Published in final edited form as:

J Struct Funct Genomics. 2014 December ; 15(4): 215–222. doi:10.1007/s10969-014-9189-7.

Structural characterization of the putative ABC-type 2 transporter from *Thermotoga maritima* MSB8

Ekaterina V. Filippova^{1,2,§,a}, Karolina L. Tkaczuk^{1,2,§}, Maksymilian Chruszcz^{1,2,b}, Xiaohui Xu^{2,3}, Alexei Savchenko^{2,3}, Aled Edwards^{2,3}, and Wladek Minor^{1,2,*}

¹Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA

²Midwest Center for Structural Genomics , USA

³Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, ON M5G 1L6, Canada

Abstract

This study describes the structure of the putative ABC-type 2 transporter TM0543 from *Thermotoga maritima* MSB8 determined at a resolution of 2.3 Å. In comparative sequence-clustering analysis, TM0543 displays similarity to NatAB-like proteins, which are components of the ABC-type Na⁺ efflux pump permease. However, the overall structure fold of the predicted nucleotide-binding domain reveals that it is different from any known structure of ABC-type efflux transporters solved to date. The structure of the putative TM0543 domain also exhibits different dimer architecture and topology of its presumed ATP binding pocket, which may indicate that it does not bind nucleotide at all. Structural analysis of calcium ion binding sites found at the interface between TM0543 dimer subunits suggests that protein may be involved in ion-transporting activity. A detailed analysis of the protein sequence and structure is presented and discussed.

Keywords

putative NatAB permease; putative ABC-type transporter; nucleotide binding domain structure; *Thermotoga maritima*

Introduction

The superfamily of ATP-binding cassette (ABC) transporters comprises a large group of proteins playing different roles in a vast range of biological processes. For the majority of ABC systems, the main role of these proteins is catalysis and the transport of substrates across the membrane [18]. They are involved in the transport of nutrients, elimination of

* correspondence to: wladek@iwonka.med.virginia.edu.

§Both authors contributed equally to the project

^aEVF present address: Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, 303 E. Chicago Ave., Chicago, IL 60611, USA

^bMCh present address: Department of Chemistry and Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC 29208, USA

waste or side products from cells, energy generation, cellular signaling, and antibiotic and antifungal resistance. In humans, they are also associated with many genetic diseases caused by transport failure across the membrane, such as bleeding disorders and many liver and eye diseases [8, 15, 28, 40]. ABC transporters span a large group of paralogous protein families, and their representatives are found in all three domains of life: *Bacteria*, *Archaea* and *Eukaryota*. The typical transporter is a heterotetramer composed of two integral membrane proteins and two cytoplasmic ATPases, which drive the translocation of the substrate by the hydrolysis of ATP. Basic ABC type 2 efflux transporters are integral membrane proteins consisting of two domains: a transmembrane domain (TMD) that provides a translocation pathway, and a cytoplasmic, water-exposed nucleotide-binding domain (ATP-BD) that hydrolyzes ATP [43].

TMDs usually consist of six transmembrane (TM) α -helices. There is little sequence identity between the TMDs of different ABC transporters. ATP-BDs are far more conserved in comparison. They contain highly conserved nucleotide binding motifs, namely Walker A (GXXXXGS/T) and Walker B (hhhhD, where “h” is a hydrophobic residue) [54]. A number of isolated ATP-BDs were crystallized and their three-dimensional structures determined [4, 21, 31, 46, 59]. In addition, several complete structures of different ABC transporters have been reported [13, 19, 22, 23, 29, 39, 41]. Depending on the host organism, the ATP-BD and TMD may be present together in a single polypeptide chain (most prokaryotes), or two separate ones (most eukaryotes). Bacterial and archaeal ABC uptake systems contain an additional (fifth) domain, the “extracellular substrate-binding protein”, which binds the substrate and delivers it to the permease domain.

