PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Insights into *Mycoplasma genitalium* metabolism revealed by the structure of MG289, an extracytoplasmic thiamine binding lipoprotein

Katherine H. Sippel,[1] Balasubramanian Venkatakrishnan,[1] Susan K. Boehlein,[2] Banumathi Sankaran,[3] Jeanne G. Quirit,[1] Lakshamanan Govindasamy,[1] Mavis Agbandje-McKenna,[1] Steve Goodison,[4] Charles J. Rosser,[5] and Robert McKenna[1]*

[1] Department of Biochemistry and Molecular Biology, College of Medicine, Gainesville, Florida 32610
[2] Program in Plant Molecular and Cellular Biology and Horticultural Sciences, University of Florida, Gainesville, Florida 32610
[3] Berkeley Center for Structural Biology, Lawrence Berkeley Laboratory, Berkeley, California 94720
[4] Department of Surgery Jacksonville, Shands Health Science Center, Florida 32209
[5] Department of Urology, University of Florida, Gainesville, Florida 32610

ABSTRACT

***Mycoplasma genitalium* is one of the smallest organisms capable of self-replication and its sequence is considered a starting point for understanding the minimal genome required for life. MG289, a putative phosphonate substrate binding protein, is considered to be one of these essential genes. The crystal structure of MG289 has been solved at 1.95 Å resolution. The structurally identified thiamine binding region reveals possible mechanisms for ligand promiscuity. MG289 was determined to be an extracytoplasmic thiamine binding lipoprotein. Computational analysis, size exclusion chromatography, and small angle X-ray scattering indicates that MG289 homodimerizes in a concentration-dependant manner. Comparisons to the thiamine pyrophosphate binding homolog CypI reveal insights into the metabolic differences between mycoplasmal species including identifying possible kinases for cofactor phosphorylation and describing the mechanism of thiamine transport into the cell. These results provide a baseline to build our understanding of the minimal metabolic requirements of a living organism.**

Proteins 2011; 79:528–536.
© 2010 Wiley-Liss, Inc.

Key words: p37; CypI; substrate binding protein; extracytoplasmic lipoprotein; sexually transmitted infection; X-ray crystallography; small angle X-ray scattering.

INTRODUCTION

Mycoplasmas (class *Mollicutes*) are tiny, pleomorphic bacteria that live in association with eukaryotes either by attachment to the cell membrane of its host or intracellularly. These organisms are highly pervasive, infecting plants and animals including humans, and highly specialized to thrive within specific tissues in its host species. Mycoplasmas are believed to have evolved from low G+C content gram-positive bacteria undergoing significant genome reduction, adapting to the resources available in its host and eliminating unnecessary genes including those responsible for oxidative phosphorylation, ATP generation via pentose phosphate pathway, and the biosynthesis of amino acids, nucleotides, lipids, and cofactors.[1,2] Mycoplasmas also adapted reduced substrate specificity allowing fewer proteins to accomplish the tasks of many.[3,4]

Mycoplasma genitalium is one of the smallest organisms capable of self-replication with a genome of 580 kb which codes for ~480 proteins and is considered the starting point for identifying which genes are essential for life.[1,2] Experimental and bioinformatic methods have identified a subset of as few as 206 to as many as 387 genes that are considered "minimal" for an organism to survive and self-replicate.[1,4–7]

M. genitalium is also a human sexually transmitted infection that preferentially adheres to ciliated epithelial cells in the genitourinary and re-

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: American Cancer Society; Grant number: 58229; Grant sponsor: Director, Office of Science, Office of Basic Energy Sciences of the U.S. DOE; Grant number: DE-AC02-05CH11231; Grant sponsor: MacCHESS (US NIH); Grant number: RR001646; Grant sponsor: US DOE; Grant number: DE-FG02-97ER62443; Grant sponsor: US NSF; Grant number: DMR-0225180; Grant sponsor: National Cancer Institute; Grant number: Y1-CO-1020; Grant sponsor: National Institute of General Medical Science; Grant number: Y1-GM-1104; Grant sponsors: STFC (UK); NIH; NIGMS; Howard Hughes Medical Institute.
Steve Goodison's current address is Section of Urologic Oncology and Cancer Research Institute, M. D. Anderson Cancer Center Orlando, Orlando, FL 32806.
Charles J. Rosser's current address is Section of Urologic Oncology and Cancer Research Institute, M. D. Anderson Cancer Center Orlando, Orlando, FL 32806.
*Correspondence to: Robert McKenna, Department of Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, FL 32610. E-mail: rmckenna@ufl.edu
Received 7 July 2010; Accepted 3 September 2010
Published online 12 October 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22900

© 2010 WILEY-LISS, INC.

spiratory regions.[8–10] It is a cause of nongonococcal urethritis in men and several genital tract inflammatory diseases in women.[11] It has also been associated with infertility, increased HIV-1 transmission, and malignant transformation of prostate cells.[12–14] Testing for *M. genitalium* infections is not trivial and can be expensive, leaving most cases un- or misdiagnosed, leading to improper treatment and an increased instance of antibiotic resistance.[15]

MG289 is a protein from *M. genitalium*, located in an operon tentatively coding for a binding protein dependant substrate transporter. It is annotated as a putative phosphonate binding lipoprotein, followed by MG290, a nucleotide-binding membrane permease, and MG291, an integral membrane protein, with twelve predicted membrane spanning regions.[1,16,17] The gene can be disrupted nonlethally, perhaps due to the inherent redundancy of action between this and several theoretical orphaned substrate-binding lipoproteins. Because the function of this protein is critical enough for a minimal bacterium to have evolved multiple copies, it is still considered to be an "essential" gene.[5–7] It is homologous to the *M. hyorhinis* extracytoplasmic thiamine binding protein Cypl, also known as p37, though they are only 32% identical.[1,18] The crystal structure of Cypl was recently determined, identifying its primary ligand as the cofactor thiamine pyrophosphate (TPP), however, based on sequence alignments it was theorized that the Cypl-like proteins from other *Mycoplasma* species would preferentially bind to thiamine (VIB) or thiamine monophosphate due to a lack of phosphate stabilizing residues.[19,20]

