University of Groningen

An updated structural classification of substrate-binding proteins

Scheepers, Giel H.; Lycklama A Nijeholt, Jelger A.; Poolman, Bert

Published in:
FEBS Letters

DOI:
[10.1002/1873-3468.12445](https://doi.org/10.1002/1873-3468.12445)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Scheepers, G. H., Lycklama A Nijeholt, J. A., & Poolman, B. (2016). An updated structural classification of substrate-binding proteins. *FEBS Letters*, 590(23), 4393-4401. <https://doi.org/10.1002/1873-3468.12445>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

An updated structural classification of substrate-binding proteins

Giel H. Scheepers[†], Jelger A. Lycklama a Nijeholt[†] and Bert Poolman

Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, Zernike Institute for Advanced Materials, University of Groningen, The Netherlands

Correspondence

B. Poolman, Groningen Biomolecular Sciences and Biotechnology Institute, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
Fax: +31 50 3634165
Tel: +31 50 3634190
E-mail: b.poolman@rug.nl

[†]Shared first-authorship

(Received 4 August 2016, revised 2 September 2016, accepted 25 September 2016, available online 23 October 2016)

doi:10.1002/1873-3468.12445

Edited by Alfonso Valencia

Substrate-binding proteins (SBPs) play an important role in solute uptake and signal transduction. In 2010, Berntsson *et al.* classified the 114 organism-specific SBP structures available at that time and defined six protein clusters, based on their structural similarity. Since then, the number of unique SBP structures has increased almost fivefold, whereas the number of protein entries in the Protein Data Bank (PDB) nearly doubled. On the basis of the much larger dataset, we now subclassify the SBPs within the original clusters. Moreover, we propose a 7th cluster based on a small group of SBPs with structural features significantly different from those observed in the other proteins.

Keywords: ABC transporter; membrane protein; signal transduction; solute transport; structural classification; substrate-binding protein

ATP-binding cassette (ABC) transporters, forming one of the largest known protein families, are responsible for a major part of the solute uptake in bacteria and archaea [1,2]. They use substrate-binding proteins (SBPs) or covalent-linked substrate-binding domains (SBDs) to capture the solute and deliver the molecule for translocation by the transmembrane domain of the ABC transporter [3]. The transport is powered by the binding and hydrolysis of ATP by the nucleotide-binding domains on the cytoplasmic side of the membrane [4].

Although the majority of the known SBPs interact with ABC transporters, they are not the only interaction partners (Fig. 1). In bacteria and archaea, secondary transporters, named tripartite ATP-independent periplasmic (TRAP) transporters and tripartite

tricarboxylate transporters (TTTs), also employ SBPs to capture their ligand [5–7]. Furthermore, in eukaryotes, a subset of metabotropic receptors (GPCRs) and ligand-gated ion channels (LGI) are activated through ligand binding to their respective SBDs [8], and in prokaryotes, some two-component regulatory systems use SBDs for signal recognition [9]. Finally, there are SBPs that function as transcriptional regulators, and here, the ligand-binding moiety is fused to a DNA-binding domain to regulate gene expression [10].

Substrate-binding proteins have a highly conserved three-dimensional structural fold, even though they often have little or no sequence similarity and serve multiple functions. They consist of two α/β domains with a central β -sheet of four to six β -strands flanked by α -helices. The two α/β domains are connected by a

Abbreviations

2c-SK, two-component sensor kinase; ABC, ATP-binding cassette; FFAS, fold and function assignment; GPCR, G-protein-coupled receptor; LGI, ligand-gated ion channel; NPR, natriuretic peptide receptor; PDB, Protein Data Bank; SBD, substrate-binding domain; SBP, substrate-binding protein; TF, transcription factor; TRAP, tripartite ATP-independent periplasmic; TTT, tripartite tricarboxylate transporter.

