Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops

Lolkema, Julius; Sobczak, [No Value]; Slotboom, Dirk; Sobczak, Iwona

Published in:
Febs Journal

DOI:
10.1111/j.1742-4658.2005.04665.x

Publication date:
2005

the FEBS Journal

Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and *trans* re-entrant loops

Juke S. Lolkema[1], Iwona Sobczak[1] and Dirk-Jan Slotboom[2]

1 Molecular Microbiology, Biomolecular Sciences and Biotechnology Institute, University of Groningen, the Netherlands
2 Enzymology, Biomolecular Sciences and Biotechnology Institute, University of Groningen, the Netherlands

Keywords
domain structure; 2-hydroxycarboxylate transporter; inverted topology; pore-loop structure; secondary transporter

Correspondence
J. S. Lolkema, Molecular Microbiology, Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751NN Haren, the Netherlands
E-mail: j.s.lolkema@rug.nl

(Received 4 February 2004, revised 10 March 2005, accepted 16 March 2005)

doi:10.1111/j.1742-4658.2005.04665.x

The 2-hydroxycarboxylate transporter (2HCT) family of secondary transporters belongs to a much larger structural class of secondary transporters termed ST3 which contains about 2000 transporters in 32 families. The transporters of the 2HCT family are among the best studied in the class. Here we detect weak sequence similarity between the N- and C-terminal halves of the proteins using a sensitive method which uses a database containing the N- and C-terminal halves of all the sequences in ST3 and involves BLAST searches of each sequence in the database against the whole database. Unrelated families of secondary transporters of the same length and composition were used as controls. The sequence similarity involved major parts of the N- and C-terminal halves and not just a small stretch. The membrane topology of the homologous N- and C-terminal domains was deduced from the experimentally determined topology of the members of the 2HCT family. The domains consist of five transmembrane segments each and have opposite orientations in the membrane. The N terminus of the N-terminal domain is extracellular, while the N terminus of the C-terminal domain is cytoplasmic. The loops between the fourth and fifth transmembrane segment in each domain are well conserved throughout the class and contain a high fraction of residues with small side chains, Gly, Ala and Ser. Experimental work on the citrate transporter CitS in the 2HCT family indicates that the loops are re-entrant or pore loops. The re-entrant loops in the N- and C-terminal domains enter the membrane from opposite sides (*trans*-re-entrant loops). The combination of inverted membrane topology and *trans*-re-entrant loops represents a new fold for secondary transporters and resembles the structure of aquaporins and models proposed for Na^+/Ca^{2+} exchangers.

Crystal structures of the LacY and GlpT proteins support the alternating access model for the mechanism of substrate translocation catalysed by secondary transporters [1,2]. The proteins contain single binding sites for the substrate (and coion) that alternately are exposed to the two sides of the membrane during the catalytic cycle. The proteins consist of two homologous domains each containing six transmembrane segments (TMS) and with the same membrane topology. The structure of the domains is related by a pseudo twofold symmetry axis perpendicular to the plane of the membrane. The substrate is bound in a pore in between the two domains in the middle of the membrane.

Abbreviations
AAT, amino acid transporter; DAACS1, dicarboxylate/amino acid:cation symporter; 2HCT, 2-hydroxycarboxylate transporter; MFS, major facilitator superfamily; OPA, organophosphate:P_i antiporter; ST, secondary transporter; TMS, transmembrane segments.

Both LacY and GlpT belong to the major facilitator superfamily (MFS), which is the largest superfamily of secondary transporters that groups some 45 transporter families believed to have the same evolutionary origin [3]. The transporter classification system (TC system) developed by Saier and coworkers [3] lists numerous other families of secondary transporters not related to the MFS that are likely to have different folds which raises the question whether these transporters have also developed a different translocation mechanism. The multidrug transporter AcrB also consists of two homologous domains but the crystal structure shows a different helix packing when compared to the LacY and GlpT structures [1,4] which suggests convergent evolution towards a similar mechanism. On the other hand, the crystal structure of the glutamate transporter homologue Glt_{Ph} of *Pyrococcus horikoshi* reveals a completely different structure with eight TMS and no internal homology, but with pore-loops or re-entrant loops suggesting a different translocation mechanism [5].

We have classified families of secondary transporters into structural classes to discriminate between different 3D structures and possibly, different mechanisms [6]. Using a limited set of sequences, the discriminative power of family hydropathy profiles was used to identify four structural classes of secondary transporters with different folds, termed ST1, ST2, ST3, and ST4. Classes ST1 and ST4 correspond to the MFS transporters and glutamate transporters, respectively, mentioned above. More recently, all sequences in the NCBI protein database belonging to class ST3 were identified [7,8]. Structural class ST3 [1] contains over 2000 unique sequences distributed over 59 subfamilies that are in 32 families (April 2004). All sequences in the class are believed to share a common evolutionary origin and folding, even though sequence identity between many members cannot be detected anymore. Most of the transporters in ST3 transport organic and inorganic anions while a smaller fraction represents Na^+/H^+ antiporters. The bacterial 2-hydroxycarboxylate transporter (2HCT) family is one of the most extensively studied families in the class. The 2HCT family corresponds to [st326]2HCT in class ST3 and to the citrate:cation symporter family (2.A.24 CCS) in the transporter classification system.

Members of the 2HCT family transport substrates with a 2-hydroxycarboxylate motif such as citrate, malate and lactate [9–13]. Experimental studies of the Na^+-dependent citrate transporter CitS of *Klebsiella pneumoniae* have demonstrated that the proteins in the family traverse the membrane 11 times [14–16]. The N terminus resides in the cytoplasm; the C terminus is extracellular. There are two well-conserved regions in the members of the 2HCT family, the periplasmic loop between membrane-spanning segments V and VI and the region, termed Xa, comprising the cytoplasmic loop between TMS X and XI and part of TMS XI. Studies of the citrate/lactate and malate/lactate exchangers, CitP and MleP, respectively, found in lactic acid bacteria indicated that Xa is part of the substrate binding site [17]. A conserved Arg residue at the cytoplasm–TMS XI interface in region Xa interacts directly with one of the carboxylate groups on the substrates [18]. Recent studies of CitS of *K. pneunmoniae* and CimH of *Bacillus subtilis* showed that the cytoplasmic loop in region Xa forms a pore-loop or re-entrant loop [19,20] by similar criteria that were used to demonstrate the pore-loops in the glutamate transporters [21,22]. The pore-loops are believed to be crucial to the catalytic mechanism.

