Evolutionary relationship between 5+5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily

Permalink
https://escholarship.org/uc/item/8ws8m6j0

Journal
Proteins: Structure, Function and Bioinformatics, 82(2)

ISSN
0887-3585

Authors
Västermark, A
Saier, MH

Publication Date
2014-02-01

DOI
10.1002/prot.24401

Peer reviewed
Evolutionary relationship between 5+5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily

Åke Västermark and Milton H. Saier Jr*

Department of Molecular Biology, University of California at San Diego, La Jolla, California 92093-0116

ABSTRACT

Evidence has been presented that 5+5 TMS and 7+7 TMS inverted repeat fold transporters are members of a single superfamily named the Amino acid-Polyamine-organoCation (APC) superfamily. However, the evolutionary relationship between the 5+5 and the 7+7 topological types has not been established. We have identified a common fold, consisting of a spiny membrane helix/sheet, followed by a U-like structure and a V-like structure that is recurrent between domain duplicated units of 5+5 and 7+7 inverted repeat folds. This fold is found in the following protein structures: AdiC, ApcT, LeuT, Mhp1, BetP, CaiT, and SglT (all 5+5 TMS repeats), as well as UraA and SulP (7+7 TMS repeats). AdiC, LeuT and Mhp1 have two extra TMSs after the second duplicated domain, SglT has four extra C-terminal TMSs, and BetP has two extra TMSs before the first duplicated domain. UraA and SulP on the other hand have two extra TMSs at the N-terminal of each duplicated domain unit. These observations imply that multiple hairpin and domain duplication events occurred during the evolution of the APC superfamily. We suggest that the five TMS architecture was primordial and that families gained two TMSs on either side of this basic structure via dissimilar hairpin duplications either before or after intragenic duplication. Evidence for homology between TMSs 1–2 of AdiC and TMSs 1–2 and 3–4 of UraA suggests that the 7+7 topology arose via an internal duplication of the N-terminal hairpin loop within the five TMS repeat unit followed by duplication of the 7 TMS domain.

© 2013 Wiley Periodicals, Inc.

Key words: AdiC; UraA; amino acid-polyamine-organocation; APC; superfamily; hairpin; duplication.

INTRODUCTION

Hydrophilic molecules such as ions, sugars, and amino acids are transported across biological membranes via proteinaceous transport systems. Solute transporters are classified into channels, primary and secondary active transporters and group translocators. While primary active transporters use ATP or another primary source of energy such as light, electron flow or a metabolic conversion reaction to drive transport, secondary transporters use electrochemical gradients derived by the pumping activities of primary active transporters.1,2 Further, while primary active transporters are often multicomponent systems, secondary carriers most frequently consist of a single polypeptide chain.

Most secondary carriers exhibit pseudo-twofold symmetry, consisting of two repeat units of four to seven transmembrane α-helical segments (TMSs). When these proteins have four to seven TMSs, the functional unit is usually a homodimer or heterodimer, but when the protein has 8–14 TMSs, the fundamental unit is usually a monomer with two internal repeat units. These observations suggest that a basic characteristic of most secondary carriers is the presence of two symmetrical or pseudosymmetrical domains both required for function. These conclusions are consistent with high resolution data as well as biochemical mechanistic information.3

The amino acid-polyamine-organocation (APC) superfamily represents the second largest recognized superfamily of secondary carriers,4 being second in size only to the major facilitator superfamily (MFS).5 Protein members of the former superfamily usually have repeat units of five TMS while those of the latter always have repeat...
units of six TMS. While the topologies differ, an evolutionary relationship between the APC superfamily and the MFS has been proposed.5

The APC superfamily was first defined and described by Jack et al. in 2000.6 At that time, it included 10 subfamilies as members of the APC family (TC# 2.A.3), with weaker relationships to the Amino Acid/Auxin Permease (AAAP) family (TC# 2.A.18) and the Hydroxy/Aromatic Amino Acid Permease (HAAAP) family (TC# 2.A.42). These investigators reported that most proteins in the APC family display 12 TMSs, except the spore germination proteins which had 10 TMSs, and eukaryotic members of the CAT family which had 14 TMSs. Most AAAP and HAAAP family members had 11 TMSs, having lost TMS12 at their C-terminal end relative to 12 TMS APC family members.