ABC transporters are classified in the TCDB (Transport Classification Database; <http://www.tcdb.org/>) under the entry number of 3.A.1: the ATP-binding Cassette (ABC) Superfamily. This superfamily consists of more than 200 families that are further divided into sub-groups. Sequence comparisons against this database showed that TM0543 from *Thermotoga maritima* MSB8 belongs to the subfamily of the Na⁺ efflux pump NatAB of ABC-type transporters (3.A.1.115.1), represented by a *Bacillus subtilis* (GI:1663528) transporter [5]. Efflux pumps from this subfamily are typically expressed by a pair of genes collocated in the same operon: one encoding ATP-binding proteins that are either components of an ABC-transport system or a traffic ATPase (*natA*), and one encoding permease substrate-binding proteins (*natB*). However, in the case of TM0543, both the NatA-like and NatB-like domains are encoded by a single gene.

Here, the structure of the predicted nucleotide-binding domain (ATP-BD) of the putative ABC type 2 transporter from *Thermotoga maritima* MSB8 (TM0543) is reported, which was determined by the single-wavelength anomalous dispersion (SAD) method and refined at a resolution of 2.3 Å. A model of the transmembrane domain (TMD) of TM0543 was also constructed using homology modeling. Structural and sequential comparisons of this protein to other known, well-described ABC transporters provide detailed information regarding conserved residues and other characteristic features of TM0543 from *Thermotoga maritima* MSB8.

Materials and methods

Cloning, expression and purification

The recombinant *Thermotoga maritima* MSB8 (TM0543) gene product, including a 6-residue polyhistidine (His6) tag at the N-terminus, was expressed in *E. coli* using the p15T₇ vector. The cloning, expression and purification procedures were done as previously described [60]. Crystals of TM0543 were grown by the hanging-drop vapor-diffusion method at 295 K. For crystallization, a 3 μ L drop consisting of 1.5 μ L protein solution (10 mg/mL Se-Met-TM0543 in 10 mM HEPES buffer, pH 7.5 and 500 mM NaCl) and 1.5 μ L well solution (28% v/v PEG400, 0.2 M CaCl₂ and 0.1 M HEPES buffer pH 7.5) was equilibrated against 100 μ L of well solution. Crystals selected for data collection were transferred to paratone-N oil and flash-cooled in liquid nitrogen at 77 K.

Data collection, structure determination and refinement

Diffraction data were collected from a single crystal of TM0543 at the 19-BM beamline [42] of the Advanced Photon Source (APS) at Argonne National Laboratory. The diffraction data were processed and scaled with the HKL-2000 program suite [36]. X-ray data-collection statistics are summarized in Table 1. The TM0543 structure was solved by single-wavelength anomalous dispersion (SAD). Initial phases were calculated with 2.4 Å resolution data using the program HKL-3000 [33]. The HKL-3000 suite interacts with SHELXD, SHELXE [45], MLPHARE [35], DM [7], CCP4 [57], ARP/wARP [38], and COOT [10]. Subsequent electron density map modification followed by initial model building was done using HKL-3000. The rest of the model was built manually with COOT. The experimental model, comprising residues 47–191 of TM0543, was then refined using REFMAC5 [34]. Solvent atoms were initially built with ARP/wARP, and later added or removed by manual inspection. The final R_{work} and R_{free} (calculated with a randomly selected 5% of reflections omitted from refinement) were 20.5% and 22.7%, respectively. The crystal structure contains one polypeptide chain per asymmetric unit. The refinement statistics and characteristics of the atomic models for TM0543 are given in Table 1. The model was validated with ADIT [58], MOLPROBITY and KING [30]. The atomic coordinates and structure factors for TM0543 have been deposited at the Protein Data Bank (PDB) with accession code 3CNI.

Sequence analyses and clustering

PSI-BLAST [1] was used to search for similar sequences, and multiple sequence alignments were constructed using MUSCLE [9]. The alignments were used to generate a set of profile HMMs (Hidden Markov Models) using HHmake from the HHsearch package [47, 48] and these HMMs were used to search a database of profile-HMMs corresponding to alignments of protein families obtained from the COG, KOG [50], and PFAM [3] databases. This comparison of profile-HMMs (including protein sequences and consensus secondary structure) was done using HHsearch 1.5.0.1 with its default parameters (a maximum of 8 PSI-BLAST iterations, with activated modes of alignment and secondary structure scoring). Transmembrane regions were predicted using HMMTOP [51], TopPred [53], TMPred [17] and SOSUI [16]. Sequence motifs were determined both automatically with Motif Finder [2] and *via* manual analysis of a multiple sequence alignment.