In this study we explore the domain structure of the transporters in class ST3 by sequence analysis [7] and combine the results with the experimental data available for the 2HCT family. We demonstrate that the proteins consist of two homologous domains and in combination with the experimental data of the 2HCT family a new structural model is proposed that shares similarities with the structures of aquaporins. The new model represents a completely different structure than observed for the glutamate transporter Glt_{Ph}, but the presence of two pore-loops entering the membrane embedded part of the proteins from opposite sides, suggests a similar mechanism.

Results

The domain database

Structural class ST3 in the MemGen database contains 59 subfamilies of secondary transporters divided over 32 families [7]. The sequences in the different families are only distantly related but are believed to share the same overall fold. A local domain database was constructed of the N- and C-terminal halves of the amino acid sequences of the proteins in structural class ST3 (see Experimental procedures for details). In addition, the N- and C-terminal halves of three unrelated control families of secondary transporters, one from structural classes ST1, ST2, and ST4 each [6], were included in the domain database: The Organophosphate:Pi Antiporter (OPA) from ST1, the Amino Acid Transporter (AAT) family from ST2, and the Dicarboxylate/Amino Acid:Cation Symporter (DAACS1) from ST4. In the Transporter Classification system [3], the OPA and AAT families are found in the MFS and the

Table 1. The domain database.

Structural class[a]	Subfamily[b]	Domains[c]	Sequence Identity (%)	Length distribution	
				N-domain	C-domain
ST1	OPA	2 × 40	16–60	185–279	200–246
ST2	AAT	2 × 43	19–60	212–297	212–278
ST3	32	2 × 533	17–61[d]	141–306	146–306
ST4	DAACS1	2 × 67	19–60	240–298	146–200

[a] MemGen structural classes as defined [6]. [b] OPA, Organophosphate:P_i Antiporter Family (TC 2.A.1.4); AAT, Amino Acid Transporter Family (TC 2.A.3.1); 32 out of the 59 subfamilies in ST3 (Supplementary Appendix S1; Table A); DAACS, dicarboxylate/amino acid:cation (Na^+ or H^+) symporter family (TC 2.A.23). [c] Each sequence was split in a N-terminal and C-terminal domain. The total number of domains is 1366. [d] Per subfamily.

Amino Acid/Polyamine/Choline superfamily, respectively. The DAACS1 family was recently defined in the MemGen classification system [8]. The sequences in the control families are unrelated to ST3 sequences but represent secondary transporters of similar length and amino acid composition. The domain database contained 1366 sequences, about 80% of which originate from ST3 families and 20% from the control families (Table 1). Highest sequence identity between the sequences in the database is about 60%.

Hits between the N- and C-terminal halves

Each of the sequences in the domain database was used as query in a BLAST search against all domains in the database (all against all). Figure 1 shows the analysis of the results for three groups of hits: hits between N-terminal halves of ST3 (NN), between C-terminal halves of ST3 (CC) and between an N-terminal half and a C-terminal half of ST3 (NC). The cumulative distribution of the hits over the *E*-values were grouped in three categories according to the evolutionary distance of query and hit in the MemGen database, termed the scope of the hit [8]. The hit may be between sequences in the same subfamily (scope: subfamily), between sequences in the same family, but not the same subfamily (scope: family) or between sequences in the same class, but not the same family (scope: class). The distribution shows that the BLAST algorithm detects 100% of the links between the N-terminal sequences of ST3 with scope subfamily, 77% with scope family, and 11% of the links with scope class (Fig. 1, NN). The distribution of the hits between the C-terminal domains was very similar (Fig. 1, CC) with corresponding values of 100, 66 and 13%, suggesting that on average the homology between the original sequences is evenly distributed over the N- and C-terminal halves of the proteins. In agreement, both distributions are similar to the distribution of the hits between the full sequences reported before [8].

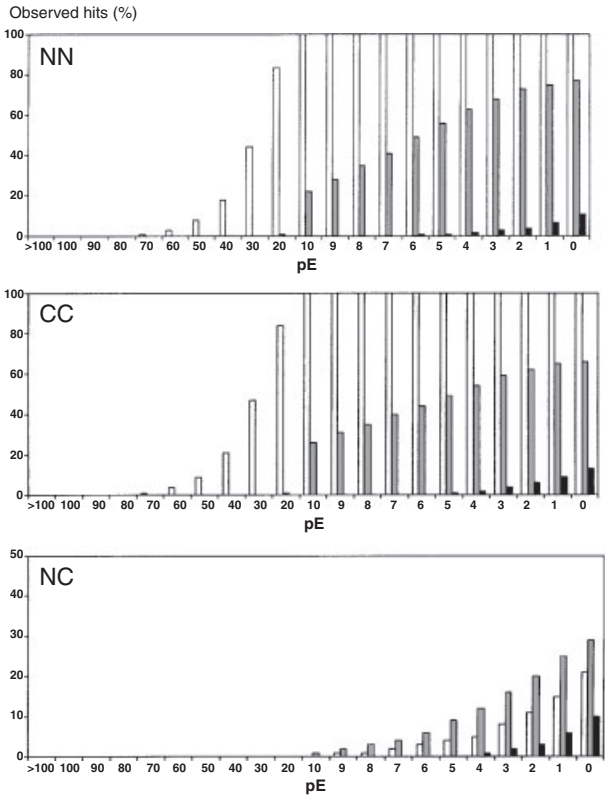


Fig. 1. Cumulative distribution of hits between N-terminal halves (NN), C-terminal halves (CC) and N- and C-terminal halves (NC) of ST3 sequences over *E*-values. Observed hits were reported as the percentage of the possible hits. The hits were categorized according to the evolutionary distance between query and hit (the scope) in the MemGen classification system: scope 'subfamily' (open bars), scope 'family' (grey bars), and scope 'class' (filled bars). The pE intervals were defined in the Experimental procedures section. BLAST searches were performed unfiltered.

