

This is a repository copy of *Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): From uptake to pathogenicity.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/127518/>

Version: Accepted Version

Article:

Rosa, Leonardo T., Bianconi, Matheus E., Thomas, Gavin Hugh orcid.org/0000-0002-9763-1313 et al. (1 more author) (2018) Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): From uptake to pathogenicity. *Frontiers in Microbiology*. 33. ISSN: 1664-302X

<https://doi.org/10.3389/fcimb.2018.00033>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): From uptake to pathogenicity

Leonardo T. Rosa¹, Matheus E. Bianconi², Gavin H. Thomas³, David J. Kelly^{1*}

¹Department of Molecular Biology and Biotechnology, University of Sheffield, United Kingdom, ²Animal and Plant Sciences, University of Sheffield, United Kingdom, ³Biology, University of York, United Kingdom

Submitted to Journal:

Frontiers in Cellular and Infection Microbiology

Article type:

Review Article

Manuscript ID:

324356

Received on:

25 Oct 2017

Frontiers website link:

www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

LTR: reviewed the literature, co-wrote the manuscript and co-analysed the bioinformatic data

MEB: generated the bioinformatics data and co-analysed it.

GT: edited and commented on the drafts.

DJK: Conceived the idea and focus of the review, co-wrote and edited the paper.

Keywords

solute transport, Periplasmic Binding Proteins, secondary transporter, high-affinity, Carboxylic Acids

Abstract

Word count: 312

The ability to efficiently scavenge nutrients in the host is essential for the viability of any pathogen. All catabolic pathways must begin with the transport of substrate from the environment through the cytoplasmic membrane, a role executed by membrane transporters. Although several classes of cytoplasmic membrane transporters are described, high-affinity uptake of substrates occurs through Solute Binding-Protein (SBP) dependent systems. Three families of SBP dependant transporters are known; the primary ATP-binding cassette (ABC) transporters, and the secondary Tripartite ATP-independent periplasmic (TRAP) transporters and Tripartite Tricarboxylate Transporters (TTT). The TRAP and TTT systems were last subject to review in 2011 and 2003, respectively. Far less well understood than the ABC family, the TRAP transporters are found to be abundant among bacteria from marine environments, and the TTT transporters are the most abundant family of proteins in many species of β -proteobacteria. In this review, recent knowledge about these families is covered, with emphasis on their physiological and structural mechanisms, relating to several examples of relevant uptake systems in pathogenicity and colonisation, using the SiaPQM sialic acid uptake system from *Haemophilus influenzae* and the TctCBA citrate uptake system of *Salmonella typhimurium* as the prototypes for the TRAP and TTT transporters, respectively. High-throughput analysis of SBPs has recently expanded considerably the range of putative substrates known for TRAP transporters, while the repertoire for the TTT family has yet to be fully explored but both types of systems most commonly transport carboxylates. Specialised spectroscopic techniques and site-directed mutagenesis have enriched our knowledge of the way TRAP binding proteins capture their substrate, while structural comparisons show conserved regions for substrate coordination in both families. Genomic and protein sequence analyses show TTT SBP genes are strikingly overrepresented in some bacteria, especially in the β -proteobacteria and some α -proteobacteria. The reasons for this are not clear but might be related to a role for these proteins in signalling rather than transport.

Funding statement

The present work was accomplished with funding from Brazilian funding agency CNPQ (National Council for Scientific and Technological Development), though PhD studentships in the remit of "Science Without Borders" program to LR (248597/2013-2) and MB (201873/2014-1)

1 Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite
2 tricarboxylate transporters (TTT): From uptake to pathogenicity
3

4 **Leonardo T Rosa¹, Matheus Bianconi², Gavin H Thomas³, David J Kelly^{1*}**
5

6 ¹ Department of Molecular Biology and Biotechnology, The University of Sheffield,
7 Sheffield, UK.

8 ² Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, UK.

9 ³ Department of Biology, University of York, York, UK

10 ***Correspondence**

11 Professor David Kelly

12 d.kelly@sheffield.ac.uk

13 **Keywords:** solute transport; periplasmic binding-proteins; secondary transporter; high-
14 affinity; carboxylic acids

15 **Running Title:** Secondary SBP-dependent transporters in pathogenicity
16
17
18

19 **Number of words (Main text only):** 9022

Number of figures/tables: 9

20

21

22

23 Abstract

24 The ability to efficiently scavenge nutrients in the host is essential for the viability of any
25 pathogen. All catabolic pathways must begin with the transport of substrate from the
26 environment through the cytoplasmic membrane, a role executed by membrane transporters.
27 Although several classes of cytoplasmic membrane transporters are described, high-affinity
28 uptake of substrates occurs through Solute Binding-Protein (SBP) dependent systems. Three
29 families of SBP dependant transporters are known; the primary ATP-binding cassette (ABC)
30 transporters, and the secondary Tripartite ATP-independent periplasmic (TRAP) transporters
31 and Tripartite Tricarboxylate Transporters (TTT). The TRAP and TTT systems were last
32 subject to review in 2011 and 2003, respectively. Far less well understood than the ABC
33 family, the TRAP transporters are found to be abundant among bacteria from marine
34 environments, and the TTT transporters are the most abundant family of proteins in many
35 species of β -proteobacteria. In this review, recent knowledge about these families is covered,
36 with emphasis on their physiological and structural mechanisms, relating to several examples
37 of relevant uptake systems in pathogenicity and colonisation, using the SiaPQM sialic acid
38 uptake system from *Haemophilus influenzae* and the TctCBA citrate uptake system of
39 *Salmonella typhimurium* as the prototypes for the TRAP and TTT transporters, respectively.
40 High-throughput analysis of SBPs has recently expanded considerably the range of putative
41 substrates known for TRAP transporters, while the repertoire for the TTT family has yet to be
42 fully explored but both types of systems most commonly transport carboxylates. Specialised
43 spectroscopic techniques and site-directed mutagenesis have enriched our knowledge of the
44 way TRAP binding proteins capture their substrate, while structural comparisons show
45 conserved regions for substrate coordination in both families. Genomic and protein sequence
46 analyses show TTT SBP genes are strikingly overrepresented in some bacteria, especially in
47 the β -proteobacteria and some α -proteobacteria. The reasons for this are not clear but might
48 be related to a role for these proteins in signalling rather than transport.

49

50

51 1. Solute Binding-Protein (SBP) dependant secondary transporters: The TRAP and
52 TTT systems

53

54 Solute Binding-Protein (SBP) dependent transport systems contain, in addition to the
55 membrane proteins, a soluble extra-cytoplasmic protein, located either free in the periplasm
56 or anchored to the membrane in the case of Gram-positive bacteria, which binds the substrate
57 with high affinity and specificity, allowing uptake even in very low concentrations of ligands.
58 Three families of SBP dependant transporters are currently known, the composition of which
59 are summarized in Fig. 1. ATP-binding cassette (ABC) transporters use the free energy of
60 ATP binding and hydrolysis to move substrates across the membrane against a concentration
61 gradient. First described in the early 1970's (Kalckar 1971, Willis and Furlong 1974), this
62 family is by far the best investigated SBP-dependant transporter family, with the maltose and
63 vitamin B₁₂ uptake systems as the most thoroughly studied models, and was subject of
64 several reviews over recent years (Jones and George 2004, Davidson *et al.* 2008, Rice *et al.*
65 2014, Maqbool *et al.* 2015, Wilkens 2015, Locher 2016).

66 The Tripartite ATP-independent periplasmic (TRAP) transporters (TC: 2.A.56) and Tripartite
67 Tricarboxylate Transporters (TTT) (TC: 2.A.80), on the other hand, use ion-electrochemical

68 gradients to move substrates in a symporter mechanism, thus being defined as secondary
69 transporters. These two families are significantly less well understood than ABC systems but
70 share a similar overall protein composition and topology, as well as genomic organization. In
71 addition to the SBP's ("P" subunit in TRAP systems, "C" subunit in TTT), each system is
72 comprised of two transmembrane proteins, one well-conserved 12 transmembrane (TM)
73 domain protein ("M" subunit in the TRAP systems, "A" subunit in TTT) and one poorly
74 conserved 4 TM domain protein ("Q" subunit in TRAP systems, "B" subunit in the TTT)
75 (Forward *et al.* 1997, Winnen *et al.* 2003, Thomas *et al.* 2006, Hosaka *et al.* 2013) (Fig 1).
76 However, no sequence similarity is found between the corresponding proteins in these
77 families, thus representing either a case of convergent evolution (Fischer *et al.* 2010) or very
78 ancient orthology and divergence (Winnen *et al.* 2003).

79 Regardless of their lack of sequence similarity, the SBPs from these two families show very
80 similar tertiary structures. They are folded in a 'Venus fly-trap' shape, with two wings
81 composed of one β -sheet containing four to six strands, surrounded by α -helices and
82 connected by a hinge. Opened in the *apo* form, the wings close around the substrate in a very
83 specific manner, binding the substrate tightly in a cleft formed between the two domains. The
84 enclosure of the substrate then allows the protein to interact with the transmembrane domains
85 (Herrou *et al.* 2007). It is suggested that these two wings were generated by a duplication
86 event in early TTT (and other SBP dependent) transporters (Winnen *et al.* 2003).
87 Classification of SBPs into related clusters has been proposed, based on their secondary and
88 tertiary structural patterns and their substrate specificities, with the first classification into
89 three distinct types proposed by Fukami-Kobayashi *et al.* (1999). With the exponential
90 increase in new entries for SBPs in genomic databases due to new sequencing capabilities, it
91 became clear that the separation into three types was too simplistic to comprise SBP
92 diversity, and thus a new model was presented by Berntsson *et al.* (2010) and recently revised
93 by (Scheepers *et al.* 2016). Both TRAP and TTT SBP's are contained within the Type II
94 group in the first classification, and inside Cluster E in the latter. This review summarises the
95 evidence of a relationship between these two classes of secondary high-affinity uptake
96 systems and pathogenicity. Additionally, it adds an evolutionary perspective regarding the
97 expansion of the TTT family in some pathogens.

98 **2. The TRAP Transporter family**

99 The first characterization and naming of TRAP transporters was described in *Rhodobacter*
100 *capsulatus* by Forward *et al.* (1997), when a solute binding-protein encoding gene was found
101 adjacent to two genes encoding transmembrane proteins of 12 (DctM) and 4 (DctQ) predicted
102 helices. Functional studies showed symport of C4-dicarboxylic acids apparently energized by
103 the proton motive-force. Subsequent studies showed that these systems can transport a variety
104 of substrates under different contexts. Detailed reviews about this family were provided by
105 Kelly and Thomas (2001) and Mulligan *et al.* (2011), and the following sections will focus on
106 more recent insights.

107 **2.1. Substrate diversity of the TRAP family and roles in pathogenicity**

108 The best studied TRAP system is undoubtedly SiaPQM from *Haemophilus influenzae* (Fig 1
109 and Fig 2a), discovered by Severi *et al.* (2005) to be involved in the uptake of sialic acid.
110 Sialic acid is a generic name for a class of 9-carbon sugar acids used by most eukaryotic cells
111 in the form of cell surface glycoproteins. For this reason, many pathogens evolved to mimic
112 these surface structures in their own cell envelope, constituting an important virulence factor
113 which improves evasion of the human immune system (Bouchet *et al.* 2003). In *H.*
114 *influenzae*, absence of SiaPQM causes loss of sialic acid uptake and lack of incorporation in
115 the lipo-oligosaccharide (Allen *et al.* 2005), and a subsequent study showed an increased

116 susceptibility of this pathogen to human serum and decreased virulence in the chinchilla otitis
117 model (Jenkins *et al.* 2010). Systems homologous to SiaPQM were subsequently found to be
118 involved in the uptake of sialic acid in several pathogens, such as *Vibrio cholerae*,
119 *Fusobacterium nucleatum* and *Vibrio vulnificus* (Severi *et al.* 2005), the latter being shown to
120 transport sialic acid in 67 clinical isolates (Lubin *et al.* 2012). Signature-tagged mutagenesis
121 studies in *Pasteurella multocida*, an opportunistic pathogen of livestock, showed that
122 disruption of genes related to sialic acid metabolism resulted in decrease of virulence in mice
123 models (Fuller *et al.* 2000). A subsequent study showed that a SiaP homolog was involved in
124 the uptake of sialic acid in this bacterium (Steenbergen *et al.* 2005), and that disruption of
125 sialic acid uptake resulted in decreased virulence in a turkey model (Tatum *et al.* 2009).
126 Severi *et al.* (2007) provides a review of how sialic acid uptake and metabolism is used as a
127 virulence factor in different pathogens, and Vimr *et al.* (2004) provides a more general
128 review about sialic acid metabolism. Thomas (2016) gives a recent overview of the different
129 uptake strategies and transport systems used by different pathogens for the uptake of sialic
130 acid.

131 In *Bordetella pertussis*, the causative agent of whooping cough, two DctP homologs are
132 encoded in the vicinity of virulence-related operons modulated by the BvgA/BvgS two-
133 component system. The two proteins were crystalized by Rucktooa *et al.* (2007) with a
134 pyroglutamic acid bound in the substrate cleft. One of these proteins is highly expressed in *B.*
135 *pertussis*, although the membrane components of the system seem to be mutated and non-
136 functional. Although it is unclear what physiological role pyroglutamic acid would have, this
137 amino-acid is present in the filamentous hemagglutinin produced by *B. pertussis*, and it was
138 speculated it could serve as a glutamate reserve. In fact, BugE, an abundantly expressed SBP
139 from the TTT family also was shown to bind glutamate (Huvent *et al.* 2006), suggesting
140 glutamate metabolism might play an important role in the pathophysiology of this bacterium.

