Trans reentrant loop structures in secondary transporters

Dobrowolski, Adam Jan

Trans reentrant loop structures in secondary transporters

Adam J. Dobrowolski

Cover design: Adam J. Dobrowolski
Cover: The *twelve angle stone*, in the Hatun Rumiyoc street of Cuzco, Peru. Part of the palace of Inca Roca. Picture taken by author.

The work described in this thesis was carried out in the Molecular Microbiology group of the Groningen Biomolecular Sciences and Biotechnology Institute (GBB) of the University of Groningen, The Netherlands, and was financially supported by a grant from the Dutch Organization for Scientific Research (NWO-ALW).

ISBN: 978-90-367-5318-0
ISBN: 978-90-367-5319-7 (electronic version)

Printed by: JAKS Wrocław, Poland
Printing of this thesis was supported by generous contribution from the University of Groningen and the Groningen Biomolecular Sciences and Biotechnology Institute (GBB).

RIJKSUNIVERSITEIT GRONINGEN

Trans reentrant loop structures in secondary transporters

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 2 april 2012
om 12.45 uur

door

Adam Jan Dobrowolski

geboren op 1 mei 1980
te Zielona Góra, Polen

Promotor:	Prof. dr. A.J.M. Driessen
Copromotor:	Dr. J.S. Lolkema
Beoordelingscommissie:	Prof. dr. K.J. Hellingwerf
	Prof. dr. B. Poolman
	Prof. dr. B. Dijkstra

Rodzicom

Contents

Chapter I

General introduction.
2-domain secondary transport proteins

Adam Dobrowolski and Juke S. Lolkema

Chapter I

Abstract

The growing number of reported X-ray crystal structures of membrane proteins has revealed that many of them consist of two homologous domains, or structural repeats, that share a similar fold. This structural motif is especially widespread in the group of secondary transporters. A plausible model for the evolution of two-domain membrane proteins with an internal structural repeat involves a duplication of a primordial gene followed by a fusion, thus resulting in a single gene encoding a protein with two homologous domains. The two domains have the same (parallel) or opposite (antiparallel) orientation in the membrane. The evolution of the antiparallel two-domain membrane proteins requires an ancestral protein with so-called “dual-topology” that inserts into the membrane with random orientation. In this chapter 3D structures of secondary transporters are described with emphasisis on the two-domain structure. Remarkably, secondary transporters that are classified into a large number of different families by sequence similarity (TC system), show a much smaller number of specific folds.

Chapter I

1. Introduction

All living cells are surrounded by biological membranes, which consist of a phospholipid bilayer and a variety of proteins, called membrane proteins. The membrane proteins are responsible for the contact between cell and its environment. Biomembranes are the interface of two compartments, such as the intra- and extracellular environment or mitochondrial matrix and cytosol. Membrane proteins interact closely with both water and lipids in their environment. They are synthesized as other proteins at the ribosome and, at the same time, targeted to the different membrane localization within a cell. Most membrane proteins are involved in transport and signaling, or they are key components in energy transduction, such as converting the chemical energy in ATP into electrochemical energy, or, in reverse, in ATP synthesis.

Knowledge about the structure of membrane proteins is an important source of information in molecular biology. Together with functional studies, the structures provide insight into the molecular mechanism by which they work. This knowledge helps in the search for improved medicines, since many targets are membrane proteins responsible for signaling or transport at the cell membrane (69).

Recently, the availability of 2D and 3D structures obtained by tools such as electron microscopy (EM) and X-ray crystallography combined with theoretical analysis has significantly increased our understanding of the molecular functionality of membrane proteins. Integral membrane proteins form a considerable part of all proteins in a cell. It was shown that on average about 30% of genes encoded in the genome of a cell code for integral membrane proteins (65, 155). Despite the fact that nowadays almost 300 unique structures of membrane proteins are available in the Protein Data Bank (PDB), this number is still limited in comparison to the many structures published for soluble proteins (6). The low number of crystal structures is related to the difficulties encountered in obtaining crystals that give a good diffraction pattern. Membrane proteins are very hydrophobic and can be only purified in a detergent solubilized state. The hydrophobic surface of the protein provides very few contact points to form a crystal lattice. Nevertheless, in the last decade, great progress has been made in obtaining 3D structures of membrane proteins. In spite of this great progress, it should be noted that transport across membranes, by definition, is a dynamic process and it is difficult to fully

understand the mechanism exclusively on the basis of static pictures, even if they are of high resolution.

In spite of a great diversity in functions, integral membrane proteins share a similar architecture. Membrane proteins follow some relatively simple rules governed by their hydrophobic nature and restrictions imposed by the lipid bilayer. Membrane proteins consist of either α-helical bundles or β-barrels (Fig.1). The fully satisfied backbone hydrogen bonding found in these secondary structures avoid unfavorable interactions of backbone amide or carbonyl groups with the hydrophobic environment of the lipid bilayer. With few exceptions, α-helical bundles are found in cytoplasmic and subcellular compartment membranes and β-barrels in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts (170). The polypeptide chain of membrane proteins composed of a transmembrane α-helical bundle or β-barrel crosses the membrane in a zig-zag fashion. The parts of a chain that connect the transmembrane segments (TMSs) are termed loops. Consequently, a membrane protein contains both intracellular and extracellular loops.