The BLAST algorithm detected a surprisingly high number of links between N- and C-terminal halves of ST3 sequences (Fig. 1, NC). Within the same subfamilies, 21% of the possible hits were actually observed and for scope family this was 29%. Even more remarkable, the percentage of observed hits with scope class was 10% which is almost as high as observed for the percentage of hits between the N-terminal domains or the C-terminal domains alone. The results strongly suggest that the N- and C-terminal halves of the ST3 sequences share sequence homology.

Significance of the homology

In a previous analysis of the full sequences in structural classes ST3 and ST4 it was concluded that 'filtering' of the BLAST searches strongly improves the specificity of the search (the ratio of true over false hits) but at the expense of much of the sensitivity of the search (the number of observed hits). In fact, it was concluded that the filters were too stringent for this type of sequences ([8], see also [23]). To determine the significance of the hits between the N- and C-terminal halves of ST3 observed above, the following analysis was done on filtered BLAST searches.

The observed hits between the N- and C-terminal halves of the unrelated families included in the database (OPA, AAT, DAACS1; see above) and the N- and C-terminal halves of the ST3 families were analysed (Table 2). As a measure of the distribution of the hits over the *E*-values the ratio of the hits at pE = 2 and pE = 0 in the cumulative distributions were calculated (compare Fig. 1). Clearly, the percentage of observed hits with the N- and C-terminal halves of these unrelated families is very low and only very few of the hits have *E*-values in the range of pE ≥ 2 (Table 2). Importantly, the percentage of observed hits and their distribution over the *E*-values between these unrelated sequences of different origin is more or less the same, strongly suggesting that these are the random hits obtained between sequences of this type.

Table 3 shows the same analysis for the hits between the N- and C-terminal halves and vice versa within the OPA family, the AAT family, the ST3 families and the DAACS1 family. In case of the AAT and DAACS1 families, the percentage observed hits and the distribution are as observed for the random hits indicating no evolutionary relationship between the N-and C-terminal halves in these families. For Glt_{Ph} of *P. horikoshii* in the DAACS1 family a crystal structure is available which indeed shows that the N- and C-terminal halves have completely different folds. In contrast, the percentage of observed hits between the N- and C-terminal halves of the OPA family is an order of magnitude higher and 25% of the hits have *E*-values of pE ≥ 2. The OPA family is in the MFS whose members consist of two homologous domains [24], which was clearly shown in the 3D structures of LacY and GlpT [2,3] (the latter is a member of the OPA family). Also, the N- and C-terminal halves of the ST3 sequences score significantly higher than random with 2.6% observed hits and over 30% of hits with pE ≥ 2. In comparison with the OPA family, it should be noted that the ST3 analysis includes many families that are very distantly related with very little detectable sequence identity, whereas OPA is a single family in which all members share significant sequence identity with one another. The complete distribution of the hits between the ST3 domains is shown in Fig. 2. It is concluded that the N- and C-terminal halves of the ST3 proteins share significant sequence homology.

Table 2. Observed hits in filtered BLAST searches between the N- and C-terminal halves of the sequences in the control (Query) and ST3 families (Subject).

Query/Subject	Query half	Subject half	Observed hits (%)	pE distribution pE(2)/pE(0) (%)
OPA/ST3	N	N	0.45	1.0
	N	C	0.51	1.9
	C	C	0.30	0
	C	N	0.38	0
AAT/ST3	N	N	0.45	0
	N	C	0.30	0
	C	C	0.34	0
	C	N	0.32	1.4
DAACS1/ST3	N	N	0.29	2.9
	N	C	0.27	1.1
	C	C	0.61	0.9
	C	N	0.48	0.6
OPA + AAT + DAACS1/ST3	N + C	N + C	0.39	0.9

Table 3. Observed hits between the N- and C-terminal domains in the control and ST3 families.

Query/Subject	Query domain	Subject domain	Observed hits (%)	pE distribution pE(2)/pE(0) (%)
OPA/OPA	N	C	5.9	25.3
	C	N	8.1	22.3
AAT/AAT	N	C	0.54	0
	C	N	0.38	0
ST3/ST3	N	C	2.6	31.3
	C	N	2.6	33.1
DAACS1/DAACS1	N	C	0.11	0
	C	N	0.38	0

Localization of sequence identity

The lowest *E*-values of local alignments between N- and C-terminal domains in ST3 were surprisingly of scope class involving sequences of the [st302]ArsB and the [st303]AIT families. These *E*-values were as low as e-13 and should represent a reliable relationship between query and hit (Fig. 1). The local alignments showed sequence identities of about 30%, contained few gaps and covered most of the sequence lengths (> 75%; Supplementary Appendix S1, Table B). The latter suggests that the homologous domains are

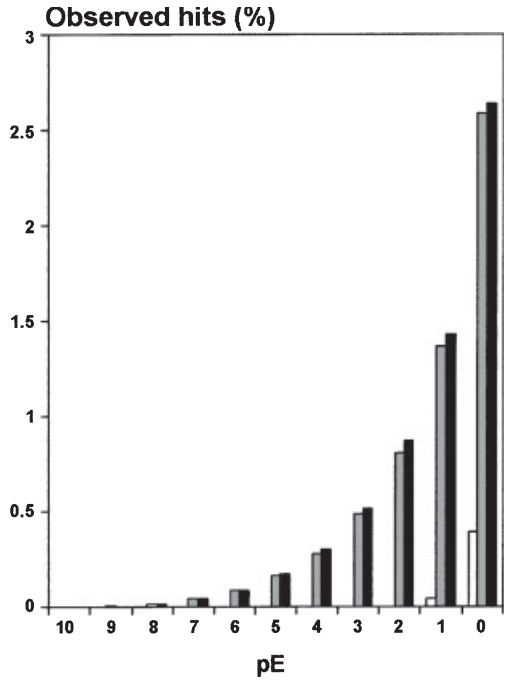


Fig. 2. Cumulative distribution of hits between N- and C-terminal halves over Expect values in filtered BLAST searches. The distributions were shown for the hits between the N- and C-terminal halves of the OPA, AAT, and DAACS families and the N- and C-terminal halves of ST3 (open bars), between the N-terminal halves of ST3 and the C-terminal halves of ST3 (grey bars), and between the C-terminal halves of ST3 and the N-terminal halves of ST3 (filled bars). The latter two are different because the BLAST filters only mask the query sequence and not the subject.

formed by the complete N- and C-terminal halves of the proteins.