In 2012, Wong et al. expanded the APC superfamily to its present state, including 11 families, tabulated in Table 1.4 The Alanine or Glycine:Cation Symporter (AGCS) family (TC# 2.A.25) and the Cation-Chloride Cotransporter (CCC) family (TC# 2.A.30) had been identified as members of the APC superfamily prior to the work of Wong et al., 2012 (unpublished results). The new 5+5 topology additions were: The Betaine/Carnitine/Choline Transporter (BCCT) family (TC# 2.A.15), the Solute:Sodium Symporter (SSS) family (TC# 2.A.21), the Neurotransmitter:Sodium Symporter (NSS) family (TC# 2.A.22) and the Nucleobase:Cation Symporter-1 (NCS1) family (TC# 2.A.39), While the five previously recognized families mainly transported amino acids and amino acid derivatives, the new members, which proved to be more sequence divergent, displayed a more varied substrate repertoire. Two topologically distinct families were also included: the Sulfate Permease (SulP) family (TC# 2.A.53), and the Nucleobase:Cation Symporter-2 (NCS2) family (TC# 2.A.40), both of which are believed to have a 7+7 TMS repeat unit topology.7

Since the first publication of the APC superfamily,6 high resolution three-dimensional X-ray structures of several members of the currently recognized APC superfamily have been published. These include: ApcT,8 LeuT,9 Mhp1,10 BetP,11 CaiT,12 and SglT13 representing constituent families displaying a common five TMS repeat unit fold with different constellations of “extra” TMSs, and UraA with a less common seven TMS repeat. In this report, we focus on AdiC and UraA, to understand the differences in their topologies. AdiC is an arginine:agmatine antiporter in the APC family while UraA is a uracil permease of the NCS2 family. The 2.8 Å resolution crystal structure of UraA7 revealed the 7 TMS inverted repeat with a “spiny” secondary structural element located near the substrate translocation site in equivalent positions of each repeat unit. The 3 Å resolution crystal structure of AdiC14 revealed structural modifications in TMS6 in the 5+5 TMS inverted repeat structure.

We here report a shared feature, the spiny secondary structural element found in UraA and the symmetry related TMSs 1 and 6 in AdiC. We establish a nomenclature to facilitate the identification, description, and comprehension of a common fold shared by these structures and compared our approaches and results with those of others. AdiC has two extra TMSs at the C-terminus, whereas UraA has two extra TMSs at the N-terminus of each domain duplicated unit. Other members of the APC superfamily may have two extra N-terminal TMSs (BCCT; TC# 2.A.15) or four extra C-terminal TMSs (SSS; TC# 2.A.21). We postulate that multiple hairpin and domain duplication events were responsible for the variations of extra TMSs, and we show that the basic fold is recurrent between all members of the 5+5 and 7+7 TMS architectures.

METHODS

RMSD measurements

The RMSD values reported here were obtained using Chimera 1.7, using the RMSD map function, part of the “morph” function. The values were confirmed using SuperPose 1.0 (http://wishart.biology.ualberta.ca/SuperPose/). They always refer to the RMSD, in units of Å, of all α-carbons present in the comparison. In some cases, when such values are tabulated in PDB under the 3D similarity tab, RMSD values for domain comparisons using only a subset of well-aligned residues are given as well.

Topology prediction, modeling

For topology prediction, the best current methods, TOPCONS and SPOCTOPUS were used.15 When results from TOPCONS are presented, we present an aggregation of any positive predictions, rather than the consensus, so that if even only one method found a TMS, it is considered as a positive finding. Homology modeling was done using MODELLER 9.11, 2012/08/29, r8834, using default parameters. We used the realign with MAC option.

Secondary structure matching

SSM (secondary structure matching)16 was used as implemented in the software PDBe Fold v2.55 to compare the model of rat prestin with vSGLT (chain A).

HMM comparisons

HHsuite (hhsuite-2.0.16) was used for HMM:HMM comparisons, and HHMAKE (HHmake version 2.0.15) was used to train models representing the first and second halves of the APC and SulP families. We then used the -M 50 flag, for “FASTA columns with fewer than X%
Table I
Proteins and Families Within the APC Superfamily Considered in This Report.