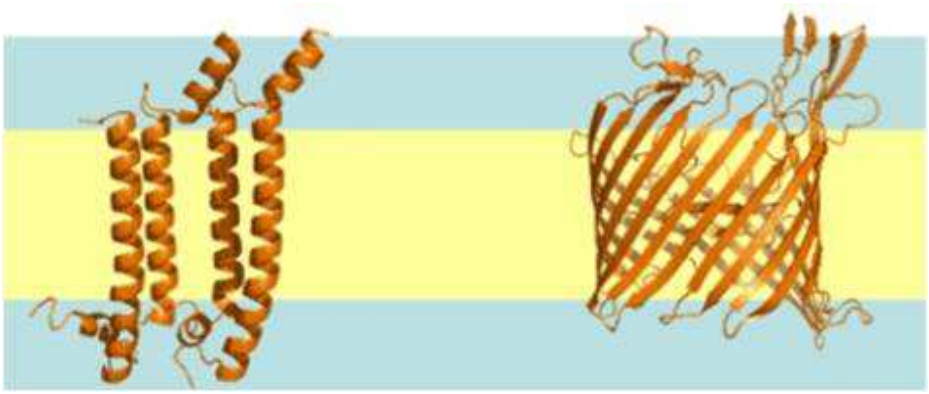


Figure 1. Two types of membrane protein structures. The α-helical (left) and β-barrel (right) membrane proteins. (http://www2.warwick.ac.uk/alumni/services/eportfolios/msrfas/research/)

The recent presentation of quite a number of 3D structures of membrane proteins has revealed another feature of membrane proteins that is found in many of them. Many membrane proteins consist of two homologous domains that result in structural repeats and two-fold symmetry in the structure. Structural repeats are found in membrane

proteins of different functions. They have been observed in channel proteins, the translocon of the protein excretion machinery, complex I of the respiratory chain and are especially frequent in secondary transporters. The structural repeats are believed to originate from ancient gene duplication events. This introductory chapter gives an overview of the fold of secondary transporters with a two-domain structure.

2. Structural repeats in membrane proteins

A common feature shared by many membrane proteins and especially by secondary transporters is the presence of structural repeats (10, 38). Each repeat typically comprises several contiguous TMSs, which assume a specific three-dimensional fold that is repeated two times (or three in case of ADP/ATP carrier). Structural repeats are thought to have arisen from gene duplications followed by fusion resulting in two homologous domains after translation. Evolution towards two-domain membrane proteins increases the asymmetry in the transporter protein, which may widen the spectrum of substrates. Broad substrate specificity may have provided the evolutionary pressure for obtaining these proteins. In many cases, the structural repeats were not detected before the 3D structure was solved, because strong sequence divergence between the two domains did not allow the detection of significant sequence similarity (85).

The duplicated domains can be arranged in the membrane with parallel or antiparallel topology (Fig. 2). Structural repeats with parallel orientation are related by a pseudo-symmetry axis perpendicular to the membrane plain (e.g Major Facilitator Superfamily, MFS). Opposite, so-called ‘inverted’ structural repeats with antiparallel orientation of domains in the membrane are related by a pseudo-symmetry axis parallel to the membrane (e.g. The Amino Acid-Polyamine-Organocation Superfamily, APC). The orientation of the two domains is determined by an even (parallel) or odd (antiparallel) number of TMSs in the repeats. Interestingly, carriers such as the zinc transporter YiiP (95) and the multidrug-resistance protein EmrE (17, 153) probably represent ancient functional homodimers that preceded gene duplication.

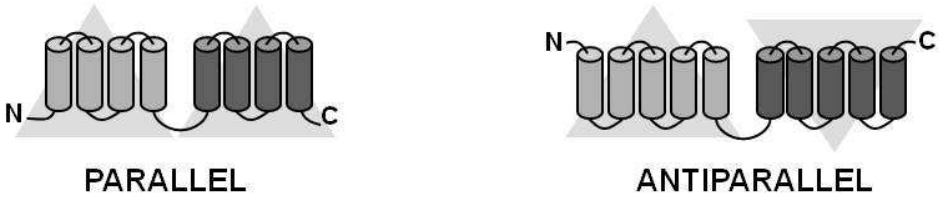


Figure 2. Two types of 2-domain membrane proteins with internal repeats. The two domains have the same (parallel - left) or opposite (antiparallel - right) orientation in the membrane depending on an even (left) or odd (right) number of transmembrane segments (TMSs) per domain. Cylinders represents transmembrane segments, triangles in the background emphasize the relative orientation of the domains in protein.

3. Evolution of two-domain membrane protein

With the rising number of membrane proteins structures obtained by X-ray crystallography it has become clear that many membrane proteins consist of two homologous domains. The repeat was not obvious from the amino acid sequences, probably because the sequences diverged too far. A plausible path for the evolution of membrane proteins with internal repeats has been proposed before (11, 124, 125, 128, 129). The hypothesis concentrates on the duplication of a primordial gene followed by a fusion, thus resulting in a single gene encoding a protein with two homologous domains (Fig. 3). As was mentioned above the two domains have the same (parallel) or opposite (antiparallel) orientation in the membrane. A protein with parallel orientation of two domains could evolve by duplication and fusion of a gene encoding a protein forming a homodimer with the same orientation of the subunits (Fig. 3A). Examples of this class are the members of the Major Facilitator Superfamily (see section 4.1). To account for the antiparallel orientation of the two domains, the ancestral membrane protein was hypothesized to be “dual topology”; i.e., it would insert with a random orientation into the membrane (Fig. 3B). Following duplication, the two dual topology proteins would adopt fixed but opposite orientations by genetic drift, driven by the introduction of positively charged amino acid residues in cytoplasmic loops (according to the “positive-inside rule” introduced by von Heijne (47)). The first gene on the chromosome may encode one orientation or the other; at the protein level, this has no consequences for the antiparallel heterodimer that is formed. However, in the fused state, this results in two different proteins with the N-terminus either inside or outside the cell. Therefore, the