The percentage of observed hits between N- and C-terminal halves within the same subfamily (scope: subfamily) differed significantly for the different subfamilies in ST3 ranging from 0% for [st303]AIT5 to 76% for [st315]AITE (unfiltered BLAST searches). Figure 3A shows the best scoring alignments for the N- and C-terminal halves of REUT3304rmet of *Ralstonia metallidurans* in [st315]AITE. The N-terminal half (left, blue) hits a C-terminal fragment (red) of NP939194cdip of *Corynebacterium diphtheriae* with an *E*-value of e-7. Sequence identity in the local alignment is 25%. The C-terminal half (right, blue) hits an N-terminal fragment (red) of NP811005bthe of *Bacteroides thetaiotaomicron* with an *E*-value of e-8 (sequence identity: 23%). The hydropathy profile representations of the local alignments in Fig. 3A clearly show that the homologous parts involve the complete membrane embedded parts of the REUT3304rmet protein.

Membrane topology inversion

The membrane topology of the transporters in the [st326]2HCT family in ST3 has been determined experimentally using a variety of techniques. The transporters consist of 11 TMS with the N terminus in the cytoplasm [14,16]. The hydropathy profile of the [st326]2HCT family is shown in Fig. 3B (red) and the positions of the membrane spanning segments were indicated below the profile (I-XI). The N- and C-terminal halves contain six and five TMSs, respectively. Unfortunately, the N- and C-terminal domains of the members of the [st326]2HCT family are not sufficiently similar to allow unambiguous alignment of the complete domains as the best local alignment obtained by the BLAST algorithm covers a stretch of 77 residues only (see below). Therefore, the hydropathy profile of the [st326]2HCT family was aligned with the profile of the [st315]AITE family (Fig. 3B). The N- and C-terminal domains of the latter allow unambiguous alignment of the entire domain based on the amino acid sequences (Fig. 3A) and the positions of the homologous repeats were boxed in Fig. 3B. The alignment shows that TMS I of [st326]2HCT is not present in [st315]AITE. The homologous N- and C-terminal domains of [st315]AITE consist of 5 TMSs each corresponding to TMS II-VI and TMS VII-XI of [st326]2HCT, respectively. Because the number of TMSs in the domains is an odd number, the two domains must have the opposite orientations in the membrane; the N terminus of the N-terminal domain is extracellular, while the N terminus of the C-terminal domain is intracellular. In the [st326]2HCT family, TMS I is not part of the two-domain structure and forms an additional domain by itself.

Evidence for the membrane topology inversion can also be obtained from the smaller homologous regions detected by BLAST within the [st326]2HCT family itself. The sequence in [st326]2HCT with the best scoring local alignment between N- and C-terminal domain is BH0400bhal of *B. halodurans* (sequence identity 33%). The overall sequence identity between BH0400bhal and CitSkpne of *K. pneumoniae*, the transporter for which the membrane topology was determined, is 36%. Figure 4A shows the hydropathy profile of the C-terminal fragment of the local alignment projected on the N-terminal domain of BH0400bhal (left) and the profile of the N-terminal fragment on the C-terminal domain (right). The local alignment corresponds to TMS V and the loop between TMS V/VI, termed Vb, in the N-terminal domain and TMS X and the loop between TMSs X and XI, termed Xa, in the C-terminal domain. It follows that TMS V corresponds to TMS X and loop Vb