<table>
<thead>
<tr>
<th>Family</th>
<th>Name</th>
<th>Abbreviation</th>
<th>SLC#</th>
<th>Fold</th>
<th>Topology</th>
<th>Protein: TC#</th>
<th>Name</th>
<th>Abbreviation</th>
<th>PDB#</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.A.53</td>
<td>Sulfate Permease</td>
<td>SulP</td>
<td>SLC26</td>
<td>7+7</td>
<td></td>
<td>2.A.53.1.1</td>
<td>sulfate permease</td>
<td>SuIP</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>2.A.42</td>
<td>Hydroxy/Aromatic AminoAcid Permease</td>
<td>HAAAP</td>
<td>b</td>
<td>2+5+2+5</td>
<td>TOPCONS prediction for 2.A.42.1.1: 11 TMS</td>
<td>2.A.42.1.1</td>
<td>tyrosine permease</td>
<td>TyrP</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>2.A.40</td>
<td>Nucleobase: Cation Symporter-2</td>
<td>NCS2</td>
<td>SLC23</td>
<td>7+7</td>
<td></td>
<td>2.A.40.1.1</td>
<td>uracil permease</td>
<td>UraA</td>
<td>3QE7</td>
<td>2.78 Å</td>
</tr>
<tr>
<td>2.A.30</td>
<td>Nucleobase: Cation Symporter-1</td>
<td>NCS1</td>
<td>b</td>
<td>5+5</td>
<td></td>
<td>2.A.30.1.1</td>
<td>benzyl-hydantoin: cation symporter-1</td>
<td>Mhp1</td>
<td>2JLN</td>
<td>2.85 Å</td>
</tr>
<tr>
<td>2.A.30</td>
<td>Cation-Chloride Cotransporter</td>
<td>CCC</td>
<td>SLC12</td>
<td></td>
<td>prediction for 2.A.30.1.1: 13 TMS</td>
<td>2.A.30.1.1</td>
<td>NaCl/KCl symporter</td>
<td>NKCC2</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>2.A.25</td>
<td>Alanine or Glycine: Cation Symporter</td>
<td>AGCS</td>
<td>b</td>
<td>2+5+2+5</td>
<td>TOPCONS prediction for 2.A.25.1.1: 11 TMS</td>
<td>2.A.25.1.1</td>
<td>Ala/Gly:Na⁺ symporter</td>
<td>DagA</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>2.A.22</td>
<td>Neuro-transmitter: Sodium Symporter</td>
<td>NSS</td>
<td>SLC6</td>
<td>5+5</td>
<td></td>
<td>2.A.22.4.2</td>
<td>amino acid (leucine):2 Na⁺ symporter</td>
<td>LeuT</td>
<td>2Q72</td>
<td>1.70 Å</td>
</tr>
<tr>
<td>2.A.18</td>
<td>Amino Acid/Auxin Permease</td>
<td>AAAP</td>
<td>SLC32, 36, 38</td>
<td>5+5+2+2</td>
<td>prediction for 2.A.18.1.1: 12 TMS</td>
<td>2.A.18.1.1</td>
<td>galactose: Na⁺ symporter</td>
<td>SgIT</td>
<td>2X02, 3DH4</td>
<td>2.70 Å</td>
</tr>
<tr>
<td>2.A.15</td>
<td>Betaine/Carnitine/ Choline Transporter</td>
<td>BCCT</td>
<td>b</td>
<td>2+5+5</td>
<td></td>
<td>2.A.15.1.10, 2.A.15.2.1</td>
<td>glycine betaine transporter, carnitine:γ- butyrobetaine antiporter</td>
<td>BetP, CaiT</td>
<td>4AIN, 2WSW</td>
<td>3.10 Å, 2.29 Å</td>
</tr>
<tr>
<td>2.A.3a</td>
<td>Amino Acid-Polyamine-Organocation</td>
<td>APC</td>
<td>SLC7</td>
<td>5+5</td>
<td></td>
<td>2.A.3.2.5</td>
<td>arginine: agmatine antiporter</td>
<td>AdiC</td>
<td>3L1L</td>
<td>3 A</td>
</tr>
</tbody>
</table>

*For the APC family (2.A.3), the structures of GadC and ApcT are also available.

*No SLC number has been assigned for these families since they are not represented in humans.

*Not applicable. An x-ray structure is not available.
gaps” that are match states, and HHsearch (2.0.15), to compare the domain halves of the APC and SulP families using a feature file from the TMHMM Server 2.0 and Jalview 2.8. For each comparison, we recorded the percentage probability of homology. Using the same programs, we trained HMMs on MAFFT v7.023b alignments of SulP and APC (as well as the other 5 + 5 families) and compared them against each other.

AlignMe

Using AlignMe17 (http://www.bioinfo.mpg.de/AlignMe/), LeuT-like folds were compared employing alignments that weigh in hydropathy profiles. Because the methodology is based on the Needleman-Wunch algorithm, there may be a bias towards aligning the ends of sequences if they have different lengths. We took AdiC (TMSs 1–5) and UraA (TMSs 1–7) and used AlignMe with the PST settings for distantly related proteins. The program generates secondary structure predictions using PSIPRED 3.2, PSSMs, a transmembrane topology using OCTUPUS and the alignment. This exercise was repeated for AdiC (TMSs 6–10) and UraA (TMSs 8–14).

InterCompare

InterCompare18 was used to compare alignments of the APC and NCS2 families, comparing aligned positions 300–400 (corresponding to TMSs 1–2) with 300–400 (corresponding to TMSs 3–4), respectively.