The cloning, expression, purification and preliminary X-ray analysis of MG289 were reported previously.[18] Presented here is the structure solution and refinement to 1.95-Å resolution. The MG289 model is mixed α/β forming two domains with a binding cleft, similar to Cypl and the Type II periplasmic binding proteins of gram-negative bacteria.[19,21,22] As predicted, the ligand found in the binding cleft was VIB. MG289 crystallizes with two molecules in the asymmetric unit. Experimental and computational analysis was undertaken to identify whether this dimerization was crystallographic or biological in nature. The discrepancy between the binding partners of Cypl and MG289 are addressed and a hypothesis regarding the phosphorylation of nucleotides and cofactors in *M. genitalium* discussed.

MATERIALS AND METHODS

Protein expression and purification

MG289 was expressed and purified as described previously.[18] Briefly, a plasmid containing an N-terminally truncated coding sequence (residues 26–386) was transformed into BL21(DE3)pLysS *E. coli* cells. The transformation was used to inoculate 1-L LB media with 100 mg L^{-1} ampicillin and cultured at 37°C until the $OD_{600\ nm}$ was 0.7–1.0. Cells were induced using 1 mL of 1*M* isopropyl β-D-1 thiogalactopyranoside and allowed to express for 3 h. Cells were lysed via French press in 20 m*M* sodium phosphate buffer, pH 7.95. Lysate was centrifuged at 40,000*g* for 20 min at 4°C. A volume of supernatant equivalent to 125 mg of total protein was applied to a 50 mL Bio-Rad Q anion-exchange column attached to a 5 mL Bio-Rad Econo-Pac S cation-exchange column equilibrated in 20 m*M* sodium phosphate buffer, pH 7.95. Flow-through was collected and adjusted to pH 6.1 with acetic acid, then applied to a 5 mL Bio-Rad Econo-Pac S cation-exchange column equilibrated in 20 m*M* sodium acetate, pH 6.1. MG289 was isocratically eluted using 15% 1*M* sodium chloride, 20 m*M* sodium acetate, pH 6.1. Purity was confirmed by 10% SDS-PAGE stained with Coomassie Blue. Concentrations were calculated by absorbance at 280 nm using a calculated extinction coefficient of 54,620 M^{-1} cm^{-1}.

Crystallization, X-ray analysis, structure solution, and refinement

The purified MG289 was buffer exchanged into 50 m*M* Tris-HCl pH 7.5 and concentrated to ~8 mg mL^{-1} using a Centriprep 10 spin column (Millipore, Bedford, MA). Crystallization was performed using hanging-drop vapor diffusion.[23] Crystallization conditions were as follows: 5 μL protein, 5 μL precipitant solution (0.2*M* ammonium sulfate, 0.1*M* sodium acetate trihydrate, pH 4.6, and 15% PEG 4000) and 1 μL Silver Bullet Bio condition 49 (0.16% w/v thiamine monophosphate chloride dihydrate, acetylsalicylic acid, cholic acid, 1,2,3-heptanetriol, vanillin, *N*-acetyl-D-mannosamine in 20 m*M* HEPES sodium, pH 6.8; Hampton Research, Aliso Viejo, CA) equilibrated against 1 mL of precipitant solution. Useful crystals appeared in 3 weeks.

A single crystal was directly flash-cooled in liquid nitrogen without any additional cryoprotection. X-ray diffraction data was collected on beamline 5.0.2 at the Advanced Light Source (λ = 1.0 Å). Two hundred useable images were collected with data to a maximum resolution of 1.95 Å. The data was indexed and scaled using HKL-2000.[24] Data statistics are listed in Table I.

Molecular replacement was performed in MOLREP implemented in the CCP4i suite using the structure of *M. hyorhinis* Cypl (PDB 3eki) as a search model.[19,25,26] Refinement was carried out using alternating rounds of computation using PHENIX and manual model building in COOT.[27,28] Positive residual density in the binding cleft and the monomer interface was modeled as thiamine (VIB) and acetate (ACT) respectively. For a more detailed description of the structure solution refer to Supporting Information. Final refinement statistics are listed in Table I.

Table I
MG289 Data Collection and Refinement Statistics (PDB ID 3myu)

Data statistics	
Space group	$P2_12_12_1$
Cell dimensions (Å)	$a = 49.3$, $b = 90.4$, $c = 175.4$
Resolution (Å)	45.18–1.95 (2.02–1.95)[a]
R_{sym} (%)[b]	7.6 (56.0)
$I/\sigma I$	22.0 (2.7)
Completeness (%)	90.9 (71.0)
Redundancy	7.2 (6.5)
Refinement statistics	
No. reflections	52,800 (4049)
R_{cryst}[c]/R_{free}[d]/$R_{all\ data}$ (%)	22.0/25.5/22.2
Residues in model (chain id)	32–280, 284–367 (A); 32–280, 283–367 (B)
No. of atoms	
Protein (dimer)	5470
VIB (two molecules)	36
ACT (two molecules)	8
Water	247
B-factors (Å^2)	
Protein main-/side-chain	37.1/40.8
VIB/ACT	31.2/46.1
Water	38.6
Ramachandran Plot (%)	
Favored	97.0
Allowed	3.0
Disallowed	0.0
RMSD[e]	
Length (Å)	0.006
Angle (°)	0.938

[a]Values in parenthesis are for highest resolution shell.
[b]$R_{sym} = (\Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle) \times 100$.
[c]$R_{cryst} = (\Sigma||F_{obs}| - |F_c||/\Sigma|F_{obs}|) \times 100$.
[d]R_{free} is calculated the same as R_{work}, except it uses 5% of reflection data omitted from refinement.
[e]The root mean squared deviation from ideal values.

