PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A novel TctA citrate transporter from an activated sludge metagenome: Structural and mechanistic predictions for the TTT family

Ramón Alberto Batista-García,[1] Ayixon Sánchez-Reyes,[1] César Millán-Pacheco,[1] Víctor Manuel González-Zuñiga,[2] Soledad Juárez,[2] Jorge Luis Folch-Mallol,[3] and Nina Pastor[1]*

[1] Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, 62209 Morelos, México

[2] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, 62210 Morelos, México

[3] Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, 62209 Morelos, México

ABSTRACT

We isolated a putative citrate transporter of the tripartite tricarboxylate transporter (TTT) class from a metagenomic library of activated sludge from a sewage treatment plant. The transporter, dubbed TctA_ar, shares ~50% sequence identity with TctA of *Comamonas testosteroni* (TctA_ct) and other β-*Proteobacteria*, and contains two 20-amino acid repeat signature sequences, considered a hallmark of this particular transporter class. The structures for both TctA_ar and TctA_ct were modeled with I-TASSER and two possible structures for this transporter family were proposed. Docking assays with citrate resulted in the corresponding sets of proposed critical residues for function. These models suggest functions for the 20-amino acid repeats in the context of the two different architectures. This constitutes the first attempt at structure modeling of the TTT family, to the best of our knowledge, and could aid functional understanding of this little-studied family.

Proteins 2014; 00:000–000.

Key words: metagenomic; TctA; molecular modeling; docking; gene phylogeny; citrate.

INTRODUCTION

There are at least 10^6–10^8 species of microorganisms in the world,[1] and approximately 99% of the microorganisms that inhabit ecosystems cannot be cultivated in the laboratory.[2,3] Metagenomic tools provide a new avenue to study these microbial populations and their genes, in the prospection for novel functions. Screening methods are critical in the search, isolation, and characterization of new proteins from metagenomics constructions. Among the most used currently are function-based and sequence-based screenings.[4]

Looking for a novel esterase that degrades organophosphates in a metagenomic library derived from activated sludge, we isolated a clone with three complete open reading frames (ORFs). Two of these code for a putative thioesterase, most likely responsible for the parathion degradation, and a dihydratase, which will be described elsewhere. The remaining ORF codes for a putative citrate transporter of the tripartite tricarboxylate transporter (TTT) class.[5] Given that these transporters can be recruited to transport aromatic substrates,[6] these three ORFs reside in the same clone, and they may be functionally coupled, we analyzed and modeled the putative transporter.

Additional Supporting Information may be found in the online version of this article.

Abbreviations: TctA_ar, novel TctA protein from activated sludge metagenome; TctA_ct, TctA protein from *Comamonas testosteroni* E6; TTT, tripartite tricarboxylate transporter

Grant sponsor: CONACyT; Grant number: CB-153789-Q.

Ramón Alberto Batista-García and Ayixon Sánchez-Reyes have contributed equally to this work.

*Correspondence to: N. Pastor, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, Cuernavaca, 62209 Morelos, México. E-mail: nina@uaem.mx

Received 9 October 2013; Revised 14 January 2014; Accepted 28 January 2014
Published online 3 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24529

The TctABC transporter system includes three proteins: TctA, TctB, and TctC.[7] All these proteins belong to the TTT family (TC 2.A.80).[5] TctB and TctC are a small protein with four transmembrane helices and a periplasmic tricarboxylate-binding receptor, respectively, and do not seem to be essential for function. TctAs are large proteins, with 12 putative transmembrane α-helical segments. They have 500 amino acids on average, ranging between 416 and 637 amino acids in length. A short repeated sequence, $(Hy)_6$-G-$(Hy)_3$-G*-$(Hy)_3$-G*-$(Hy)_2$-P*G*-Hy, where Hy is a hydrophobic residue and * is a highly conserved residue, was presented[5] as evidence of an ancient duplication event.

Crystallized structures of this family of proteins do not exist to date, and functional studies have only been carried out for TctA proteins of the genera *Salmonella*,[7,8] *Comamonas*,[9] *Corynebacterium*,[10,11] and *Bordetella*.[12] The proposed transport mechanism is the symport of citrate with sodium, although divalent cations also play an important role in binding citrate to TctC,[10] eventually bringing citrate to TctA. To the best of our knowledge, the only reported mutation for this family was isolated in ciprofloxacin-stressed *Salmonella*; it involves the mutation of G109 to serine and eliminates citrate transport.[8]

Citrate transport can be achieved with a large variety of secondary transporters, both symporters (reviewed in Ref. 10) and antiporters (reviewed in Ref. 13), as can be seen in the Transporter Classification Database (TCDB).[14] Currently, there are crystal structures for various members of the MFS, IT, APC, and MOP superfamilies. Some superfamilies, such as MFS, can harbor antiporters and symporters with similar transporter architectures (reviewed in Ref. 15), and similar architectures can be accommodated by sequences with very low similarity, as reported recently in the case of two citrate transporters from TC classes 2.A.11 and 2.A.47.[16] TctA proteins display less than 15% identity with any of the crystallized transporters deposited in the Protein Data Bank (PDB).[17] Therefore, any exercise in structure modeling must consist of threading the sequence of the protein over plausible structural templates. We propose two possible structures for this novel protein, and a list of critical residues for function, that can be assayed by site-directed mutagenesis in future studies.