To visualize pairwise similarities between and within protein families, CLANS (CLuster ANALysis of Sequences) [12], a Java utility that applies a version of the Fruchterman–Reingold graph layout algorithm, was used. CLANS creates a three-dimensional graph where each vertex represents a sequence and each edge the pairwise similarity of two sequences. CLANS uses the p -values of high-scoring segment pairs (HSPs) obtained from an $N \times N$ BLAST search to compute attractive and repulsive forces between each sequence pair in a user-defined dataset. The attractive force between each pair of sequences is proportional to $-\log p$ for the pairwise interaction, so more similar pairs attract more strongly. To start, sequence vertices are randomly seeded in 3-D space, and iteratively moved within this environment by the algorithm according to the force vectors resulting from all pairwise interactions until the system converges.

Homology modeling

Both the full-length and individual domains of the target sequence were submitted to the GeneSilico MetaServer [25] to identify the TM0543 fold and to select the best template for homology modeling of the fragments of the protein missing from the crystal structure, namely the TMD. The localization of transmembrane regions (TM, residues 192–412) was predicted based on the amino acid sequence as described in the "sequence analyses and clustering" section. The alignments between the sequence of TM regions reported by the metasever and the structures of selected templates, members of the aquaporin-like fold (SCOP f.19.1.1) as identified by Pcons, were used as a starting point for modeling the regions of TM0543 missing tertiary structure: the N-terminal fragment (residues 1–46) and the C-terminus. The two missing regions were modeled separately and the hybrid model was constructed using the "Frankenstein's Monster" method [24]. This was comprised of cycles of model building by MODELLER [11] and SwissModel, evaluation by MetaMQAPII [37], manual realignment of poorly scored regions and merging of the best scoring fragments (GDT_TS: 40.408, RMSD: 4.512). The quality of the final model (i.e. to which extent it approximates the real, unknown structure) was predicted using MetaMQAP and PROQ [55]. The TMD was docked to the ATP-BD using GRAMM [52].

Structure analysis

Structures were manipulated and modeled with SwissPDBViewer [14, 44]. Visualization and structure figures were generated with CCP4MG [32]. Structural similarity searches and superimpositions were done with DALI [20] and ProFunc [26].

Results and discussion

Multiple sequence analysis of TM0543

A multiple sequence alignment (MSA) was calculated for all sequences homologous to TM0543 collected by a database search. However, the MSA exhibited poor quality due to significant divergence between different paralogous families. In order to identify a group of more closely related (and thus more confidently alignable) members, all collected sequences of TM0543 homologs were clustered using CLANS.

The clustering analysis is presented in Figure 1. Most of the “typical” ABC transporters described in the literature by Reizer *et al.* [40] and Lee *et al.* [27] cluster together (green). There are a number of other clusters as well. COG1511 (yellow) contains predicted membrane proteins. COG0842 (brown) contains multidrug transport system permease proteins. PF01061 (purple) contains bacterial ABC type 2 membrane transporters that export drugs and carbohydrates. COG1668 (blue) contains NatAB Na⁺ efflux pumps, permease components and a putative ABC transporter domain from *Porphyromonas gingivalis* W83 (PDB structure 2P0S). The last three listed COGs also belong to the conserved domain superfamily cl00725 (ABC2 membrane superfamily in NCBI CDD database). Members of COG0842 form part of a defense mechanism that includes COG1668, PF01061, COG1277 (the permease component of the NosY – ABC– type transport system involved in multi–copper enzyme maturation) and COG1682 (TagG, ABC–type polysaccharide/polyol phosphate export systems—a permease component involved in carbohydrate transport and metabolism and cell envelope biogenesis in the outer membrane). The remaining cluster (teal) collects annotated NatAB-like sequences and the TM0543 sequence. This cluster includes other bacterial proteins, including the previously described *Bacillus subtilis* NatB [5]. The NatAB/TM0543 cluster is most closely connected to COG1511, PF01061 and COG1668.