Fig. 1. Schematic overview of SBP-dependent (membrane) proteins. From left to right: ABC; an ABC importer with the SBP in the periplasm (in Gram-negative prokaryotes), ABC*; an ABC importer with two SBDs fused to the TMD, yielding four SBDs per transporter complex, TRAP/TTT; a tripartite ATP-independent periplasmic or tripartite tricarboxylate transporter with periplasmic SBP, LGI; a ligand-gated ion channel, based on the ionotropic glutamate receptors (tetrameric structures), with at the top the ATD domains involved in the oligomerization of the protein, NPR; a natriuretic peptide receptor with SBD, single transmembrane helix and an intracellular domain, GPCR; a G-protein-coupled receptor with a cytoplasmic domain, based on the metabotropic glutamate receptors, 2c-SK; a two-component sensor kinase and TF; a transcription factor with SBD and DNA-binding domain. In the cartoons, the SBPs or SBDs are indicated in gray and with a ligand bound (different colors). Below the cartoons, we indicate the clusters to which the SBP/SBD of the corresponding system belongs. Type I and Type II refers to different structural and functional classes of ABC transporters, and they clearly are also separate in the type SBP they interact with.

hinge region, which comprises one to three interconnecting strands. The ligand is bound in the region between the two domains, and the binding causes the SBP to close around the ligand much like a ‘Venus Fly-trap’ [11]. We have recently shown that SBDs belonging to different structural classes all bind their ligand via a special type of induced fit rather than conformational selection [12]. The rate of closing increases with ligand concentration as in induced fit, but the proteins can also close in the absence of ligand, which has implications for the transport activity.

In 1999, Fukami-Kobayashi *et al.* divided SBPs into two different classes based on sequence similarity and topological arrangements of the β -sheets, to which later on a third class was added [13,14]. Eleven years later, Berntsson *et al.* [15] performed a reclassification based on the alignments of the available 3D structures of the proteins, which resulted in six different classes, which were described as clusters A to F.

The number of protein structures in the Protein Data Bank has almost doubled since the classification of Berntsson *et al.*, in 2010, and now, there are more than 112 000 protein entries in the PDB. As membrane transporters, channels, and signal transducers involving SBPs are under intense study, we felt it was opportune to redo the structure-based classification on the much larger dataset available in 2016.

Methods

To expand on the 114 SBPs found by Berntsson *et al.* [15,16], the FFAS server was searched against the PDB using the 114 SBPs as input. The FFAS server first carries out a multiple sequence alignment using five PSI-BLAST iterations, and it then calculates a profile from the sequences found in the PSI-BLAST. This profile is compared to the PDB database, and an alignment score for each found structure is calculated [17].

The new PDB codes with a sequence identity to the query of less than 70% were converted to Uniprot IDs after duplicates were removed. These IDs were matched against the Uniprot IDs of the original 114 SBPs so that multiple structures of the same protein could be filtered out. This resulted in about 500 unique PDB codes, which were manually checked to verify that they are indeed associated with transporters, ion channels, signal transducers, or transcription factors. This resulted in the elimination of about 100 structures, such as artificial proteins (synthetic fusions with SBPs) and some enzymes.

The remaining 387 PDB codes were added to the original 114, all of these were pairwise structurally aligned using the PDBeFold server [18], and the RMSD values were determined. The resulting list of RMSD values was converted to a file suitable for input in the

KITSCH program of the PHYLIPS package [19], which was used to build a structural distance tree. The online tool iTOL was then used to visualize the tree [20].

To verify the clusters and classes evident in the tree, protein structures were visually checked, using PYMOL [21], and data regarding ligand specificity were taken from Uniprot and the PDB or primary literature. The full dataset can be found in Table S1.

Results and Discussion

The new structural distance tree is based on almost five times as many structures as the original tree described by Berntsson *et al.* [15] and contains 501 SBP-type structures (Fig. 2). The previous search did not include transcriptional regulators, but the 18 found in our analysis are included in the distance tree.

Interestingly, almost all new structures seem to fall within the six previously described clusters. The number of structures per cluster has increased to varying degrees, which allowed for a more detailed subclassification in most cases (Table 1). Clusters B and E seem to have grown the most, with eightfold to ninefold more structures, and this allowed us to further classify B and E into five and two subclusters, respectively. Below we describe the different clusters and the corresponding changes compared to the previous analysis.