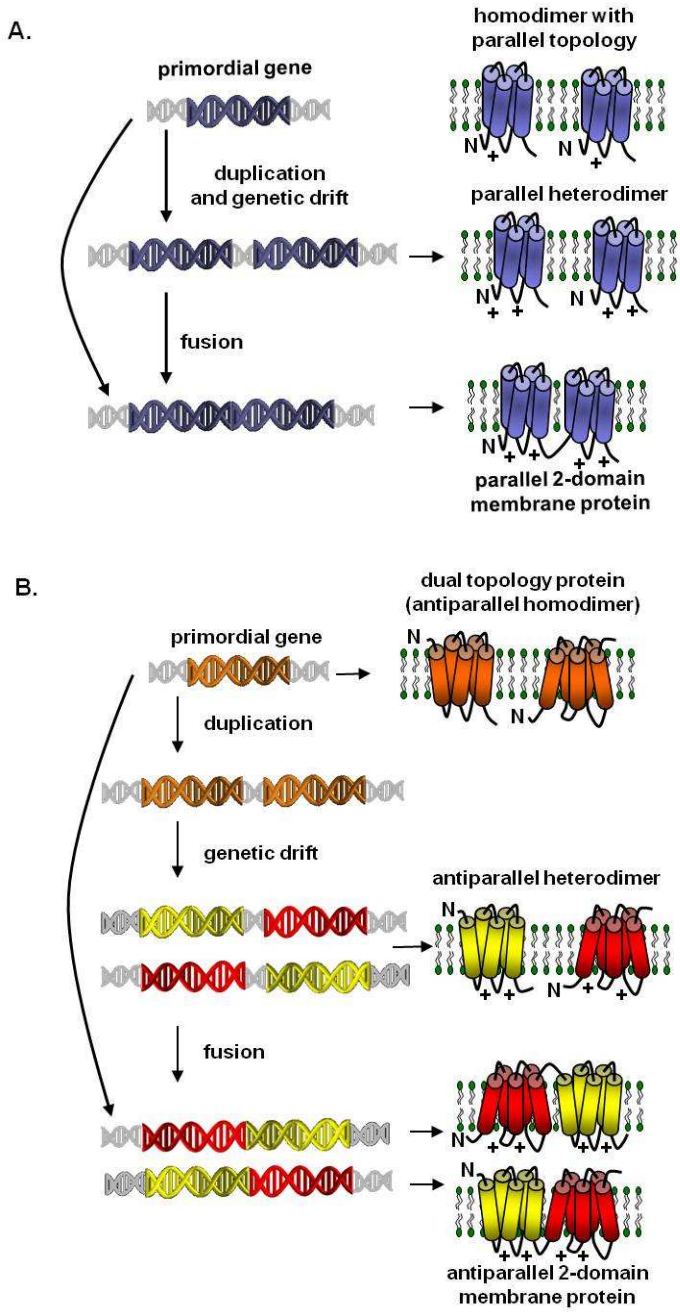


Figure 3. Model for the evolution of two-domain membrane proteins. The left column shows genetic states, the right column shows the encoded proteins embedded in the membrane. (A) Parallel two-domain membrane proteins. A primordial gene (blue) encoding a protein (blue - homodimer) with fixed orientation in the membrane, results, after duplication and genetic drift, in an operon of two genes (blue) encoding homologous proteins (blue - heterodimer) with a parallel orientation. Finally, a fusion of the two genes results in the parallel two-domain membrane protein. (B) Antiparallel two-domain membrane proteins. A primordial gene (orange) encoding a dual topology protein (orange - homodimer) results, after duplication and a genetic drift, in an operon of two genes (yellow and red) encoding homologous proteins (yellow and red - heterodimer) with a fixed but opposite orientation. Finally, fusion of the two genes results in the antiparallel two-domain membrane protein with two possible orders of the domains.

evolutionary pathway has two outcomes representing proteins that differ in the order of the two domains in the primary structure. In the course of evolution, one of these outcomes may be selected at random or because of a selective advantage.