RESULTS

Relationship between 5 + 5 and 7 + 7 TMS inverted repeat folds

We downloaded the uracil permease (UraA; 3QE7) structure of the NCS2 family and the arginine/agmatine antiporter (AdiC; 3L1L) structure of the APC family from the Protein Databank (PDB). UraA, with 14 transmembrane α-helices (TMSs), can be subdivided into two clear halves (from residues #1–200 and 201–410). These halves are superimposable (global α-C RMSD 9.04 Å), showing that these two halves share the same fold, containing seven TMSs each, with no “extra” TMSs. The third TMS is discontinuous in both halves, appearing incomplete. In TOPCONS,15 for the first half, there is a clear prediction of seven TMSs. For the second half, a weak prediction for the 2nd TMS was obtained, and there appeared to be only six TMSs due to a fused prediction of TMSs 9 and 10.

The structure of AdiC can also be divided in two halves: #7–175 and 176–380 with a 5 + 5 TMS basic structure and two extra C-terminal TMSs (381–440), possibly due to duplication of a terminal hairpin structure. TMSs 1 and 6 are discontinuous and annotated as incomplete in the PDB file, even though they span the membrane, physically similar to TMS3 in UraA. We observed a small helix located between TMSs 2 and 3, similar to a helix located between TMSs 4 and 5 in UraA. We introduced a simple nomenclature, where the symbol “I” represents an incomplete helix (or β sheet in the case of UraA19) that spans the membrane; the symbol “V” represents two helices at an angle to each other with a short connecting loop; and the symbol “U” represents two parallel TM helices separated by an extramembranous α-helix, holding them apart. The I corresponds to the first half of the “bundle” motif, the U corresponds to second half of the “bundle” and the first half of the “hash,” and the V corresponds to the second half of the “hash” and the “arm” motif, described by Perez and Ziegler (2013).19 We also use the symbol V’ (V prime) to indicate a pair of N-terminal helices, with features similar to V but probably not derived from V. We also use a short dash to indicate space between two secondary structural elements that are not part of the same domain duplicated unit. Using this makeshift nomenclature, we could describe the domain duplicated structure of UraA as two consecutive units of V’UUV: V’ - I – U – V || V’ - I – U – V. Similarly, AdiC would be I – U – V || I – U – V - V, having an extra V after the domain duplicated unit. These features are conserved in all of the proteins for which three-dimensional structures are available.

When we excised the first two TMSs of UraA, focusing on residues #65–200 (TMSs 3–7), and compared this with residues #7–175 of AdiC (TMSs 1–5), we recognized the same fold of I – U – V. The RMSD between 24 atom pairs was 0.93 Å, while the listed UraA-AdiC domain distance in PDB using the jFATCAT-rigid algorithm was 6.90 Å, using selected atoms. The 5th and corresponding 10th TMSs pointed in opposite directions relative to each other from the rest of the structure, but otherwise, they showed the same fold. What speaks for this interpretation is not only the overall RMSD, but also the matching up of irregularities, including firstly the spiny TMS1/β1-sheet, and secondly, the “U,” including the small non-membrane helix that separates the two parts of the “U.” There are no TMS irregularities that do not match up. As a control, we compared region I – U – V || V” in UraA with the end of AdiC, I – U – V - V. UraA has a long membrane-parallel arm connecting residues #202 and 224. A comparison of #65–280 in UraA to the second half of AdiC (untruncated, with the two terminal extra TMSs present) resulted in a poor superposition (15.18 Å), rejecting this latter possibility.

AdiC, residues #7–175, could be compared with UraA, residues #65–200. This scored an RMSD of 9.46 Å. AdiC residues #176–380 and UraA residues #65–200, resulted in a similar matching, displaying an RMSD of 9.37 Å. On the other hand, the first two TMSs from AdiC (residues #19–67) and UraA (residues #5–63), a V-formed pair of helices, could not be superimposed; they pointed in different directions, casting doubt on the possibility
that TMSs 1 and 2 are equivalent between UraA and AdiC. However, UraA residues #113–200 could be superimposed on AdiC residues #68–170, that is, the last three TMSs in the seven and five TMS units including half of the “U” and all of the “V.” This is consistent with a domain duplication model. Furthermore, superimposing the “U” in AdiC (TMSs 2–3; residues #41–113) onto TMSs 4–7 in UraA (U – V; residues #40–153), resulted in matching between the U-like structures (i.e., TMS 4–5 in UraA).

The RMSD between the first and second halves of AdiC, removing the terminal two helices, resulted in an overall RMSD for α-carbons of 7.67 Å. We hypothesize that residues #65–200 of UraA and residues #7–175 of AdiC represent the domain duplicated unit of I – U – V in both structures. This comparison resulted in an overall α-carbon RMSD of 9.03 Å (the listed distance in PDB is 6.90 Å) for five TMSs, indicating a common fold. This possibility was confirmed by corresponding locations of TMS irregularities between the structures, including an incomplete helical prediction/β segment for the first transmembrane secondary structural element, and the equivalent positioning of a small helix located between TMSs 3 and 4. The interpretation of these comparisons between UraA and AdiC are illustrated in Figures 1 and 2 and Supporting Information Figure S1.