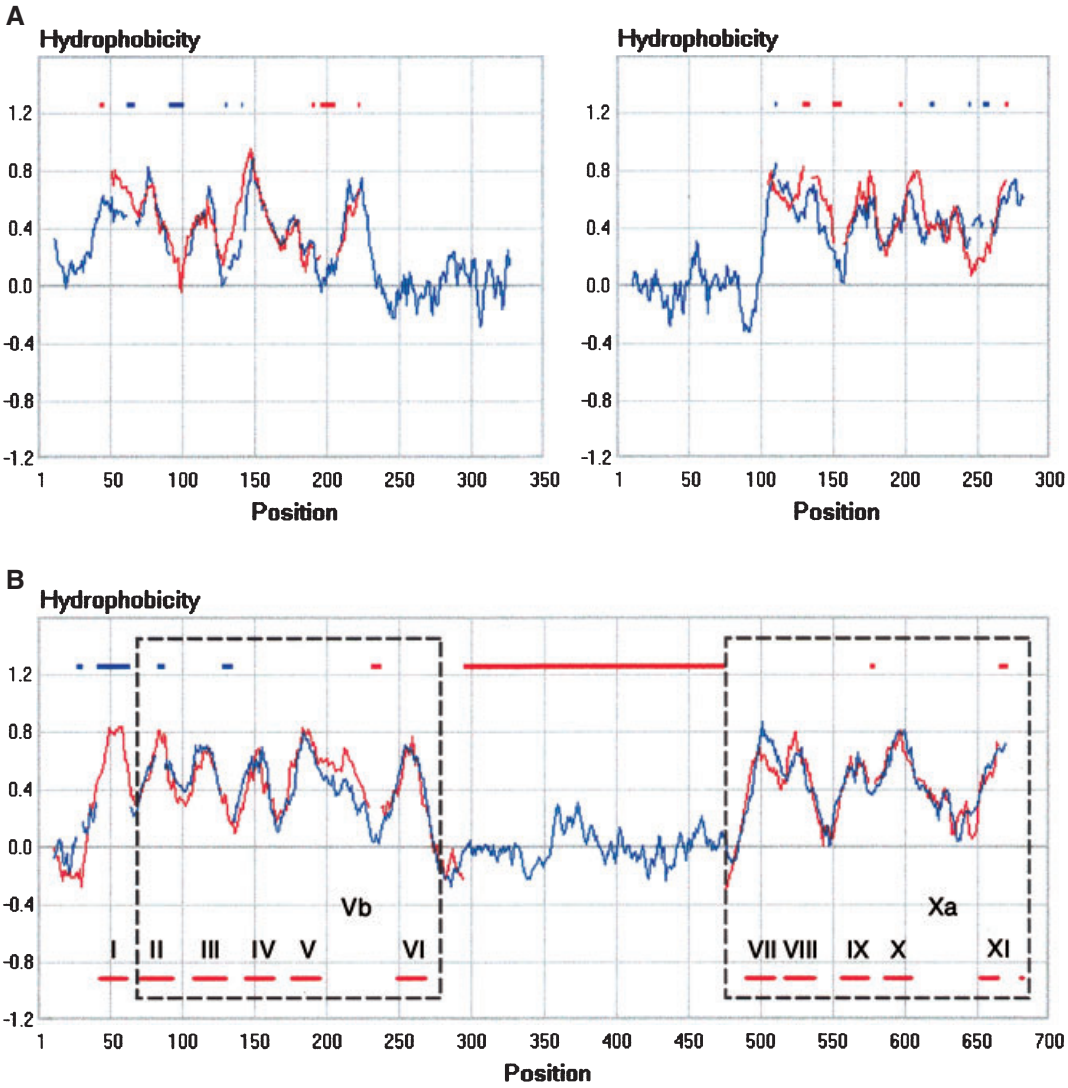
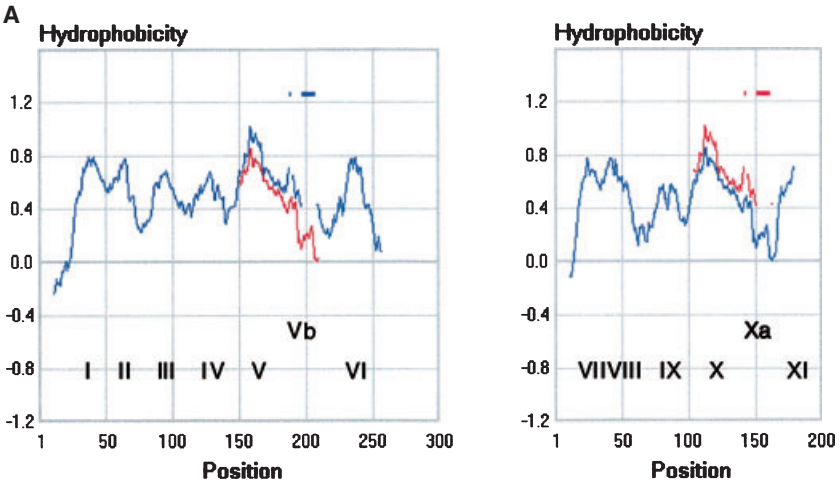


Fig. 3. Local alignments between N- and C-terminal domains in the [st315]AITE1 family (top) and analysis of the membrane topology of the two domains (bottom). Top: the left panel shows the hydropathy profiles of the N-terminal half (blue) of REUT3304rmet of *R. metallidurans* and the fragment from the local alignment with the C-terminal half of NP939194cdip of *C. diphtheriae* (red). The profiles were aligned based on the local alignment. The right panel shows in a similar way the local alignment between the C-terminal half of the REUT3304rmet sequence (blue) and the N-terminal half of NP811005bthe of *Bacteroides thetaiotaomicron* (red). The three sequences are in the [st315]AITE1 family. The bars above indicate the positions of gaps in the local alignments. Bottom: alignment of the family hydropathy profiles of the [st315]AITE1 (blue) and [st326]2HCT1 (red) subfamilies. The similarity test (*S*-test) of the alignment was 0.701, indicative of very similar profiles [6,7]. Red bars below indicate the positions of TMS I-XI in the [st326]2HCT family. Loops Vb and Xa were also indicated. The dashed boxes mark the positions of the N-terminal and C-terminal domains.

to loop Xa. Loops Vb and Xa of CitSkpne have been determined to be positioned at opposite sides of the membrane and are of particular importance to the function of the transporter (see Discussion).

Loops Vb and Xa

Loops Vb and Xa are the best conserved regions in the [st326]2HCT family. Conserved residues in the N- and C-terminal halves were indicated above and below the local alignment in Fig. 4B, respectively. In a pairwise sequence identity profile which measures the all-against-all pairwise sequence identity at each position in the multiple sequence alignment averaged over a window, peaks were observed at the position of the Vb and Xa loops with 55 and 63% identical pairs, respectively (Supplementary Appendix S1; Table C and Fig. A). Moreover, both regions contain



B

```
                  [V                    V]*
N-term: 140 LLIKVGVRFFLPIIGAIIGAIVIAAIIGSLVGFSVQDAILVMTL 183
            +L  +  ++ L + G ++GAI+ AA++G LVGF   +A +   L
C-term: 347 VLEALSYQYVLTVFGVVLGAILGAALLGKLVGFYPIEAAITGGL 390
            ~     [X  ~     ~          X]     *~~ ~*

            * ~  ** * *           *~*  ~
N-term: 184 PIMG-GGMGAGAV----------PMSQIYSELLG 206
             +   GG G  AV          P +QI S L G
C-term: 391 CMANMGGTGDVAVLSAAKRMQLMPFAQISSRLGG 424
                 **~* ~ ~* ~  *~ ~~ ~**~  *~ *
```