Support for the evolutionary model comes from bioinformatics analyses of protein families and from experiments. The different states in the model may still be recognized in a few families of membrane proteins existing to date. The DUF606 family, a family of membrane proteins of unknown function, is especially rich in evolutionary states proposed in the pathway: single genes that would code for dual topology homodimeric proteins, paired genes coding for homologues proteins with fixed but opposite orientation in the membrane, that would form heterodimers, and fused genes that encode antiparallel two-domain fusion proteins (85). Bioinformatics analysis of the DUF606 family revealed a total of nine independent duplication events, five of which resulted in paired genes, and four resulted in fused genes. The general conclusion from these studies was that antiparrallel two-domain proteins are the direct result of a gene duplication event rather than a sequential mechanism in which fusions evolve from a pair of genes. Further support for the evolutionary model comes from the identification of two subfamilies of the NhaC family, a family of putative Na^+/H^+ antiporters, the members of which show the two possible orders of the homologous domains in the antiparallel fusion proteins (Fig. 3B; (87, 159)). The N-terminal half's of the sequence of the proteins in one subfamily share significant sequence identity with the C-terminal half's of the proteins in another subfamily. Experimental evidence for the two membrane topologies has been presented (159). Experimental support for the properties of the ancestor proteins in the evolutionary pathway comes from studies of members of the Small Multidrug-Resistant (SMR) transporter family. The best-studied SMR protein is EmrE from *E. coli*, an inner-membrane drug efflux pump containing four TMSs, which is coded by a single gene and believed to form an antiparallel homodimer (128, 129, 166) which implies that the EmrE protein inserts in both orientations in the membrane ("dual topology"). With few exceptions, a selection of SMR proteins encoded by a single gene was shown to be dual topology as well (72). It should be noted that the antiparallel orientation of the subunits in the EmrE dimer are still under debate and data has been presented that favors a parallel homodimer (141, 142, 152, 153). In the same SMR family, the EbrA and EbrB proteins

from *Bacillus subtilis* are encoded in an operon and the gene products form a heterodimer with antiparrallel oriented subunits providing strong evidence for the antiparallel orientation in general (67). Similarly, a selection of SMR proteins encoded in pairs of genes revealed in all cases opposite orientations of the two proteins in the membrane (72). Further evidence follows from experiments in which the protein is changed from one state to the other by genetic manipulation and rational design based on the positive-inside rule as the determinant of the orientation in the membrane. The positive-inside states that the loops that connect the TMSs at the cytoplasmic side of the membrane contain a surplus of positively charged arginine and lysine residues relative to the loops at the opposite side of the membrane (47). The EbrA and EbrB proteins of *B. subtilis* could both be mutated back to dual topology proteins (68) and, in an elegant series of experiments with EmrE, an evolutionary path connecting a dual topology protein to a pair of oppositely oriented homologues could be emulated (129). One step further was taken in the re-routing of the evolution of the glutamate transporter GltS, an antiparallel two-domain membrane protein that is at the end of its evolution with respect to the orientation in the membrane. The *gltS* gene was split in the two halves encoding the two domains and put back together again with the two halves in the opposite order. Functional assays showed that rerouting the pathway to the alternative output (Fig. 3B) by swapping the domains does not significantly affect its biogenesis and function. These experiments are discussed in chapter 5 of this thesis (24).

4. Secondary transporters.

Secondary transporters are ubiquitously distributed molecular machines found in every cell. In *Escherichia coli* 10,8 % of all chromosomal genes code for membrane transport proteins and one of the largest functional category are secondary transporters (122). They accept a broad range of solutes, including ions, neurotransmitters, nutrients, and numerous drugs. Secondary transports use the free energy stored in ion or solute gradients to drive the transport of a solute across the cytoplasm or internal membranes of biological cells. Accumulation of the solute at one side of the membranes is achieved by coupling the translocation of the solute to the translocation of one or more ions (H^+ or Na^+) that move down their own gradients, the proton motive force and/or Na^+ -ion motive

force, respectively (86). Secondary transporters are commonly classified in three groups based on their mode of energy coupling: (i) uniporters catalyze the translocation of a single solute across the membrane, (ii) symporters couple the translocation of a solute to the translocation of a co-ion(s) in the same direction, and (iii) antiporters couple the translocation of a solute and a co-ion(s) in opposite directions (Fig. 4). Many antiporters couple the translocation of one solute to the translocation of another solute rather than a co-ion. They exchange a substrate at one side of the membrane for another substrate at the other side of the membrane. The different modes of energy coupling enable transporters to play an important role in different aspects of the physiology of the cell. Thus, the symport and uniport mechanisms allow the cell to take up nutrients from the medium, while antiporters may function in the excretion of end products or in defense mechanisms by removing harmful compounds from the cell. Antiporters may combine the uptake of a nutrient from the environment and the excretion of a metabolic end product (149).

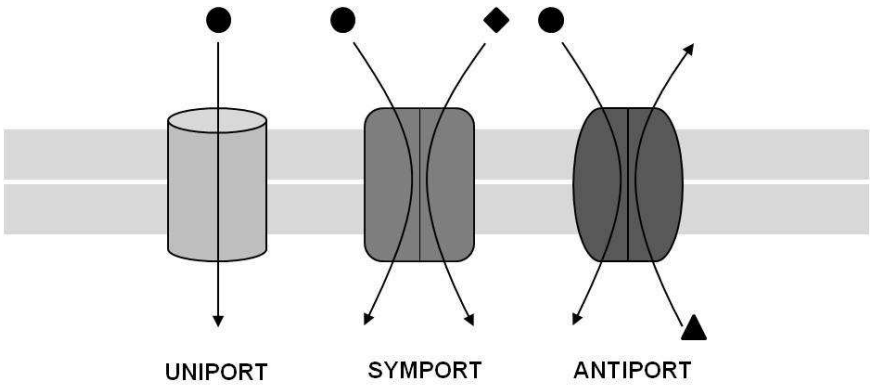


Figure 4. Schematic representation of the three modes of transport of secondary transporters, uniport, symport and antiport. Circles represent solutes, diamonds the co-ions, and triangles the co-ions or solutes in exchange.