Structures of other 5+5 architecture inverted repeat folds

To compare AdiC to other 5+5 fold members in the APC superfamily, members, we considered the outward-facing conformation of LeuT (2Q72) of the NSS family.21 Residue #230 was used as the arbitrary point subdividing the structure into the domain duplicated halves. The last two helices (the “extras”) were removed, taking the sequence for the second domain up to #430 (#230–430). The basic I – U – V fold found in AdiC was present. TMS1 was divided into two halves (TMS1a and TMS1b), forming the “spiny” helix. However, the pairwise RMSD between the halves was ~15 Å. What affected the RMSD negatively was the presence of a small, non-TMS helix in a loop region between TMSs 3 and 4 which was only present in the first of the two domain duplicated units. In the second five TMS repeat, the connector between the two arms of the ‘U’ was split into two small helices. Furthermore, the arm connecting the two domain duplicated units contained a small helix that was not found in AdiC. These features added up to a higher RMSD since these anomalies did not match between the two copies. As the length of the new helical sequence was smaller in the second duplicated domain, it scored better against residues #7–175 of AdiC (displaying an RMSD of 9.07 Å), comparable to the score obtained between UraA and AdiC.

The outward facing conformation of Mhp1 (2JLN) of the NCS1 family10 contains an intradomain linker peptide (residues #191–209). The structure is AdiC-like, starting with the I – U – V fold, followed by the peptide linker, a second set of I – U – V and a terminal V. To remove the two helices, we truncated the sequence at residue #385. A peptide tail at the N-terminus was removed as well, and hence domain #1 starts at residue 26. The listed distance of Mhp1 and UraA in PDB using the jFATCAT-rigid algorithm is 7.78 Å. The two units of the I – U – V fold produced a self-superimposition, scoring an RMSD of 6.34 Å, supporting the conclusion of the five TMS duplication.

The first half of Mhp1 superimposed on the first half of AdiC both contain an identical broken helix in TMS1, and U- and V-like structures of the same length and shape. For the second halves of both proteins superimposed on each other, two specific features are conserved, showing the close similarity: (1) the helical structure of...
the second leg of the U is briefly interrupted in the equivalent position, and (2) the V-like structures have a tiny insert between the arms, making it slightly U-like (see Table II for details).

For the intermediate conformation of BetP (4AIN) of the BCCT family,

we used chain A. Sequences #57–273 and 274–546 were used, removing helical segments at the beginning and end of the full sequence which were outside the membrane. The sequence was truncated at position #490, assuming (based on the AdiC configuration) that the last two TMSs were the additional ones. However, we noted that in both domain duplicated units, the first helices were not “spiny” as expected from AdiC, that the U was unexpectedly wide, that its second leg appeared interrupted, and that the supposed V was U-like. These halves were superimposable on each other, giving an RMSD of 9.88 Å, but they were not superimposable on the two halves of AdiC. Assuming instead that the N-terminal V was the irregularity and removing it, in that case starting at position #130 and proceeding to position #324, the RMSD was 7.85 Å for the self-comparison with these portions. Beginning at position #324, the structure resembled the intra-domain arm seen in AdiC. Consequently, residues #130–324 (the I – U – V fold) against residues #7–175 of AdiC, and obtained an RMSD of 9.16 Å. The same structures have a listed RMSD of 3.96 in PDB using selected atoms. Thus, BetP can be described, using our nomenclature, as: V' - I – U – V - I – U – V. The fact that the two halves of BetP are more similar to each other than they are to AdiC suggests that duplication events occurred multiple times during the evolution of the superfamily (see Discussion).

In CaiT (2WSW) of the BCCT family, which is very similar to BetP, the intra-domain connector was identified, and the structure was divided into two halves, #1–280 and 281–500. For the first half, the elements V' - I – U – V were clearly visible. We removed #280–300, the domain linker, from the second half. Otherwise, the second half contained only the I – U – V fold. Like BetP, CaiT had two extra TMSs at its N-terminus. We removed residues #1–85 (the initial V) to get comparable structures and determined similarity for the comparison of residues #80–280 and 300–500, displaying an RMSD of 8.82 Å. In addition, CaiT residues #80–280 superimposed on residues #7–175 of AdiC displayed an RMSD of 9.37 Å.