141 TRAP systems are very important also in environmental organisms and in biotechnologically
142 relevant processes. In *Halomonas elongata*, Grammann *et al.* (2002) showed that the
143 TeaABC operon was responsible for accumulation of the compatible solute ectoine, in
144 response to osmolarity stress, and that this transporter was osmoregulated. In *Rhodobacter*
145 *capsulatus* a TRAP system was shown to be involved in the import of several
146 monocarboxylic 2-oxo-acids involved in amino-acid biosynthesis (Thomas *et al.* 2006). Chae
147 and Zylstra (2006) showed the involvement of TRAP transporters in the degradation of
148 several benzoate derivatives, including toxic chlorinated aromatics. In *Rhodopseudomonas*
149 *palustris* the TarPQM system was shown to be involved in the degradation of lignin-derived
150 aromatic compounds, in a redundant function also executed by an ABC transporter in the
151 same gene cluster (Salmon *et al.* 2013). GaaPQM from *Agrobacterium tumefaciens* was
152 described to be involved in plant virulence (Zhao and Binns 2016). Maimanacos *et al.* (2016)
153 showed that TRAP transporters are found in the vicinity of arylmalonate decarboxylases
154 (AMDases) and recently, Meinert *et al.* (2017) showed a TRAP system involved with the
155 uptake of five different sugars in *Advenella mimigardefordensis*, but only after they have
156 been converted to their respective sugar acids in the periplasm (Thomas 2017).

157 Vetting *et al.* (2015) published a highly significant study, which multiplied several times our
158 understanding about substrate specificity in TRAP systems. 8240 SBP's were used to build a
159 sequence similarity network, grouping them into several clusters. From these, 304
160 representatives of non-characterized groups were then screened, coupling differential
161 scanning fluorescence, crystallography and mass spectrometry of co-purified ligands. The
162 methodology shows the importance of using complementary methods and proposes an
163 efficient strategy for the study of SBP's. As a result, 71 of the isofunctional clusters had a

164 ligand assigned; 69 high-resolution crystal structures were obtained; previously known
165 ligands were assigned to non-characterized clusters and several new ligands were found to be
166 captured by TRAP transporters, such as D-glucuronate/D-galacturonate, 6-carbon aldonic
167 acids, cell-wall constituents, lipopolysaccharide components, glycerol-3-phosphate/
168 diglycerol-phosphate, 2-acetolactate, orotic acid, indole acids, pantoate/D-erythronate, and
169 ethanolamine, this last being a particular surprise due to its positive charge in contrast to the
170 typical negatively charged carboxylates of most other TRAP transporter substrates. This work
171 was done as part of the Enzyme Function Initiative (EFI), a network aiming to characterize
172 the biochemical and physiological function of different classes of enzymes, among which are
173 soluble binding proteins, through high-throughput sequence/structure based strategies
174 (<http://www.enzymefunction.org/>).

175 **2.2. The neglected group: TAXI-TRAP transporters**

176 It was observed by Kelly and Thomas (2001) that, in some cases, the SBPs associated with
177 the DctQM subunits in the genome showed very limited sequence similarity to DctP, forming
178 a distinct group, TRAP associated extracytoplasmic immunogenic (TAXI) proteins, named
179 after an immunogenic protein of unknown function from the pathogen *Brucella* (Mayfield *et al.*
180 *al.* 1988). A previous study by Rabus *et al.* (1999) had found some similarity between TAXI
181 proteins and the *E. coli* glutamate binding protein, and the only structure available for a
182 TAXI protein, generated by Takahashi *et al.* (2004), reinforced these initial findings, as it was
183 described as a glutamate/glutamine binding protein. However, the deletion of a TAXI protein
184 from *Psychrobacter arcticus* was shown by Bakermans *et al.* (2009) to affect growth also in
185 other dicarboxylic acids such as acetate, butyrate and fumarate. TAXI-TRAP systems usually
186 have the DctQM subunits fused (Fig 1 and Fig 2A) and, because they are found also in many
187 Archaea species, it is believed that this system is an ancient form of TRAP transporter.
188 Although Mulligan *et al.* (2011) provided a brief speculation about potential function of
189 TAXI-TRAP systems based on their genomic context, a complete characterisation of this
190 group is still to be generated.

191 **2.3. Two is not enough: The TPAT system**

192 In addition to the classical TRAP and the TAXI-TRAP transporters, a third class of TRAP
193 system was characterised by Deka *et al.* (2012) in the pathogen *Treponema pallidum*. *T.*
194 *pallidum* is the causative agent of syphilis, a disease which continues to be a challenge in
195 global health. This organism is an obligatory pathogen, which lacks many vital biosynthetic
196 pathways for nucleotides, lipids and most amino-acids, relying on transport systems to obtain
197 these vital requirements from the human host (Radolf *et al.* 2016). Deka *et al.* (2012)
198 observed the existence of a single operon encoding a TRAP system in *T. pallidum* genome,
199 composed of three genes, as shown in Fig. 2a. One *dctP* and one *dctQM* homologue, named
200 *tatP_T* and *tatQM*, and a third gene of unknown function, named *tatT*. The biochemical and
201 crystallographic characterisation of TatT showed this soluble protein was formed by 13 α -
202 helices and one small helix, structured around a central hydrophobic pore which opened to
203 both ends of the structure. Some of these helices were homologous to a tetratricopeptide
204 motif (TPR), normally involved in protein-protein interactions (D'Andrea and Regan 2003),
205 which gave the name for this group of TRAP transport systems as TPR-protein associated
206 transporters (TPAT). Using cross-linking, western blotting, analytical ultracentrifugation and
207 computational modelling, it was shown that TatT formed a trimer, which in turn interacted
208 with three subunits of the DctP homolog *TatP_T* (Deka *et al.* 2012). A later study by
209 Brautigam *et al.* (2012) confirmed these predictions through crystallization of the TatT and
210 *TatP_T* complex. In these structures, it was shown that the substrate cleft from *TatP_T* was
211 aligned to the C-terminal side of the pore in TatT, with minor structural changes happening
212 upon complexation, the main one being the displacement of one loop from *TatP_T* domain 2 in

213 contact with the binding cleft, called a “cleft-finger”. The hydrophobicity observed both in
214 TatT pore and TatP_T cleft, together with the presence of a linear hydrophobic molecule
215 crystalized in the TatT pore, suggested that this system is involved in the uptake of
216 hydrophobic molecules (Deka *et al.* 2012). As both proteins are found in vivo as lipoproteins,
217 anchored to the membrane, it was suggested as a mechanism that this interaction created a
218 chaperone environment for the transport of lipids through the periplasmic hydrophilic
219 environment, where TatT would receive the lipid from the host, anchored in the outer
220 membrane, and transfer it to TatP_T, anchored in the inner membrane, which in turn would
221 deliver it to the TatQM subunit (Brautigam *et al.* 2012). TPAT systems were found in 35
222 other species, among other spirochaetes and also among free-living proteobacteria. In this
223 latter group, it was mostly found in species capable of degrading hydrocarbons, reinforcing
224 the potential role in aliphatic transport this distinct group of TRAP transporters might have.

225

226 **2.4. Biochemical and functional studies of the DctQM subunits**

227 Unlike the ABC transporters, no crystal structures have been obtained to date regarding the
228 membrane components of TRAP systems, however some mechanistic information is
229 available particularly regarding energy-coupling. In many systems, such as SiaPQM from
230 *Haemophilus influenzae*, the DctM and DctQ membrane units are not expressed separately,
231 but fused in one only protein containing 17 transmembrane helices, one more than expected
232 due to an additional helix that connects the cytoplasmic C-terminal part of DctQ with the
233 periplasmic N-terminal part of DctM (Fig 2a) (Mulligan *et al.* 2009). Even when expressed
234 separately, DctM and DctQ were shown to form a tight complex with a 1:1 stoichiometry
235 during the folding procedure, and attempts to separate the two proteins resulted in disruption
236 of function (Mulligan *et al.* 2012). While DctM is believed to form a translocation channel
237 and is a member of the ion transporter superfamily (Rabus *et al.* 1999), the role of DctQ has
238 not been established yet and has a much more variable sequence. It is known that it is
239 essential for transporter function and it was suggested that DctQ might act to mediate
240 interactions between DctM and DctP, chaperoning DctM and stabilising it in the membrane
241 or participating in energy coupling (Wyborn *et al.* 2001). Mulligan *et al.* (2009) performed a
242 series of experiments showing that the presence of Na⁺ ions was required for sialic acid
243 transport via SiaPQM in *H. influenzae*. Replacement of Na⁺ for Li⁺ ions did not result in
244 uptake activity, and although neither ΔpH or Δψ alone resulted in transport in absence of Na⁺,
245 the gradients were able to promote substrate uptake when Na⁺ was present in equal
246 concentrations in both sides of the membrane. These results show that substrate uptake in
247 TRAP transporters is Na⁺ dependent and characterised as an eletrogenic process, where at
248 least two Na⁺ ions are co-transported. Not surprisingly, the TRAP family is widely found in
249 bacteria living in saline environments, using the naturally provided Na⁺ gradient to provide
250 substrate uptake, as discussed by Mulligan *et al.* (2007). In addition, Mulligan *et al.* (2009)
251 showed that in opposition to conventional secondary transporters such as the ones from the
252 MFS family, the transport in the TRAP family is unidirectional. The substrate transporter
253 exposes the binding cavities alternatively in the cytoplasm and the periplasm, but because in
254 TRAP transporters the exposure in the periplasmic side only occurs when in interaction with
255 the SBP, movement in the opposite direction is blocked, even when gradients are inverted.
256 The only condition in which contrary movement was observed was in the presence of an
257 excess of un-liganded SiaP in the periplasm, but these conditions are not physiologically
258 relevant. In addition, it was shown that replacement of the SiaP in *Haemophilus influenzae*
259 (HiSiaP) by an homolog from *Vibrio cholerae* (VcSiaP) did not complement its function,
260 suggesting that the interactions between DctP proteins and the membrane counterparts are
261 specific in each case, rather than promiscuous among the family (Mulligan *et al.* 2009).

262 Mulligan *et al.* (2012) performed these same transport assays and characterisation of the
263 SiaQM subunits in the homologous system from *Vibrio cholerae*, which comprises a true
264 tripartite system instead of the fused subunits. The results from this study were very similar to
265 the *H. influenzae* fused SiaQM system.

266

267 **2.5 Crystal structure and dynamics of TRAP SBP's**

268 The first crystal structure of a TRAP SBP was the SiaP protein from *Haemophilus influenzae*
269 (Muller *et al.* 2006). TRAP SBP's have wings very similar to the Type II proposed structure
270 by Fukami-Kobayashi *et al.* (1999), but with a remarkably large single β -strand, which
271 connects both domains and participates in both β -sheet domains (Fig 3A). In addition, this
272 family contains a long α -helix, which spans both domains and kinks upon ligand binding.
273 These features characterise the TRAP transporters in Cluster E of the division proposed by
274 Scheepers *et al.* (2016). The hinge-bending upon ligand-binding was estimated by Muller *et al.*
275 *et al.* (2006) to be $\sim 30\text{\AA}$ based on comparison between unliganded and ligand protein crystals.

276 Although ligand positioning inside the binding pocket is conserved, the hydrogen bonds and
277 hydrophobic interactions for each molecule coordination vary, making substrate prediction
278 difficult for this family. A conserved arginine residue in domain 2, however, turns out to be
279 crucial for ligand interaction (Fig 3C), as discussed by Fischer *et al.* (2010, 2015). Localised
280 in β -strand 6, which is in a stable β -sheet, the side chain of this highly conserved residue
281 (96.8% of 6142 sequences searched) points towards the binding cavity and, unusually, is
282 stabilised through a hydrophobic patch and a hydrogen bond, with a bending in C β which
283 allows the side chain to reach the pocket (Fig 3A and Fig 3C). In the presence of ligand, it
284 makes a salt bridge with the ligand carboxylate group, believed to be the first step in ligand
285 coordination. This interaction is believed to be critical for proper functioning of most SBPs
286 from the TRAP family as high-affinity binding proteins, although it is not essential for the
287 coordination of domain closure upon ligand binding (Fischer *et al.* 2015). The TatP_T
288 homologs, believed to be involved in the uptake of aliphatic substrates, mostly lack this
289 residue, having it substituted for an alanine (Deka *et al.* 2012). Mutations of this arginine
290 residue in SiaP were shown by Johnston *et al.* (2008) to disrupt sialic acid uptake in *H.*
291 *influenzae* and recently Fischer *et al.* (2015) showed that replacing it by a lysine decreased
292 the binding affinity for sialic acid by SiaP from 0.14 μM to 38.7 μM , and mutating it to an
293 alanine resulted in no binding. Crystallization of these two mutant proteins, however, showed
294 minor differences in ligand coordination, where in place of the missing N atoms, coordinated
295 water molecules bridged the carboxylic group of the ligand to the protein, dissipating the
296 negative charge. Subsequent growth experiments showed that cell growth could be restored
297 in the presence of high external concentrations of sialic acid, as the higher concentration
298 would compensate for the weaker affinity; In the same study, it is shown also that this water
299 coordination enables a higher promiscuity in the binding pocket, allowing it to coordinate an
300 analogue ligand containing an amide group in place of the carboxylic acid. In addition,
301 PELDOR spectroscopy analysis recently performed by Glaenger *et al.* (2017) showed that no
302 intermediate state of VcSiaP is observed in solution upon ligand binding, which can only be
303 in an open or closed conformation. Moreover, the protein does not alternate to closed
304 conformation unless the ligand is present, a fact that supports the current model in which the
305 SBP will only return to the open conformation upon interaction with the membrane
306 components, avoiding unproductive opening and closing of the binding protein (Mulligan *et al.*
307 *et al.* 2011).

308 In addition, the variable positioning of helix 3 across different proteins seem to be
309 responsible for the adaptation of the binding pocket for different ligand sizes, given by
310 structural changes in regions flanking this helix (Lecher *et al.* 2009). In some cases, generally
311 for smaller TRAP ligands, cation atoms are also required for ligand coordination. However,
312 as shown by Akiyama *et al.* (2009), these cations are usually non-specific, and have a
313 structural role to bridge the interaction with the protein chain, and are not necessary when the
314 ligands are capable of filling the respective space and interact directly (Fischer *et al.* 2010).

315 Although most TRAP SBPs are found to act as monomers, there is evidence that some of
316 them might require dimerization for function. Gonin *et al.* (2007) showed that TakP, a
317 pyruvate binding protein from *Rhodobacter sphaeroides* crystalized as a dimer, and the
318 functional importance of dimerization was validated by tryptophan fluorescence quenching,
319 gel filtration and cross-linking experiments. The dimerization is believed to occur through a
320 kinked C-terminal helix, which swaps its position with the same portion of the dimer
321 counterpart. Additionally, Akiyama *et al.* (2009) crystalized a lactate binding protein from
322 *Thermus thermophilus* which interacts back-to-back in a dimerization process stabilised by
323 hydrophobic interactions in the C-terminal region of the protein. Finally, Cuneo *et al.* (2008)
324 confirmed also the dimerization state of a TRAP protein from *Thermotoga maritima* through
325 gel filtration analysis and X-ray scattering. It remains unclear, however, how this
326 dimerization process would promote or interfere with the transport mechanism when
327 interacting with the DctQM subunits.