According to the transport classification system (TC system), which is based on sequence and functional similarities, secondary transporters are represented by over 100 families (13, 134, 138)(http://www.tcdb.org/). However, many of these families are evolutionary and structurally related (12, 87) suggesting that a few membrane protein folds have

successfully evolved to provide a transport mechanism and widely divergent substrate specificities (114, 161). The distant evolutionary relationships and structural similarity of secondary transporters from different gene families was predicted by Lolkema and Slotboom by identifying distant evolutionary relationships by hydropathy profile alignment of the amino acid sequences (88-91) (MemGen classification system; see also section 4.3 of this chapter). The unexpected finding, that transporters unrelated in sequence share a similar structure, emphasizes the remarkable plasticity of transporters, which allows them to use a common scaffold to translocate different substrates (10).

A central concept in the mechanism of secondary transporters is the alternating access mechanism proposed by Jardetzky (57). According to this mechanism, the carrier or transporter isomerizes between an outward-facing state with a substrate-binding site accessible from the external solution, and inward-facing state with the site accessible from the cytoplasm (106, 121). Two conceptual models prevail: the “rocker-switch” and the “gated pore”. The former involves rocking movements of protein domains pivoting at the substrate-binding site. The latter involves local motions of external and internal gates flanking the substrate-binding site, and undergoing alternating openings and closings. Recent studies suggest that the transport mechanisms of secondary transporters probably integrate both modes of action, thereby including an isomerization between the inward- and outward-facing states, as well as local opening and closure of gates that partially or completely occlude bound substrates.

To date, over twenty 3D structures of individual proteins from different families of secondary transporters have been determined (Table 1). Many of them contain two homologous domains in either parallel or antiparallel orientation. Below, the different folds of such proteins are discussed, mainly based on proteins with known 3D structure.

4.1 MFS fold (ST[1])

The Major Facilitator Superfamily (MFS) represents the largest group of secondary transporters. It is predicted that about 25% of all known transport proteins in prokaryotes belong to this superfamily (15, 119, 136). The members of MFS, like secondary transporters in general, are found ubiquitously in all kingdoms of life. The MFS members

Table 1. Secondary transporter structures containing two homologous domains or structural repeats.

	"Core" fold [1]	Family/TC#[3]	Protein	Function	PDB code[2]	Ref.
1	↑↑6+6	MFS/2.A.1	GlpT	Glycerol-3-phosphate/ phosphate antiport	1PW4	(55)
2			LacY	H^+/sugar symport	1PV7	(1)
3			EmrD	H^+/drug antiport	2GFP	(181)
4			FucP	H^+/fucose symport	3O7Q	(20)
5			$PepT_{SO}$	H^+/oligopeptide symport	2XUT	(112)
6	↓↑5+5	NSS/2.A.22	LeuT	Na^+ /leucine symport	2A65	(179)
7		SSS/2.A.21	vSGLT	Na^+ /glucose symport	3DH4	(34)
8		NCS1/2.A.39	Mhp1	Na^+ /hydantoin symport	2JLN	(176)
9		NCS2/2.A.40	UraA	H^+/uracil symport	3QE7	(94)
10		BCCT/2.A.15	BetP	Betaine-glycine/Na^+ symport	2WIT	(130)
11			CaiT	L-carnitine/γ-butyrobetaine antiport	3HFX	(143)
12		APC/2.A.3	ApcT	H^+ /amino acid symport	3GIA	(147)
13			AdiC	Arginine/agmatine antiport	3LRB	(42)
14	↓↑5+5	2HCT/2.A.24	CitS	Na^+/citrate symport	-	(149)
15		ESS/2.A.27	GltS	Na+/glutamate symport	-	(26)
16	↓↑3+3+2+2	DAACS/2.A.23	Glt_{Ph}	Na^+/aspartate symport	1XFH	(180)
17	↑↑6+6	RND/2.A.6	AcrB	H^+/drug antiport	1IWG	(109)
18			CusA	H^+/metal ion antiport	3K07	(93)
19			MexB	H^+/drug antiport	2V50	(146)
20	↑↑6+6	MATE/2.A.66.1	NorM	H^+/drug antiport	3MKT	(46)
23	↓↑3+3	FNT/2.A.44	FocA	H^+/formate antiport	3KCU	(172)
24	↓↑8+8	ClC/2.A.29	StClC	H^+/Cl^- antiport	1KPL	(28)
25			EcClC	H^+/Cl^- antiport	1OTS	(29)
26			CmClC	H^+/Cl^- antiport	3ORG	(36)
27			SYClC	H^+/Cl^- antiport	3ND0	(58)
21	↓↑5+5	NhaA/2.A.33	NhaA	Na^+/H^+ antiport	1ZCD	(56)
22		BASS/2.A.28	$ASBT_{NM}$	Na^+/bile acid symporter	3ZUY	(54)
28	↑↑2+2+2	MCF/2.A.29	AAC1	ADP/ATP antiport	1OKC	(123)

The table lists secondary transporters containing two homologous domains. The proteins are grouped according to core structure, based on reported X-ray structures in the PDB database (http://www.pdb.org/pdb/home/home.do) or by biochemical and bioinformatics studies. Names of families and TC numbers correspond to the Transporter Classification Database (TCDB - http://www.tcdb.org/)