The NatAB family includes all members of COG1668 and members of the family are thought to be the ABC-type Na⁺ efflux pump permease component implicated in energy production and conversion as well as the inorganic ion transport and metabolism. A multiple sequence alignment of TM0543 homologs from NatAB family members is presented in Supplementary Figure S1. Sequence analysis suggests that the most similar homolog of the TM0543 protein is the previously described NatAB system from *Bacillus subtilis*, and is considered to be a part of the ABC type 2 transport system (local sequence identity of 24% and similarity of 46%, as calculated by BLAST pairwise sequence alignment) [56]. Additionally, analysis of potential sequence motifs using the amino acid sequence only predicted the investigated *Thermotoga maritima* target to be a putative ABC type 2 transporter. Motif Finder predicted regions that correspond to 156–395 as an ABC type 2 transporter (PROSITE PROFILE method) domain. Residues 139–364 were assigned by PFAM to PF01061 with an expectation value (*e*-value) of 4.4×10^{-5} and the region comprising residues 173–325 were assigned to the same protein family (PF01061) by the PFAM_FS method with an *e*-value of 9.4×10^{-10} .

According to previously published data [49], the ATP-binding domain of NatAB family members shows a greater degree of conservation than the membrane spanning domain.. Typically ATP-BD domains contain a classical Walker A motif (G-X-X-X-X-G-K-S/T) and Walker B motif (ending in h-h-h-h-D, where h represents any hydrophobic residue), which together comprise the presumed nucleotide binding fold [54]. However, the multiple sequence alignment of TM0543 with its homologs from NatAB family members (Supplementary Figure S1) suggests that the C-terminal TMD domain of the full-length TM0543 has a higher degree of conservation in than N-terminal ATP-BD domain. The region corresponding to the Walker A motif in the predicted ATP-BD domain of TM0543 is only weakly conserved. The Walker B motif is more strongly conserved (see Figure S1),

which is defined only by a conserved aspartic acid (D55 in TM0543) and an V/L/I-X-V/L/I repeat.

Structure of the putative ABC transporter TM0543

The crystal structure of the putative ABC transporter TM0543 from *Thermotoga maritima* MSB8 was obtained by SAD at 2.3 Å resolution. The asymmetric unit of the crystal contains one molecule of the putative N-terminal domain of TM0543. The refined model of the putative domain consists of 145 residues (residues: 47–191) out of 156 presented in the purified protein construct (residues: 37–193). Thirty-three N-terminal residues and one C-terminal residue were omitted in the structure due to the lack of corresponding electron density. The putative TM0543 structure is composed of an 6-stranded antiparallel β -sheet comprising β 1 (48–53), β 2 (72–77), β 3 (94–98), β 4 (112–120), β 5 (163–171), and β 6 (174–176); surrounded by five α -helices and two short helical segments: α 3 (58–70), α 4 (80–91), α 5 (102–109), α 6 (125–143), α 7 (181–188), α 8 (146–150), α 9 (157–160) and α 10 (161–165). Figure 2A shows the overall structure of the putative TM0543 domain.

The crystal packing analysis suggests that putative domain of TM0543 forms a tightly associated dimer with symmetry equivalent molecule (Figure 2B). The total interface area of the surface between dimer subunits (1129 Å²) and predicted interaction energy ($\Delta G = -23.2$ kcal/mol) calculated by PDBePISA server (www.ebi.ac.uk/pdbe/pisa) suggests that TM0543 dimer is most likely biological and important for the function of the protein. The interface area of the dimer is formed by helices α 6, α 8 and α 9. There are in total 32 interface residues that form 5 hydrogen bonds.

The electrostatic potential molecular surface reveals that a cavity with volume of 504 Å³ is formed at the interface in the center of the TM0543 dimer (Figure 2C). There are two symmetry related lysine residues (K142) that create a small positively charged area inside the large negatively charged pocket. The inner surface of the cavity is lined by residues I116, W118, I134, S135, E139, K142, D161, Q167, S180, P181, and E182. The residues I116, I134, S180, and P181 are conserved among TM0543 homologs (Figure S1). The negative potential of the cavity suggests that the TM0543 protein may prefer to bind small positively charged ligands or ions. Two calcium ions related by 2-fold rotation axis were found inside the cavity of TM0543 dimer. Oxygen atoms of the side chains of S135, E182 and four water molecules coordinate calcium ion inside the cavity.