328 **3. The TTT family**

329 As with the TRAP transporters, systems in the TTT family are composed of a conserved 12
330 TM protein, (TctA homologs) believed to act as a symport protein energized by an
331 electrochemical ion-gradient (although this has not been experimentally determined) and
332 poorly-conserved 4 TM protein (TctB homologs) with unknown function, in combination
333 with an SBP (TctC homologs) which binds the substrate with high affinity (Fig. 1) (Sweet *et*
334 *al.* 1979, Winnen *et al.* 2003). However, the TTT family has not been subject to many
335 experimental studies and knowledge about this family is still scarce; the topic being last
336 reviewed by Winnen *et al.* (2003).

337

338 **3.1. Substrate diversity of the TTT family: Role of the solute binding-protein and** 339 **occurrence in pathogens**

340 As the prototype for the TTT family, the Tct citrate transporter was first described by Sweet
341 *et al.* (1979) in the pathogen *Salmonella typhimurium*, one of the most important causative
342 agents of food-borne gastrointestinal infections and a growing problem due to the recent
343 emergence of multidrug resistant strains (Hur *et al.* 2012). TctC was found to be involved in
344 the uptake of the tricarboxylic acid citrate with low- μ M affinity, with citrate uptake severely
345 reduced in this organism upon disruption of the *tctC* gene. This function gave the name to the
346 family as tricarboxylate transporters (Sweet *et al.* 1979, Somers *et al.* 1981). A series of 36
347 tricarboxylate and di-carboxylate metabolites were later shown by Sweet *et al.* (1984) to
348 inhibit citrate binding to TctC to varying extents, suggesting that the substrate range for this
349 protein might not be restricted to citrate. Genetic mapping studies initiated by Somers and
350 Kay (1983) and finished by Widenhorn *et al.* (1988) showed that downstream of the *tctC*
351 locus there were two more encoded proteins, of 19 kDa and 45 kDa, corresponding,
352 respectively, to the transmembrane proteins TctB and TctA (Fig. 1 and Fig. 2b). The gene
353 arrangement of *tctCBA*, is similar to that found for the majority of TRAP transporters
354 (Mulligan *et al.* 2011). Encoded in the opposite direction, a fourth gene, *tctD*, was shown by

355 Widenhorn *et al.* (1989) to encode a transcription regulator of the *tctCBA* operon, which was
356 found to be repressed when *tctD* was deleted or in the presence of glucose in the medium.
357 Homologous systems to TctCBA are found in many bacteria, mainly Proteobacteria, and
358 citrate uptake is the commonest identified role for the few other TTT systems experimentally
359 characterised to date (Antoine *et al.* 2003, Brocker *et al.* 2009, Hosaka *et al.* 2013, Graf *et al.*
360 2016). Citrate has been shown to act in some cases as an iron chelator for different transport
361 systems (Yancey and Finkelstein 1981, Braun 2001, Luck *et al.* 2001, Banerjee *et al.* 2016),
362 and although the potential role of TctC acting as an iron transport protein has not been
363 investigated to date, experiments performed with *Salmonella typhimurium* (Sweet *et al.* 1979)
364 showed that citrate binding to TctC is improved in the presence of Na⁺ Ca²⁺, Mn²⁺ and Fe²⁺,
365 while partially inhibited by Mg²⁺, Ni²⁺, Zn²⁺ and Co²⁺. In addition, growth experiments
366 performed by Brocker *et al.* (2009) using a homologous TctCBA system from
367 *Corynebacterium glutamicum* showed that this system was able to uptake citrate in the
368 presence of Ca²⁺ and Mg²⁺, but not Sr²⁺.

369 After the characterisation of TctC, all proteins homologous to the TctCBA systems in newly
370 released genomes were annotated either as unknown proteins or citrate uptake systems, and
371 this family was neglected for over a decade, until one TctC homologue was found by Antoine
372 *et al.* (2000) to be encoded upstream of the pertussis toxin (PTX) virulence island, one of the
373 most important toxins produced by the causative agent of whooping cough, *Bordetella*
374 *pertussis*. This gene was found to be conserved in this locus for different *Bordetella* species
375 and was named *bugT*, standing for ‘*Bordetella* uptake gene’. Although a relationship between
376 the BugT protein and the production of PTX was not confirmed, regions coding for other
377 virulence factors, such as the adenylate cyclase toxin (AC) and the dermonecrotic toxin
378 (DNT) also contained *bugT* homologs, while two *bug* homologs were negatively regulated by
379 the BvgAS two-component system, responsible for the activation of virulence factor
380 production (Antoine *et al.* 2000, Antoine *et al.* 2003). Recently, a single-nucleotide-
381 polymorphism (SNP) in one *bug* gene was consistently identified in an Australian epidemic
382 strain of *B. pertussis* (Safarchi *et al.* 2016). As discussed in the next sections, further searches
383 in the *B. pertussis* genome found homologs of Bug proteins to be extensively
384 overrepresented, with 76 genes encoding distinct homologs (Antoine *et al.* 2000). In contrast,
385 only two sets of genes coding for transmembrane proteins homologous to *tctAB* were found
386 in *B. pertussis*, and most of the BugT homologs showed no obvious membrane counterparts
387 encoded in their genomic vicinity, hence the designation of them as “orphan” proteins.

388 The only complete operonic encoded TTT system in *Bordetella pertussis*, encoded by
389 *bctCBA*, contains *bug4* as the *tctC* homolog, and was found to be the equivalent of *tctCBA*
390 from *S. typhimurium*, as expression was upregulated by citrate and gene disruption resulted in
391 lower citrate uptake rates (Antoine *et al.* 2003). As shown by Antoine *et al.* (2005), upstream
392 of the *bctCBA* operon is encoded the two-component system *bctDE*, transcribed in the same
393 direction but forming a separate operon (Fig 2B), which showed a basal expression level
394 independent of citrate. When this two-component system was deleted, expression of *bctCBA*
395 was not detected, showing that *bctDE* was in fact activating transporter gene expression.
396 Disruption of the *bctBA* components, on the other hand, increased operon expression, due to
397 an accumulation of citrate in the periplasm to be directed to signalling purposes (Antoine *et*
398 *al.* 2005). Finally, when *bctC* was deleted, *bctBA* expression was reduced to basal levels even
399 in high citrate concentrations, inferring that the two-component system is enough to maintain
400 a basal level expression, but is not enough to enhance expression in the presence of citrate.
401 Together, those data showed that citrate-bound BctC was required for both transport and
402 signalling, interacting either with BctE or BctA; a model confirmed in the same study by
403 bacterial two-hybrid assays, showing unprecedented evidence that TctC homologs can be

404 involved also in regulatory processes (Antoine *et al.* 2005). The presence of citrate
405 responsive regulatory genes and two-component systems adjacent to *tctCBA* operons is not
406 uncommon, as shown for *S. typhimurium* by Widenhorn *et al.* (1989), for *Comamonas sp.* by
407 Hosaka *et al.* (2013), for *Advenella mimigardefordensis* by Wubbeler *et al.* (2014); and in the
408 genomic searches provided by Antoine *et al.* (2003). Brocker *et al.* (2009) also characterised
409 a citrate-responsive two-component system controlling *tctCBA* expression in *Corynebacterium*
410 *glutamicum*, although in this case the regulatory proteins were adjacent to another transport
411 system. Interestingly, some of the TTT systems are found in the genome with the *tctB* subunit
412 downstream of *tctA*, such as the *slcHFG* systems from *Roseovarius nubinhibens* (Denger *et al.*
413 *et al.* 2009) and *Chromohalobacter salexigens* (Fig 2B) (Denger and Cook 2010). This feature
414 is also observed in some TRAP systems, and systems with this genomic organization are
415 thought to be more similar among them (Mulligan *et al.* 2011).

416 The study of two abundantly expressed Bug proteins led to the first crystal structures for TTT
417 family SBPs. These proteins, discussed in detail in the next sections, were fortuitously
418 crystalized with substrates in their binding pocket. BugD contained an aspartate molecule
419 (Huvent *et al.* 2006) and BugE contained a glutamate molecule (Huvent *et al.* 2006). Amino-
420 acids are the most important carbon and nitrogen sources for *B. pertussis*, which is incapable
421 of metabolism of substrates through the glycolytic pathway (Huvent *et al.* 2006); As the Bug
422 proteins are highly expressed, they might play a crucial role in uptake of core metabolic
423 pathways. Herrou *et al.* (2007) characterised one of the Bug proteins (Bug27), found to be
424 overexpressed in the presence of nicotinic acid, an essential vitamin and a negative modulator
425 of *B. pertussis* virulence. It was shown that this protein binds, with an affinity lower than 1
426 μM , not only to nicotinate, but also nicotinamide, citrate, benzoate and quinaldic acid. This
427 protein generated also the first TTT SBP crystal structure in an unliganded conformation
428 (Herrou *et al.* 2007). The binding of Bug27 to nicotinic acid/nicotinamide might suggest it
429 plays a role in virulence modulation, either by interacting with a membrane signal protein or
430 simply transporting nicotinic acid to the cytoplasm. Interestingly, Brickman *et al.* (2017)
431 suggested that another Bug protein, Bug69, might also be related to the uptake of nicotinic
432 acid and related compounds.

433 Although not the focus of this review, the potential of the TTT family as a new source for
434 biotechnology relevant uptake systems was also exposed by the genomic analysis performed
435 by Antoine *et al.* (2003), where it was observed that in many organisms, the *bug* homologs
436 were located near operons that conferred specific abilities to each strain, such as catechol
437 degradation, showing that the importance of this family is wider than the suggested so far and
438 that its diversity might correlate with the metabolic versatility and adaptability of an
439 organism. In addition, a genomic search regarding arylmalonate decarboxylases (AMDases)
440 by Maimanacos *et al.* (2016) found several members of the TTT family in the vicinity of
441 these enzymes for five of the eight predicted AMDase clusters, either as orphan proteins, in
442 the case of β -proteobacteria; or complete systems, in the case of α -proteobacteria, suggesting
443 the TTT proteins might act to import the carboxylated substrates for subsequent catalysis by
444 the AMDases. Other biotechnologically relevant discoveries include sulfolactate metabolic
445 pathways in *Roseovarius nubinhibens* (Denger *et al.* 2009) and *Chromohalobacter*
446 *salexigens*, which contain a TTT uptake system for this substrate (Denger and Cook 2010)
447 named *slcHFG*; a TTT system from *Comamonas sp.*, TpiBA and TphC, able to uptake
448 terephthalate (Hosaka *et al.* 2013); the TctCBA from *Advenella mimigardefordensis* able to
449 uptake the synthetic molecule disulfide 3,3'-dithiodipropionic acid (DTDP), a precursor for
450 synthetic polythioesters (Fig 2B) (Wubbeler *et al.* 2014); a *tctA* homolog genetically
451 proximal to genes coding to esterase enzymes that degrade organophosphates and potentially
452 related to aromatic compound degradation (Batista-Garcia *et al.* 2014); the TctABC system

453 from *Halomonas* involved in galactarate/glucarate metabolism (Leyn *et al.* 2017); and a
454 recent discovery from our group of AdpC, an “orphan” SBP from *Rhodopseudomonas*
455 *palustris* which binds medium chain-length dicarboxylic acids ranging from adipate (C6) to
456 azelate (C9) (Rosa *et al.* 2017). Searching the Enzyme Function initiative (EFI) database, it
457 was observed that 19 homologs of TctC were in their library, however only one of them, a
458 TctC homolog from *Polaromonas sp.* was crystalized in the open apo conformation (PDB
459 accession code 4X9T). Table 1 summarises the known range of characterised TTT systems
460 with their respective ligands. Initially believed to bind exclusively to citrate, the substrate
461 range for the TTT family is clearly much broader and new substrates are continually being
462 found. With the exception of nicotinic acid in Bug27 (Herrou *et al.* 2007), all substrates
463 characterised so far seem to have two carboxylic groups, or other functional groups such as
464 sulphate and amide, and further studies will show if this is indeed a required property for
465 substrates in the TTT family.

466 **3.1.Properties and function of the TctAB subunits**

467 The TTT systems are predicted to contain two membrane proteins, homologous to TctA and
468 TctB. Although crystal structures of these proteins have not been elucidated, information
469 from the primary and secondary sequences of these subunits were studied, in addition to
470 some physiological characterisations, in an attempt to understand the energetic and structural
471 mechanisms of the TTT family.

472 Winnen *et al.* (2003) showed that while TctB and TctC showed only 27% and 31% identity
473 on average between family members, respectively, TctA orthologues suggested 42% identity
474 and 53% similarity in similar comparisons. Topology predictions suggested the number of
475 transmembrane helices in TctA homologs might vary in different systems, ranging from 9 to
476 12 in bacteria, and 7 to 11 in archaea. In either group, the N-terminal side was predicted to be
477 in the cytoplasm, and large hydrophilic loops between helices 2 and 3 are conserved among
478 all organisms analysed, suggesting this region must have an important role in protein
479 function. The motif G-Hy₃-*G-Hy₃-*G-Hy₂-*P-G-Hy-G, where Hy is an aliphatic
480 hydrophobic residue and * means a fully conserved residue, are found to be highly conserved
481 both in TM1 and TM7, suggesting the 12TM protein originated from a duplication in a 6TM
482 ancestor. TctB homologs were predicted to have between 4 and 5 transmembrane domains,
483 also with predicted cytoplasmic N-termini, and are very poorly conserved among bacteria and
484 were not observed in archaeal sequences (Winnen *et al.* 2003). In a systems biology study
485 comparing a wild type and a multi-drug resistant strain of *Salmonella enterica*, Ricci *et al.*
486 (2012) observed a G109S SNP in TctA. Although shown to be unrelated to the antimicrobial
487 resistance, this mutation compromised growth on several carbon sources, and curiously
488 seemed to confer a delay in the production of Reactive Oxygen Species (ROS) under stressful
489 conditions. Our genome searches revealed for the first time, that at least one case, in
490 *Paraburkholderia caribensis*, a fusion between the two membrane subunits can also be
491 observed, forming a single TctAB protein (Fig 2b), similar to what was described for some
492 DctQM proteins from the TRAP family, in particular the ones constituting TAXI-TRAP
493 systems (Mulligan *et al.* 2011).