MATERIALS AND METHODS

Library construction

DNA extraction from duplicate samples of activated sludge, collected from an industrial and urban sewage treatment plant of Cuernavaca, Morelos, México, was performed using the UltraClean®Microbial DNA Isolation Kit, MO BIO Laboratories. DNA was characterized using agarose gel electrophoresis and nanodrop readings. We amplified fragments of the genes coding for the 16S and fungal 18S RNA with PCR.[18,19]

We fragmented the isolated DNA mechanically by nebulization with nitrogen. The result was confirmed using agarose gel electrophoresis and quantitation in nanodrop. Nebulized DNA was repaired with T4 DNA polymerase to generate blunt ends, using standard procedures. DNA was ligated to vector pJET1.2 and subsequently transformed by electroporation of *E. coli* DH5α. The library was amplified and preserved in glycerol (20%) at −70°C.

Selection and characterization of clones

The functional scrutiny consisted in testing for esterase activity.[20] Plasmid extraction was performed for the positive clones using GeneGet Plasmid Miniprep Kit|Thermo Scientific. The size of the inserts was estimated with a *BglII* digestion, and the inserts were sequenced.

Bioinformatic analysis

We used the SnapGene Viewer 2.1 program and the site of NCBI to predict ORFs. Protein–protein BLASTs were carried out to ascertain sequence homology with the nonredundant protein set at the NCBI site.

Sequence alignment between functionally studied TctA proteins and our clone was performed with Clustal Omega Multialign Software and ESPript 2.2 servers. The binary comparison scores are expressed as percent identity.

Phylogenetic analysis was performed online with the server Phylogeny.fr. BLAST searches on the nonredundant protein collection at the National Center for Biotechnology Information (NCBI) were done, using as queries our clone and the TctAs of *Salmonella, Comamonas, Bordetella*, and *Corynebacterium*. A nonredundant set was constructed from the highest 100 hits for each search, and the TctAs reported in Ref. 5, complemented with five TctA homologs of archaebacteria.

Molecular modeling

The predicted amino acid sequence for the putative TctA protein (TctA_ar hereon), and TctA from *Comamonas testosteroni* E6 (TctA_ct hereon), were submitted to the I-TASSER server[21] without constraints. Selected templates that were not transporters were forbidden in subsequent modeling attempts. We generated models with the same template for both TctA_ar and TctA_ct.

Each model was inserted into a dimyristoylphosphatidylcholine (DMPC) bilayer with 25% cholesterol and solvated with water and 0.15*M* KCl, using the CHARMM-GUI server. The systems were structurally relaxed by energy minimizations for 10,000 conjugate gradient steps and a short molecular dynamics run (25,000 steps at 30°C), using the CHARMM36 force field in NAMD.[22] This procedure eliminates steric clashes and regularizes

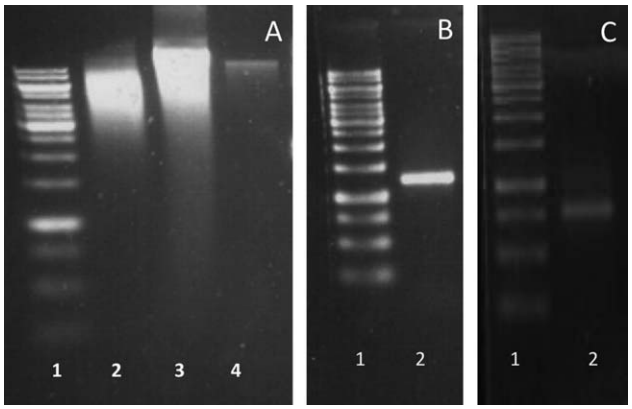

Figure 1

(A) Metagenomic DNA isolation from activated sludge samples. Lane 1: 1 kb ladder, lanes 2 and 3: activated sludge metagenomic DNA, lane 4: soil metagenomic DNA. (B) PCR to amplify approximately 1300 bp of the gene coding for 16S RNA. Lane 1: 1 kb ladder, lane 2: band corresponding to gene fragments coding for 16S RNA. (C) PCR to amplify approximately 750 bp of the gene coding for the 18S RNA. Lane 1: 1 kb ladder, lane 2: band corresponding to gene fragments coding for 18S RNA.

the stereochemistry of the models, but does not alter the trace of the main chain of the protein.

We docked citrate to the relaxed models using AUTODOCK/VINA[23] as a functional criterion. We used the default AUTODOCK/VINA parameters, allowing for flexibility in citrate but not in the proteins. Structural alignments between the model structure of TctA_ar and TctA_ct and the closest structural homolog detected by I-TASSER were used to identify putative relevant functional residues.

RESULTS AND DISCUSSION

Characterization of metagenomic DNA library

We isolated 10 μg of DNA from the sample on average. Because the kit we used is recommended for extraction of soil metagenomic DNA, we also isolated DNA from a natural soil sample, as a positive control. For the control sample, we obtained a defined band of high molecular weight (Fig. 1A).