For the inward-facing conformation of SglT (vSGLT; 2XQ2) of the SSS family, we first selected chain A. What was immediately striking was that there were missing
backbone atoms in several places and that there were more extra TMSs than seen for other 5+5 TMS-structured homologues. The structure started with a membrane helix, which appeared to be external to the domain duplicated unit (possibly external to the membrane), followed by a nick in the determined backbone trace. The “spiny I” started at position #50; after that, there was a normal “U,” but also a cut in the backbone trace in the V. Then, there was an uninterrupted domain linker peptide, and finally it continued directly with the spiny “I” in the next domain duplicated unit, ending at residue position #240. The second domain’s U contained a break, but the U-like structure was clearly recognizable followed by a V. We compared residues #50–212 (prior to the domain linker) and residues #240–450. After #450, two more “V” structures were observed that were not part of the domain duplicated unit. In total, we found 1 extra N-terminal helix, the I–U–V-I–U–V repeat fold, and a C-terminal VV. The self-superimposition of the two halves displayed an RMSD of 9.17 Å. The listed UraA-SglT distance in PDB using the jFATCAT-rigid algorithm is 7.83 Å.

Homology modeling of rat prestin

Rat prestin (SLC26A5; TC# 2.A.53.2.5), a member of the SulP family, which functions as an anion transporter in the human ear,25 has 744 residues and contains a region (#80–510) that can be homology-modeled on UraA’s 7+7 inverted repeat fold. SPOCTOPUS and TOPCONS predicted that it had only 13 TMSs, but a homology model using MODELLER, based on UraA as the template, generated a model of rat prestin which displayed an RMSD of 3.75 Å to UraA (henceforth referred to as our “SulP model”), where all 14 TMSs matched up. This model was subsequently compared with a closely related representative of the 5+5 architecture, vSGLT, using chain A of PDB entry 3DH4, not 2XQ2 as previously discussed. It appeared that the additional helix, prior to the “spiny” structural element, but partly located in an undetermined region, was in fact part of the spiny α-helical segment. The alignment of the SSS family revealed that the two extra C-terminal Vs are not well conserved between SSS family members, most of which display only a single pair of terminal Vs.

The SSM results showed that TMSs 1–2 of rat prestin matches TMSs 1–2 of vSGLT, and that the same is true for UraA and AdiC. SGLTs 3–4 in the 7+7 structure are unmatched. The final three TMSs in both the 5+5 and 7+7 topologies match. In the second domain duplicated unit, however, it is TMSs 3–4 of the 7+7 topology that correlate with TMSs 1–2 of the 5+5 topology, implying asymmetry between the domain duplicated units. These results are consistent with our suggestion that the first two hairpin structures in each of the seven TMS repeat units resulted from intragenic duplication of the N-
terminal hairpin structure in each 5+5 repeat and are homologous (see next sections and Discussion).

**HMM comparisons**

To resolve the differences between the superimposition- and SSM-based results which suggested that TMSS 1–2 in vSGLT were equivalent to TMSS+ 1–2 in SulP/rat prestin in the first domain duplicated unit, we used HMM:HMM comparisons of AdiC and UraA to determine if sequence similarity could be found to support the SSM-based model of the TMS correspondences between the homology model of rat prestin and vSGLT.

The first and second domain duplicated halves of the APC alignment scored a 0.8% chance of homology to each other, despite the common fold. However, the first and second domain duplicated halves of the SulP alignment scored a significant 41.7% chance of homology to each other, possibly due in part to the fact that SulP is a smaller and consequently less diverse family than APC.

For the comparison between UraA and AdiC, we obtained an even more significant 61% chance of homology between TMSS2 (in the 5+5 topology) and TMSS2 (in the 7+7 topology), the primary similarity between these divergent but related families. The matching segments were residues #46-81 in 2.A.40.3.2 of the SulP family, containing a single TMS, and the segment from #45–79 in 2.A.3.11.1 of the APC family, containing a single TMS.

One explanation is that a hairpin loop duplicated at the N-terminus, meaning that TMSSs 1–2 and 3–4 are homologous. It is possible that the sequence similarity between TMSSs 2 in AdiC and UraA (as detected by HMMs) is partly independent of the structures, but that it still biased the SSM result for the whole structure comparison.

**Use of AlignMe**

The AlignMe results are shown in Supporting Information Figure S2A. The first two TMSSs of UraA in both domain duplicated units are unmatched, indicating that they could be considered the "extra" TMSSs, contradicting the SSM results for the whole structure comparison.