Fig. 4. Local alignments between N- and C-terminal domains of BH0400bhal in [st326]2HCT1. Top: hydropathy profiles of the N- and C-terminal halves (left and right, respectively) of BH0400bhal are shown in blue. The fragments from the local alignment between the two halves were indicated in red. Bars on top indicate the positions of gaps in the local alignment. The position of the TMSs I-XI and loops Vb and Xa were indicated at the bottom. Bottom: local alignment between the N- and C-terminal of BH0400bhal. In between, identical (*) and similar (~) residues were indicated. Above and below the sequences, the identical and similar residues in the multiple sequence alignment of the [st326]2HCT1 family in the N-terminal and C-terminal halves are indicated. The position of TMS V and TMS X is in brackets.

an above average fraction of the residues Gly, Ser, and Ala, residues with small side chains. In a window of 20 residues, these three residues are found at 48 and 50% of the positions in regions Vb and Xa, respectively, while the average in the complete alignment is 26%. Both features, pairwise sequence identity and composition, were analysed in the Vb and Xa regions in the other families of ST3. The positions of the Vb and Xa regions were determined from the optimal hydropathy profile alignment of each family and the [st326]2HCT family (Fig. 3B for [st315]AITE1 [6]);. In most subfamilies peaks in both types of profiles were found in the two regions (Supplementary Appendix S1; Table C and Fig. A). It follows that both features are conserved between the different families.

Discussion

The loop termed Xa has been studied extensively in members of the [st326]2HCT family where it resides between TMSs X and XI. Xa is believed to be part of the translocation site and to form a pore-loop like structure, i.e., the loop folds back in between the transmembrane segments. Firm experimental evidence has been presented that localizes the loop at the cytoplasmic side of the membrane [16,19,20,25]. Nevertheless, cysteine residues in the loop were shown to be accessible for (small) membrane impermeable thiol reagents from both sides of the membrane, in line with the properties of the binding site in the alternate access model for translocation [19,20]. Access from the extracellular side was restricted effectively in the coion-bound state of the Na^+-citrate transporter CitS of *K. pneumoniae* and partially in the substrate bound state [20,25]. Moreover, Arg425 in the citrate/lactate exchanger CitP of *Leuconostoc mesenteroides* situated at the interface between loop Xa and TMS XI, was shown to bind one of the carboxylate groups of the substrate [18] and mutation of Cys398 to Ser in loop Xa in CitS of *K. pneumoniae* reduced the affinity for the coion by one order of magnitude [20]. The functional importance of the region is in line with the high degree of conservation of the amino acid sequence, while the presence of a high proportion of residues with small side chains may reflect the folding of the loop in between the TMSs. The latter property was also observed for the pore-loop structure detected in the unrelated glutamate

transporter family (DAACS [22]);. The conservation of the two features in the other ST3 families suggests that the pore-loop will be common to all members of ST3.

The homology between the N- and C-terminal domains in ST3 indicates that loop Vb corresponds to loop Xa, strongly suggesting that the former will also form a pore-loop structure. Several lines of evidence support this view. Like Xa, the amino acid sequence in the Vb loop is highly conserved and contains a high proportion of Gly, Ala, and Ser residues, features that are conserved throughout the structural class. The Vb region is moderately hydrophobic and for many members, including CitS of *K. pneumoniae*, secondary structure prediction algorithms (e.g., TMHMM [26]); predict the presence of a TMS at this position. The prediction was falsified by experiment [14] and the moderate hydrophobicity, which is also observed for the Xa region, is an additional feature shared by both regions throughout the structural class (data not shown). It is in line with the membranous disposition of the loops. Functional relevance of the Vb loop is supported by the effect of mutations in the loop on the affinity of the substrate, again in the CitS protein [27].

Figure 5 shows the structural model for the transporters in the [st326]2HCT family that follows from the present study. The proteins consist of two homologous domains that are connected by a loop that is situated in the cytoplasm. Each domain consists of five TMSs and the orientation of the two domains in the membrane is opposite (topology inversion). The charge distribution in the loops follows the positive inside rule. The loop in between the fourth and fifth TMS of each domain forms a pore-loop structure that in the N-terminal domain enters from the extracellular side of the membrane and in the C-terminal domain from the cytoplasm (*trans* pore-loops). It is likely that in the 3D structure the two pore-loops interact. The transporters in the [st326]2HCT family have an additional TMS segment at the N terminus, TMS I. Most other families in structural class ST3 do not have this segment (e.g. [7]), but additional segments at the N terminus, the C terminus, or in between the two domains are observed in other (sub)families as well.

The topology model for the transporters in class ST3 resembles the model that has been proposed for two families of sodium/calcium exchangers (the CAX and NCX families [28,29]);. The proteins are believed to consist of two homologous domains that are connected by a large cytoplasmic loop and that have opposite orientations in the membrane. Like in the ST3 model, two re-entrant loops (parts of the α_1 and α_2 repeats) enter the membrane from opposite sides of the membrane. However, the position of the loops seems to be different than in the ST3 model. The re-entrant loops would be in between the second and third TMS in each domain rather than between the fourth and fifth TMS. Experimental evidence was presented that the re-entrant loops are in close vicinity in the structure [30].

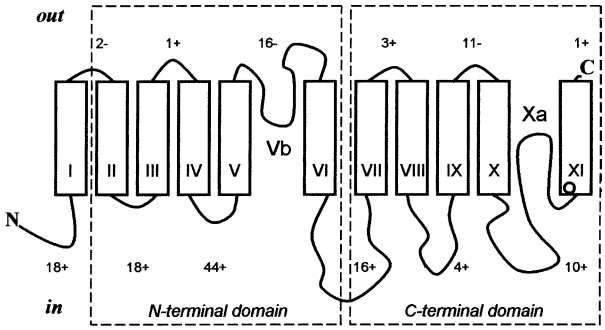


Fig. 5. Model of the structure of the transporters of the 2HCT family in class ST3. See text for explanation. The charge distribution gives the sum of the positive and negative charges in the loop regions of a multiple sequence alignment of 16 representative members of the 2HCT family.