We ascertained DNA concentration and quality by reading absorbance of the sludge and soil samples at 260 and 280 nm in nanodrop. For the sludge the ratio (A260/280) was 1.8, and for the soil sample the ratio (A260/280) was 2.01, confirming that DNA purity is suitable for metagenomic construction. As another quality criterion, we performed PCR to amplify fragments of the genes coding for 16S and fungal 18S RNA [Fig. 1(B,C)]. These PCRs served to confirm the presence of genomic material of bacteria and fungi in our samples. It is beyond the interest of this work to describe the structure of the microbial communities present in the activated sludge sample; hence, the PCRs are just taken as qualitative criteria of the purity of the metagenomic DNA, and presence of bacterial and fungal organisms in the activated sludge samples.

For the construction of the metagenomic DNA library, fragments between 6 and 10 kb are the target, and were achieved with nebulization at 8 psi for 5 s. These fragments were ligated in pJET1.2/blunt and transformed in *E. coli* DH5α, to obtain a metagenomic library of 20,000 clones. The title of DNA was approximately 90 mb, and we estimated that 93% of the clones have an insert with an average size of approximately 8.5 kb.

Bioinformatic analysis of the insert

A clone was identified positive for methyl parathion degradation. Restriction analysis with *BglII* defined the presence of an insert of about 6 kb. We found a complete ORF (1311 bp) that encodes a protein of 45.1 kDa theoretical molecular weight (463 amino acids) with homology to TctA proteins, part of the TTT family; the gene was named *tctA_ar*. TctA consist of between 416 and 637 amino acids,[5] which brackets the size of TctA_ar. The sequence determined in this study was deposited in the NCBI database under Accession No. KF570143.

To produce an educated guess regarding the origin of our cloned transporter, we selected a set of 119 TctA sequences that include those described in Ref. 5, and five archaeal TctAs (see Methods). With this set, we built a phylogenetic tree using Phylogeny.fr software (Supporting Information Fig. S1). It clearly shows that TctA_ar is not archaeal, and probably originated in a β-*Proteobacteria*, given its nearest neighbors in the tree. Given the possibility of horizontal transfer amongst bacteria, yeast, and plants, it is currently impossible to determine the host of the protein, and should definitely be considered a clone from a soil sample. This search selected a TctA homolog in a plant (gi 224156094, from *Populus trichocarpa*), the first report of a eukaryotic TctA.

The amino acid sequence of TctA_ar exhibited ~50% identity with the TctA protein of *C. testosteroni*, and with other TctA proteins of β-*Proteobacteria*, as summarized in Table I. Given that the TctA of *Comamonas* has been functionally characterized, we chose it over that of *Verminephrobacter* as a control protein for our modeling efforts.

Conserved regions in the TctA sequences from Table I were identified with Clustal. The alignment in Figure 2 shows high conservation in the first half of the protein, reaching 85% pairwise identity. Two repeated 20-amino acid signatures $((Hy)_6\text{-}G\text{-}(Hy)_3\text{-}G^*\text{-}(Hy)_3\text{-}G^*\text{-}(Hy)_2\text{-}P^*G^*\text{-}Hy)$[5] are located between amino acids L20 and G40 in the N-terminal repeat, and between D251 and

Table I
Pairwise Identity Matrix Between the Closest Neighbor and Functionally Characterized TctAs and TctA_ar

	1	2	3	4	5	6
1. *Verminephrobacter eiseniae* (gi 121610934)	100.00	88.47	73.20	54.13	40.60	37.85
2. *Comamonas testosteroni* (gi 264676989)	88.47	100.00	72.20	51.61	39.00	37.85
3. *Bordetella holmesii* (gi 491160357)	73.20	72.20	100.00	47.58	40.80	39.00
4. TctA_ar (KF570143)	54.13	51.61	47.58	100.00	28.18	25.06
5. *Salmonella enterica* (gi 446304178)	40.60	39.00	40.80	28.18	100.00	35.06
6. *Corynebacterium glutamicum* (gi 489959448)	37.85	37.85	39.00	25.06	35.06	100.00

R271 in the C-terminal repeat (boxed in full lines in Fig. 2; TctA_ar numbering). A highly conserved region, located between amino acids 61 and 111 of these sequences, is also apparent (boxed in broken black lines in Fig. 2). In the absence of directed mutagenesis studies of these transporters, the functional or structural relevance of these sequence signatures is unknown. The only reported mutant for this family, G109S, in *Salmonella enterica*[8] is marked with an asterisk.

Structural modeling and functional inferences

Given the negligible sequence similarity of TctAs to transporter structures deposited in the PDB, we chose I-TASSER[21] as a modeling tool, in view of its success in Critical Assessment of protein Structure Prediction (CASP). To have a measure of the robustness and possible sense of our structural predictions, we generated models for both TctA_ar and TctA_ct. We reasoned that the ~50% sequence identity shared by these two proteins should be reflected in the conservation of structural features.