494 Evidence of cation dependence for transport in the TTT family was first provided by Sweet *et al.*
495 (1979), showing that binding of citrate to TctC was enhanced in the presence of Na⁺ Ca²⁺,
496 Mn²⁺ and Fe²⁺, while Mg²⁺, Ni²⁺, Zn²⁺ and Co²⁺ inhibited uptake. Similarly, Brocker *et al.*
497 (2009) demonstrated that citrate transport by a TctCBA system of *Corynebacterium*
498 *glutamicum* was enhanced by Ca²⁺ and Mg²⁺, but not Sr²⁺. Later studies by Hosaka *et al.*
499 (2013), however, showed that both the addition of protonophores and an alkaline pH
500 disrupted terephthalate uptake by the TpiBA system in *Comamonas sp.*, while deletion

501 experiments in the same study showed that both subunits were essential for substrate uptake.
502 Furthermore, terephthalate uptake was not disrupted when *Comamonas sp* was grown in
503 absence of Na⁺, suggesting that at least this particular process was more dependent on the
504 proton-motive force rather than a sodium gradient, distinct from SiaPQM sialic acid TRAP
505 transporters (Mulligan *et al.* 2009).

506 Discovering a *tctA* homolog next to esterase genes in metagenomics searches, named *tctA_ar*,
507 Batista-Garcia *et al.* (2014) attempted to build a structural model for TctA_ar based on its
508 primary sequence, given that this protein showed no considerable homology to any other
509 secondary transporter available in the database. The TctA homolog of *Comamonas sp.*,
510 characterised in previous studies (Hosaka *et al.* 2013), was used as a control and called
511 TctA_ct. Two templates were used against each sequence (PDB codes 3VVN and 4K1C),
512 resulting in 4 models in total. In addition to the already described duplicated motif, a high
513 degree of identity was observed between residues 60 and 110 in TctA homologs. The
514 computational models agreed with the predicted 12TM domain protein, and it was proposed
515 in the 3VVN-based models that the 20 residue conserved motif were in the vicinity of the
516 predicted binding pocket of the protein. In addition, G106, mutated in previous studies (Ricci
517 *et al.* 2012), would be located between the two repeats, being involved in the beginning of the
518 translocation pathway. In the 4K1C model, the two copies of the domain would be in contact
519 with each other in a helix, which would facilitate conformational changes during the transport
520 cycle. Moreover, given that the TTT transporters were initially known for the transport of
521 citrate, docking of this molecule into the two models was attempted, in addition to the
522 modelling of a Na⁺ binding pocket, suggested to be necessary for citrate transport (Sweet *et*
523 *al.* 1979). It was shown in the 3VVN models that the potential Na⁺ binding pockets were
524 located in C-terminal variable regions, while the citrate-binding pocket differs in each of the
525 two proteins. The TctA_ar showed a unique pocket for citrate, while TctA-ct showed several
526 along the predicted channel, which might act as different steps in the translocation pathway.
527 In the 4K1C models, several binding pockets were predicted for citrate in both proteins, with
528 at least one positively charged residue to interact with the ligand in each of them. These
529 models are an important step towards understanding of translocation mechanisms in the TTT
530 family, although biological confirmation of the transport mechanisms in this family are
531 clearly essential.

532 Evolutionary studies on the TTT family performed by Winnen *et al.* (2003) suggested that the
533 TctA protein is the original core transport protein, and that the TctB homolog worked as an
534 accessory protein. This claim is supported by the presence of TctA homologs in Archaea,
535 added to the absence of TctB or TctC homologs, and reinforced by the model studies
536 performed by Batista-Garcia *et al.* (2014). However, the findings on the TRAP transporters,
537 where the DctQ subunit is also observed in archaea, and was found essential to function
538 challenge this hypothesis (Mulligan *et al.* 2011, Mulligan *et al.* 2012). Because the archaea
539 harbouring TctA homologs were found in extreme environments, amongst groups of
540 methanogens, hyperthermophiles and halophiles, it was suggested that the TctA homologs
541 would be involved in a range of different specific metabolic niches (Winnen *et al.* 2003), but
542 experimental evidence is still lacking to show if this is the case.

543

544 **3.2. Crystal structure and substrate coordination in TctC homologs**

545 The binding proteins from the TTT family, homologous to TctC (Sweet *et al.* 1979), show a
546 conserved size (ranging from 29kDa to 33kDa), topology and secondary structure
547 organization, but differ considerably in the primary sequence, where among them an identity
548 around 30% is observed. Consequently, a big difference in overall pI is also seen, ranging

549 from 5 to 9.6 (Antoine *et al.* 2000). At the time of writing, only six structures of TTT SBPs
550 have been deposited in the Protein Data Bank (PDB), four of them with a substrate in the
551 binding pocket; it is already possible to identify, however, some common features among
552 them. The average of 300 amino-acid residues comprises the mature form of the proteins
553 (without signal peptide), separated into two globular domains. Domain one is usually formed
554 by residues 1~100 and 230~300 from the N and C termini, forming a β -sheet of five strands,
555 with topological arrangement β 2- β 1- β 3- β 9- β 4, surrounded by \sim 6 α -helices. Domain 2 is
556 comprised of residues 100~229, forming also a central β -sheet of 5 strands with topological
557 arrangement β 6- β 5- β 7- β 4- β 8, surrounded by \sim 4 helices. Domain 2 sometimes contains a
558 disulphide bridge between cysteine residues located in α 5 and β 7, but this is not a feature
559 common to all proteins (Huvent *et al.* 2006, Rosa *et al.* 2017). The junction of the two
560 domains is formed by two β -strands, S4 and S9, which are part of domain 1 but extend up to
561 domain 2, and hydrogen bonds between the two domains are scarce. All these features show
562 that TTT SBP characterized so far can be classified into the Type II binding protein group,
563 according to the scheme of Fukami-Kobayashi *et al.* (1999), or cluster E-II, accordingly to
564 the new division proposed by Scheepers *et al.* (2016). Fig. 3B shows the crystal structure of
565 Bug D as a representative for the TctC homologs. Upon binding to the substrate, it was
566 estimated that the two domains close in an angle of 24.7° , based on the structure of the
567 unliganded nicotinic acid binding protein Bug27 (Herrou *et al.* 2007). Although TctC
568 proteins most commonly seem to bind to molecules containing carboxylic groups, curiously
569 there is usually a slight overall negative charge in the binding pocket, likely dissipated by the
570 water molecules or dipole effects of the surrounding helices (Rosa *et al.* 2017). Two β -turns,
571 between β 1 and α 1; and β 7 and α 7, form a “pincer-like” structure important in substrate
572 coordination, closing around one carboxylic group of the ligand, while the remainder of it is
573 buried in the pocket (Fig 3). The residues present in the loops characterise distinguishing
574 signatures for proteins of this family, with the motif [P*-F-X-A-G*-G*-X-X-D*] in domain 1
575 being almost ubiquitous among the protein sequences, where X means any residue and *
576 means a very conserved residue. The backbone atoms of residues in this region seem to make
577 hydrogen bonds with two water molecules, present in all substrate-containing structures and
578 also very well conserved in position. These water molecules bridge hydrogen bonds between
579 the protein main chain and the proximal carboxylic group in the substrate. This pattern is
580 observed in the coordination of adipate by AdpC (Rosa *et al.* 2017), aspartate in BugD
581 (Huvent *et al.* 2006) and glutamate in BugE (Huvent *et al.* 2006). In some Bug protein
582 sequences, although these residues are not conserved, they are substituted by others where the
583 side-chain would contain a hydroxyl group, potentially maintaining hydrogen bonds in
584 similar position to what would be expected of the water molecules (Huvent *et al.* 2006). As
585 shown by Herrou *et al.* (2007), the two β -loops which form the “pincer-like” structure and the
586 two water molecules are not well defined when the ligand is not present. The coordination of
587 the ligand’s distal carboxylate group, buried in the pocket, is much less conserved, with α 3
588 and α 5 helices apparently varying in position to accommodate each substrate (Huvent *et al.*
589 2006, Rosa *et al.* 2017). Although not conserved in sequence or topology, the involvement of
590 water molecules in the coordination of the distal carboxylate groups was observed in all
591 cases. In some proteins, hydroxyl groups from threonine and serine residues also form
592 hydrogen bonds with the carboxylate in the substrate, but their positions vary. The carbon
593 chain of the substrate, on the other hand, is stabilized by much more conserved hydrophobic
594 interactions, such as Phe14 in AdpC, which seem to act as a docking site for the substrate,
595 and two glycines (Gly18 and Gly163 in AdpC). As a dynamic model, Herrou *et al.* (2007)
596 suggested that the unliganded form of the SBP would be in an open conformation, with the
597 “pincer-like” structures flexible. Substrate would then bind to domain 1, which would cause a
598 conformational change that would bring the water molecules and domain 2 together. A

599 comparison between the five available TctC homologs available in the PDB (only one AdpC
600 structure was used) is shown in Table 2. The root mean square deviations (RMSD), as
601 expected, show a bigger difference when comparing closed and Apo structures, and smaller
602 RMSDs when comparisons were made between two liganded structures. Taken together, the
603 crystal structures presented to date give a good general mechanism for ligand coordination in
604 the TctC homologs, and further studies will enable us to validate this model and detail the
605 potential differences for substrates containing different functional groups, such as
606 nicotinamide (Herrou *et al.* 2007) and sulfolactate (Denger and Cook 2010).

607

608 **3.3. Present in abundance: The overrepresentation of TctC homologs in some bacteria**

609 The limited number of sequences available at the time resulted in a bias in the studies
610 performed by Winnen *et al.* (2003), which suggested that the TTT systems were mostly
611 present in α -proteobacteria, and that most other bacterial groups had few or no homologs of
612 these proteins. Genomic searches following the discovery of BugT by Antoine *et al.* (2000),
613 however, revealed the *bug* genes to be very overrepresented in *Bordetella pertussis*, with 79
614 BugT homologs, making this family the most abundant in the genome. Following this
615 discovery, Antoine *et al.* (2003) performed a wider genome analysis, showing that this
616 overrepresentation was extended to several *Bordetella* species, and that some of the Bug
617 proteins were also among the most abundant in cell protein extracts in *B. pertussis*. As stated
618 in previous sections, the numbers of TTT transmembrane components did not follow the
619 same process, being found in small numbers and consequently configuring most BugT
620 homologs as “orphan proteins”, with no obvious transmembrane counterparts. The existence
621 of orphan *bug* homologs was also observed by Antoine *et al.* (2003) in the genomes of
622 several other bacteria genera, although in that study the only *bacterium* shown to have as
623 many representatives as the *Bordetella* species was the β -proteobacterial relative *Cupriavidus*
624 *metallidurans*. At the time that search was performed, there were around 200 complete
625 bacterial genomes available in the databases, and more recent genome releases showed that at
626 least two other β -proteobacteria genera, *Advenella* and *Cupriavidus* also showed an
627 overrepresentation of *tctC* homologs (Wubbeler *et al.* 2014). For this review, we reassessed
628 the distribution of TTT systems using the 8049 fully assembled genomes in Genbank, of
629 2323 different species, to provide an updated analysis of the presence of TTT systems in
630 bacterial genomes.

631 A total of 2,323 complete bacterial genomes retrieved from the NCBI database were screened
632 for TctA and TctC homologs using the TBLASTN tool. Searches were performed against the
633 coding sequence (CDS) database of each species using lists of protein sequences of either
634 TctA or TctC homologs as queries rather than single sequences, in order to avoid query bias.
635 For the TctC search, the queries were: TctC from *S. enterica* (Sweet *et al.* 1979); BugD,
636 BugE and Bug27, from *B. pertussis* (Herrou *et al.* 2007), TphC from *Comamonas sp.* (Hosaka
637 *et al.* 2013) and AdpC from *R. palustris* (Rosa *et al.* 2017). For the TctA search, we used the
638 TctA from *S. enterica* (Sweet *et al.* 1979), *A. mimigardefordensis* (Wubbeler *et al.* 2014), and
639 *C. glutamicum* (Brocker *et al.* 2009); the BctA from *B. pertussis* (Antoine *et al.* 2005); the
640 TpiA from *Comamonas sp.* (Hosaka *et al.* 2013) and the SlcF from *R. nubinhibens* (Denger *et al.*
641 *et al.* 2009). Due to the poor sequence conservation among the TctB proteins, our searches with
642 this subunit proved to be unsuccessful. TBLASTN reports were obtained for a range of E-
643 values from 1 to 10^{-15} in order to determine the best threshold to avoid spurious hits, while
644 still retaining distant paralogs. The complete table with number of hits for both proteins with
645 an e-value of 10^{-15} is presented in supplementary Table S1.

646

647 Our searches revealed that, in accordance with the findings of Antoine *et al.* (2003) and
648 further reinforced by Wubbeler *et al.* (2014), the most extreme examples of
649 overrepresentation of this group of proteins are found among β -proteobacteria, especially
650 among *Bordetella* species, as shown in Table 3. However, this phenomenon is not restricted
651 to this group, but extends also to species in the α -proteobacteria phyla (Table 3). In fact, the
652 genome of the environmental α -proteobacterium *Rhodoplanes sp.* encodes 434 TctC
653 homologs in its 8.2 Mbp genome, more than double that of some *Bordetella* species. A more
654 detailed investigation of the expansion in *Rhodoplanes* will be reported elsewhere
655 (manuscript in preparation). Although this analysis shows that the overrepresentation of TctC
656 homologs is mostly found in proteobacteria, a deeper phylogenetic analysis is still required in
657 order to clarify whether this feature found in different subgroups originates from duplications
658 in a common ancestor or were independent events resulting from convergent evolution and
659 independent multiplication events. A search for TctA homologs, on the other hand, as shown
660 in Table 4, suggests that the genomes containing the largest numbers of homologs are found
661 among α and γ -proteobacteria, with only one β -proteobacteria showing 8 or more homologs.
662 In this search, the top hits are no higher than 21 per genome, and are usually associated with a
663 similar number of TctC homologs, possibly forming complete tripartite systems. In this case,
664 organisms outside of the class of proteobacteria, such as clostridia, spirochaetes and bacilli
665 are also observed to harbour these homologs. An overview of the number of genomes
666 encoding different numbers of TctC and TctA homologs are shown in Fig. 4, and the full
667 complement of genomes analysed is shown in Supplementary Table S1.