Inverted topology or antiparallel architecture has been observed in the crystal structures of several membrane proteins and may be a recurring theme in membrane protein structure. Aquaporins, ClC chloride channels, the SecY subunit of the protein secretion machinery, the ammonia channel AmtB and the membrane subunit BtuC of the ABC transporter for vitamin B_{12} all contain domains that are weakly related by sequence alignments but have clearly similar folds related by a pseudo twofold symmetry axis in the plane of the membrane [31–37]. Gene fusion analysis indicates that that Na^+/Ca^{2+} exchangers are also likely to consist of two homologous domains with inverted topology [38] and sequence analyses predict a similar structural arrangement for a family of drug/metabolite transporters [39]. Inverted topology arrangements introduce symmetry in a protein with respect to the two sides of the membrane and may be particularly suited for exchangers that transport substrates in both directions during their transport cycle. Many transporters in ST3 are indeed exchangers or antiporters. It should be noted though that there are also different ways to build antiporters as is demonstrated by the structure of the glycerolphosphate-phosphate exchanger GlpT [2]. Similar to some families in ST3, the structure of the ammonia channel AmtB contains an additional TMS that is not part of the two-domain structure.

The two re-entrant loops in the model of the [st326]2HCT transporters that enter the membrane from opposite sides resemble the situation observed in the 3D structures of aquaporins [31–33] and the glutamate transporter homologue Glt_{Ph} [5]. With aquaporins, the resemblence is even higher because of the two-domain structure with topology inversion mentioned before. The two re-entrant loops of aquaporins are in close proximity at the centre of the membrane embedded part of the channel where they interact with the substrates (water or glycerol). Similarly, in the glutamate transporter structure, the tips of both reentrant loops are in close vicinity and the substrate is suspected in between them, suggesting that the re-entrant loops function as gates to the translocation pathway through the protein. A similar role of the re-entrant loops in ST3 transporters seems likely as mutagenesis experiments have shown that re-entrant loop Xa in the citrate transporters CitS and CitP and the malate transporters MleP and CimH in [st326]2HCT are involved in substrate and/or coion recognition [18–20,25].

Experimental procedures

Construction of the database

Structural class ST3 is stored in the MemGen database that may be browsed at our web site (http://molmic35.biol.rug.nl/memgen/main.htm). The sequences in structural class ST3 were dissected into the N- and C-terminal halves at the position of the central hydrophilic loop that for many subfamilies is easily recognized in the average hydropathy profile of the subfamily. Subfamilies for which the position of the central loop was ambiguous were not included in a first approach, nor were the sequences in family [st312]NHAC included for reasons discussed in reference [7] (for a list of included subfamilies, see Supplementary Appendix S1; Table A). The proteins of the OPA family from ST1 are known to contain a central hydrophilic loop that connects the N- and C-terminal domains. The sequences were dissected at this position. The sequences in the AAT family from ST2 and the DAACS1 family from ST4 were split right after the sixth putative transmembrane segment.

Blast searches

BLAST searches [40] were run locally against the database containing the N- and C-terminal sequences using the BLASTPGP executable. The database was formatted using the FORMATDB program. Both executables are freely available from the NCBI at their ftp site (ftp://ftp.ncbi.nih.gov/BLAST/). The size of the database was 311 254 letters and the database contained a total of 1366 sequences. BLAST searches were performed with 'low complexity filtering' and 'composition based statistics' set to 'off' (unfiltered) or 'on' (filtered) as indicated. All hits with *E*-values equal to or less than 10 were retrieved. Distributions of hits over *E*-values are reported in ranges of *E*-values indicated by the pE intervals as follows: a pE of 9 represents all values between 1e-9 and 9e-9, a pE of 40 all values between 1e-49 and 9e-40, etc.

Other methods

Methods and algorithms involved in the classification schemes and alignment of family hydropathy profiles have been described before [6–8] and can be found at our web site (http://molmic35.biol.rug.nl/memgen/main.htm). Multiple sequence alignments were done using the command line version of CLUSTAL W [41] for the WINDOWS XP platform that was downloaded from ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/and was used with the default settings.

Acknowledgements

This work was supported by the Netherlands Organization for Scientific Research, NWO-CW (JSL), European Commission grant QLK1-CT-2002-02388 (JSL), and a long-term fellowship of the International Human Frontier Science Program (DJS).

References

1 Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR & Iwata S (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. Science **301**, 610–615.
2 Huang Y, Lemieux MJ, Song J, Auer M & Wang DN (2003) Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616–620.
3 Saier MH Jr (2000) A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**, 354–411.
4 Murakami S, Nakashima R, Yamashita E & Yamaguchi A (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593.
5 Yernool D, Boudker O, Jin Y & Gouaux E (2004) Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431**, 811–818.
6 Lolkema JS & Slotboom DJ (1998) Estimation of structural similarity of membrane proteins by hydropathy profile alignment. *Mol Membr Biol* **15**, 33–42.
7 Lolkema JS & Slotboom DJ (2003) Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis. *J Mol Biol* **327**, 901–909.