Both TctA_ar and TctA_ct are predicted to be mostly α helical by I-TASSER, with many long and broken helices, in agreement with published hydropathy plots.[5] In our first attempt at modeling of TctA_ar, I-TASSER selected as templates HEAT repeat proteins. Given the information in Table I and Supporting Information Figure S1 that strongly suggests TctA_ar to be a citrate transporter, we discarded these HEAT repeat templates as biologically irrelevant. The following rounds of modeling denied I-TASSER the possibility of choosing previously identified templates and their homologs with up to 50% sequence identity. Subsequent modeling attempts selected a variety of transporters, from different families. These are collected and summarized in Table II.

TctA_ar and TctA_ct selected different templates. To have both proteins referred to the same template, we included some of the transporter templates already chosen by one protein in the forbidden list, so I-TASSER was forced to try new templates. To speed up this process, we also specified a particular template (4K1C for TctA_ct). We obtained six possible structures for each TctA sequence, based on the following PDB templates: 4F35, 1KPL, 4K1C, 3VVN, 3DH4, and 4IKV, representing all the superfamilies in Table II. We used TctA-specific features to select the most plausible models from this set. For each template, we located the 20-amino acid repeat on the structures of both proteins. Successful templates placed these repeats at equivalent positions in both modeled proteins. Furthermore, we required that the location of G107 (the equivalent of G109 in *S. enterica*) be close by in the models of both proteins. Only two templates complied with these conditions, 3VVN (MATE, MOP superfamily[24] and 4K1C (CDF superfamily[25]), as shown in Figure 3. The overall quality of the models, as indicated by their C-scores and TM-scores is as follows: TctA_ar with the 3VVN template (−3.22, 0.35 ± 0.13), TctA_ar with the 4K1C template (−2.76, 0.40 ± 0.13), TctA_ct with the 3VVN template (−1.68, 0.51 ± 0.15), and TctA_ct with the 4K1C template (−2.81, 0.39 ± 0.13). As expected from the low sequence similarity to the deposited structures in the PDB, C-scores are low, and TM-scores are also low. The MATE transporters have 12 transmembrane helices, in a 6 + 6 topology, whereas the CDF transporters have a 5 + 5 topology. The two templates provide different solutions for the ~70 amino acid length difference between TctA_ar and TctA_ct: in the context of 3VVN, insertions are spread throughout the whole structure; in the context of 4K1C, most of the difference is pushed to the C-terminus, concentrated in a helical hairpin that lies outside the core of the transporter (see Fig. 4). Although both MATE and CDF transporters described to date are antiporters, the exact mechanism for translocation has not been elucidated, and there is no *a priori* reason to forbid symport in these architectures.

A possible role of the 20-amino acid repeat can be proposed from these structures. In those derived from 3VVN, this segment lies in two symmetry-related helices (helices 1 and 7, as envisioned in Ref. 5) that line the ample binding pocket of the transporter. In the 4K1C-derived structures, the first repeat lies in the MRb helix, apposed to helix 6, where the other repeat lies; glycine-rich helices can slide easily against other helices and may therefore be involved in the transport cycle conformational switch. Furthermore, as MRb helices are not conserved in the CDF superfamily (Supporting Information