[1]) arrows represent a parallel or antiparallel orientation of domains (or structural repeats) in the membrane, numbers give the number of TMSs per domain (or structural repeats)

[2]) PDB code of first protein with specific structure in the database

[3]) abbreviations: MFS – The Major Facilitator Superfamily
NSS – The Neurotransmitter Sodium Symporter Family
SSS – The Solute Sodium Symporter Family
NCS1 – The Nucleobase-Cation-Symport-1 Family
NCS2 – The Nucleobase:Cation Symporter-2 Family
BCCT – The Betaine/Choline/Carnitine Transporter Family
APC – The Amino Acid/Polyamine/Organocation Superfamily
2HCT – The 2-Hydroxycarboxylate Transporter Family
ESS - The Glutamate/Na^+ Symporter (ESS) Family
DAACS - The Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter Family
RND - The Resistance-Nodulation-Cell Division Superfamily
MATE - The Multi Antimicrobial Extrusion Family
FNT- The Formate-Nitrite Transporter Family
ClC - The Chloride Carrier/Channel Family
NhaA - The NhaA Na^+/H^+Antiporter Family
BASS – The Bile Acid/Na^+ Symporter Family
MCF – The Mitochondrial Carrier Family

represent all types of energy coupling, symport, antiport and uniport. Some of the individual members within the MFS show a stringent specificity, yet as a group the superfamily accepts an enormous diversity of substrate types like sugars, polyols, drugs, neurotransmitters, Krebs cycle metabolites, phosphorylated glycolytic intermediates, amino acids, peptides, osmolites, siderophores (efflux), iron-siderophores (uptake), nucleosides, organic anions, inorganic anions, and others.

Before the first 3D structure of a membrane protein was obtained, sequence analysis and topology studies revealed that most transporters had a uniform topology of 12 TMSs connected by hydrophilic loops, and with both N and C termini located in the cytoplasm (119, 136). Often, the N-terminal half of the proteins (TM1-TM6) displayed weak homology to the C-terminal half (TM7-TM12) of the proteins, suggesting that the molecule may have arisen from a gene duplication/fusion event (80, 135). This prediction had implications regarding an underlying structural symmetry of two domains. The studies were confirmed in 2003 with the simultaneously reported high-resolution 3D crystal structures of the glycerol-3P/P_i antiporter GlpT (55) and the lactose/H^+ symporter LacY (1), both of *E. coli*, and later by structures of the drug/H^+ antiporter EmrD (181), the fucose/H^+ symporter FucP (20), again both of *E. coli* and the oligopeptide/H^+ symporter $PepT_{SO}$ of *Shewanella oneidensis* (112) (Table 1) and a lower-resolution structure of the oxalate/formate exchanger OxlT of *Oxalobacter formigenes* (49, 50). MFS transporters consist of a core of two homologous domains each consisting of 6 TMSs inserted into the membrane in the same orientation (parallel topology; Fig. 5A). The interface between two domains creates a central hydrophilic cavity that forms a substrate-binding cleft. The structure supports the alternating access model for substrate translocation in which two domains move relatively to one another thereby opening the hydrophilic cleft containing the substrate alternately to the two sides of the membrane (rocker-switch mechanism (38, 80)).