Additionally, we have found that a well-ordered calcium ion is also bound near the Walker B motif on the surface of the putative TM0543 protein. This calcium ion binds to the residue E61 and T56 from Walker B motif through water molecule. The Walker A and Walker B motifs correspond to ATP binding pocket in known structures of NatAB family member[54]. . Therefore, the presence of calcium ion in our structure could occupy the position of divalent metal required for ATP hydrolysis. Unfortunately, the region containing the Walker A motif is omitted in our structure. Therefore, it is difficult to predict the nucleotide-binding site and interactions between ATP and a predicted ATP-BD domain in the TM0543 putative structure. The overall fold of the predicted ATP-BD domain of TM0543 is different from that of any known protein from NatAB family. Due to the relative lack of sequence and structure similarity between ATP-BD domains of TM0543 homologs,

it is not clear if TM0543 is capable of binding nucleotide. It is also possible that N-terminal domain of TM0543 protein serves a completely different function.

Structural comparison of TM0543 against the PDB database using the secondary structure matching program of the ProFunc server identified the structure of a zinc transporter from *Thermus thermophilus* HB8 [6] with PDBID 3BYP (a Z-score of 2.8 and RMSD 3.5 Å, as calculated by ProFunc) as having the most similar crystal structure (Figure 2D).

Additionally, a search for structural homologs with the DALI server revealed another structure of a putative ABC transporter domain (Figure 2E) from *Porphyromonas gingivalis* W83 with PDBID 2P0S (a Z-score of 5.3 and RMSD 3.0 Å, as calculated by DALI).

However, the target protein does not show significant sequence similarity to either of these transporters, and variations in the secondary structure elements at the N- and C-terminus between superimposed structures can be observed (Figure 2D, 2E). In the superimposed structures three β strands (β 1- β 3) and helix α 6 display significant structural conservation. Additionally, the structure of the TM0543 dimer shows significant similarity in dimer architecture to an uncharacterized ABC transporter domain from *Porphyromonas gingivalis* W83, as compared to the dimer of the zinc transporter from *Thermus thermophilus* HB8. All proteins have different topology in each of their putative ligand binding cavities that are most likely related to their specific functions.

Model structure of the Transmembrane Binding Domain

To date, no (full length) crystal structure of a NatAB-like homolog highly similar to TM0543 has been deposited in the PDB. Since the TMD (residues 192–412) was not present in the crystallized protein, we tried to reconstruct a complete structure of TM0543, which might have provided guidelines for domain assembly of the TM0543 transporter. Sequence search analysis for the TMD domain of the TM0543 protein (residues: 192–412) predicted that it is composed almost entirely of transmembrane (TM) helices (see Materials and Methods). Predictions performed with the GeneSilico metaserver showed that the best modeling templates (according to the Pcons consensus method) are members of the aquaporin-like fold composed of a core of 8 helices (2 short helices surrounded by 6 long transmembrane helices; SCOP f.19.1.1). Therefore, the alignments between the sequence of TM regions of TM0543 and the structures of selected templates were used as a starting point for homology modeling of the ternary structure of the C-terminal region of TM0543 (as described in Materials and Methods). The resulting model of TMD was composed of six α -helices (TM1 (193–203), TM2 (225–250), TM3 (289–311), TM4 (320–339), TM5 (348–364), and TM6 (374–400) which together form an anti-parallel bundle (Figure S2). In the model the TM1 is the “linking” helix between the two domains (ATP-BD and TMD). Interactions between the two domains are mostly formed with helices α 1 (residues 3–11) and α 2 (17–41) of the predicted ATP-BD domain (Figure S2), which were not present in the crystallographic structure. These helices were also modeled.