8 Lolkema JS & Slotboom DJ (2004) Sequence and hydropathy profile analysis of two classes of secondary transporters. *Mol Membr Biol*, in press.
9 Bandell M, Ansanay V, Rachidi N, Dequin S & Lolkema J (1997) Membrane potential generating malate (MleP) and citrate (CitP) transporters of lactic acid bacteria are homologous proteins. Substrate specificity of the 2-hydroxy-carboxylate transporter family. *J Biol Chem* **272**, 18140–18146.
10 Kawai S, Suzuki H, Yamamoto K & Kumagai H (1997) Characterization of the 1-malate permease gene (maeP) of *Streptococcus bovis* ATCC 15352. *J Bacteriol* **179**, 4056–4060.
11 Wei Y, Guffanti AA, Ito M & Krulwich TA (2000) *Bacillus subtilis* YqkI is a novel malic/Na+-lactate antiporter that enhances growth on malate at low protonmotive force. *J Biol Chem* **275**, 30287–30292.
12 Krom BP, Aardema R & Lolkema JS (2001) *Bacillus subtilis* YxkJ is a secondary transporter of the 2-hydroxycarboxylate transporter family that transports 1-malate and citrate. *J Bacteriol* **183**, 5862–5869.
13 Kastner CN, Schneider K, Dimroth P & Pos KM (2002) Characterization of the citrate/acetate antiporter CitW of *Klebsiella pneumoniae*. *Arch Microbiol* **177**, 500–506.
14 van Geest M & Lolkema JS (1996) Membrane topology of the Na^+-dependent citrate transporter of *Klebsiella pneumoniae*. Evidence for a new structural class of secondary transporters. *J Biol Chem* **271**, 25582–25589.
15 van Geest M, Nilsson I, von Heijne G & Lolkema JS (1999) Insertion of a bacterial secondary transport protein in the ER membrane. *J Biol Chem* **274**, 2816–2823.
16 van Geest M & Lolkema JS (2000) Membrane topology of the Na^+-dependent citrate transporter of *Klebsiella pneumoniae* by insertion mutagenesis. *Biochim Biophys Acta* **1466**, 328–338.
17 Bandell M & Lolkema JS (2000) The conserved C-terminus of the citrate (CitP) and malate (MleP) transporters of lactic acid bacteria is involved in substrate recognition. *Biochemistry* **39**, 13059–13067.
18 Bandell M & Lolkema JS (2000) Arg425 of the citrate transporter CitP is responsible for high affinity binding of di- and tricarboxylates. *J Biol Chem* **275**, 39130–39136.
19 Krom BP & Lolkema JS (2003) Conserved residues R420 and Q428 in a cytoplasmic loop of the citrate/malate transporter CimH of Bacillus subtilis are accessible from the external face of the membrane. *Biochemistry* **42**, 467–474.
20 Sobczak I & Lolkema JS (2004) Alternate access and a pore-loop structure in the Na^+-dependent citrate transporter CitS of *Klebsiella pneumoniae*. *J Biol Chem* **279**, 31113–31120.
21 Grunewald M & Kanner BI (2000) The accessibility of a novel reentrant loop of the glutamate transporter GLT-1 is restricted by its substrate. *J Biol Chem* **275**, 9684–9689.
22 Slotboom DJ, Sobczak I, Konings WN & Lolkema JS (1999) A conserved serine-rich stretch in the glutamate transporter family forms a substrate sensitive reenetrant loop. *Proc Natl Acad Sci USA* **96**, 14282–14287.
23 Hedman M, Deloof H, von Heijne G & Elofsson A (2002) Improved detection of homologous membrane proteins by inclusion of information from topology predictions. *Protein Sci* **11**, 652–658.
24 Henderson PJF & Maiden MC (1990) Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **326**, 391–410.
25 Sobczak I & Lolkema JS (2003) Accessibility of cysteine residues in a cytoplasmic loop of CitS of *Klebsiella pneumoniae* is controlled by the catalytic state of the transporter. *Biochemistry* **42**, 9789–9796.
26 Krogh A, Larsson B, Von Heijne G & Sonnhammer EL (2003) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Membr Biol* **305**, 567–580.
27 Kästner CN, Schneider K, Dimroth P & Pos KM (2002) The Na+-dependent citrate carrier of *Klebsiella pneumoniae*: high-level expression and site-directed mutagenesis of asparagine-185 and glutamate-194. *Arch Microbiol* **177**, 500–506.
28 Quednau BD, Nicoll DA & Philipson KD (2004) The sodium/calcium exchanger family – SLC8. *Pflugers Arch – Eur J Physiol* **447**, 543–548.
29 Cai X & Lytton J (2004) The cation/Ca^{2+} exchanger superfamily: phylogenetic analysis and structural implications. *Mol Biol Evol* **21**, 1692–1703.
30 Qiu Z, Nicoll DA & Philipson KD (2001) Helix packing of functionally important regions of the cardiac Na(+)-Ca(2+) exchanger. *J Biol Chem* **276**, 194–199.
31 Savage DF, Egea PF, Robles-Colmenares Y, Iii JD & Stroud RM (2003) Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin z. *Plos Biol* **1**, E72.
32 Fu D, Libson A, Miercke LJ, Weitzman C, Nollert P, Krucinski J & Stroud RM (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486.
33 Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A & Fujiyoshi Y (2000) Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605.
34 Dutzler R, Campbell EB, Cadene M, Chait BT & MacKinnon R (2002) X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294.

35 Van den Berg B, Clemons WM Jr, Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44.
36 Khademi S, O'Connell J, Ird Remis J, Robles-Colmenares Y, Miercke LJ & Stroud RM (2004) Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science* **305**, 1587–1594.
37 Locher KP, Lee AT & Rees DC (2002) The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098.
38 Saaf A, Baars L & von Heijne G (2001) The internal repeats in the Na+/Ca2+ exchanger-related *Escherichia coli* protein YrbG have opposite membrane topologies. *J Biol Chem* **276**, 18905–18907.
39 Jack DL, Yang NM & Saier MH Jr (2001) The drug/metabolite transporter superfamily. *Eur J Biochem* **268**, 3620–3639.
40 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
41 Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Supplementary material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/EJB/EJB4665/EJB4665.htm

Appendix S1.

Fig A. Family profiles of hydrophobicity, pairwise sequence identity, and frequency of residues Gly, Ala, Ser of the [st326]2HCT1 (A), [st316]NHAD1 (B), and [st325]GLTS1 (C) subfamilies. The position of loop regions Vb and Xa were indicated. Red bars indicate the positions of gaps in any sequence in the multiple sequence alignment. The window was 20 positions.

Table A. Subfamilies included in the domain database.

Table B. Hits between N- and C-terminal domains in ST3 with pE ≥ 10 in filtered BLAST searches.

Table C. Properties of regions Vb and Xa in ST3 subfamilies.