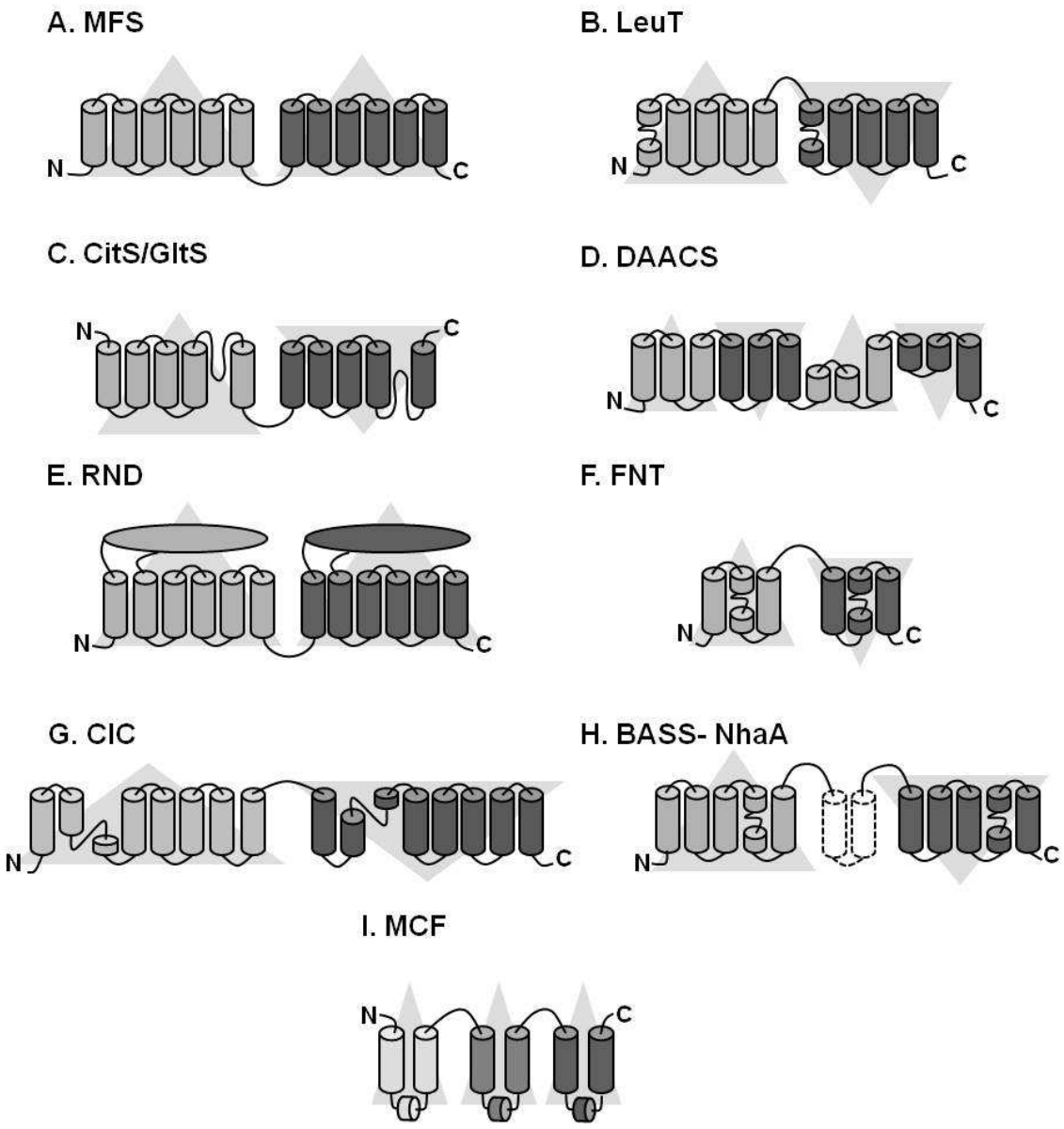


Figure 5. Membrane topology of the "core structure" of secondary transporters with parallel and antiparallel structural repeats. The topologies are shown for (A) the MFS transporters-ST[1], (B) LeuT like transporters – ST[2], (C) CitS/GltS like transporters – ST[3], (D) DAACS transporters - ST[4], (E) RND transporters, (F) FNT transporters, (G) ClC transporters, (H) BASS – NhaA transporters, and (I) MCF transporters. The structural repeats are highlighted by triangles, which emphasize the relative orientation and different intensities of gray. Cylinders represent transmembrane segments (TMSs); non-helical regions in the middle of the TMS are shown as lines; top, extracellular side of the membrane, bottom, cytoplasmic side of membrane.

Exceptions to the 12 TMSs rule exist. A number of MFS protein families consist of 14 TMSs and one other family of 24 TMSs. Examples of families with 14 α-helical spanner proteins are the POT family containing the $PepT_{SO}$ transporter mentioned above (112) and **t**he DHA2 Family containing drug/H^+ antiporters. TetL, the Me^{2+}-tetracycline/H^+ antiporter of *Bacillus subtilis* (60) is in the latter family. The extra two helices are inserted in between the two domains into the central cytoplasmic loop. The core structure of 2x6 TMSs with twofold symmetry is still conserved. Interestingly, the TetL Me^{2+} tetracycline/H^+ antiporter of *B. subtilis* can be converted to a monovalent cation (Na^+, K^+, H^+) antiporter that lacks tetracycline transport activity by deletion of TMSs 7 and 8, the two extra TMSs (161). The 24 TMS MFS transporter, NarK, of *Paracoccus pantotrophus* consists of two 12 TMS domains, NarK1 and NarK2, both of which are required for normal nitrate uptake. NarK1 catalyzes NO_3^-/H^+ symport, dependent on the pmf, while NarK2 catalyzes NO_3^-/NO_2^- antiport, independent of the pmf (177). Thus, the protein is a fusion protein of two homologous but distinct MFS permeases. All members of the MFS superfamily are grouped in structural class ST[1] in the MemGen classification system (see section 4.3) that is based on similarity of family hydropathy profiles..

4.2 LeuT fold (ST[2])

The name "LeuT fold" for this group of secondary transporters comes from the bacterial leucine/alanine transporter LeuT in the Neurotransmitter Sodium Symporter (NSS) family, which was the first crystallized protein with this structure in 2005 (179). By now, a similar fold was found for seven other transporters (Table 1): the Na^+/galactose symporter vSGLT (34) from *Vibrio parahaemolyticus* of The Solute/Sodium Symporter (SSS) family, the benzyl-hydantoin transporter Mhp1 (176) from *Microbacterium liquefaciens* of the nucleobase/cation symporter (NCS1) family, the H^+/uracil symporter UraA (94) from *E. coli* from nucleobase/cation symporter (NCS2) family, the Na^+/glycine betaine symporter BetP (130) from *Corynebacterium glutamicum* and the L-carnitine/γ-butyrobetaine antiporter CaiT (143, 158) from *E. coli* or *Proteus mirabilis*, both members of the betaine-choline-carnitine transporter (BCCT) family, the H^+-coupled amino acid transporter ApcT (147) from *Methanocaldococcus jannaschii* as well as in the arginine/agmatine antiporter AdiC (35, 42) from *E. coli*, both of which are

