MetQ (Figure 7). In MetQ, G194 and A104 of GmpC are replaced with threonine (T202) and tyrosine (Y112), respectively, and physically prevent GlyMet binding. According to this modeling, we can infer that SA0771 is in fact the true MetQ homologue, i.e., the SBP responsible for binding and transporting D-methionine.

CONCLUSIONS

In this study, we report the crystal structure of a dipeptide SBP (GmpC) purified in the ligand-bound state. Upon solving the crystal structure of this complex, we found that the substrate was unique and consisted of the glycylmethionine dipeptide. The structure also revealed that despite the low level of sequence homology, the overall fold of GmpC assigns the protein to the MetQ family of proteins, a class of SBPs that bind the amino acid D-methionine. A comparative analysis of the substrate binding pockets clearly shows that in GmpC the pocket has been extended to accommodate a methionine-containing dipeptide. Changes in the binding cavity allow for the specific and high-affinity binding of glycylmethionine by GmpC. Bacteria such as *S. aureus* encode both MetQ and GmpC proteins. It is unclear whether GmpC, like MetQ, contributes merely to methionine transport and homeostasis. It seems equally plausible that GmpC may fulfill additional unknown functions. Clearly, the findings presented here suggest that GmpC may bind a single dipeptide and that it may represent the first characterized member of a novel class of dipeptide SBPs.

ACKNOWLEDGMENT

We thank all members of the Structural Biology Center at Argonne National Laboaratory for their help conducting diffraction experiments, E. Skaar for the gift of the α-IsdG antibody, and O. Schneewind for critical reading of the manuscript.

REFERENCES

1. Tam, R., and Saier, M. H., Jr. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria, *Microbiol. Rev. 57*, 320−346.
2. Manson, M. D., Blank, V., Brade, G., and Higgins, C. F. (1986) Peptide chemotaxis in *E. coli* involves the Tap signal transducer and the dipeptide permease, *Nature 321*, 253−256.
3. Detmers, F. J., Lanfermeijer, F. C., and Poolman, B. (2001) Peptides and ATP binding cassette peptide transporters, *Res. Microbiol. 152*, 245−258.
4. Higgins, C. F. (1992) ABC transporters: From microorganisms to man, *Annu. Rev. Cell Biol. 8*, 67−113.
5. van der Heide, T., and Poolman, B. (2000) Osmoregulated ABC-transport system of *Lactococcus lactis* senses water stress via changes in the physical state of the membrane, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7102−7106.
6. Quiocho, F. A., and Ledvina, P. S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: Variation of common themes, *Mol. Microbiol. 20*, 17−25.
7. Hou, Y. X., Riordan, J. R., and Chang, X. B. (2003) ATP binding, not hydrolysis, at the first nucleotide-binding domain of multidrug resistance-associated protein MRP1 enhances ADP·Vi trapping at the second domain, *J. Biol. Chem. 278*, 3599−3605.
8. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The *E. coli* BtuCD structure: A framework for ABC transporter architecture and mechanism, *Science 296*, 1091−1098.
9. Davidson, A. L. (2002) Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters, *J. Bacteriol. 184*, 1225−1233.
10. Lanfermeijer, F. C., Detmers, F. J., Konings, W. N., and Poolman, B. (2000) On the binding mechanism of the peptide receptor of the oligopeptide transport system of *Lactococcus lactis*, *EMBO J. 19*, 3649−3656.
11. Lanfermeijer, F. C., Picon, A., Konings, W. N., and Poolman, B. (1999) Kinetics and consequences of binding of nona- and dodecapeptides to the oligopeptide binding protein (OppA) of *Lactococcus lactis*, *Biochemistry 38*, 14440−14450.
12. Sanz, Y., Toldra, F., Renault, P., and Poolman, B. (2003) Specificity of the second binding protein of the peptide ABC-transporter (Dpp) of *Lactococcus lactis* IL1403, *FEMS Microbiol. Lett. 227*, 33−38.
13. Dunten, P., and Mowbray, S. L. (1995) Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis, *Protein Sci. 4*, 2327−2334.
14. Nickitenko, A. V., Trakhanov, S., and Quiocho, F. A. (1995) 2 Å resolution structure of DppA, a periplasmic dipeptide transport/chemosensory receptor, *Biochemistry 34*, 16585−16595.
15. Tame, J. R., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F., and Wilkinson, A. J. (1994) The structural basis of sequence-independent peptide binding by OppA protein, *Science 264*, 1578−1581.
16. Sleigh, S. H., Tame, J. R., Dodson, E. J., and Wilkinson, A. J. (1997) Peptide binding in OppA, the crystal structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine, *Biochemistry 36*, 9747−9758.
17. Lever, J. E. (1972) Purification and properties of a component of histidine transport in *Salmonella typhimurium*. The histidine-binding protein J, *J. Biol. Chem. 247*, 4317−4326.
18. Rosen, B. P., and Vasington, F. D. (1971) Purification and characterization of a histidine-binding protein from *Salmonella typhimurium* LT-2 and its relationship to the histidine permease system, *J. Biol. Chem. 246*, 5351−5360.
19. Rahmanian, M., Claus, D. R., and Oxender, D. L. (1973) Multiplicity of leucine transport systems in *Escherichia coli* K-12, *J. Bacteriol. 116*, 1258−1266.
20. Adams, M. D., Wagner, L. M., Graddis, T. J., Landick, R., Antonucci, T. K., Gibson, A. L., and Oxender, D. L. (1990) Nucleotide sequence and genetic characterization reveal six essential genes for the LIV-I and LS transport systems of *Escherichia coli*, *J. Biol. Chem. 265*, 11436−11443.
21. Duthie, E. S., and Lorenz, L. L. (1952) Staphylococcal Coagulase: Mode of Action and Antigenicity, *J. Gen. Microbiol. 6*, 95−107.
22. Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R., and Joachimiak, A. (1999) Taking MAD to the extreme: Ultrafast protein structure determination, *Acta Crystallogr. D55* (Part 6), 1168−1173.
23. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
24. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new

software suite for macromolecular structure determination, *Acta Crystallogr. D54* (Part 5), 905−921.
25. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
26. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
27. Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*, *Lancet 357*, 1225−1240.
28. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V. (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res. 29*, 22−28.
29. Holm, L., and Sander, C. (1995) Dali: A network tool for protein structure comparison, *Trends Biochem. Sci. 20*, 478−480.
30. Hu, Y., Rech, S., Gunsalus, R. P., and Rees, D. C. (1997) Crystal structure of the molybdate binding protein ModA, *Nat. Struct. Biol. 4*, 703−707.
31. Oh, B. H., Pandit, J., Kang, C. H., Nikaido, K., Gokcen, S., Ames, G. F., and Kim, S. H. (1993) Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand, *J. Biol. Chem. 268*, 11348−11355.
32. Merlin, C., Gardiner, G., Durand, S., and Masters, M. (2002) The *Escherichia coli* metD locus encodes an ABC transporter which includes Abc (MetN), YaeE (MetI), and YaeC (MetQ), *J. Bacteriol. 184*, 5513−5517.
33. Zhang, Z., Feige, J. N., Chang, A. B., Anderson, I. J., Brodianski, V. M., Vitreschak, A. G., Gelfand, M. S., and Saier, M. H., Jr. (2003) A transporter of *Escherichia coli* specific for L- and D-methionine is the prototype for a new family within the ABC superfamily, *Arch. Microbiol. 180*, 88−100.
34. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis 18*, 2714−2723.
35. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
36. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 33−38.
37. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 27−28.

BI048877O