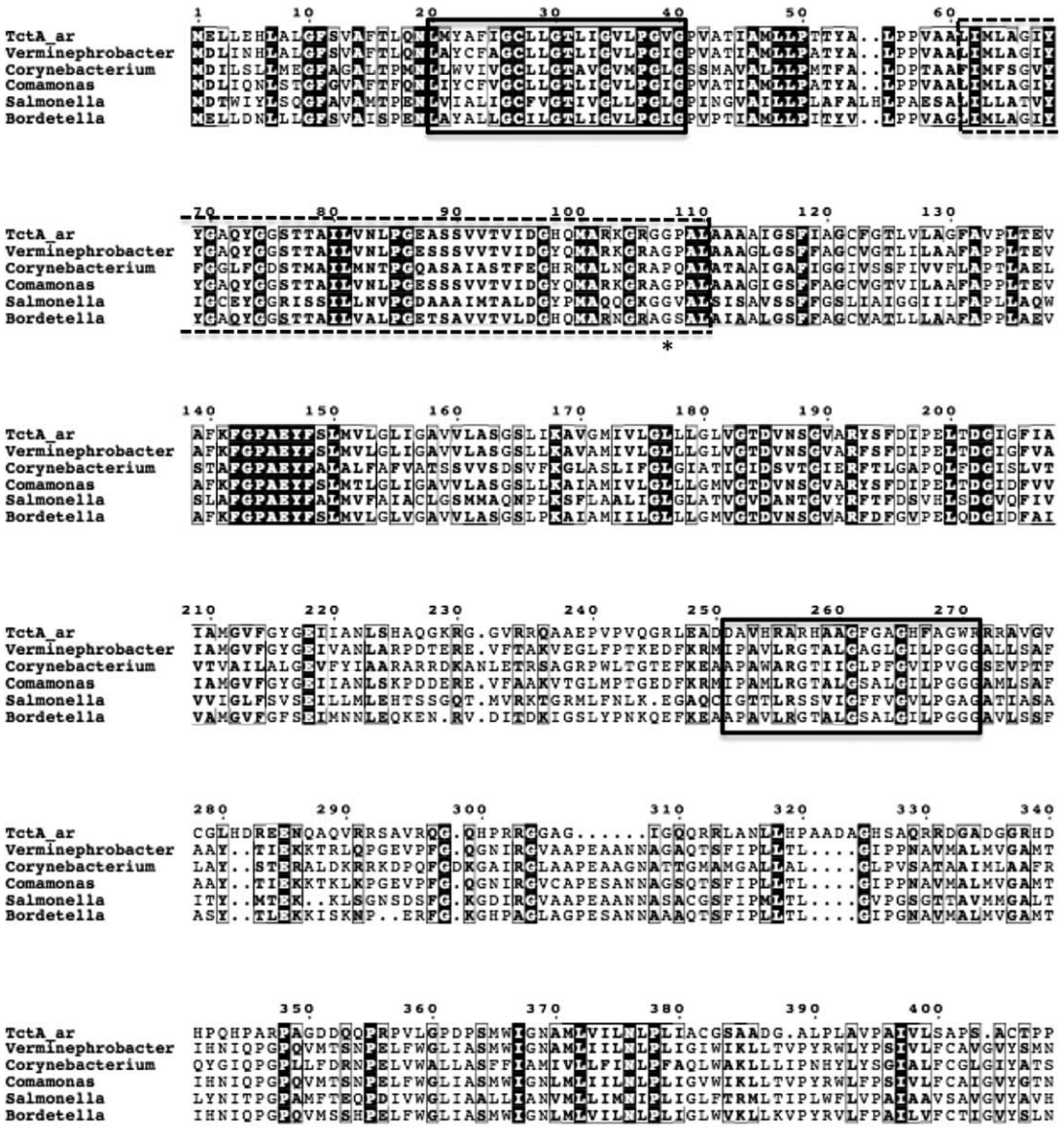

Figure 2

Conserved sequence blocks in the multiple sequence alignment between TctA_ar and functionally characterized TctAs. Boxes indicate conserved sequences, and similar sequences are indicated by darker background. The thick full line boxes indicate the repeated 20-amino acid blocks. The thick broken line box indicates a highly conserved region in the N-terminal repeat of TctA proteins. The asterisk marks the mutated G that eliminates transporter function in *Salmonella*.

Fig. 2 in Ref. 25), this model could also accommodate TctA sequences with truncated N-termini. With regard to G107, in 3VVN-derived structures it lies at the beginning of helix 3, facing the connecting loop between the two repeats, and could be involved in the translocation cycle. In 4K1C-derived structures, it lies also at the beginning of helix 3, facing the loop between helices 8 and 9, at the opposite end of the acidic helix. Its functional role in this template is harder to glean.

The TM scores to the templates range from 0.856 in the best case (TctA_ar in the 3VVN template) to 0.700 in the worst case (TctA_ct in the 4K1C template), with

Table II
Transporter Templates Selected by I-TASSER for TctA_ar and TctA_ct

PDB ID	Family	Transported species	TC number	Superfamily
4F35	DASS	Divalent anion:Na^+ symport	2.A.47.5.2	IT
3Q17, 3ORG, 1KPL, 3ND0	CLC	Cl^-:H^+ antiport	2.A.49.5	–
4K1C,	VCX	Ca^{2+}:H^+ antiport	2.A.19	CDF
4KPP, 4KJS	CAX			
3V5U	NCX	Ca^{2+}:Na^+ antiport		
3VVN, 3MKT, 3HUK	MATE	Substrate:H^+ antiport Substrate:Na^+ antiport	2.A.66.1	MOP
2XQ2, 3DH4	SSS	Galactose:Na^+ symport	2.A.21.3	APC
2JLN	NCS1	Benzyl-hydantoin:cation symport	2.A.39.3	APC
3QE7	NCS2	Nucleobase:cation symport	2.A.40.1.1	APC
3GIA	APC	Amino acid:H^+ symport	2.A.3.6.3	APC
2WIT	BCCT	Glycine-betaine:Na^+ symport	2.A.15.1	APC
4IU8	NNP	Nitrate:H^+ symport; nitrate/nitrite antiport	2.A.1.8.10	MFS
1PW4	OPA	Glycerol-phosphate:phosphate antiport	2.A.1.4.3	MFS
4APS, 4IKV, 2XUT, 4LEP	POT	Peptide:H^+ symport	2.A.17.1	MFS
4J05	PHS	Phosphate:H^+ symport	2.A.1.9	MFS
2GFP	DHA1	Grug:H^+ antiport	2.A.1.2	MFS
4GBY	SP	Xylose:H^+ symport	2.A.1.1	MFS

coverages ranging from ~70% to ~90%. TM scores above 0.5 are good indicators of structural compatibility, and our models are clearly above this threshold. We stress that this does not mean that the models are correct; it only suggests their plausibility. Furthermore, TM scores above 0.5 are not common when the trial sequence cannot be successfully threaded into a template, an expected outcome for protein sequences that have negligible homology to reported structures in the PDB; we found many instances of TM scores below 0.3.