Cluster A

Although cluster A has expanded considerably, the two subclusters identified 6 years ago are still clearly visible in the structural distance tree. The SBPs in cluster A are identified by a single α -helix functioning as the hinge region, and they are all part of ABC import systems. The substrates are either metal ions for subcluster A-I and siderophores or hemes for subcluster

A-II [22–24]. While the overall folds and structural features that determine the clusters are the same between A-I and A-II, the main difference lies in the substrate-binding cavity. The loops lining the cavity display considerable variation to accommodate either the small metal ions or the much larger chelator complexes. An exception to this classification is the MolA protein (PDB: 3PSA), which binds the oxanion molybdate or tungstate [25–27]. The overall structure undeniably classifies the protein as a cluster A member (subcluster A-II), but the substrate-binding cavity shows an intermediate size between A-I and A-II. In the future as more structures appear, it is likely that MolA-type proteins will form a third subcluster.

Cluster B

The more than ninefold increase in structures in cluster B allowed us to distinguish between certain structural elements and subdivide B into five subclusters. SBPs in cluster B are identified by the presence of three interconnecting strands between the two α/β domains. The main determinants for the subdivision are the respective substrates and the absence or presence of one or two α -helices interrupting the second strand (Fig. 3). Subcluster B-I SBPs, where this α -helix is missing, mainly bind sugars or sugar alcohols and autoinducer 2. These SBPs are part of ABC transporters and two-component regulatory systems or function as a transcriptional regulator.

The substrates of subcluster B-II and B-III SBPs are amino acids and aromatic compounds, respectively, and are part of ABC transporters. The defining structural motif for these subclusters is one α -helix for B-II, which is complemented with an additional small α -helix for B-III. Cluster B-IV SBPs are part of the natriuretic peptide receptors (NPR) and are involved in signal transduction. The shift of the α -helix, interrupting the hinge in cluster B-IV, seems to be caused by the presence of a glycan moiety as a result of glycosylation. On the basis of research on a NPR dimer, it has been proposed that the α -helix acts as a spring to constrain the SBP in the closed state [28]. Subcluster B-V shows two small α -helices interrupting the hinge, and the second helix is located more in front of the substrate-binding site. The SBPs of B-V interact with metabotropic glutamate receptors, which belong to the GPCR family, so it is possible that the extra α -helix stabilizes the closed state and thereby enhances the sensitivity of the receptor.

The 18 transcription factors found in our analysis are part of subcluster B-I and D-II. Thirteen of the

Table 1. The number of SBP structures found in 2010 (Berntsson *et al.* [15]) and in 2016 (this paper).

Cluster	2010	2016	Fold increase
A	16	55	3.4
B	14	129	9.2
C	8	30	3.8
D	27	108	4.0
E	8	70	8.6
F	38	106	2.8
G	0	3	NA
Unclassified	3	0	NA
Total	114	501	4.4

Fig. 2. Structural distance tree containing 501 SBP structures. The seven clusters are represented in different colors, and the subclusters are indicated by curved lines.

regulators found in B-I are of the *lac*-repressor type. They bind carbohydrates and play a role in the regulation of sugar metabolism. Interestingly, five transcription factors are now assigned to D-II, while previously all the DNA regulators were proposed to be in cluster B. The DNA-binding SBPs found in cluster D are of the *lysR*-type transcriptional regulators (LTTR) and are structurally different from the *lac*-repressor type found in cluster B. They are involved in the

upregulation of genes under starvation conditions or in the production of virulence factors [29–31].

Cluster C

This cluster contains SBPs with an additional domain in between the two lobes and has molecular masses ranging from 53 to 70 kDa, which is similar to that of cluster G proteins and much larger than that of all

Fig. 3. Cluster B is divided into five subclusters on the basis of the cognate substrate and the location and form of the α -helix that interrupts the second interconnecting strand. The distinctive structural feature is indicated in orange. Class B-IV shows a glycan moiety (red) as the result of glycosylation. The PDB codes of the structures shown here are 3T95, 4NQR, 3SG0, 1JDN, 3KS9, respectively [28,45,46].