Size exclusion chromatography

Purified MG289 and Cypl in 50 m*M* Tris-HCl pH 7.5 were each diluted to a concentration of 2.5 mg mL^{-1}. About 150 µL (~400 µg of purified protein) was loaded onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare) equilibrated with 150 m*M* sodium chloride, 50 m*M* sodium phosphate, pH 7.2. The flow rate was 0.4 mL min^{-1} and peaks were detected by absorbance at 280 nm. Molecular weight estimates were calibrated using Bio-Rad Gel Filtration Standard.

Small angle X-ray scattering

Purified MG289 was buffer exchanged into 50 m*M* sodium phosphate buffer pH 7.5 and concentrated to 0.5 mg mL^{-1}. The SAXS data was collected at the G1 beamline at the Cornell High-Energy Synchrotron Source (CHESS). The wavelength used was 1.296 Å with a detector distance of 1210 mm. The data was collected between angular displacement values (s) of 0.001 and 0.275 Å^{-1}. The images were processed for intensity and s values using DATASQUEEZE.[29] GNOM was used to compute the radii of gyration and the pairwise distribution function. DAMMIN was used to generate the 3D *ab initio* model. Ten DAMMIN simulations were averaged using the DAMAVER program to generate a final model. This model was converted to a SITUS volume map for docking purposes and the MG289 X-ray dimer was docked manually into the map using the CHIMERA program.[30,31] Theoretical SAXS curves from the crystal structures of the monomer and the dimer were computed using the CRYSOL. GNOM, DAMMIN, DAMAVER, and CRYSOL are all located within the ATSAS suite.[32]

RESULTS

Overall structure of MG289

The structure of MG289 was solved to a resolution of 1.95 Å with an R_{work} and R_{free} of 22.0 and 25.5%, respectively. Data and refinement statistics are listed in Table I. Solution and refinement details are described in the Experimental Methods section.

The MG289 monomer is a mixed α/β fold split into two globular domains (Domains I and II) connected by a hinge region formed by two nonsequential strands and the C-terminal helix [Fig. 1(A)]. The two domains form a deep cleft with thiamine (VIB) tightly bound between them. There are two MG289 monomers in the crystallographic asymmetric unit (ASU) [Fig. 1(B)]. The overall shape of the protein is an irregular prolate ellipsoid with the approximated dimensions of 70 × 40 × 30 Å^3 [Fig. 1(C)]. Domain I is slightly smaller than Domain II. Domain I includes residues 32–115 and 308–352. It has six strands flanked by five helices with β4 crossing the hinge region into Domain II. Domain II includes residues 116–307. Residues 281–283 in chain A and 281–282 in chain B were not modeled due to lack of density. Domain II contains six strands flanked by eight helices with β10 terminating at the hinge junction. The C-terminal helix, α14 (residues 353–367), lies parallel to the hinge region sitting nearly equally between Domains I and II.

The distribution of temperature factors (B-factors) in the structure is consistent with existing protein structures, with the core β-sheets having the lowest B-factors that increase as the residues become more solvent exposed [Fig.1(C)]. Domain I has ~20% lower temperature factors than Domain II partly owing to a densely packed hydrophobic core that imposes high order. This is also the result of an ~7% loss of helical structure with Domain II being composed of long semi-ordered loops forming 42% of the total secondary structure. An additional consideration is that there is a difference in the average temperature factor between the two chains in the ASU. Because of the high resolution of the data the refinement was carried out without noncrystallographic