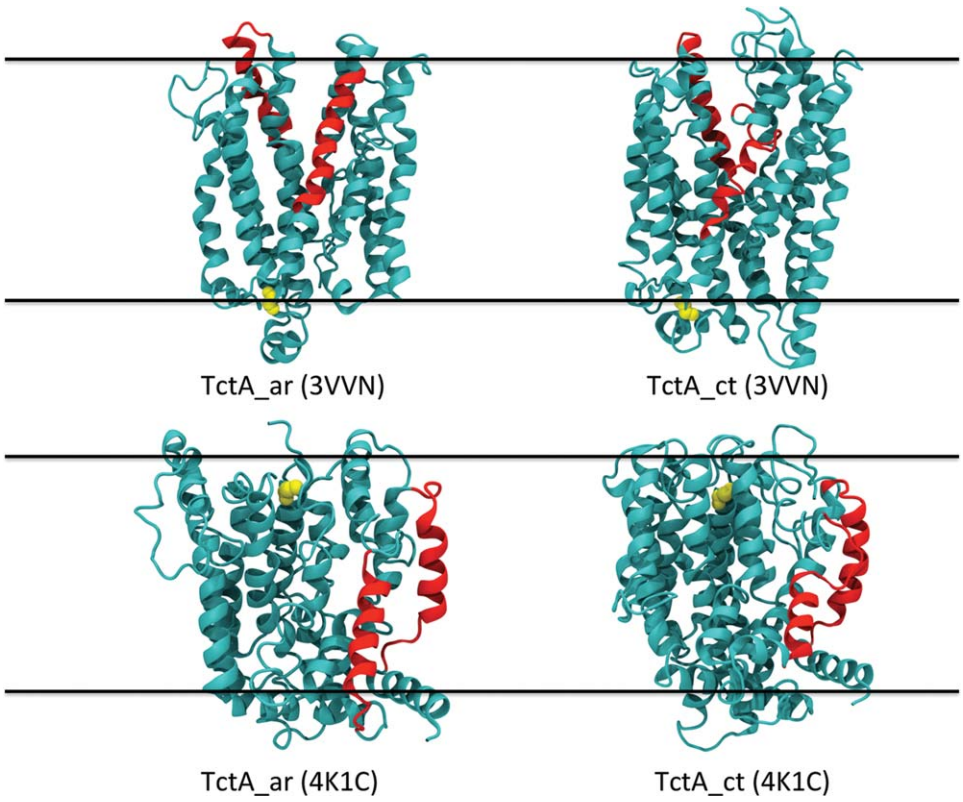

Figure 3
Model structures for TctA_ar and TctA_ct derived from 3VVN (top) and 4K1C (bottom). The structures are shown in cyan ribbons, highlighting in red ribbons the 20-amino acid repeat. G107 is shown as yellow spheres. The horizontal lines mark the approximate location of a lipid bilayer. The extracellular/vacuolar space lies at the top of the figure.