Fig. 4. Cluster E is subdivided into two different classes based on their structural dissimilarity. The distinctive structural features are indicated in orange (domain spanning α -helix) and green (α -helices on the backside of the SBP). The PDB codes of the structures shown here are 3FXB (Class E-I) and 2DVZ (Class E-II) [36,47].

other types of SBPs. Cluster C proteins bind a considerable range of substrates like (oligo)peptides, cellobiose, opines, and nickel ions, and they all interact with ABC transporters. The large variety in size of these substrates is reflected in the size of the binding cavity. However, the overall structure remains the same, and due to the absence of clearly defined motifs or features linked to cavity size, we could not confidently assign subclusters.

Cluster D

The previous analysis divided cluster D into four subclusters based on the branching in the structural distance tree and the respective substrates. Although

the number of structures has increased fourfold, the four subclusters are still valid.

Subcluster D-III, which Berntsson *et al.* classified as tetrahedral oxanion-binding proteins, is now split in two parts: D-III-a and D-III-b. Although the original classification was based on the cognate substrate, the further subclassification is based on an extra structural element in D-III-b, namely a β -sheet located laterally on the surface of the protein [32,33].

A large part of cluster D forms a separate branch in the distance tree and does not fall under the previously assigned classes. We could not find a clear common motif or substrate and could not confidently assign new subclusters.

Cluster E

All the SBPs found in this cluster associate with either TRAP or TT transporters, which use an electrochemical ion gradient to drive transport. The majority of structures in class E-I (46–67) is the result of an extensive screen for new substrates of TRAP-associated SBPs [34], including ectoine, pyroglutamate, and several uronic, aldarc, dicarboxylic, keto, and sialic acids.

Fig. 5. Cluster G contains three different SBPs; AlgQ1 (PDB: 1Y3P, yellow), AlgQ2 (PDB: 1J1N, green), and an unidentified SBP from *Bifidobacterium longum* (PDB: 3OMB, blue). The bound calcium and magnesium are shown as an orange and red sphere, respectively. The permease component of the alginate transporter is shown in gray (PDB: 4TQU) [40].

Importantly, two SBPs, BugD and BugE [35,36], which Berntsson *et al.* assigned to cluster F, now show a high structural similarity with cluster E proteins. Both proteins have the typical long interconnecting β -strand, which is part of both domains. However, the domain spanning α -helix is shortened to the point where it no longer connects the domains (Fig. 4). Additionally, the α -helices on the backside of the SBP are arranged differently. This prompted us to subdivide BugE and BugD into a separate subcluster, E-II. Also, these SBPs do not seem to interact with TRAP, but with the related TT transporters. Although our analysis did not classify Bug27 (PDB: 2QPQ) in E-II, visual inspection of the structure suggests it also belongs to this subcluster [37].

Cluster F

This cluster showed the least growth in new structures and those found are clearly distributed amongst the four subclusters previously identified. Although the subcluster F-IV seems to have two different branches, closer examination did not reveal discernible structural features nor do they bind different ligands, hence further subdivision was not necessary.

Cluster G

The two SBPs that did not group with any of the clusters in our previous analysis also now do not fall in cluster A-F. The early branching of these two alginate-

Table 2. Correlation of ligand specificity with SBP cluster type and receptor or transport mechanism.

Cluster	Type of ligands	Associated with
A-I	Metal ions	Type II ABC transporter
A-II	Siderophores, hemes, vitamin B12, molybdate, tungstate	Type II ABC transporter
B-I	Sugars, sugar alcohols, autoinducer 2	Type I ABC transporter, 2 component sensor kinase, transcription factor
B-II	Amino acids	Type I ABC transporter
B-III	Aromatic acids	Type I ABC transporter
B-IV	Natriuretic Compounds	Natriuretic peptide receptor
B-V	Glutamate	G-protein-coupled receptor
C	Di- and Oligopeptides, cellobiose, opine, nickel	Type I ABC transporter
D-I	Sugars, sugar alcohols, spermidine/putrescine	Type I ABC transporter
D-II	Thiamine, spermidine/putrescine	Type I ABC transporter, transcription factor
D-IIIa	Phosphate, molybdate, tungstate	Type I ABC transporter
D-IIIb	Molybdate, tungstate	Type I ABC transporter
D-IV	Iron ions	Type I ABC transporter
E-I	Organic acids, amino alcohols, dipeptides, glycerol-3-phosphate	TRAP transporter
E-II	Amino acids	TT transporter
F-I	Thiamine, pyrimidines, sulfonates, bicarbonate	Type I ABC transporter
F-II	Methionine	Type I ABC transporter
F-III	Compatible solutes	Type I ABC transporter
F-IV	Amino acids	Type I ABC transporter, ligand-gated ion channel, G-protein-coupled receptor