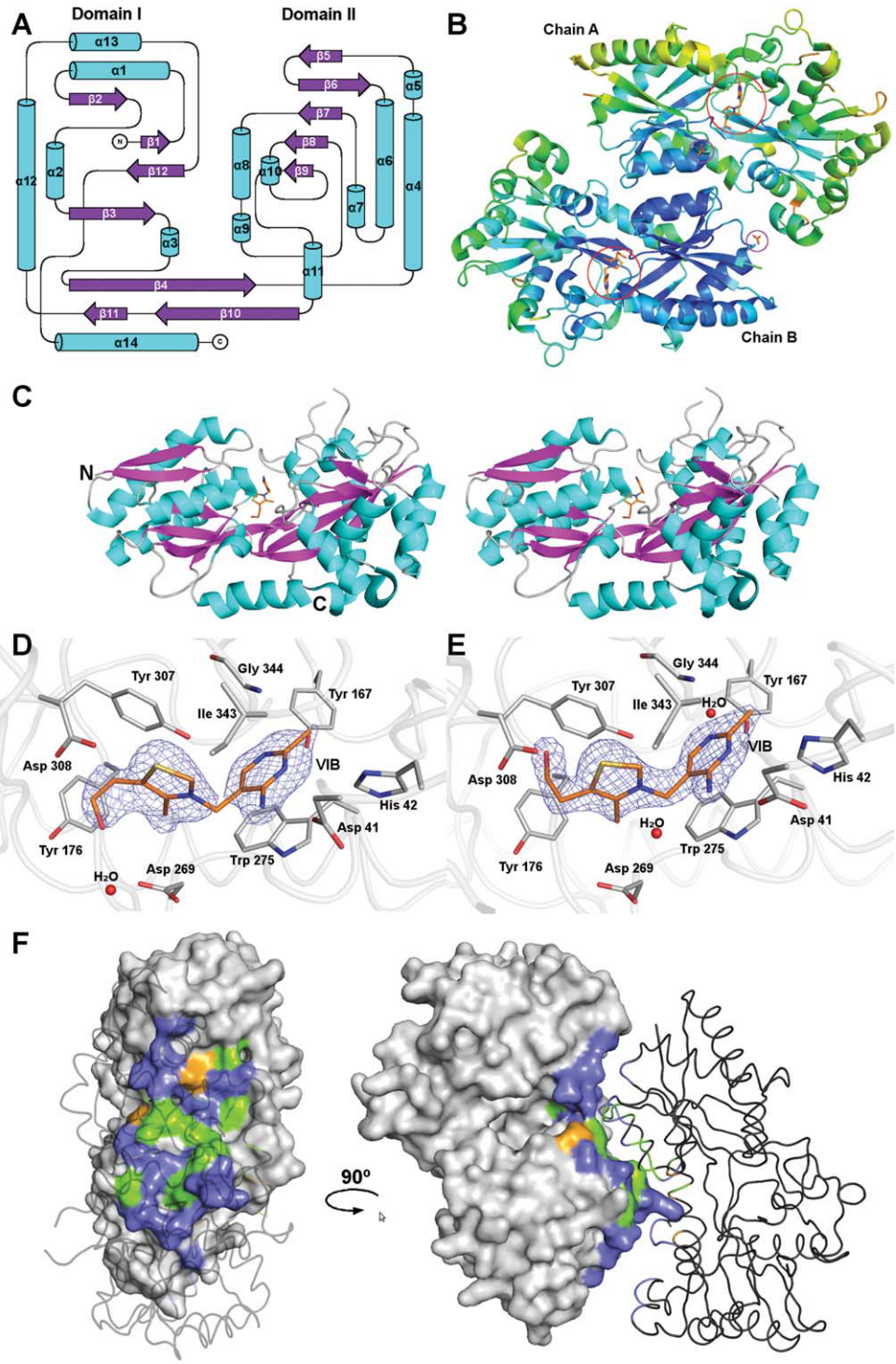

Figure 1

The structure of MG289. (**A**) Secondary structural topology of the MG289 monomer. Helices are colored cyan and strands, magenta. Figure generated in Topdraw.[33] (**B**) A Cα ribbon trace of the MG289 dimer colored from low to high (blue to red) temperature factor. Ligands are shown as sticks with carbon colored orange; oxygen, red; nitrogen, blue; and sulfur, gold. VIB binding sites are indicated as open red circles. Open purple circles indicate ACT binding sites. (**C**) Stereo image of the MG289 monomer with VIB ligand bound. Protein is colored as described in (A), and VIB as described in (B). (**D**) The VIB binding site of Chain A. The $2F_o–F_c$ average kick map (mesh) is contoured to 1.0σ.[34] Protein carbons are colored grey; ligand carbons, orange; oxygen, red; nitrogen, blue; sulfur, gold. Waters are represented as red spheres. (**E**) The VIB binding site of Chain B. The $2F_o–F_c$ average kick map (mesh) is contoured to 1.8σ. Figure colored as described in (D). (F) The dimer interface. Chain A is shown as a molecular surface, while Chain B is a cartoon. Noninterface residues are colored grey; nonpolar, green; polar uncharged, violet, and polar charged, orange. All figures generated in PYMOL unless stated otherwise.[35]

symmetry restraints resulting in an average B of 44.5 Å^2 for Chain A and 33.6 Å^2 for Chain B. This 24.5% increase in B-factor may be correlated to a 34.4% loss of crystal contact surface area in Chain A as compared to Chain B.[36,37]

Thiamine binding site

Thiamine (VIB) was located at the base of the hinge region at the junction between the two domains in both MG289 monomers in the ASU. In both chains the bulk of the residue interactions are hydrophobic [Fig 1(D,E)]. The methyl pyrimidine ring π stacks with Trp 275, and has Van der Waals interactions with Gly 344, Ile 343, His 42, and Tyr 167. The VIB also has a direct hydrogen bond of ∼3 Å to the OD2 of Asp 41 and there is a water-mediated hydrogen bond to the OH of Tyr 167 in Chain B. Also additional water is bound to N4A at a distance of 3.2 Å in Chain B. The thiazole ring engages in exclusively hydrophobic interactions binding Tyr 307, Asp 269, and Ile 343. The ethanol moiety of VIB exhibits two conformations between the two monomers of the ASU. In Chain A the ethyl portion of the tail lies parallel to the phenol ring of Tyr 176 and the hydroxyl hydrogen bonds through water to Asp 269, and in Chain B, C7 interacts hydrophobically with Tyr 176 and Asp 308, and has a direct hydrogen bond to the N of Asp 308 at a distance of 3.0 Å. Therefore, it appears structurally that either conformation is equally favored.

Dimer interface

The overall accessible surface area of the MG289 monomer, as calculated by the PISA server, is 15940.0 Å^2, of which 1041 Å^2 (9.6%) is buried in the interface between the two molecules in the ASU.[37] There is an estimated gain of −4.2 kcal mol^{-1} upon complexation. The composition of the interface residues is as follows: 12 hydrophobic residues, 18 polar uncharged residues, and 2 polar charged residues for Chain A and 12, 17, and 3 residues respectively for Chain B [Fig. 1(F)]. There are 10 hydrogen bonds in the interface, mostly between Asn, Gln, and Thr residues, and no disulfides, covalent interactions, salt bridges or large continuous hydrophobic patches were detected. The overall complexation significance score (CSS) of the interface was 0.113. CSS ranges from 0 to 1 as the biological relevance of the interaction increases. This indicates that the interface may be a very weak interaction that would only play an auxiliary role in complexation.

Size exclusion chromatography

In an effort to confirm the computational results, experimental methods were employed to identify the quaternary structure of MG289 in solution. Size exclusion chromatography (SEC) was used to derive an approximate molecular weight of the soluble species. As the shape of a protein is a factor in SEC, in addition to the globular molecular weight standards, the MG289 homolog Cypl was used for comparative purposes. Cypl is highly similar in shape to MG289 and known to be monomeric [Fig. 2(A)].[19] The MG289 has a computed molecular weight of 39.5 kDa and has an elution peak at 15.9 mL, which corresponds to ∼45 kDa. This is compared to Cypl with a computed molecular weight of 43.4 kDa. Cypl elutes at 15.7 mL corresponding to a molecular weight of 49 kDa. There was a small peak at ∼14 mL; however SDS-PAGE of the fraction showed that it was contaminating protein (data not shown).