Summary

In conclusion, through sequence-clustering analysis, we have determined that TM0543 is closely related to a family of ABC-type transporters of the sodium efflux pump permease component, in particular to members of the NatAB family. The multiple sequence alignment

of TM0543 with its homologs shows that the predicted ATP-BD domain has lower sequence conservation than TMD domain. We have determined the crystal structure of the putative ATP-BD domain from TM0543 protein from *Thermotoga maritima* MSB8. The structure characterization shows that predicted ATP-BD domain of TM0543 has a different structural fold than any other NatAB family members of known structure. The structure of TM0543 ATP-BD also displays a unique dimer architecture among known ABC transporters and has a large negatively charged cavity formed at the interface between dimer subunits. The electrostatic surface analysis suggests that protein could be involved in transport of positively charged ions such as calcium ions, one of which was found inside the cavity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work described in this paper was supported by NIH PSI grants GM74492 and GM094585. The results shown in this report are derived from work performed at Argonne National Laboratory, at the Structural Biology Center of the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. We would also like to thank Dr Matthew D. Zimmerman for critically reading and correcting the manuscript.

List of abbreviations

TMD	transmembrane domain
ATP-BD	nucleotide-binding domain
TM	transmembrane α -helices
ABC Superfamily	ATP-binding Cassette Superfamily
TCDB	Transport Classification Database (http://www.tcdb.org/)
PDB	Protein Data Bank

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25(17):3389–3402. [PubMed: 9254694]
2. An J, Nakama T, Kubota Y, Sarai A. 3DinSight: an integrated relational database and search tool for the structure, function and properties of biomolecules. *Bioinformatics.* 1998; 14(2):188–195. [PubMed: 9545451]
3. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res.* 2004; 32(Database issue):D138–141. [PubMed: 14681378]
4. Chen J, Lu G, Lin J, Davidson AL, Quirocho FA. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol Cell.* 2003; 12(3):651–661. [PubMed: 14527411]
5. Cheng J, Guffanti AA, Krulwich TA. A two-gene ABC-type transport system that extrudes Na⁺ in *Bacillus subtilis* is induced by ethanol or protonophore. *Mol Microbiol.* 1997; 23(6):1107–1120. [PubMed: 9106203]

6. Cherezov V, Hofer N, Szebenyi DM, Kolaj O, Wall JG, Gillilan R, Srinivasan V, Jaroniec CP, Caffrey M. Insights into the mode of action of a putative zinc transporter CzcB in *Thermus thermophilus*. *Structure*. 2008; 16(9):1378–1388. [PubMed: 18786400]
7. Cowtan KD, Zhang KY. Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol*. 1999; 72(3):245–270. [PubMed: 10581970]
8. Davidson AL, Maloney PC. ABC transporters: how small machines do a big job. *Trends Microbiol*. 2007; 15(10):448–455. [PubMed: 17920277]
9. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004; 5:113. [PubMed: 15318951]
10. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. Pt 12 Pt 1. [PubMed: 15572765]
11. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol*. 2003; 374:461–491. [PubMed: 14696385]
12. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004; 20(18):3702–3704. [PubMed: 15284097]
13. Gerber S, Comellas-Bigler M, Goetz BA, Locher KP. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science*. 2008; 321(5886):246–250. [PubMed: 18511655]
14. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009; 30(Suppl 1):S162–173. [PubMed: 19517507]
15. Higgins CF. ABC transporters: physiology, structure and mechanism--an overview. *Res Microbiol*. 2001; 152(3-4):205–210. [PubMed: 11421269]
16. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. 1998; 14(4):378–379. [PubMed: 9632836]
17. Hofmann K, Stoffel W. TMbase - A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler*. 1993; 166
18. Holland IB, Blight MA. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol*. 1999; 293(2):381–399. [PubMed: 10529352]
19. Hollenstein K, Frei DC, Locher KP. Structure of an ABC transporter in complex with its binding protein. *Nature*. 2007; 446(7132):213–216. [PubMed: 17322901]
20. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol*. 1993; 233(1):123–138. [PubMed: 8377180]
21. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*. 1998; 396(6712):703–707. [PubMed: 9872322]
22. Hvorup RN, Goetz BA, Niederer M, Hollenstein K, Perozo E, Locher KP. Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science*. 2007; 317(5843):1387–1390. [PubMed: 17673622]
23. Kadaba NS, Kaiser JT, Johnson E, Lee A, Rees DC. The high-affinity *E. coli* methionine ABC transporter: structure and allosteric regulation. *Science*. 2008; 321(5886):250–253. [PubMed: 18621668]
24. Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. A "Frankenstein's monster" approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. *Proteins*. 2003; 53(Suppl 6):369–379. [PubMed: 14579325]
25. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res*. 2003; 31(13):3305–3307. [PubMed: 12824313]
26. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*. 2005; 33(Web Server issue):W89–93. [PubMed: 15980588]
27. Lee SJ, Bohm A, Krug M, Boos W. The ABC of binding-protein-dependent transport in Archaea. *Trends Microbiol*. 2007; 15(9):389–397. [PubMed: 17764951]
28. Linton KJ. Structure and function of ABC transporters. *Physiology (Bethesda)*. 2007; 22:122–130. [PubMed: 17420303]