668 At an e-value of 10^{-15} in BLAST searches, the TctC homologs outnumber the TctA homologs
669 in 176 genomes, as shown in Table 5. As already discussed, in these cases it might be that
670 one TctA interacts with more than one TctC, or that the latter are involved in processes other
671 than transport, such as signalling and chemotaxis (Antoine *et al.* 2003, Piepenbreier *et al.*
672 2017). These latter suggestions are reinforced by the fact that in 36 genomes, one *tctC*
673 homolog was found, but no *tctA* homologs, although the hypothesis that the binding proteins
674 might interact with transmembrane domains of other transporter classes cannot be excluded.
675 In our initial searches, we found that 210 genomes showed an excess of *tctA* homologs in
676 relation to *tctC*, an unprecedented observation to the best of our knowledge. In order to see
677 whether these observations were due to too strict threshold, our searches were repeated with
678 different e-values, shown in table 5. As shown, using an e-value of 10^{-9} , the number of
679 genomes where this situation occurs is reduced to 66, and to 39 in 10^{-6} . At the latter
680 threshold, 6 genomes indicated the presence of TctA homologs, but no TctC homologs.
681 Investigating these 6 genomes individually, we found that 4 of them contained a truncated
682 *tctC* homolog in the vicinity of the *tctA* gene and in one the *tctA* gene was clearly mutated.
683 The single remaining genome, *Mageeibacillus indolicus*, indeed seems to have no indication
684 of any periplasmic binding proteins in the vicinity of the *tctA* gene. If the existence of a *tctA*
685 gene without any *tctC* is not a search artefact, one possible explanation would be that SBPs of
686 other types of transport systems could be capable of interacting with the TTT transmembrane
687 subunits. An alternative is that such rare orphan TctA proteins in bacteria are, like in archaea,
688 capable of functioning without the involvement of an SBP (Winnen *et al.* 2003), although
689 both the TTT systems characterised so far and the experiments with the TRAP transporters
690 suggest otherwise (Brocker *et al.* 2009, Mulligan *et al.* 2009, Hosaka *et al.* 2013). In another
691 316 genomes, the numbers of *tctC* and *tctA* genes match exactly, suggesting all proteins
692 would be involved in transport through tripartite systems. Finally, about half of the genomes
693 searched (1621) contained no homologous proteins to any of the queries used, suggesting,
694 given TctA homologs are also found in archaea (Winnen *et al.* 2003), that TTT systems were
695 lost during evolution in these phylogenetic branches.

697 The reason for the overrepresentation of the Bug proteins in *Bordetella* species and other
698 Proteobacteria remains unclear. Antoine *et al.* (2003) suggested that the few transmembrane
699 domains of TTT systems evolved to be poorly specific, being able to interact with several
700 TctC homologs and thus be required for the uptake of different substrates. This hypothesis
701 was also suggested by Hosaka *et al.* (2013), but no evidence for this mechanism is yet
702 available. One could hypothesize that perhaps many of these proteins have similar binding
703 functions, but are expressed differentially during the infection cycle in *Bordetella* species in
704 order to evade the immune system more efficiently. However, the fact that many
705 environmental Proteobacteria also have this expansion of Bug proteins suggests instead that it
706 is an earlier evolutionary trait. The genome of *Ralstonia eutropha* containing 154 homologs
707 of *tctC*, reveals that the majority of them (64.1%) have in their vicinity a regulatory protein,
708 suggesting that most of these proteins are associated with regulatory mechanisms rather than
709 transport (Pohlmann *et al.* 2006). In this sense, the nomenclature of these SBPs as “uptake
710 genes” might not reflect their actual role in the cell. Piepenbreier *et al.* (2017) provides a
711 good review of how transporters from different classes can act as first agents in signalling
712 pathways, and further studies will enrich our understanding to whether this is the case for the
713 TTT family.

714

715 **4. Concluding Remarks**

716 In this review, we showed how transport systems from the TRAP and TTT families can play
717 important roles in bacteria with a focus on pathogenicity and colonisation. Recent high-
718 throughput studies increased substantially the range of substrates known for the TRAP
719 family, while the TTT family is still understudied with a more limited known substrate range,
720 being unravelled in individual studies. In addition, while a lot has been elucidated regarding
721 the binding mechanisms, energetics and kinetics of the TRAP family, very few equivalent
722 studies exist for the TTT family, where especially the energy-coupling mechanisms are yet to
723 be elucidated properly. For both families, a crystal structure of the complete tripartite systems
724 would greatly increase our understanding of the transport process across the membrane,
725 perhaps with potential applications as new drug targets in pathogenic bacteria, given the
726 absence of these transporters in eukaryotic cells.

727

728 **Non-standard abbreviations**

729 SBP – Solute Binding-Proteins; TTT – Tripartite Tricarboxylate Transporters; TRAP –
730 Tripartite ATP-independent periplasmic transporter

731 **Conflicts of interest**

732 The research was conducted in the absence of any commercial or financial relationships that
733 could be construed as a potential conflict of interest.

734 **Author contributions**

735 LR reviewed the literature, co-wrote the manuscript and co-analysed the bioinformatic data.
736 MB generated the bioinformatics data and co-analysed it. GT edited and commented on the
737 drafts. DK Conceived the idea and focus of the review, co-wrote and edited the paper, and
738 provided supervisory guidance.

739 **Funding**

740 The present work was accomplished with funding from Brazilian funding agency CNPQ
741 (National Council for Scientific and Technological Development), through PhD studentships
742 in the remit of “Science Without Borders” program to LR (248597/2013-2) and MB
743 (201873/2014-1). May this manuscript reinforce the importance of prioritizing investment in
744 science and research by the Brazilian government.

745 **Acknowledgements**

746 We thank Dr. Pascal-Antoine Christin and Daniel Wood for providing suggestions on the
747 phylogenetic analyses.

748

In review

749 **References**

750

- 751 Akiyama, N., K. Takeda and K. Miki (2009). "Crystal structure of a periplasmic substrate-
752 binding protein in complex with calcium lactate." J Mol Biol **392**(3): 559-565.
- 753 Allen, S., A. Zaleski, J. W. Johnston, B. W. Gibson and M. A. Apicella (2005). "Novel sialic
754 acid transporter of *Haemophilus influenzae*." Infect Immun **73**(9): 5291-5300.
- 755 Allen, S., A. Zaleski, J. W. Johnston, B. W. Gibson and M. A. Apicella (2005). "Novel sialic
756 acid transporter of *Haemophilus influenzae*." Infect Immun **73**.
- 757 Antoine, R., I. Huvent, K. Chemlal, I. Deray, D. Raze, C. Locht, *et al.* (2005). "The
758 Periplasmic Binding Protein of a Tripartite Tricarboxylate Transporter is Involved in Signal
759 Transduction." Journal of Molecular Biology **351**(4): 799-809.
- 760 Antoine, R., I. Huvent, K. Chemlal, I. Deray, D. Raze, C. Locht, *et al.* (2005). "The
761 periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal
762 transduction." J Mol Biol **351**(4): 799-809.
- 763 Antoine, R., F. Jacob-Dubuisson, H. Drobecq, E. Willery, S. Lesjean and C. Locht (2003).
764 "Overrepresentation of a gene family encoding extracytoplasmic solute receptors in
765 *Bordetella*." J Bacteriol **185**(4): 1470-1474.
- 766 Antoine, R., D. Raze and C. Locht (2000). "Genomics of *Bordetella pertussis* toxins." Int J
767 Med Microbiol **290**(4-5): 301-305.
- 768 Bakermans, C., R. E. Sloup, D. G. Zarka, J. M. Tiedje and M. F. Thomashow (2009).
769 "Development and use of genetic system to identify genes required for efficient low-
770 temperature growth of *Psychrobacter arcticus* 273-4." Extremophiles **13**(1): 21-30.
- 771 Banerjee, S., S. Paul, L. T. Nguyen, B. C. H. Chu and H. J. Vogel (2016). "FecB, a
772 periplasmic ferric-citrate transporter from *E. coli*, can bind different forms of ferric-citrate as
773 well as a wide variety of metal-free and metal-loaded tricarboxylic acids." Metallomics **8**(1):
774 125-133.
- 775 Batista-Garcia, R. A., A. Sanchez-Reyes, C. Millan-Pacheco, V. M. Gonzalez-Zuniga, S.
776 Juarez, J. L. Folch-Mallol, *et al.* (2014). "A novel TctA citrate transporter from an activated
777 sludge metagenome: structural and mechanistic predictions for the TTT family." Proteins
778 **82**(9): 1756-1764.
- 779 Berntsson, R. P., S. H. Smits, L. Schmitt, D. J. Slotboom and B. Poolman (2010). "A
780 structural classification of substrate-binding proteins." FEBS Lett **584**(12): 2606-2617.
- 781 Bouchet, V., D. W. Hood, J. Li, J. R. Brisson, G. A. Randle, A. Martin, *et al.* (2003). "Host-
782 derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a
783 major virulence factor in experimental otitis media." Proc Natl Acad Sci U S A **100**(15):
784 8898-8903.
- 785 Braun, V. (2001). "Iron uptake mechanisms and their regulation in pathogenic bacteria."
786 International Journal of Medical Microbiology **291**(2): 67-79.
- 787 Brautigam, C. A., R. K. Deka, P. Schuck, D. R. Tomchick and M. V. Norgard (2012).
788 "Structural and thermodynamic characterization of the interaction between two periplasmic
789 *Treponema pallidum* lipoproteins that are components of a TPR-protein-associated TRAP
790 transporter (TPAT)." Journal of Molecular Biology **420**(1-2): 70-86.
- 791 Brickman, T. J., R. J. Suhadolc, P. J. McKelvey and S. K. Armstrong (2017). "Essential role
792 of *Bordetella* NadC in a quinolinate salvage pathway for NAD biosynthesis." Mol Microbiol
793 **103**(3): 423-438.
- 794 Bocker, M., S. Schaffer, C. Mack and M. Bott (2009). "Citrate Utilization by
795 *Corynebacterium glutamicum* Is Controlled by the CitAB Two-Component System through

796 Positive Regulation of the Citrate Transport Genes *citH* and *tctCBA*." J Bacteriol **191**(12):
797 3869-3880.

798 Chae, J. C. and G. J. Zylstra (2006). "4-Chlorobenzoate uptake in *Comamonas sp.* strain DJ-
799 12 is mediated by a tripartite ATP-independent periplasmic transporter." J Bacteriol **188**(24):
800 8407-8412.

801 Cuneo, M. J., A. Changela, A. E. Miklos, L. S. Beese, J. K. Krueger and H. W. Hellinga
802 (2008). "Structural Analysis of a Periplasmic Binding Protein in the Tripartite ATP-
803 independent Transporter Family Reveals a Tetrameric Assembly That May Have a Role in
804 Ligand Transport." The Journal of Biological Chemistry **283**(47): 32812-32820.

805 D'Andrea, L. D. and L. Regan (2003). "TPR proteins: the versatile helix." Trends Biochem
806 Sci **28**(12): 655-662.

807 Davidson, A. L., E. Dassa, C. Orelle and J. Chen (2008). "Structure, function, and evolution
808 of bacterial ATP-binding cassette systems." Microbiol Mol Biol Rev **72**(2): 317-364, table of
809 contents.

810 Deka, R. K., C. A. Brautigam, M. Goldberg, P. Schuck, D. R. Tomchick and M. V. Norgard
811 (2012). "Structural, bioinformatic, and in vivo analyses of two *Treponema pallidum*
812 lipoproteins reveal a unique TRAP transporter." J Mol Biol **416**(5): 678-696.

813 Denger, K. and A. M. Cook (2010). "Racemase activity effected by two dehydrogenases in
814 sulfolactate degradation by *Chromohalobacter salexigens*: purification of (S)-sulfolactate
815 dehydrogenase." Microbiology **156**(Pt 3): 967-974.

816 Denger, K., J. Mayer, M. Buhmann, S. Weinitschke, T. H. Smits and A. M. Cook (2009).
817 "Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via
818 sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfolyase." J Bacteriol **191**(18): 5648-
819 5656.

820 Fischer, M., A. P. Hopkins, E. Severi, J. Hawkhead, D. Bawdon, A. G. Watts, *et al.* (2015).
821 "Tripartite ATP-independent Periplasmic (TRAP) Transporters Use an Arginine-mediated
822 Selectivity Filter for High Affinity Substrate Binding." The Journal of Biological Chemistry
823 **290**(45): 27113-27123.

824 Fischer, M., Q. Y. Zhang, R. E. Hubbard and G. H. Thomas (2010). "Caught in a TRAP:
825 substrate-binding proteins in secondary transport." Trends in Microbiology **18**(10): 471-478.

826 Forward, J. A., M. C. Behrendt, N. R. Wyborn, R. Cross and D. J. Kelly (1997). "TRAP
827 transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM*
828 genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria."
829 Journal of Bacteriology **179**(17): 5482-5493.

830 Fukami-Kobayashi, K., Y. Tateno and K. Nishikawa (1999). "Domain dislocation: a change
831 of core structure in periplasmic binding proteins in their evolutionary history." J Mol Biol
832 **286**(1): 279-290.

833 Fuller, T. E., M. J. Kennedy and D. E. Lowery (2000). "Identification of *Pasteurella*
834 *multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis."
835 Microbial Pathogenesis **29**(1): 25-38.

836 Glaenger, J., M. F. Peter, G. H. Thomas and G. Hagelueken (2017). "PELDOR Spectroscopy
837 Reveals Two Defined States of a Sialic Acid TRAP Transporter SBP in Solution." Biophys J
838 **112**(1): 109-120.