**InterCompare results**

We compared TMSS 1–2 of the APC family (TC# 2.A.3) with either TMSS 1–2 or 3–4 of the NCS2 family (TC# 2.A.40). The best scoring comparison was TMSSs 1–2 of 2.A.3.4.5 (Tpo5p) and TMSSs 3–4 of 2.A.40.4.4 (YbbY), scoring 65 (far better than the comparison of TMSSs 1–2 in the two families, which gave 11% identity and a maximal comparison score of 3 standard deviations, S.D.). The exact aligning regions were: (302 GSIV-TMSs 1–2 in the two families, which gave 11% identity (YbbY), scoring 65 (far better than the comparison of TMSSs 1–2 of 2.A.40.7.6 (180–300) and 2.A.40.4.2 (300–400). We then used GASAT to make a binary comparison of these, giving a Z-score of 5 S.D. In summary, the trend we observed with the InterCompare scores, the %-identity in the binary comparisons, and the Z-scores, was that the highest ranked comparison is for TMSSs 1–2 in the APC family and TMSSs 3–4 in the NCS2 family.

Further analyses with InterCompare revealed that TMSSs 9–10 (residues 1120–1200) and TMSSs 11–12 (residues 1400–1480) of the APC family scored ~55 in InterCompare, the best pair of 2.A.3.9.2 and 2.A.3.7.3 displaying a Z-score of 6 S.D., constituting the strongest evidence for a hairpin loop duplication in the 2.A.3 family. TMSSs 9–10 and 11–12 (residues 510–600 and 605–670) of 2.A.21 scored up to 45 in InterCompare, TMSSs 1–2 and 3–4 in 2.A.15 (residues 60–120 and 140–220) scored only up to 40 in InterCompare, while TMSSs 8–9 and 10–11 (630–680 and 680–720) of 2.A.40 scored up to 39, unlikely to and in the cases we checked, not translating to high comparison scores.

**DISCUSSION**

Membrane proteins constitute <1% of the known high resolution 3D protein structures, but restraints imposed by the membrane compress the fold space for these proteins considerably.26 Contact prediction methods27 suggest that the fold space becomes small for membrane proteins having 8+ TMSSs. This consideration restricts the potential conformations a transmembrane protein can assume and renders structural properties of transmembrane transport proteins more similar to each other than is observed for cytoplasmic proteins which lack these restrictions.

Lu et al. (2011) expressed uncertainty about whether the UraA fold conforms to the LeuT fold. Four subsequent relevant structural studies since the publication of the UraA structure have appeared, but the authors of these papers did not specifically investigate the fold similarities and differences between UraA and 5+5 TMSS type APC family homologues.28–31 One of these studies,28 related UraA/XanQ to NCS1 (Mhp1) and other nucleobase transporters, including UPS/NBUT (TC# 2.A.7.19) and PUP/POP (TC# 2.A.7.14) which are members of the drug/metabolite transporter (DMT) superfamily,32 and AzgA of the NCS2 family (TC# 2.A.40.7). Of these, only the NCS2 family is represented in humans, and these proteins are not known to transport nucleobases. The shared substrate specificities of NCS1 and NCS2 family
members could reflect a common fold, but the suggestion of Frillingos (2012) that these proteins are related to the DMT superfamily has not been substantiated. Indeed, the 3D structures of known proteins in these two superfamilies are strikingly dissimilar (ZENK of SLC30A9).

A multitude of studies of LeuT-like transporters have supported the “rocking bundle” transport paradigm. The model is supported by the observation that different states of the structurally related transporters are similar. Two alternate states may not be required for transport.\(^3\)

The fold has been described with the terms “bundle” (TMSs 1–2 and 6–7), “hash” (TMSs 3–4 and 8–9) and “arm” (TMSs 5 and 10).\(^34\) Conserved sodium binding sites are important in some LeuT fold homologues.\(^20\)

However, whereas animal NCS2 transporters function by symporting sodium, the bacterial NCS2 transporters use protons instead of sodium.\(^7\) One paper suggests that the LeuT fold may have its evolutionary origins in DedA,\(^17\) but the evidence is weak.

Both LeuT and UraA share a domain duplicated inverted repeat structure, a feature shared with many transporters.\(^35\) Discontinuous (spiny) secondary structural elements that are important for transport have been described in several different types of transporters, including LeuT-like fold porters\(^19\) and UraA.\(^7\) It is known that TMSs 1 and 6 in the LeuT fold are important for function,\(^36\) and even though the fold is different, there could be an evolutionary relationship.\(^37\)

Wong et al. (2012) established that TMSs 1–8 in LeuT correspond to TMs 1–8 in AdiC/ApcT, in agreement with the structural information presented in this report. Furthermore, this same report established that TMSs 4–7 in BetP of the BCCT family are equivalent to TMSs 2–5 in Mhp1 of the NCS1 family. This is explicable in light of the structural findings that BetP/CaiT differ from AdiC/ApcT, Mhp1 and LeuT, in having a V' - I – U – V - I – U – V structure, rather than an I – U – V - I – U – V - V - V structure. UraA has a V' - I – U – V - I – U – V structure.