Small angle X-ray scattering (SAXS)

Given the deviation of the SEC estimated molecular weight of MG289 from computed values, a second method of experimental verification was employed. The estimated radius of gyration (R_g) of the monomer and the dimer were 22.4 and 31.8 Å, respectively, indicating that SAXS would be amenable to addressing the question of monomer versus dimer. The experimental R_g of MG289 was calculated to be 33.7 Å [Fig. 2(B)]. The overall shape of the dimer is a slightly twisted disk, whereas the monomer is dumbbell shaped. The shape of the pairwise distribution curve can describe this information. The curve of the experimental data strongly resembles the dimer [Fig. 2(C)]. The experimental data was used to derive a low resolution *ab initio* model of MG289 in solution. The overall shape and size of this model is consistent with a dimer (Supporting Information Fig. S1).

Comparison of MG289 to homolog Cypl

When the overall structure of MG289 is compared to the *M. hyorhinis* homolog Cypl, there are strong similarities between to two proteins [Fig. 3(A)]. The RMSD of the Cα is 1.36 Å for the 365 comparable atoms. As mentioned previously the two share ∼32% sequence identity with high conservation in the ligand binding site residues.[19] The largest structural variation is in the truncation of two surface loops in MG289 as compared to Cypl, one along the exterior of the binding cleft and one located proximal to the dimer interface [Fig. 3(A)]. The binding cleft loop is shortened in MG289 bringing it closer to the ligand binding site, which leads to several additional residues stabilizing the methyl pyrimidine ring in VIB that do not have equivalent amino acids in Cypl [Fig. 3(B)].

Comparison of the Cypl and MG289 ligand binding sites show that, as predicted, several of the ligand binding residues are highly conserved, at least for those interacting with the methyl pyrimidine and thiazole ring (see [Fig. 3(B)]).[19] The Trp, Gly, and three Tyr residues are nearly superimposable. The conservative substitution

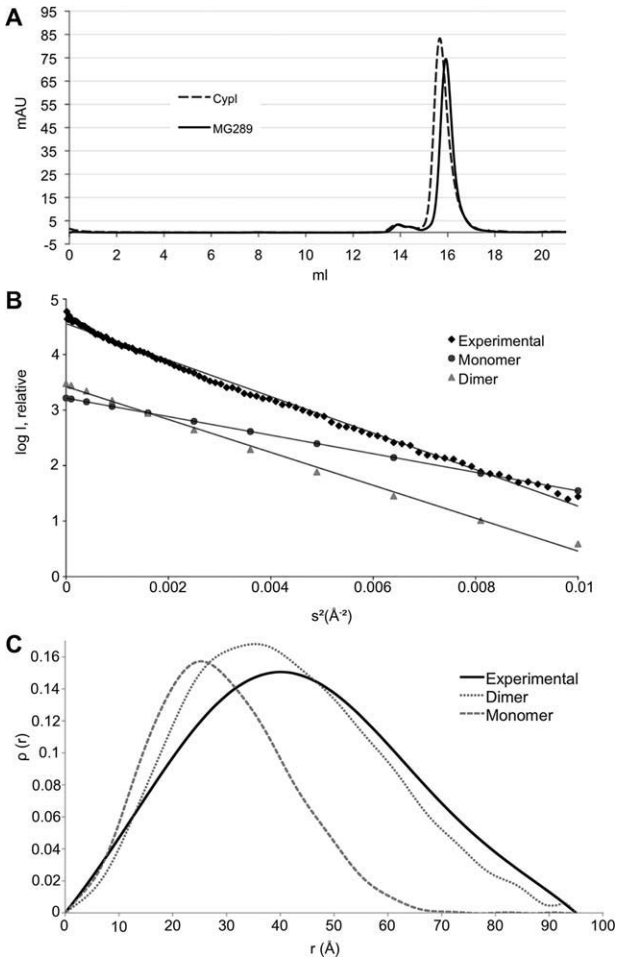

Figure 2

The quaternary structure of MG289. (**A**) Size exclusion chromatograph of MG289 (solid line) and Cypl (dotted line) with UV absorbance (mAU) at 280 nm plotted against elution volume (mL). (**B**) Guinier plot of the squared angular displacement values (s^2) versus the intensity (I). Experimental data (◆) is charted against the scaled theoretical plots of the monomer (●) and the dimer (▲). (**C**) Pair distribution function [$\rho(r)$] versus the distance from the center of gravity (r). Experimental data (solid line) charted against the scaled theoretical plots of the monomer (dashed line) and the dimer (dotted line). $\rho(r) = \frac{r}{2\pi^2}\int_0^\infty I(q)q\sin(qr)\mathrm{d}q$, where $q = 4\pi\sin(\theta)/\lambda$ with λ = wavelength in E and the 2θ = scattering angle. The theoretical scale factor = [Σ (theoretical y-values)/no. points]/[Σ (experimental y-value)/no. observations].

of Ile for Leu also causes very little change in the pyrimidine ring orientation. Residues 41 and 42 of MG289, which interact with the methyl pyrimidine ring, have no equivalent amino acids in Cypl because they are located on the shortened binding cleft loop. The greatest structural differences are located in the tail region. TPP has an ethylene bridged pyrophosphate tail, whereas VIB merely has the ethanol group. In Cypl the TPP is stabilized by three Ser and two Lys residues, whereas in MG289, one of the Ser is replaced by an Ala, one Lys by a Gln, and the other Lys by an Arg.