839 Gonin, S., P. Arnoux, B. Pierru, J. Lavergne, B. Alonso, M. Sabaty, *et al.* (2007). "Crystal
840 structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and
841 closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding."
842 BMC Struct Biol **7**: 11.

843 Graf, S., C. Broll, J. Wissig, A. Strecker, M. Parowatkin and G. Uden (2016). "CitA (citrate)
844 and DcuS (C4-dicarboxylate) sensor kinases in thermophilic *Geobacillus kaustophilus* and
845 *Geobacillus thermodenitrificans*." Microbiology **162**(1): 127-137.

846 Grammann, K., A. Volke and H. J. Kunte (2002). "New type of osmoregulated solute
847 transporter identified in halophilic members of the bacteria domain: TRAP transporter
848 TeaABC mediates uptake of ectoine and hydroxyectoine in *Halomonas elongata* DSM
849 2581(T)." J Bacteriol **184**(11): 3078-3085.

850 Herrou, J., C. Bompard, R. Antoine, A. Leroy, P. Rucktooa, D. Hot, *et al.* (2007). "Structure-
851 based mechanism of ligand binding for periplasmic solute-binding protein of the Bug
852 family." J Mol Biol **373**(4): 954-964.

853 Hosaka, M., N. Kamimura, S. Toribami, K. Mori, D. Kasai, M. Fukuda, *et al.* (2013). "Novel
854 tripartite aromatic acid transporter essential for terephthalate uptake in *Comamonas sp.* strain
855 E6." Appl Environ Microbiol **79**(19): 6148-6155.

856 Hur, J., C. Jawale and J. H. Lee (2012). "Antimicrobial resistance of *Salmonella* isolated
857 from food animals: A review." Food Research International **45**(2): 819-830.

858 Huvent, I., H. Belrhali, R. Antoine, C. Bompard, C. Loch, F. Jacob-Dubuisson, *et al.* (2006).
859 "Crystal structure of *Bordetella pertussis* BugD solute receptor unveils the basis of ligand
860 binding in a new family of periplasmic binding proteins." J Mol Biol **356**(4): 1014-1026.

861 Huvent, I., H. Belrhali, R. Antoine, C. Bompard, C. Loch, F. Jacob-Dubuisson, *et al.* (2006).
862 "Structural analysis of *Bordetella pertussis* BugE solute receptor in a bound conformation."
863 Acta Crystallogr D Biol Crystallogr **62**(Pt 11): 1375-1381.

864 Jenkins, G. A., M. Figueira, G. A. Kumar, W. A. Sweetman, K. Makepeace, S. I. Pelton, *et*
865 *al.* (2010). "Sialic acid mediated transcriptional modulation of a highly conserved
866 sialometabolism gene cluster in *Haemophilus influenzae* and its effect on virulence." BMC
867 Microbiology **10**: 48-48.

868 Johnston, J. W., N. P. Coussens, S. Allen, J. C. D. Houtman, K. H. Turner, A. Zaleski, *et al.*
869 (2008). "Characterization of the N-Acetyl-5-neuraminic Acid-binding Site of the
870 Extracytoplasmic Solute Receptor (SiaP) of Nontypeable *Haemophilus influenzae* Strain
871 2019." Journal of Biological Chemistry **283**(2): 855-865.

872 Jones, P. M. and A. M. George (2004). "The ABC transporter structure and mechanism:
873 perspectives on recent research." Cellular and Molecular Life Sciences CMLS **61**(6): 682-
874 699.

875 Kalckar, H. M. (1971). "The Periplasmic Galactose Binding Protein of *Escherichia coli*."
876 Science **174**(4009): 557-565.

877 Kelly, D. J. and G. H. Thomas (2001). "The tripartite ATP-independent periplasmic (TRAP)
878 transporters of bacteria and archaea." FEMS Microbiology Reviews **25**(4): 405-424.

879 Lecher, J., M. Pittelkow, S. Zobel, J. Bursy, T. Bönig, S. H. J. Smits, *et al.* (2009). "The
880 Crystal Structure of UehA in Complex with Ectoine—A Comparison with Other TRAP-T
881 Binding Proteins." Journal of Molecular Biology **389**(1): 58-73.

882 Leyn, S. A., Y. Maezato, M. F. Romine and D. A. Rodionov (2017). "Genomic
883 Reconstruction of Carbohydrate Utilization Capacities in Microbial-Mat Derived Consortia."
884 Frontiers in Microbiology **8**: 1304.

885 Locher, K. P. (2016). "Mechanistic diversity in ATP-binding cassette (ABC) transporters."
886 Nat Struct Mol Biol **23**(6): 487-493.

887 Lubin, J.-B., J. J. Kingston, N. Chowdhury and E. F. Boyd (2012). "Sialic Acid Catabolism
888 and Transport Gene Clusters Are Lineage Specific in *Vibrio vulnificus*." Applied and
889 Environmental Microbiology **78**(9): 3407-3415.

890 Luck, S. N., S. A. Turner, K. Rajakumar, H. Sakellaris and B. Adler (2001). "Ferric Dicitrate
891 Transport System (Fec) of *Shigella flexneri* 2a YSH6000 Is Encoded on a Novel
892 Pathogenicity Island Carrying Multiple Antibiotic Resistance Genes." Infection and
893 Immunity **69**(10): 6012-6021.

894 Maimanakos, J., J. Chow, S. K. Gaßmeyer, S. Güllert, F. Busch, R. Kourist, *et al.* (2016).
895 "Sequence-Based Screening for Rare Enzymes: New Insights into the World of AMDases

896 Reveal a Conserved Motif and 58 Novel Enzymes Clustering in Eight Distinct Families." Frontiers in Microbiology **7**(1332).

897

898 Maqbool, A., R. S. Horler, A. Muller, A. J. Wilkinson, K. S. Wilson and G. H. Thomas

899 (2015). "The substrate-binding protein in bacterial ABC transporters: dissecting roles in the

900 evolution of substrate specificity." Biochem Soc Trans **43**(5): 1011-1017.

901 Mayfield, J. E., B. J. Bricker, H. Godfrey, R. M. Crosby, D. J. Knight, S. M. Halling, *et al.*

902 (1988). "The cloning, expression, and nucleotide sequence of a gene coding for an

903 immunogenic *Brucella abortus* protein." Gene **63**(1): 1-9.

904 Meinert, C., J. Senger, M. Witthohn, J. H. Wubbeler and A. Steinbuchel (2017).

905 "Carbohydrate uptake in *Advenella mimigardefordensis* strain DPN7T is mediated by

906 periplasmic sugar oxidation and a TRAP-transport system." Mol Microbiol.

907 Muller, A., E. Severi, C. Mulligan, A. G. Watts, D. J. Kelly, K. S. Wilson, *et al.* (2006).

908 "Conservation of structure and mechanism in primary and secondary transporters exemplified

909 by SiaP, a sialic acid binding virulence factor from *Haemophilus influenzae*." J Biol Chem

910 **281**(31): 22212-22222.

911 Mulligan, C., M. Fischer and G. H. Thomas (2011). "Tripartite ATP-independent periplasmic

912 (TRAP) transporters in bacteria and archaea." FEMS Microbiology Reviews **35**(1): 68-86.

913 Mulligan, C., E. R. Geertsma, E. Severi, D. J. Kelly, B. Poolman and G. H. Thomas (2009).

914 "The substrate-binding protein imposes directionality on an electrochemical sodium gradient-

915 driven TRAP transporter." Proc Natl Acad Sci U S A **106**(6): 1778-1783.

916 Mulligan, C., D. J. Kelly and G. H. Thomas (2007). "Tripartite ATP-independent periplasmic

917 transporters: application of a relational database for genome-wide analysis of transporter gene

918 frequency and organization." J Mol Microbiol Biotechnol **12**(3-4): 218-226.

919 Mulligan, C., A. P. Leech, D. J. Kelly and G. H. Thomas (2012). "The Membrane Proteins

920 SiaQ and SiaM Form an Essential Stoichiometric Complex in the Sialic Acid Tripartite ATP-

921 independent Periplasmic (TRAP) Transporter SiaPQM (VC1777-1779) from *Vibrio*

922 *cholerae*." Journal of Biological Chemistry **287**(5): 3598-3608.

923 Piepenbreier, H., G. Fritz and S. Gebhard (2017). "Transporters as information processors in

924 bacterial signalling pathways." Molecular Microbiology **104**(1): 1-15.

925 Pohlmann, A., W. F. Fricke, F. Reinecke, B. Kusian, H. Liesegang, R. Cramm, *et al.* (2006).

926 "Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha*

927 H16." Nat Biotechnol **24**(10): 1257-1262.

928 Rabus, R., D. L. Jack, D. J. Kelly and M. H. Saier, Jr. (1999). "TRAP transporters: an ancient

929 family of extracytoplasmic solute-receptor-dependent secondary active transporters."

930 Microbiology **145 (Pt 12)**: 3431-3445.

931 Radolf, J. D., R. K. Deka, A. Anand, D. Smajs, M. V. Norgard and X. F. Yang (2016).

932 "*Treponema pallidum*, the syphilis spirochete: making a living as a stealth pathogen." Nat

933 Rev Microbiol **14**(12): 744-759.

934 Ricci, V., N. Loman, M. Pallen, A. Ivens, M. Fookes, G. C. Langridge, *et al.* (2012). "The

935 TCA cycle is not required for selection or survival of multidrug-resistant *Salmonella*."

936 Journal of Antimicrobial Chemotherapy **67**(3): 589-599.

937 Rice, A. J., A. Park and H. W. Pinkett (2014). "Diversity in ABC transporters: Type I, II and

938 III importers." Critical Reviews in Biochemistry and Molecular Biology **49**(5): 426-437.

939 Rosa, L. T., S. R. Dix, J. B. Rafferty and D. Kelly (2017). "Structural basis for high-affinity

940 adipate binding to AdpC (RPA4515), an orphan periplasmic binding protein from the

941 Tripartite Tricarboxylate Transporter (TTT) family in *Rhodopseudomonas palustris*." The

942 FEBS Journal **In Press**: DOI 10.1111/febs.14304.

943 Rucktooa, P., R. Antoine, J. Herrou, I. Huvent, C. Loch, F. Jacob-Dubuisson, *et al.* (2007).

944 "Crystal Structures of two *Bordetella pertussis* Periplasmic Receptors Contribute to Defining

945 a Novel Pyroglutamic Acid Binding DctP Subfamily." Journal of Molecular Biology **370**(1):
946 93-106.

947 Safarchi, A., S. Octavia, S. Z. Wu, S. Kaur, V. Sintchenko, G. L. Gilbert, *et al.* (2016).
948 "Genomic dissection of Australian *Bordetella pertussis* isolates from the 2008–2012
949 epidemic." Journal of Infection **72**(4): 468-477.

950 Salmon, R. C., M. J. Cliff, J. B. Rafferty and D. J. Kelly (2013). "The CouPSTU and
951 TarPQM transporters in *Rhodopseudomonas palustris*: redundant, promiscuous uptake
952 systems for lignin-derived aromatic substrates." PLoS One **8**(3): e59844.

953 Scheepers, G. H., J. A. Lycklama a Nijeholt and B. Poolman (2016). "An updated structural
954 classification of substrate-binding proteins." FEBS Letters **590**(23): 4393-4401.

955 Severi, E., D. W. Hood and G. H. Thomas (2007). "Sialic acid utilization by bacterial
956 pathogens." Microbiology **153**(9): 2817-2822.

957 Severi, E., G. Randle, P. Kivlin, K. Whitfield, R. Young, R. Moxon, *et al.* (2005). "Sialic acid
958 transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum
959 resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter."
960 Mol Microbiol **58**(4): 1173-1185.

961 Somers, J. M. and W. W. Kay (1983). "Genetic fine structure of the tricarboxylate transport
962 (*tct*) locus of *Salmonella typhimurium*." Mol Gen Genet **190**(1): 20-26.

963 Somers, J. M., G. D. Sweet and W. W. Kay (1981). "Fluorocitrate resistant tricarboxylate
964 transport mutants of *Salmonella typhimurium*." Molecular and General Genetics MGG
965 **181**(3): 338-345.

966 Steenbergen, S. M., C. A. Lichtensteiger, R. Caughlan, J. Garfinkle, T. E. Fuller and E. R.
967 Vimr (2005). "Sialic Acid Metabolism and Systemic Pasteurellosis." Infection and Immunity
968 **73**(3): 1284-1294.

969 Sweet, G. D., C. M. Kay and W. W. Kay (1984). "Tricarboxylate-binding proteins of
970 *Salmonella typhimurium*. Purification, crystallization, and physical properties." J Biol Chem
971 **259**(3): 1586-1592.

972 Sweet, G. D., J. M. Somers and W. W. Kay (1979). "Purification and properties of a citrate-
973 binding transport component, the C protein of *Salmonella typhimurium*." Can J Biochem
974 **57**(6): 710-715.

975 Takahashi, H., E. Inagaki, C. Kuroishi and T. H. Tahirov (2004). "Structure of the *Thermus*
976 *thermophilus* putative periplasmic glutamate/glutamine-binding protein." Acta Crystallogr D
977 Biol Crystallogr **60**(Pt 10): 1846-1854.

978 Tatum, F. M., L. B. Tabatabai and R. E. Briggs (2009). "Sialic acid uptake is necessary for
979 virulence of *Pasteurella multocida* in turkeys." Microb Pathog **46**(6): 337-344.

980 Thomas, Gavin H. (2016). "Sialic acid acquisition in bacteria—one substrate, many
981 transporters." Biochemical Society Transactions **44**(3): 760-765.

982 Thomas, G. H. (2017). "On the pull: periplasmic trapping of sugars before transport." Mol
983 Microbiol.

984 Thomas, G. H., T. Southworth, M. R. Leon-Kempis, A. Leech and D. J. Kelly (2006). "Novel
985 ligands for the extracellular solute receptors of two bacterial TRAP transporters."
986 Microbiology **152**(Pt 1): 187-198.

987 Vetting, M. W., N. Al-Obaidi, S. Zhao, B. San Francisco, J. Kim, D. J. Wichelecki, *et al.*
988 (2015). "Experimental strategies for functional annotation and metabolism discovery:
989 targeted screening of solute binding proteins and unbiased panning of metabolomes."
990 Biochemistry **54**(3): 909-931.