The data presented by Wong et al. suggested that the APC family is closely related to the NSS family, consistent with members of both families having the same TMS configuration and displaying a comparison score of 16.3 S.D. for their best pair of homologues. The NSS family, in turn, proved to be closely related to the SSS family, which, in turn, appeared to be fairly closely related to members of the 7+7 topology SulP family for which only preliminary structural information exists (http://dx.doi.org/10.1016/j.bpj.2012.11.648). The 5+5 topology proteins also proved to be related to NCS2 family members which have the 7+7 structure. A weaker relationship exists between NSS and NCS1, although both display the same arrangement of I – U – V - I – U – V - V (comparison score of 13.4 S.D.), vSGLT of the SSS family displays the I – U – V - I – U – V - V - V structure, while BetP of the BCCT family displays the V' - I – U – V - I – U – V arrangement, and may be more closely related to SSS family members than to NCS1 family members (Table I). Given that several members of the APC, NSS and NCS1 families display an I – U – V - I – U – V - V architecture, that at least two members of the BCCT family display a V' - I – U – V - I – U – V structure, and that at least one member of the NCS2 family displays a V' - I – U – V - V - V' - I – U – V architecture, an evolutionary model of TMS gains and losses can be proposed. In this model, the 5+5 architecture was the original pattern, and members of certain families have gained two TMSs on one or the other side of this basic structure (Fig. 3).

The first part of the analysis revealed that all of these homologues contain a common repeat unit, located inside the 7+7 architecture, comprising TMSs 3–7 and TMSs 10–14. Although other results complicated this picture, rat prestin of the SulP family was compared to its closest pseudo- 5+5 neighbor, vSGLT of the SSS family using secondary structure matching (SSM).\(^16\) The result was that TMSs 1–2 in the 7+7 topology matched TMSs 1–2 in the 5+5 topology, but for the second repeat unit, TMSs 10–11 (not TMSs 8–9) in the 7+7 topological architecture matched TMSs 6–7 in the 5+5 topology. For several reasons, our interpretation of this observation is that there may have been a hairpin loop duplication, giving rise to TMSs 1–2 and 3–4 in the 7+7 proteins. Results from AlignMe\(^17\) supported the conclusion that the first two TMSs in both copies of the
To objectively identify regions of strong sequence similarity between AdiC and UraA, we resorted to alignment comparisons using HMM:HMM. This approach identified a region of strong similarity between UraA and AdiC in the first pair of TMSs in both alignments, agreeing with the initial results obtained with SSM. Wong et al. established that the APC superfamily includes both SulP and UraA as well as all of the 5+5 transporters discussed here. To objectively test if a hairpin loop duplication had occurred, we used InterCompare, which revealed that the best comparison scores were obtained when TMSs 1 and 2 of 5+5 were compared with TMSs 3 and 4 of 7+7 (8 S.D.), although the value of 3 S.D. was obtained when TMSs 1 and 2 of 7+7 were compared with TMSs 1 and 2 of the 5+5 architecture. Although 8 S.D. is not sufficient to establish homology, this huge difference demonstrates a higher degree of similarity between TMSs 3 and 4 of 7+7 and TMSs 1 and 2 of 5+5. In fact, when TMSs 1 and 2 were compared with TMSs 3 and 4 of 7+7, a value of 5 S.D. was obtained, substantially better than 3 S.D. These observations can best be reconciled by assuming that TMSs 1 and 2 in the 7+7 topology arose by duplication of TMSs 1 and 2 in the 5+5 topology or TMSs 3–4 in the 7+7 topology (Fig. 4).

In summary, we believe that a five TMS unit as observed in APC family proteins duplicated to create a 10 TMS protein. Additional hairpin loop duplications occurred on either side of the 10 TMS unit, creating the topological diversity of that in AdiC, LeuT, Mhp1, and BetP (Table I, Fig. 1). In the case of SglT, a second hairpin loop duplication may have occurred at the C-terminus, creating 4 “extra” TMSs after the core unit. SulP and UraA arose from duplication of a 7 TMS (2+5 TMSs) unit. We propose that the N-terminal hairpin duplicated first, and that the resultant 7 TMS unit then duplicated, forming the 7+7 TMS topology. Compared to the possibly related Major Facilitator Superfamily (MFS), where most proteins have the 6+6 architecture, but some have a 6+2+6 architecture, there is similar evidence for an internal hairpin loop duplication. This intragenic duplication of 2 TMSs appears to have occurred multiple times in different families during the evolution of the MFS. We therefore propose that during the evolution of either the APC or the MFS superfamily, hairpin and entire 5, 6, or 7 TMS bundles duplicated multiple times. Further experimentation will be required to fully establish this possibility.

ACKNOWLEDGMENTS

We thank Roberto Battistutta, Mattia Sturlese, University of Padua, and Dominik Oliver, University of Marburg for valuable discussions and insightful considerations.

REFERENCES