DISCUSSION

The overall structure of MG289 resembles a periplasmic substrate binding protein (SBP). More specifically, it resembles the Type II periplasmic binding proteins derived from domain displacement as described by Fukami-Kobayashi.[22] Recently a more extensive classification system for SBPs was developed.[38] Based on this new classification system MG289 is a Class D SBP, Subclass II based off of the VIB ligand. This assertion was confirmed by the secondary structural matching results obtained from the DALI server, as the first 200+ structure matches are Class D.[39] Since mycoplasmas lack a periplasmic space having evolved from gram-positive bacteria and combined with the fact that they have a lipoprotein moiety, which was

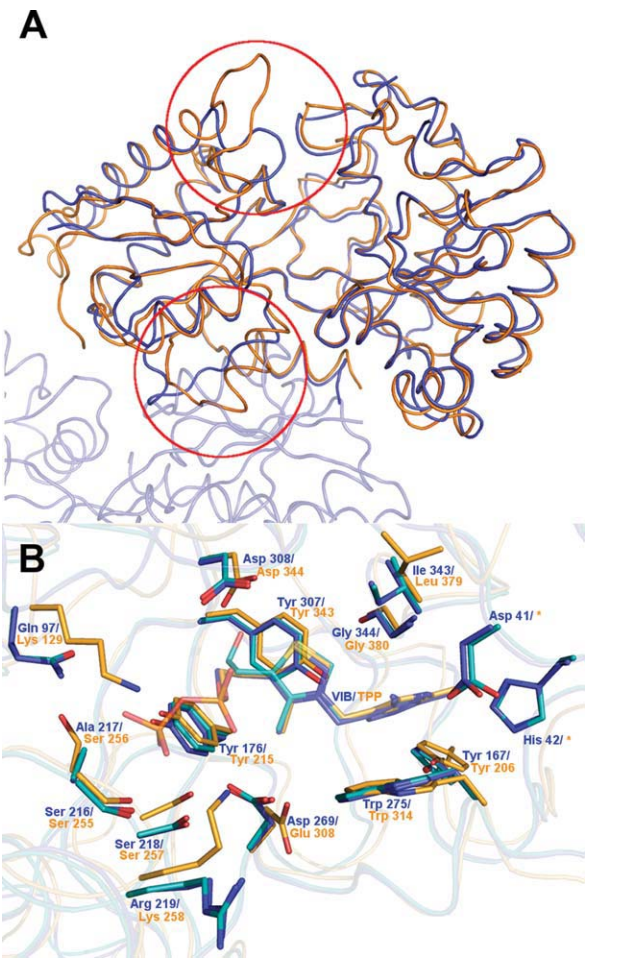

Figure 3

MG 289 versus Cypl. (**A**) Secondary structure matching alignment of MG289 (blue) Chain A and Cypl (orange). Regions of significant difference are indicated by open red circles. (**B**) Superposition of ligand binding site from MG289 and Cypl. MG289 Chain A carbons are colored blue; Chain B carbons, cyan; Cypl carbons, light orange; oxygen, red; nitrogen, blue; sulfur, gold; phosphate, orange. Blue text indicates MG289 residue numbering; Orange, Cypl numbering. * indicates no equivalent residues. Figure generated in Pymol.[35]

Table II
List of Putative *M. genitalium* Thiamine Kinases

MG locus	Gene name	Function	UniprotKB accession
038	glpK	Glycerol kinase	P47284
085	hprK	HPr kinase/phosphorylase	P47331
128	ppnK	Putative inorganic polyphosphate/ATP-NAD kinase	P47374
145	ribF	Putative riboflavin biosynthesis protein	P47391
268	n/a	Putative purine deoxynucleoside kinase	P47510
356	n/a	Putative choline/ethanolamine kinase	Q49423
382	Udk	Uridine kinase	P47622

truncated for expression purposes, this supports the idea that MG289 is probably extracytoplasmic, being presented on the exterior of the cell membrane.

The quaternary structure of MG289 has not been fully answered. The computational and SEC data indicates a monomeric protein whereas the SAXS data indicates a dimer. Computationally the dimer lacks many of the critical features of a biological interface, including a large percentage of buried surface area, an adequate number of hydrophobic residues, and any interactions stronger than a hydrogen bond.[40] The SEC experiments support the computational results. Though there was a 5 kDa difference between the calculated molecular weight and the experimental one, this can be accounted for by the distinctive shape of MG289 which would affect its mobility in a size exclusion column. Its match to the elution curve of Cypl, a monomer, confirms its quaternary structure in solution.[19] However, the SAXS data contradicts the computation and SEC data fitting to a dimeric structure. It is likely that the interaction is concentration dependant. The SAXS data was collected at a concentration of 0.5 mg ml^{-1}, whereas the SEC is 0.05 mg mL^{-1}. Prokaryotic SBPs are usually monomeric when bound to an ABC transporter.[38] Occasionally they can purify as dimers, however in those cases the SPBs would shift towards the monomeric form upon ligand binding.[41,42] In the context of the native lipid-associated form the protein must interact with its transporter to transfer its ligand. The observed interface may be the result of a "sticky" integral membrane binding region of the protein surface trying to bury itself while in concentrated solution.

When compared to the *M. hyorhinis* homolog Cypl, there are strong similarities between the two proteins. The largest distinction is in the loss of several large loops in the binding cleft and at the binding interface [Fig. 3(A)]. This interface loop might play a role in the dimeric nature of the protein in concentrated solution. Cypl crystallizes in the $P2_1$ space group with only one molecule in the ASU. Cypl has an extended loop in what would be the dimer interface region, which may account for both the monomeric structure of Cypl in solution and the lack of a necessary crystal contact that would lead to the two molecules in the ASU and the $P2_12_12_1$ space group of MG289 [Fig. 3(A)].[19,20]