binding proteins (PDB: 1Y3P and 1J1N) suggests that their structure is unique and warrants a new 7th cluster; we now found an additional uncharacterized protein that we assign to cluster G. (PDB: 3OMB) [38–40]. Although the relative large size of these proteins of 60 kDa is similar to the molecular mass of the SBPs of cluster C, they have a different tertiary arrangement. Another interesting feature of the cluster G SBPs is the metal-binding site, which is located close to the SBP–membrane domain interface, which suggests a regulatory role in docking of the SBP (Fig. 5). Taken together, this group of SBPs signifies the inclusion of the new cluster G.

When we look at the structural distance tree, seven structures do not fall clearly within any of the seven clusters. However, visual inspection allowed us to assign those structures to one of the clusters, and this information is presented in Table S1.

Conclusion

Initially, when few structures of SBPs were available, the proteins were grouped on the basis of their structure and apparent substrate specificity [41]. In our current analysis, we still observe some correlation between structural class and specificity (Table 2), but within most clusters, the diversity of substrates bound is very large; for example, cluster C proteins bind substrates that range from peptides to sugars and metal ions, and cluster D proteins have sugars, inorganic anions, vitamins, metals, and polyamines as ligands. We anticipate that as more SBP structures become available, they may be subclassified further on the basis of their binding pocket and substrate specificity, as we already noticed for cluster B. However, the overall structures of the SBPs, the basis for the main clusters, tolerate a wide variation of binding pockets with little restrictions to the substrates that can be bound (Table 2). The apparent correlation between substrate specificity and some of the main clusters may thus ultimately vanish. An exception may be the substrates that require large binding pockets, for example, as found in the members of cluster A (siderophores or vitamin B12 [42]) and cluster C (oligopeptides) [43,44]; these substrates do not readily fit in the pockets of cluster B, D–G proteins.

The more than 500 substrate-binding proteins for which crystal structures are available have been classified in seven main clusters, each having distinct structural features. The large number of structures now allowed a further subdivision of the proteins, and in several cases, additional structural elements or discriminative features of the binding pocket were identified.

Furthermore, we found that two types of transcription factors containing a DNA-binding domain are assigned to different clusters. We anticipate that the new and refined classification of SBPs aids in the characterization of this important and widespread class of proteins.

Acknowledgements

We would like to thank C.M. Punter and L. Syga for providing scripts for the conversion of the input and output files. This work was supported by a NWO TOP-Punt (#13.006) grant and an ERC-Advanced Grant (#670578 ABC-Volume) awarded to BP.

Author contributions

BP conceived the study; GS and JL designed and performed experiments, analyzed data and wrote the manuscript; BP made manuscript revisions.

References

- Higgins CF (1992) ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* **8**, 67–113.
- Locher KP (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487–493.
- Fulyani F, Schuurman-Wolters GK, Zagar AV, Guskov A, Slotboom DJ and Poolman B (2013) Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria. *Structure* **21**, 1879–1888.
- Rees DC, Johnson E and Lewinson O (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218–227.
- Fischer M, Zhang QY, Hubbard RE and Thomas GH (2010) Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends Microbiol* **18**, 471–478.
- Winnen B, Hvorup RN and Saier MH (2003) The tripartite tricarboxylate transporter (TTT) family. *Res Microbiol* **154**, 457–465.
- Maqbool A, Horler RSP, Muller A, Wilkinson AJ, Wilson KS and Thomas GH (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem Soc Trans* **43**, 1011–1017.
- O'Hara PJ, Sheppard PO, Thorgersen H, Venezia D, Haldeman BA, McGrane V, Houamed KM, Thomsen C, Gilbert TL and Mulvihill ER (1993) The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41–52.
- Sun Y, Olson R, Horning M, Armstrong N, Mayer M and Gouaux E (2002) Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253.