29. Locher KP, Lee AT, Rees DC. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*. 2002; 296(5570):1091–1098. [PubMed: 12004122]
30. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003; 50(3):437–450. [PubMed: 12557186]
31. Lu G, Westbrook JM, Davidson AL, Chen J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc Natl Acad Sci U S A*. 2005; 102(50): 17969–17974. [PubMed: 16326809]
32. McNicholas S, Potterton E, Wilson KS, Noble ME. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr*. 2011; 67:386–394. Pt 4. [PubMed: 21460457]
33. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*. 2006; 62:859–866. Pt 8. [PubMed: 16855301]
34. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997; 53:240–255. Pt 3. [PubMed: 15299926]
35. Otwinowski, Z., editor. Proceedings of the CCP4 study weekend. Daresbury Laboratory; Warrington, UK: 1991. MLPHARE: Isomorphous replacement and anomalous scattering. SERC
36. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography*. 1997; (276):307–326. Pt A.
37. Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM. MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics*. 2008; 9:403. [PubMed: 18823532]
38. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol*. 1999; 6(5):458–463. [PubMed: 10331874]
39. Pinkett HW, Lee AT, Lum P, Locher KP, Rees DC. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science*. 2007; 315(5810):373–377. [PubMed: 17158291]
40. Reizer J, Reizer A, Saier MH Jr. A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Sci*. 1992; 1(10):1326–1332. [PubMed: 1303751]
41. Reyes CL, Chang G. Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science*. 2005; 308(5724):1028–1031. [PubMed: 15890884]
42. Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang RG, Ginell SL, Duke N, Naday I, Lazarz J, Molitsky MJ, Keefe L, Gonczy J, Rock L, Sanishvili R, Walsh MA, Westbrook E, Joachimiak A. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J Synchrotron Radiat*. 2006; 13:30–45. Pt 1. [PubMed: 16371706]
43. Saier MH Jr. Paulsen IT, Sliwinski MK, Pao SS, Skurray RA, Nikaido H. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *Faseb J*. 1998; 12(3):265–274. [PubMed: 9506471]
44. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research*. 2003; 31(13):3381–3385. [PubMed: 12824332]
45. Sheldrick GM. A short history of SHELX. *Acta Crystallogr A*. 2008; 64:112–122. Pt 1. [PubMed: 18156677]
46. Smith PC, Karpowich N, Millen L, Moody JE, Rosen J, Thomas PJ, Hunt JF. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell*. 2002; 10(1):139–149. [PubMed: 12150914]
47. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005; 21(7): 951–960. [PubMed: 15531603]
48. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*. 2005; 33(Web Server issue):W244–248. [PubMed: 15980461]