The ligand binding residues show a high degree of conservation between MG289 and Cypl, though there are differences [Fig. 3(B)]. The residues of the tail region seem to show the greatest sequence variation. The three Ser and two Lys residues strongly stabilize the pyrophosphate in Cypl. In MG289, the residues are two Ser, an Ala, a Gln and an Arg that do not interact with the VIB. These differences would not preclude the possibility for low affinity binding of a phosphate or even pyrophosphate moiety, but they are obviously not the preferred ligand. There is also the addition of residues 41 and 42 from the shortened binding cleft loop in MG289. This does not seem to make a difference in the context of thiamine selection; however it may contribute to the affinity for other ligands. Further studies of the kinetics of MG289 and Cypl would be needed to fully characterize the significance. Both ligand binding sites lack strong hydrogen bonds in favor of less specific hydrophobic interactions that may allow for some of the predicted substrate promiscuity including the transport of unsynthesizable guanylate or phosphonates.[6,7,43]

When comparing MG289 and Cypl, the most obvious question that arises is why does *M. hyorhinis* have a preference towards binding TPP while *M. genitalium* binds VIB? Thiamine is a critical component of Mycoplasma minimal media.[44,45] This is because most Mycoplasmas are solely dependant on glycolysis, fermentation, and substrate phosphorylation to meet energetic requirements.[1,45] Two enzymes critical to survival are pyruvate dehydrogenase and transketolase, both of which need TPP as a cofactor.[4,45] Generation of TPP from thiamine requires its phosphorylation, usually by some form of nucleoside diphosphokinase (Ndk). In most known genomes, including prokaryotes, there is some form of highly conserved Ndk; however no such gene exists in Mollicutes. It is theorized that to compensate nonorthologous gene displacement occurred allowing other sugar kinases to perform the duties of Ndk.[3–5,43,45] From the crystallographically observed preferential binding of Cypl for TPP rather than VIB it seems likely that *M. hyorhinis* lacks the kinase necessary for TPP production inside the cell; a kinase produced by *M. genitalium*. Several possible

Ndk replacements have been suggested including 6-phosphofructokinase (MG215), phosphoglycerate kinase (MG300), pyruvate kinase (MG216) and acetate kinase (MG357) that can phosphorylate nucleotide mono- and diphosphates, a dephospho-CoA kinase (MG264) and a putative deoxyribonucleoside kinase (MG268).[3–5] We have recently sequenced ~96% of the *M. hyorhinis* genome, with the remainder being mostly repetitive, non-coding sequences. We searched the *M. hyorhinis* genome for homologs of all known or putative sugar or nucleoside kinase sequences found in *M. genitalium* (data not shown). Seven *M. genitalium* sugar or nucleoside kinases had no detectable homologs in the *M. hyorhinis* genome. These include MG268, mentioned above, as well as a choline/ethanolamine kinase (MG356), uridine kinase (MG382), glycerol kinase (MG038), HPr kinase/phosphorylase (MG085), an inorganic phosphate/ATP-NAD kinase (MG128), and a riboflavin biosynthetic protein (MG145). Full descriptions of the kinases and accession numbers are listed in Table II. MG268, MG382, and MG128 seem to have a higher probability of being the actual thiamine kinase given the nature of their "endogenous" substrate.

Another consideration is the mechanism of VIB transport into the cell. The remaining genes encoded on the operon include a nucleotide binding protein and a transmembrane protein, typical of prokaryotic ATP-binding cassette (ABC) transporters. It is theorized that the transporter may function as a VIB/proton symporter using proton motive force as a propellant, however this does not explain the nucleotide binding protein on the operon.[45] Nucleotide binding proteins or domains are equipped for substrate hydrolysis, usually ATP to power transport.[46] This implies that thiamine and probably other nucleoside transport is achieved via ABC transporters rather than symporters as previously theorized.

CONCLUSIONS

The structure of MG289 has revealed many interesting insights into mycoplasma metabolism as a whole. In the context of a minimal genome it discloses a possible mechanism for promiscuity in ligand binding through a dearth of specificity conferring hydrogen bonds. It behaves as a weak dimer in concentrated solution, which may reveal a biological relevant interface in the context of a plasma membrane and interactions with a transmembrane protein. The preference of the protein for VIB, the precursor, rather than TPP, the enzymatic cofactor, indicates that it possesses a kinase capable of processing VIB. The list of putative kinases was reduced to seven possibilities. Further studies will be necessary to identify the overall structure of the transporter, the true kinase and the extent of its binding diversity. Hopefully, this knowledge will be used to further aid our understanding about the metabolic requirements of a minimal genome and may lead to the development of better drugs for the treatment of *M. genitalium* infections.

ACKNOWLEDGMENTS

The authors thank Dr. Arthur H. Robbins and the presenters of the 2010 CCP4 workshop for assistance and advice during structure solution and refinement, Dr. Ranjana Sarma for providing the advice for the SAXS analysis, and Dr. Virginia Urquidi for her contribution to sequencing the *M. hyorhinis* genome.