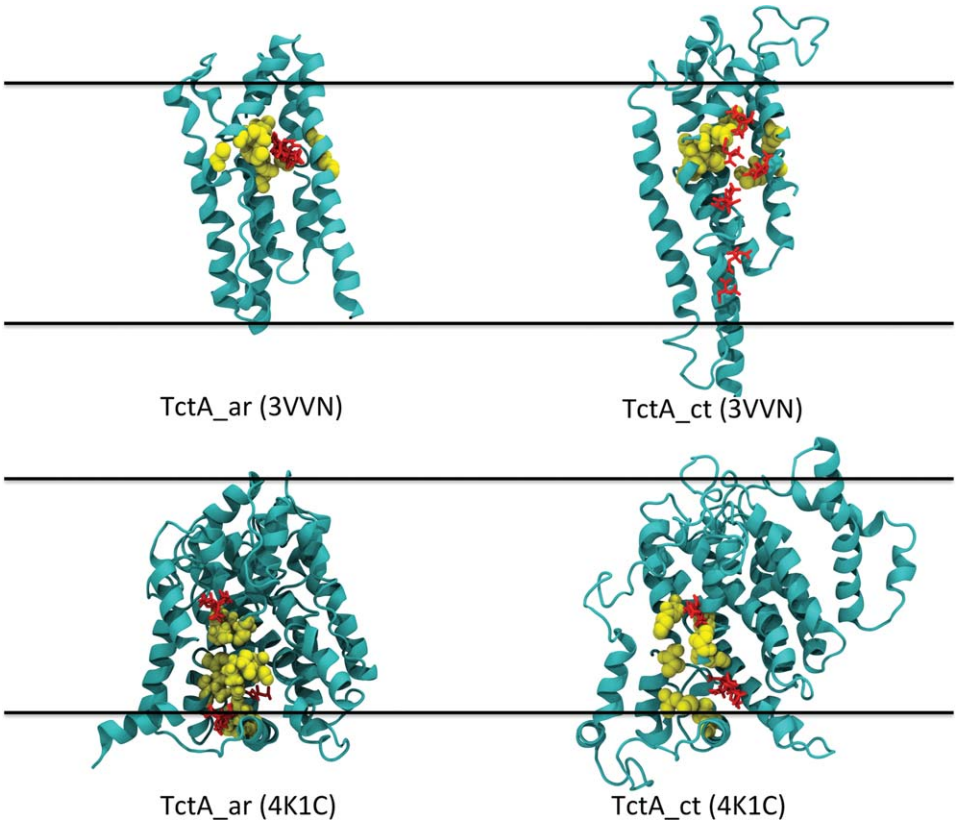

Figure 4

Citrate binding sites for TctA_ar and TctA_ct derived from 3VVN (top) and 4K1C (bottom). The models for each transporter, based on the same template, were superimposed and are presented in the same orientation, which is approximately rotated 90° with respect to that shown in Figure 3. For clarity, roughly one of the repeats has been cut away to show the citrate binding sites. The transporters (cyan ribbons) are immersed in a bilayer (horizontal lines), with the bound citrate molecules (red sticks), and the substrate binding residues (yellow spheres) translated from the template. The extracellular/vacuolar space lies at the top of the figure. The acidic helix in 4K1C-derived structures points toward the viewer and lies at the center bottom of the structures. The extra helical hairpin at the C-terminus of TctA_ct lies to the right of the main structure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Docking of citrate in the TctA models

Many regions in the models display loops inserted in helical regions, and we did not attempt to fix those, as this lies beyond the scope of this work. By necessity, the models are still at a low-resolution phase, and require experimental input to be improved. Nonetheless, the models can be useful to inquire about the function of the proteins. Given that we are proposing these structures to be citrate transporters, we tested their ability to bind citrate, using AUTODOCK/VINA[23] with the standard default parameters, allowing for flexibility in citrate. We restricted the docking area to the surfaces not in contact with the DMPC bilayer and analyzed the default 10 poses returned by the program. The results are shown in Figure 4.

To guide the analysis, we identified equivalent binding site residues from the templates and their homologs in both proteins, by performing structural alignments with STAMP[26] in VMD.[27] 3VVN is a MATE transporter from *Pyrococcus furiosus*, and it uses protons as cotransported solutes. Given that TctAs are proposed to use sodium in cotransport, we used the NorM MATE homolog from *Vibrio cholerae* (3MKT and 3MKU[28] to identify the sodium-binding site. This architecture presents a large substrate-binding site; crystal structures of other NorM homologs have been crystallized with cationic drugs and peptides, binding at different depths and locations.[24,28,29]. Table III summarizes the sodium binding pocket residues in *V. cholerae* NorM (3MKU[28] and the structurally equivalent residues in TctA_ar and TctA_ct. The counterion for sodium is D371, and it is conserved in our models at the appropriate location. All of these residues lie in the C-terminal half of the transporter, where sequence variability is greater. TctA_ar binds citrate at only one place, adjacent to the putative sodium-binding site (Fig. 4, upper left), whereas TctA_ct binds citrate at that location and at various others (Fig. 4, upper right). As these binding sites are found along a line perpendicular to the membrane plane, and along the dyad axis of the transporter, we find them suggestive of a possible translocation pathway, which can also be tested experimentally.

Table III
Ligand Binding Residues in Templates and Model Transporters

3MKU	TctA_ar	TctA_ct	4K1C	TctA_ar	TctA_ct
E255	A259	A263	G102	G86	V82
F259	G263	L274	V105	S89	P85
S285	R291	Q315	E106	S90	G86
F288	A294	F318	G298	A328	T323
Y367	V358	V413	A301	R331	P327
D371	D362	D415	E302	D332	N329
Y398	G388	L445	E83	I67	M63
F429	L414	R470	E230	R235	V213
L433	G417	A474	D234	E239	G217

4K1C is a calcium/proton antiporter that belongs to a family that can also use sodium as cotransported ion.[25,30–32]. The available crystal structures indicate at least two possible binding sites for calcium, which in our case would correspond to citrate. These two sites, also listed in Table III, are occupied, as seen in Figure 4 by the close proximity of citrate molecules to the extrapolated binding site residues. The same architecture has been shown to transport organic and inorganic ions, as happens in the APC family,[33] so this proposal is reasonable. At each of the binding sites there is at least one positive side chain, which binds citrate directly.

From the data presented above, it is currently impossible to choose one template over the other. Table III constitutes a set of direct functional hypotheses, susceptible to exploration with site-directed mutagenesis. As the critical residues are not shared between the two architectures, it may be possible to validate or discard the models in a straightforward approach.

CONCLUSIONS

The modeling attempt presented here rationalizes the 20-amino acid repeats that are a hallmark of TctA transporters, and also suggests why the G109S mutation in TctA of *S. enterica* abrogates function. Furthermore, we suggest two disjoint sets of putative citrate and/or sodium-binding residues, which are presented as concrete functional hypotheses to be tested by mutagenesis in future studies. After experimental validation of any of the two proposed architectures, this would be a relevant starting point for understanding substrate specificity and the transport mechanism for this understudied transporter family.