991 Vimr, E. R., K. A. Kalivoda, E. L. Deszo and S. M. Steenbergen (2004). "Diversity of
992 microbial sialic acid metabolism." Microbiol Mol Biol Rev **68**.

993 Widenhorn, K. A., J. M. Somers and W. W. Kay (1988). "Expression of the divergent
994 tricarboxylate transport operon (*tctI*) of *Salmonella typhimurium*." J Bacteriol **170**(7): 3223-
995 3227.

996 Widenhorn, K. A., J. M. Somers and W. W. Kay (1989). "Genetic regulation of the
997 tricarboxylate transport operon (*tctI*) of *Salmonella typhimurium*." J Bacteriol **171**(8): 4436-
998 4441.

999 Wilkens, S. (2015). "Structure and mechanism of ABC transporters." F1000Prime Reports **7**:
1000 14.

1001 Willis, R. C. and C. E. Furlong (1974). "Purification and Properties of a Ribose-binding
1002 Protein from *Escherichia coli*." Journal of Biological Chemistry **249**(21): 6926-6929.

1003 Winnen, B., R. N. Hvorup and M. H. Saier, Jr. (2003). "The tripartite tricarboxylate
1004 transporter (TTT) family." Res Microbiol **154**(7): 457-465.

1005 Wubbeler, J. H., S. Hiessl, J. Schuldes, A. Thurmer, R. Daniel and A. Steinbuchel (2014).
1006 "Unravelling the complete genome sequence of *Advenella mimigardefordensis* strain DPN7T
1007 and novel insights in the catabolism of the xenobiotic polythioester precursor 3,3'-
1008 dithiodipropionate." Microbiology **160**(Pt 7): 1401-1416.

1009 Wyborn, N. R., J. Alderson, S. C. Andrews and D. J. Kelly (2001). "Topological analysis of
1010 DctQ, the small integral membrane protein of the C4-dicarboxylate TRAP transporter of
1011 *Rhodobacter capsulatus*." FEMS Microbiol Lett **194**(1): 13-17.

1012 Yancey, R. J. and R. A. Finkelstein (1981). "Assmilation of iron by pathogenic *Neisseria sp.*"
1013 Infection and Immunity **32**(2): 592-599.

1014 Zhao, J. and A. N. Binns (2016). "Involvement of *Agrobacterium tumefaciens* Galacturonate
1015 Tripartite ATP-Independent Periplasmic (TRAP) Transporter GaaPQM in Virulence Gene
1016 Expression." Applied and Environmental Microbiology **82**(4): 1136-1146.

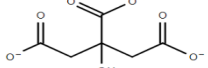
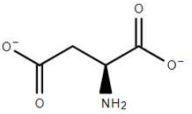
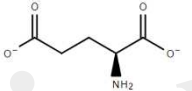
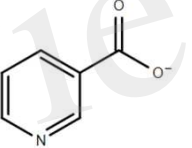
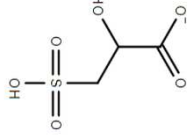
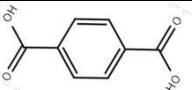
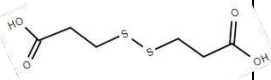
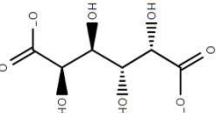
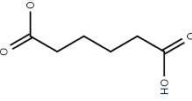
1017

1018

1018

1019 **Table 1.** Experimentally characterised TTT transporters and SBPs

1020

Name(s)	Organism(s)	Ligand	PDB code	Reference
TctCBA BctCBA TctCBA TctCBA TctCBA	<i>Salmonella typhimurium</i> <i>Bordetella pertussis</i> <i>Corynebacterium glutamicum</i> <i>Commamonas sp.</i> <i>Geobacillus thermodenitrificans</i>	Citrate		Sweet <i>et al.</i> (1979) Antoine <i>et al.</i> (2003) Brocker <i>et al.</i> (2009) Hosaka <i>et al.</i> (2013) Graf <i>et al.</i> (2016)
BugD	<i>Bordetella pertussis</i>	Aspartate		2F5X Huvent <i>et al.</i> (2006)
BugE	<i>Bordetella pertussis</i>	Glutamate		2DVZ Huvent <i>et al.</i> (2006)
Bug27	<i>Bordetella pertussis</i>	Nicotinic acid <i>et al.</i>		2QPQ Herrou <i>et al.</i> (2007)
SlcHFG SlcHFG	<i>Roseovarius nubinhibens</i> <i>Chromohalobacter salexigens</i>	Sulfolactate		Denger <i>et al.</i> (2009) Denger and Cook (2010)
TpiBa/ TphC	<i>Commamonas sp.</i>	Terephthalate		Hosaka <i>et al.</i> (2013)
TctCBA	<i>Advenella mimigardefordensis</i>	disulfide 3,3'-dithiodipropionic acid (DTDP)		Wubbeler <i>et al.</i> (2014)
TctC	<i>Polaromonas sp.</i>	Unknown		4X9T
TctABC	<i>Halomonas sp.</i>	Galactarate/ glucarate		Leyn <i>et al.</i> (2017)
AdpC	<i>Rhodopseudomonas palustris</i>	Adipate <i>et al.</i>		5OEI Rosa <i>et al.</i> (2017)

1021

1022

1022

1023

1024 **Table 1.** Comparison between the TctC homolog structures available in the PDB. Root Mean
1025 Square Deviations (RMSD) and Sequence Identity (SI) were generated using EMBL 3D
1026 alignment tool available at <http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver>. PDB codes
1027 for each of the structures is shown below the protein names in the horizontal line.

RMSDs (Å)	BugD (B)	BugE	Bug27 (B)	TctC	AdpC
SI (%)	2F5X	2DVZ	2QPQ	4X9T	5OEI
BugD (B)	NA	1.33	3.00	3.89	1.75
BugE	34	NA	2.62	3.56	1.54
Bug27 (B)	25	32	NA	2.2	2.75
TctC	15	19	28	NA	3.96
AdpC	30	33	27	18	NA

1028

1029

In review

1030 **Table 2.** Number of TctC and TctA homologs per accession. Only accessions with more than
 1031 50 TctC homologs are presented. An e-value of 10^{-15} was used.

Species	Phyla	tctC homologs	tctA homologs
<i>Achromobacter denitrificans</i>	β-proteobacteria	99	6
<i>Achromobacter insolitus</i>	β-proteobacteria	177	4
<i>Achromobacter xylosoxidans</i>	β-proteobacteria	202	5
<i>Acidovorax avenae</i>	β-proteobacteria	66	4
<i>Acidovorax citrulli</i>	β-proteobacteria	55	4
<i>Advenella kashmirensis</i>	β-proteobacteria	116	1
<i>Advenella mimigardefordensis</i>	β-proteobacteria	129	7
<i>Alicyclophilus denitrificans</i>	β-proteobacteria	143	3
<i>Bordetella bronchialis</i>	β-proteobacteria	195	5
<i>Bordetella bronchiseptica</i>	β-proteobacteria	182	4
<i>Bordetella flabilis</i>	β-proteobacteria	214	6
<i>Bordetella genomosp</i>	β-proteobacteria	139	4
<i>Bordetella hinzii</i>	β-proteobacteria	105	5
<i>Bordetella holmesii</i>	β-proteobacteria	56	6
<i>Bordetella parapertussis</i>	β-proteobacteria	142	4
<i>Bordetella pertussis</i>	β-proteobacteria	81	2
<i>Bordetella petrii</i>	β-proteobacteria	107	5
<i>Bordetella pseudohinzii</i>	β-proteobacteria	86	5
<i>Bordetella trematum</i>	β-proteobacteria	100	5
<i>Comamonas serinivorans</i>	β-proteobacteria	87	1
<i>Comamonas testosteroni</i>	β-proteobacteria	100	2
<i>Cupriavidus basilensis</i>	β-proteobacteria	155	5
<i>Cupriavidus gilardii</i>	β-proteobacteria	98	6
<i>Cupriavidus metallidurans</i>	β-proteobacteria	122	5
<i>Cupriavidus necator</i>	β-proteobacteria	190	3
<i>Delftia acidovorans</i>	β-proteobacteria	157	3
<i>Delftia tsuruhatensis</i>	β-proteobacteria	145	3
<i>Hydrogenophaga sp. PBC</i>	β-proteobacteria	64	4
<i>Polaromonas sp. JS666</i>	β-proteobacteria	96	4
<i>Pseudorhodoplanes sinuspersici</i>	α-proteobacteria	99	8
<i>Pusillimonas sp. T7-7</i>	β-proteobacteria	52	3
<i>Ralstonia eutropha</i>	β-proteobacteria	156	5
<i>Ramlibacter tataouinensis</i>	β-proteobacteria	75	3
<i>Rhodoferax sp. DCY110</i>	β-proteobacteria	90	3
<i>Rhodoplanes sp. Z2</i>	α-proteobacteria	434	9
<i>Variovorax paradoxus</i>	β-proteobacteria	135	8
<i>Verminephrobacter eiseniae</i>	β-proteobacteria	130	3

1032

1033

1034 **Table 3.** Number of TctC and TctA homolog per accession. Only accessions with 8 or more
 1035 TctA homologs are presented. An e-value of 10^{-15} was used.

Specie	Class	<i>tctA</i> homologs	<i>tctC</i> homologs
<i>Antarctobacter heliothermus</i>	α -proteobacteria	12	9
<i>Bradyrhizobium icense</i>	α -proteobacteria	8	43
<i>Chelativorans</i> sp. BNC1	α -proteobacteria	13	15
<i>Chelatococcus</i> sp. CO-6	Chlorophyceae	11	10
<i>Chromohalobacter salexigens</i>	γ -proteobacteria	8	9
<i>Defluviimonas alba</i>	α -proteobacteria	9	9
<i>Desulfovibrio fairfieldensis</i>	δ -proteobacteria	9	9
<i>Ensifer sojae</i>	α -proteobacteria	9	6
<i>Geosporobacter ferrireducens</i>	Clostridia	8	9
<i>Granulosicoccus antarcticus</i>	γ -proteobacteria	13	13
<i>Halomonas chromatireducens</i>	γ -proteobacteria	8	8
<i>Halomonas huangheensis</i>	γ -proteobacteria	17	14
<i>Kushneria konosiri</i>	γ -proteobacteria	9	9
<i>Kushneria marisflavi</i>	γ -proteobacteria	9	9
<i>Lachnoclostridium</i> sp. YL32	Clostridia	8	5
<i>Marinobacterium aestuarii</i>	γ -proteobacteria	13	11
<i>Marinomonas</i> sp. MWYL1	γ -proteobacteria	9	7
<i>Marinovum algicola</i>	α -proteobacteria	14	12
<i>Martelevella mediterranea</i>	α -proteobacteria	21	18
<i>Martelevella</i> sp. AD-3	α -proteobacteria	10	9
<i>Oligotropha carboxidovorans</i>	α -proteobacteria	8	8
<i>Paenibacillus naphthalenovorans</i>	Bacilli	16	19
<i>Pelagibacterium halotolerans</i>	α -proteobacteria	8	7
<i>Pseudorhodoplanes sinuspersici</i>	α -proteobacteria	8	99
<i>Rhodoplanes</i> sp. Z2	α -proteobacteria	9	434
<i>Sediminispirochaeta smaragdinae</i>	Spirochaetes	9	9
<i>Sinorhizobium fredii</i>	α -proteobacteria	8	7
<i>Sphaerochaeta globosa</i>	Spirochaetes	9	6
<i>Sphaerochaeta pleomorpha</i>	Spirochaetes	9	7
<i>Starkeya novella</i>	α -proteobacteria	15	17
<i>Sulfitobacter pseudonitzschiae</i>	α -proteobacteria	8	7
<i>Variovorax paradoxus</i>	β -proteobacteria	8	135
<i>Yangia</i> sp. CCB-MM3	α -proteobacteria	8	6

1036

1037

1038

1039 **Table 4.** Number of genomes showing different patterns in terms of numbers of *tctC* and *tctA*
1040 homologs, using different e-values as thresholds. ^{a2} is also counted inside of ^{a1}, and ^{b2} is also
1041 counted inside ^{b1}.

1042

e-value	10 ⁻¹⁵	10 ⁻¹²	10 ⁻⁹	10 ⁻⁶	10 ⁻³	10 ⁰
<i>tctC</i> > <i>tctA</i> ^{a1}	176	196	212	220	255	2157
<i>tctA</i> > <i>tctC</i> ^{b1}	210	100	66	47	39	64
<i>tctC</i> without <i>tctA</i> ^{a2}	36	43	54	56	79	338
<i>tctA</i> without <i>tctC</i> ^{b2}	115	33	13	6	8	9
<i>tctC</i> = <i>tctA</i> ≠ 0	316	415	444	457	456	97
No homologs	1620	1612	1601	1599	1573	5
Total Number of <i>tctA</i> homologs	1633	1637	1639	1641	1649	5705
Total number of <i>tctC</i> homologs	7213	7405	7495	7552	7632	25225

1043

1044

In review

1044

1045 **Figures legends**

1046

1047 **Figure 1. Overall topologies and structures of the different Solute Binding Protein (SBP**
1048 **dependant) transporter families.** The ABC transporters are represented by the *E. coli*
1049 maltose transporter MalEFGK2 (PDB 2R6G). It is composed of a solute binding protein
1050 (red), two Transmembrane (TM) domains (blue) and two Nucleotide binding domains (NBD)
1051 (orange); The secondary Tripartite ATP-independent periplasmic (TRAP) Transporters are
1052 composed of a 12 TM domain channel DctM (blue) and a 4 TM domain protein DctQ
1053 (green), which can be fused together by an additional TM domain (yellow) in a DctQM
1054 protein, and a DctP or TAXI SBP protein, represented respectively by the SiaP from *H.*
1055 *influenza* (PDB 2CEY) (light red) and the TT1099 from *T. thermophilus* (PDB 1US4) (dark
1056 red). TAXI-TRAP were always found associated with fused DctQM proteins; The Tripartite
1057 Tricarboxylate Transporter (TTT), is formed also by a 12 TM channel TctA (purple), a 4 TM
1058 protein TctB (cyan) and a TctC solute binding protein, represented by Bug27 from *B.*
1059 *pertussis* (PDB 2QPQ) (Brown). In some rare cases, TctAB proteins may be also fused.
1060 Although sharing similar topology, the TRAP and TTT systems share no sequence similarity.