REFERENCES

1. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, III, Venter JC. The minimal gene complement of *Mycoplasma genitalium*. Science 1995;270:397–403.
2. Maniloff J. Phylogeny of Mycoplasmas. In: Mycoplasmas Molecular Biology and Pathogenesis, Maniloff J, McElheney R, Finch L, Baseman J, editors. Washington, DC, USA: American Society of Microbiology Press; 1992. pp 549–559.
3. Pollack JD, Myers MA, Dandekar T, Herrmann R. Suspected utility of enzymes with multiple activities in the small genome Mycoplasma species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. OMICS 2002;6:247–258.
4. Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 1996;93:10268–10273.
5. Gil R, Silva FJ, Pereto J, Moya A. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 2004;68: 518–537.
6. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, III, Smith HO, Venter JC. Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 2006;103:425–430.
7. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. Global transposon mutagenesis and a minimal Mycoplasma genome. Science 1999;286:2165–2169.
8. Baseman JB, Dallo SF, Tully JG, Rose DL. Isolation and characterization of *Mycoplasma genitalium* strains from the human respiratory tract. J Clin Microbiol 1988;26:2266–2269.
9. Tully JG, Taylor-Robinson D, Cole RM, Rose DL. A newly discovered mycoplasma in the human urogenital tract. Lancet 1981;1: 1288–1291.
10. Tully JG, Taylorrobinson D, Rose DL, Cole RM, Bove JM. *Mycoplasma genitalium*, a new species from the human urogenital tract. Int J Syst Bacteriol 1983;33:387–396.
11. Ross JD, Jensen JS. *Mycoplasma genitalium* as a sexually transmitted infection: implications for screening, testing, and treatment. Sex Transm Infect 2006;82:269–271.
12. Manhart LE, Holmes KK, Hughes JP, Houston LS, Totten PA. *Mycoplasma genitalium* among young adults in the United States: an emerging sexually transmitted infection. Am J Public Health 2007;97:1118–1125.
13. Perez G, Skurnick JH, Denny TN, Stephens R, Kennedy CA, Regivick N, Nahmias A, Lee FK, Lo SC, Wang RY, Weiss SH, Louria DB. Herpes simplex type II and *Mycoplasma genitalium* as risk factors for heterosexual HIV transmission: report from the heterosexual HIV transmission study. Int J Infect Dis 1998;3:5–11.
14. Namiki K, Goodison S, Porvasnik S, Allan RW, Iczkowski KA, Urbanek C, Reyes L, Sakamoto N, Rosser CJ. Persistent exposure to

Mycoplasma induces malignant transformation of human prostate cells. PLo S One 2009;4:e6872.
15. Jensen JS. Single-dose azithromycin treatment for *Mycoplasma genitalium*-positive urethritis: best but not good enough. Clin Infect Dis 2009;48:1655–1656.
16. Gilson E, Alloing G, Schmidt T, Claverys JP, Dudler R, Hofnung M. Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in Mycoplasma. EMBO J 1988;7:3971–3974.
17. Dudler R, Schmidhauser C, Parish RW, Wettenhall RE, Schmidt T. A mycoplasma high-affinity transport system and the in vitro invasiveness of mouse sarcoma cells. Embo J 1988;7:3963–3970.
18. Sippel KH, Boehlein SK, Sakai Y, Quirit JG, Agbandje-McKenna M, Rosser CJ, McKenna R. Cloning, expression, purification, crystallization and preliminary X-ray analysis of *Mycoplasma genitalium* protein MG289. Acta Crystallogr Sect F Struct Biol Cryst Commun 2009;65(Part 9):910–912.
19. Sippel KH, Robbins AH, Reutzel R, Boehlein SK, Namiki K, Goodison S, Agbandje-McKenna M, Rosser CJ, McKenna R. Structural insights into the extracytoplasmic thiamine-binding lipoprotein p37 of *Mycoplasma hyorhinis*. J Bacteriol 2009;191: 2585–2592.
20. Sippel KH, Robbins AH, Reutzel R, Domsic J, Boehlein SK, Govindasamy L, Agbandje-McKenna M, Rosser CJ, McKenna R. Structure determination of the cancer-associated *Mycoplasma hyorhinis* protein Mhp37. Acta Crystallogr D Biol Crystallogr 2008;64(Part 11):1172–1178.
21. Quiocho FA, Ledvina PS. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. Mol Microbiol 1996;20:17–25.
22. Fukami-Kobayashi K, Tateno Y, Nishikawa K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. J Mol Biol 1999;286:279–290.
23. McPherson A. Preparation and analysis of protein crystals. New York: Wiley; 1982.
24. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Methods in Enzymology, Carter C, Jr, Sweet R, editors. New Haven, CT, USA: Yale University; 1997. pp 307–326.
25. CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50(Part 5):760–763.
26. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 2010;66(Part 1):22–25.
27. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(12, Part 1): 2126–2132.
28. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 2010;66(Part 2):213–221.
29. Heiney PA. Datasqueeze. v. 2.1. Philadelphia, PA: University of Pennsylvania; 2002.
30. Wriggers W, Milligan RA, McCammon JA. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. J Struct Biol 1999;125:185–195.
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25: 1605–1612.
32. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. J Appl Crystallogr 2007;40:S223–S228.
33. Bond CS. TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 2003;19:311–312.
34. Praaenikar J, Afonine PV, Guncar G, Adams PD, Turk D. Averaged kick maps: less noise, more signal and probably less bias. Acta Crystallogr D Biol Crystallogr 2009;65(Part 9):921–931.
35. DeLano WL. The Pymol molecular graphics system. Palo Alto, CA: Delano Scientific; 2002.
36. Finzel BC, Salemme FR. Anomalous temperature factor behavior and crystal lattice mobility in cytochrome C'. Biophys J 1986; 49:73–76.
37. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007;372:774–797.
38. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. A structural classification of substrate-binding proteins. FEBS Lett 2010;584:2606–2617.
39. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v. 3. Bioinformatics 2008;24: 2780–2781.
40. Janin J, Bahadur RP, Chakrabarti P. Protein–protein interaction and quaternary structure. Q Rev Biophys 2008;41:133–180.
41. Richarme G. Associative properties of the *Escherichia coli* galactose binding protein and maltose binding protein. Biochem Biophys Res Commun 1982;105:476–481.
42. Richarme G. Associative properties of the *Escherichia coli* galactose-binding protein and maltose-binding protein. Biochim Biophys Acta 1983;748:99–108.
43. Pollack JD, Jones MA, Williams MV. The metabolism of AIDS-associated mycoplasmas. Clin Infect Dis 1993;17(Suppl 1): S267–S271.
44. Smith SL, Vandemark PJ, Fabricant J. Respiratory pathways in the Mycoplasma. I. Lactate oxidation by *Mycoplasma gallisepticum*. J Bacteriol 1963;86:893–897.
45. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD. A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, iPS189. PLoS Comput Biol 2009;5:e1000285.
46. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev 2008;72:317–364.