- 10 Schumacher MA, Choi KY, Zalkin H and Brennan RG (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* **266**, 763–770.
- 11 Mao B, Pear MR, McCammon JA and Quijcho FA (1982) Hinge-bending in L-arabinose-binding protein. The ‘Venus’s-flytrap’ model. *J Biol Chem* **257**, 1131–1133.
- 12 Gouridis G, Schuurman-Wolters GK, Ploetz E, Husada F, Vietrov R, de Boer M, Cordes T and Poolman B (2015) Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat Struct Mol Biol* **22**, 57–64.
- 13 Lee YH, Dekka RK, Norgard MV, Radolf JD and Hasemann CA (1999) *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat Struct Biol* **6**, 628–633.
- 14 Fukami-Kobayashi K, Tatenos Y and Nishikawa K (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol* **286**, 279–290.
- 15 Berntsson RPA, Smits SHJ, Schmitt L, Slotboom DJ and Poolman B (2010) A structural classification of substrate-binding proteins. *FEBS Lett* **584**, 2606–2617.
- 16 Berntsson RPA, Smits SHJ, Schmitt L, Slotboom DJ and Poolman B (2010) Addendum to ‘A structural classification of substrate-binding proteins’ [FEBS Lett. 584 (2010) 2606–2617]. *FEBS Lett* **584**, 4373.
- 17 Jaroszewski L, Rychlewski L, Li Z, Li W and Godzik A (2005) FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res* **33** (Suppl. 2), 284–288.
- 18 Krissinel E and Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2256–2268.
- 19 Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA.
- 20 Letunic I and Bork P (2016) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **44**, 242–245.
- 21 Schrodinger LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.
- 22 Pederick VG, Eijkelkamp BA, Ween MP, Begg SL, Paton JC and McDevitt CA (2014) Acquisition and role of molybdate in *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **80**, 6843–6852.
- 23 Eijkelkamp BA, McDevitt CA and Kitten T (2015) Manganese uptake and streptococcal virulence. *Biomaterials* **28**, 491–508.
- 24 Rice AJ, Park A and Pinkett HW (2014) Diversity in ABC transporters: type I, II and III importers. *Crit Rev Biochem Mol Biol* **49**, 426–437.
- 25 Tirado-Lee L, Lee A, Rees DC and Pinkett HW (2011) Classification of a *Haemophilus influenzae* ABC transporter HII470/71 through its cognate molybdate periplasmic binding protein. *Mol. Structure* **19**, 1701–1710.
- 26 Rice AJ, Harrison A, Alvarez FJD, Davidson AL and Pinkett HW (2014) Small substrate transport and mechanism of a molybdate ATP binding cassette transporter in a lipid environment. *J Biol Chem* **289**, 15005–15013.
- 27 Vigonsky E, Ovcharenko E and Lewinson O (2013) Two molybdate/tungstate ABC transporters that interact very differently with their substrate binding proteins. *Proc Natl Acad Sci USA* **110**, 5440–5445.
- 28 Xi H, De C, Martick MM and Garcia KC (2001) Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. *Science* **293**, 1657–1662.
- 29 Tyrrell R, Verschuere KH, Dodson EJ, Murshudov GN, Addy C and Wilkinson AJ (1997) The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5**, 1017–1032.
- 30 Ding Y, Liu X, Chen F, Di H, Xu B, Zhou L, Deng X, Wu M, Yang C-G and Lan L (2014) Metabolic sensor governing bacterial virulence in *Staphylococcus aureus*. *Proc Natl Acad Sci USA* **111**, E4981–E4990.
- 31 Stec E, Witkowska-Zimny M, Hryniewicz MM, Neumann P, Wilkinson AJ, Brzozowski AM, Verma CS, Zaim J, Wysocki S and Bujacz GD (2006) Structural basis of the sulphate starvation response in *E. coli*: crystal structure and mutational analysis of the cofactor-binding domain of the Cbl transcriptional regulator. *J Mol Biol* **364**, 309–322.
- 32 Hollenstein K, Frei DC and Locher KP (2007) Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216.
- 33 Hollenstein K, Comellas-Bigler M, Bevers LE, Feiters MC, Meyer-Klaucke W, Hagedoorn P-L and Locher KP (2009) Distorted octahedral coordination of tungstate in a subfamily of specific binding proteins. *J Biol Inorg Chem* **14**, 663–672.
- 34 Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, Wichelecki DJ, Bouvier JT, Solbiati JO, Vu H, Zhang X *et al.* (2015) Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**, 909–931.
- 35 Huvent I, Belrhali H, Antoine R, Bompard C, Locher C, Jacob-Dubuisson F and Villeret V (2006) Crystal structure of *Bordetella pertussis* BugD solute receptor unveils the basis of ligand binding in a new family of periplasmic binding proteins. *J Mol Biol* **356**, 1014–1026.