1061

1062 **Figure 2. Examples of genetic organisation for different secondary SBP dependant**
1063 **transporters. (A) Gene organisations for TRAP systems. (B) Gene organisation for TTT**
1064 **systems. * represents an amido-hydrolase gene.**

1065

1066 **Figure 3. Comparison between TRAP and TTT crystal structures. (A) Overall structure**
1067 **of SiaP, a sialic acid binding SBP from the TRAP family in *H. influenzae* (PDB 3B50).**
1068 Domain 1 is represented in cyan (α -helix) and purple (β -sheet), and domain 2 is represented
1069 in blue (α -helix) and green (β -sheet). A sialic acid molecule is shown in the binding pocket.
1070 Arg147, important to perform a salt bridge with the carboxylic group of the substrate in most
1071 SiaP homologs is shown in red. **(B) Overall structure of BugD, a aspartate binding SBP**
1072 **from the TTT family in *B. pertussis* (PDB 2F5X).** Domain 1 is represented in green (α -
1073 helix) and orange (β -sheet), and domain 2 is represented in blue (α -helix) and brown (β -
1074 sheet). An aspartate molecule is shown in the binding pocket. Two loops, between β 1 and α 1
1075 and between β 7 and α 6, are involved in the conserved coordination of two water molecules,
1076 which bridge hydrogen bonds with the proximal carboxylic group in the substrate. These
1077 loops and waters are shown in red **(C) Binding pocket of SiaP, showing the coordination**
1078 **of the sialic acid molecule.** Arg147 perform a salt bridge with the carboxylic group in the
1079 substrate. The remaining of the molecule is coordinated by hydrogen bonds and water
1080 molecules which are variable inside the TRAP family. **(D) Binding pocket of BugD,**
1081 **showing the coordination of the aspartate molecule.** The residues Ala-15, gly-16, gly-17
1082 and Asp-20 participate in the loop between β 1 and α 1, coordinating two water molecules
1083 which bridge hydrogen bonds with the proximal carboxylic group in the substrate. This
1084 coordination is very conserved among the TTT family. The remaining substrate coordination
1085 occur through non-conserved hydrogen bonds and water bridging.

1086

1087 **Figure 4. Numbers of genomes containing different ranges of homologs for *tctC* and**
1088 ***tctA*.**

Figure 1.TIF

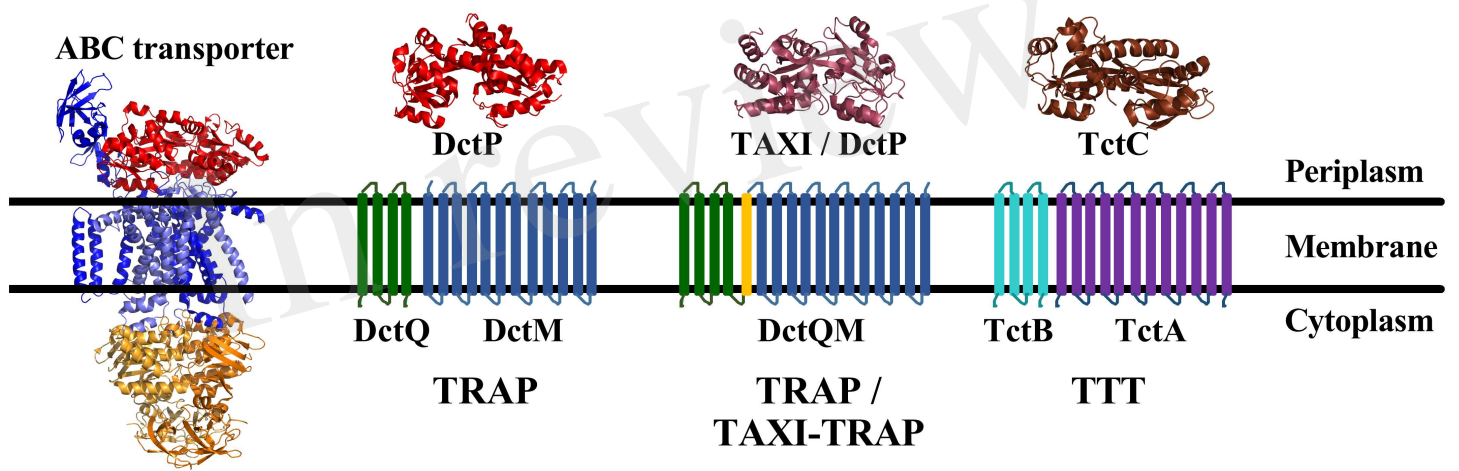
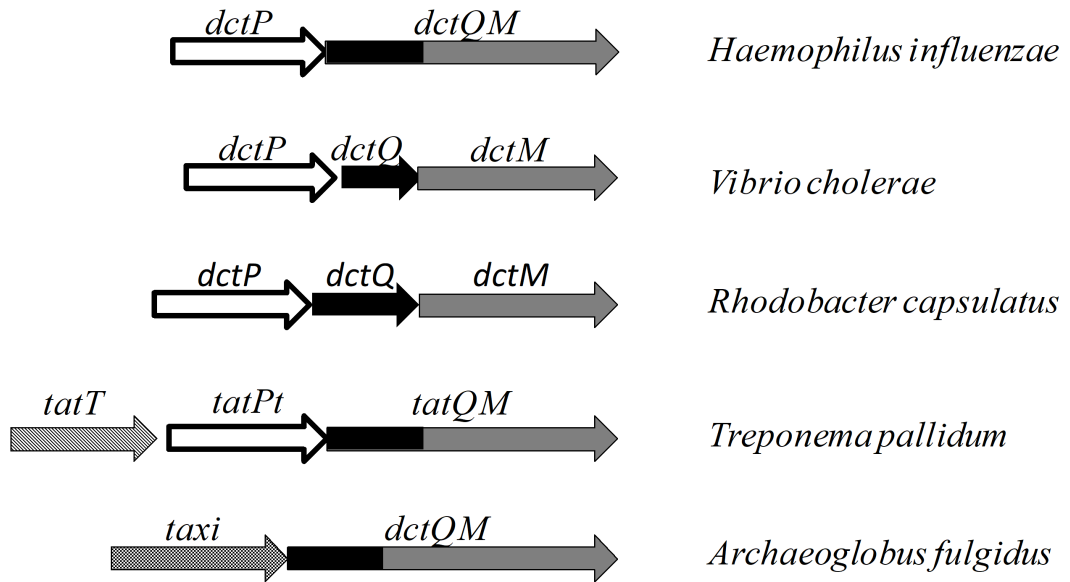


Figure 2.TIF

(A)



(B)

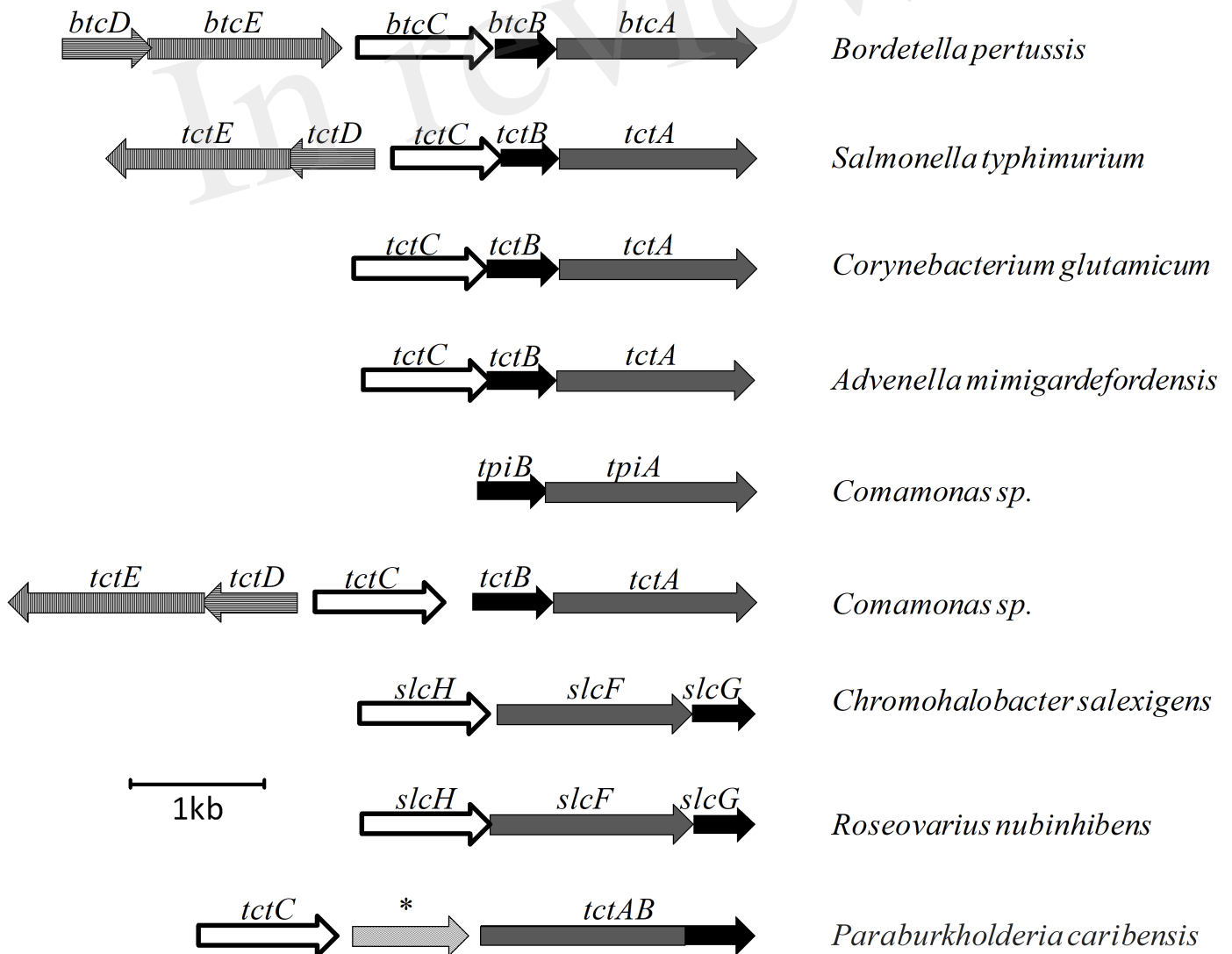


Figure 3.TIF

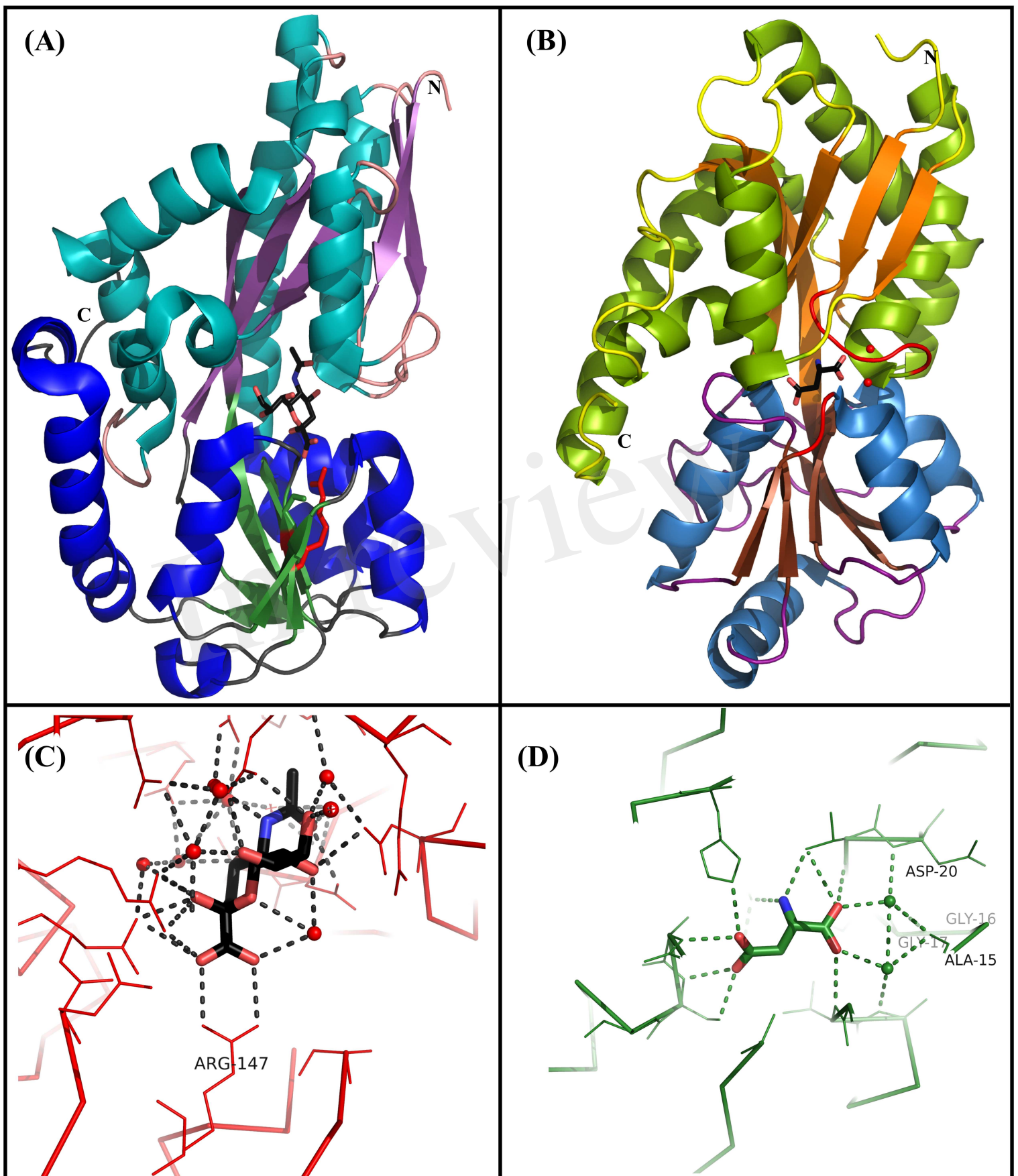


Figure 4.TIF