REFERENCES

1. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. PNAS 2002;99:10494–10499.
2. Daniel R. The soil metagenome: a rich resource for the discovery of novel natural products. Curr Opin Biotech 2004;15:199–204.
3. Schloss R, Patrick D, Handelsman J. Biotechnological prospects from metagenomics. Curr Opin Biotech 2003;14:303–310.
4. Simon C, Daniel R. Metagenomic analyses: past and future trends. Appl Environ Microbiol 2011;77:1153–1161.
5. Winnen B, Hvorup RN, Saier MH. The tripartite tricarboxylate transporter (TTT) family. Res Microbiol 2003;154:457–465.
6. Hosaka M, Kamimura N, Toribami S, Mori K, Fukuda M, Masai E. A novel tripartite aromatic acids transporter is essential for terephthalate uptake in Comamonas sp. strain E6. Appl Environ Microbiol 2013;79:6148–6155.
7. Widenhorn KA, Somers JM, Kay WW. Expression of the divergent tricarboxylate transport operon (tctI) of Salmonella typhimurium. J Bacteriol 1988;170:3223–3227.
8. Ricci V, Loman N, Pallen M, Ivens A, Fookes M, Langridge GC, Wain J, Piddock LJ. The TCA cycle is not required for selection or survival of multidrug-resistant Salmonella. J Antimicrob Chemother 2012;67:589–599.
9. Ma YF, Zhang Y, Zhang JY Chen DW, Zhu Y, Sheng H, Wang SY, Jiang CY, Zhao GP, Liu SJ. The complete genome of Comamonas testosteroni reveals its genetic adaptations to changing environments. Appl Environ Microbiol 2009;75:6812–6819.
10. Brocker M, Schaffer S, Mack C, Bott M. Citrate utilization by Corynebacterium glutamicum is controlled by the CitAB two-component system through positive regulation of the citrate transport genes citH and tctCBA. J Bacteriol 2009;191:3869–3880.
11. Polen T, Schluesener D, Poetsch A, Bott M, Wendisch VF. Characterization of citrate utilization in Corynebacterium glutamicum by transcriptome and proteome analysis. FEMS Microbiol Lett 2007;73:109–119.
12. Antoine R, Jacob-Dubuisson F, Drobecq H, Willery E, Lesjean S, Locht C. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in Bordetella. J Bacteriol 2003;185:1470–1474.
13. Magalhaes JV. How a microbial drug transporter became essential for crop cultivation on acid soils: aluminium tolerance conferred by the multidrug and toxic compound extrusion (MATE) family. Ann Bot 2012;106:199–203.
14. Saier MH, Yen MR, Noto K, Tamang DG, Elkan C. The transporter classification database: recent advances. Nucl Acids Res 2009;3:274–278.
15. Shi Y. Common folds and transport mechanisms of secondary active transporters. Ann Rev Bioph 2013;42:51–72.
16. Kebbel F, Kurz M, Arheit M, Grutter MG, Stahlberg H. Structure and substrate-induced conformational changes of the secondary citrate/sodium symporter CitS revealed by electron crystallography. Structure 2013;21:1243–1250.
17. Berman HM, Westbrook J, Feng Z. The protein data bank. Nucl Acid Res 2000;28:235–242.
18. Felske A, Engelen B, Nubel U, Backhaus. Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. Appl Environ Microbial 1996;62:4162–4167.
19. Bomeman J, Jack R. PCR primers that amplify fungul rRNA genes from environmental samples. Appl Environ Microbial 2000;66:4356–4360.
20. Cho CM, Mulchandani A, Chen W. Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. Appl Environ Microbiol 2002;68:2026–2030.
21. Zhang Y, Skolnick J. Segment assembly, structure alignment and iterative simulation in protein structure prediction. BMC Biol 2013;11:44.
22. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
23. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, #efficient loptimization, and multithreading. J Comput Chem 2010;31:455–461.
24. Tanaka Y, Hipolito CJ, Maturana AD, Ito K, Kuroda T, Higuchi T, Katoh T, Kato HE, Hattori M, Kumazaki K, Tsukazaki T, Ishitani R, Suga H, Nureki O. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. Nature 2013;496:247–251.
25. Waight AB, Pedersen BP, Schlessinger A, Bonomi M, Chau BH, Roe-Zurz Z, Risenmay AJ, Sali A, Stroud RM. Structural basis for

alternating access of a eukaryotic calcium/proton exchanger. Nature 2013;99:107–110.
26. Rusell RB, Barton GH. Multiple protein sequence alignment from tertiary structure comparison. Prot Struct Funct Genet 1992;67:309–323.
27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996: 33–38.
28. He X, Szewczyk P, Karyakin A, Evin M, Hong WX, Zhang Q, Chang G. Structure of a cation-bound multidrug and toxic compound extrusion transporter. Nature 2010;467:991–994.
29. Lu M, Symersky J, Radchenko M, Koide A, Guo Y, Nie R, Koide S. Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. PNAS 2013;110:2099–2104.
30. Abramson J, Paz A, Philipson KD. It's all in the symmetry. Science 2012;335:669–670.
31. Liao J, Li H, Zeng W, Cao H, Cui Z. Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger. Science 2012;335:686–690.
32. Nishizawa T, Kita S, Maturana AD, Furaya N, Hirata K, Kasuya G, Ogasawara S, Dohmae N, Iwamoto T, Ishitani R, Nureki O. Structural basis for the counter-transport mechanism of a H+/Ca2+ exchanger. Science 2013;341:168–172.
33. Wong FH, Chen JS, Reddy V, Day JL, Shlykov MA, Wakabayashi ST, Saier ST. The amino acid-polyamine-organocation superfamily. J Mol Microbiol Biotech 2012;22:105–113.