- 36 Huvent I, Belrhali H, Antoine R, Bompard C, Loch C, Jacob-Dubuisson F and Villeret V (2006) Structural analysis of *Bordetella pertussis* BugE solute receptor in a bound conformation. *Acta Crystallogr D Biol Crystallogr* **62** (Pt 11), 1375–1381.
- 37 Herrou J, Bompard C, Antoine R, Leroy A, Rucktooa P, Hot D, Huvent I, Loch C, Villeret V and Jacob-Dubuisson F (2007) Structure-based mechanism of ligand binding for periplasmic solute-binding protein of the Bug family. *J Mol Biol* **373**, 954–964.
- 38 Momma K, Mishima Y, Hashimoto W, Mikami B and Murata K (2005) Direct evidence for *Sphingomonas* sp. A1 periplasmic proteins as macromolecule-binding proteins associated with the ABC transporter: molecular insights into alginate transport in the periplasm. *Biochemistry* **44**, 5053–5064.
- 39 Mishima Y, Momma K, Hashimoto W, Mikami B and Murata K (2003) Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of *Sphingomonas* sp. A1, complexed with an alginate tetrasaccharide at 1.6-Å resolution. *J Biol Chem* **278**, 6552–6559.
- 40 Maruyama Y, Itoh T, Kaneko A, Nishitani Y, Mikami B, Hashimoto W and Murata K (2015) Structure of a bacterial ABC transporter involved in the import of an acidic polysaccharide alginate. *Structure* **23**, 1643–1654.
- 41 Tam R and Saier MH (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* **57**, 320–346.
- 42 Borths EL, Locher KP, Lee AT and Rees DC (2002) The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proc Natl Acad Sci USA* **99**, 16642–16647.
- 43 Detmers FJM, Lanfermeijer FC, Abele R, Jack RW, Tampé R, Konings WN and Poolman B (2000) Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor OppA of *Lactococcus lactis*. *Proc Natl Acad Sci USA* **97**, 12487–12492.
- 44 Berntsson RP-A, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink S-J, Thunnissen A-MWH, Poolman B and Slotboom D-J (2009) The structural basis for peptide selection by the transport receptor OppA. *EMBO J* **28**, 1332–1340.
- 45 Kavanaugh JS, Gakhar L and Horswill AR (2011) The structure of LsrB from *Yersinia pestis* complexed with autoinducer-2. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67** (Pt 12), 1501–1505.
- 46 Tan K, Chang C, Cuff M, Osipiuk J, Landorf E, Mack JC, Zerbs S, Joachimiak A and Collart FR (2013) Structural and functional characterization of solute binding proteins for aromatic compounds derived from lignin: p-coumaric acid and related aromatic acids. *Proteins* **81**, 1709–1726.
- 47 Lecher J, Pittelkow M, Zobel S, Bursy J, Bonig T, Smits SHJ, Schmitt L and Bremer E (2009) The crystal structure of UehA in complex with ectoine-A comparison with other TRAP-T binding proteins. *J Mol Biol* **389**, 58–73.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Overview of substrate-binding proteins available in the Protein Data Bank (PDB) as of July 2016.