49. Tam R, Saier MHJ. A bacterial periplasmic receptor homologue with catalytic activity: cyclohexadienyl dehydratase of *Pseudomonas aeruginosa* is homologous to receptors specific for polar amino acids. *Res Microbiol.* Mar-Apr;1993 144(3):165–169. [PubMed: 8210673]
50. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 2003; 4:41. [PubMed: 12969510]
51. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001; 17(9):849–850. [PubMed: 11590105]
52. Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins.* 1997; (Suppl 1):226–230. [PubMed: 9485517]
53. von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol.* 1992; 225(2):487–494. [PubMed: 1593632]
54. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J.* 1982; 1(8):945–951. [PubMed: 6329717]
55. Wallner B, Elofsson A. Prediction of global and local model quality in CASP7 using Pcons and ProQ. *Proteins.* 2007; 69(Suppl 8):184–193. [PubMed: 17894353]
56. Wei Y, Guffanti AA, Krulwich TA. Sequence analysis and functional studies of a chromosomal region of alkaliphilic *Bacillus firmus* OF4 encoding an ABC-type transporter with similarity of sequence and Na⁺ exclusion capacity to the *Bacillus subtilis* NatAB transporter. *Extremophiles.* 1999; 3(2):113–120. [PubMed: 10356997]
57. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* 2011; 67:235–242. Pt 4. [PubMed: 21460441]
58. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:1833–1839. Pt 10. [PubMed: 15388930]
59. Zaitseva J, Oswald C, Jumpertz T, Jenewein S, Wiedenmann A, Holland IB, Schmitt L. A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *Embo J.* 2006; 25(14):3432–3443. [PubMed: 16858415]
60. Zhang CT, Zhang R. A refined accuracy index to evaluate algorithms of protein secondary structure prediction. *Proteins.* 2001; 43(4):520–522. [PubMed: 11340667]

Fig. 1. Clustering analysis of ABC transporters using CLANS

COG0842 (brown) represents multidrug transport system permease components, COG1511 (yellow) predicted membrane proteins, PF01061 (red) ABC2 membrane transporters, and COG1668 (blue) NatB Na⁺ efflux pump permease components. Vertices shown in teal represent annotated NatB-like sequence and the sequence of 3CNI (the query sequence). “Typical” ABC transporters as described by Lee *et al.* [27] and Reizer *et al* [40] are shown in green.

Fig. 2. Structural representation of a putative TM0543 protein

A. Ribbon diagram of the putative domain of ABC-type transporter TM0543 from *Thermotoga maritima* MSB8. Secondary structure elements are labeled. Residues involved in calcium ion (red sphere binding are shown in cylinder representation, colored by atom type (carbon in light green and oxygen in red); B. Ribbon diagram of TM0543 dimer; C. Electrostatic potential molecular surface representation of the TM0543 dimer. The surface was created by program CCP4MG and colored by surface potential charge scaled from negative in red (−0.5 V) to positive in blue (+0.5 V); D. Superposition of TM0543 protein structure (green) with structure of zinc transporter from *Thermus thermophilus* HB8 (gold); E. Superposition of TM0543 protein structure (green) with structure of a putative ABC transporter domain from *Porphyromonas gingivalis* W83 (light blue).

Table 1

Crystallographic parameters, and data collection and refinement statistics.

	TM0543
Crystal Parameters	
Space group	P 6 ₄ 2 2
Cell dimensions:	
a=b, c (Å)	65.6, 141.2
α, β, γ, (°)	90, 90, 120
Matthews coefficient (Å ³ /Da)	2.56
Solvent Content (%)	52
Data Collection	
Wavelength (Å)	0.97934
Temperature, K	100
Resolution (Å)	56.8 - 2.3 (2.34 - 2.30)
R _{sym} (%)	5.0 (71.3)
No. of unique reflections	8587
No of unique reflections in R _{free} set	591
Mean Redundancy	13.4 (10.6)
Overall completeness (%)	99.4 (99.2)
Mean I/σ _I	65.7 (3.4)
Refinement Residuals (f 2σ_f):	
R _{free} (%)	22.3
R _{work} (%)	20.5
Completeness (%)	99.6 (99.2)
Model Quality	
RMSD bond lengths (Å)	0.022
RMSD bond angles (°)	1.8
MolProbity Ramachandran statistics [†] :	
most favored (%)	98.6
allowed (%)	1.4
B-factor from Wilson Plot (Å ²)	64.9
Mean overall B-factor (Å ²)	51.5
Model Contents	
Protomers in ASU	1
Protein residues	37-192
No. of protein atoms	1094
No. of calcium ions	2
No. of water molecules	31

Data for the highest resolution shell are given in parentheses. Ramachandran statistics were calculated with *MOLPROBITY*. Rmerge was calculated for merged Friedel pairs.

[†]Pro and Gly residues were excluded from calculation