members of the amino acid-polyamine-organocation (APC) super family. The core helices of all theses proteins from different families share a similar fold but they show little or no sequence similarity. This was noted first by Lolkema and Slotboom in 1998 using a hydropathy profile alignment approach prior to any structure determination (88, 90). All proteins with resolved X-ray structures sharing this type of fold, belong to structural class ST[2] in the MemGen classification system (see section 4.3). Although differing in total number of TMSs, all the 'LeuT fold' transporters share a common core structure of 10 TMSs (Fig. 5B) build by two structurally similar domains of 5 TMSs each. The variation in TMSs numbers, between different families of the "LeuT fold" group of secondary transporters shows 2 TMSs downstream of the core (LeuT and Mhp1), 1 TMS before, and 3 TMSs after core (vSGLT) and 2 TMSs before the core (BetP). The two domains of the core have opposite orientation (antiparallel) in the membrane. The first helix of each of the two domains contains an unwound helical region that is involved in the interaction with the substrate (74). The first two helices of each domain in LeuT come together to form a four-helix bundle that is surrounded by an outer scaffold of helices. Depending on the specific structure, this helix bundle either contains within it (e.g., for BetP), or lines one side of (e.g., for LeuT) the central translocation pathway and contains the binding sites for substrate and co-ions. Although the relative location of a substrate binding sites may be shared among the transporters, the specific interactions vary in order to accommodate the diverse range of substrates. An additional difference between the transporters is the organization of the access pathway for substrate and ion binding sites. Different TMSs participate in the formation of the extracellular and intracellular vestibule. Another difference comes from different sodium to substrate stoichiometries of transport: LeuT and BetP have a 2:1 stoichiometry; and vSGLT and Mhp1 a 1:1 stoichiometry. In the structures of LeuT, two likely sodium sites (Na1 and Na2) were identified, close to the unwound helices. In vSGLT and Mhp1 the Na1 site is absent [50]. One of most significant differences between various "LeuT fold" structures is the orientation of the bundle of 1 and 2 TMSs from each domain to the scaffold of the rest of core TMSs. The angle between the bundle axis and the vertical scaffold TMSs varies from +15º in LeuT to -14º in vSGLT (and Mhp1, BetP and ApcT at intermediate angles) (39).

4.3 CitS and GltS like fold (ST[3])

CitS of *Klebsiella pneumoniae* and GltS of *Escherichia coli* are secondary transporters catalyzing Na^+ symport, and are the main subject of the study described in this thesis. The 3D structures for these two families are still elusive, and for this reason extensive bioinformatic and biochemical studies were carried out. The two proteins belong to different families. The Na^+-citrate transporter CitS belongs to the 2-hydroxycarboxylate Transporter (2HCT) family (reviewed in (149)) while the Na^+-glutamate transporter GltS (22, 62) belongs to the Glutamate Sodium Symporter (ESS) family (Table 1). Members of the 2HCT and ESS families are found exclusively in bacteria. No sequence homology can be detected between members from the two families, but both families are found in the same structural class ST[3] of the MemGen classification system (87-91, 160).

The MemGen classification system groups membrane proteins in structural classes based on hydropathy profile analysis. The hydropathy profile of the amino acid sequence of a membrane protein is taken to be characteristic for the folding of the protein in the membrane. The hydropathy profiles, like the 3D structures of homologous proteins are much better conserved than their amino acid sequence and, therefore, they report on the global fold of the proteins in a family. The MemGen classification system is not a membrane topology prediction method *per se*, but a major consequence of the approach is that all proteins in the different families in one class share the same fold, i.e., knowing the topology of one, is knowing them all. Based on this approach the well-established membrane topology model of the 2HCT family, mostly based on studies of Na^+-citrate transporter CitS (84, 92, 149, 151, 168), was used to predict the membrane topology of Na^+-glutamate transporter GltS, a member of the ESS family. The model was verified by accessibility studies of cysteine residues introduced into the GltS protein (26). The structural model of the transporters shows a core of two homologous domains consisting of five TMSs each that are connected by a large cytoplasmic loop region (Fig. 5C). The CitS protein and all members of the 2HCT family have an additional TMS at the N-terminal end of the core structure, placing the N-terminus in the cytoplasm. Members of the ESS family including GltS do not have this additional segment and their structure corresponds to the core structure that has the N-terminus in the periplasm. Other families

in class ST[3] are characterized by additional TMSs at the C-terminal end. Recently a topology screening study (TopScreen method (160)), was described for structural class ST[3] in MemGen (159). Because of the odd number of helices in the two domains that form the core structure, they have opposite orientations in the membrane. In between the 4^{th} and 5^{th} TMS in each domain, the connecting loop folds back in between the TMSs to form a so-called 'pore loop' or 'reentrant loop'. The reentrant loop in the N-terminal domain enters the membrane from the periplasmic side, the one in the C-terminal domain from the cytoplasm. Sequence motifs GGXG present in the transporters of both the 2HCT and ESS families are at the vertex of the reentrant loops and were demonstrated to be crucial for the activity of the CitS and GltS proteins (25). Additional evidence was presented suggesting that in the 3D structure, the reentrant loops of the two domains are in close vicinity and overlapping at the interface of the two domains (23). It is believed that they form the translocation pore and that translocation proceeds through an alternate access mechanism (150, 151) involving movement of the two domains relative to one another (24). In addition to a similar 3D structure, the CitS and GltS proteins also share the same quaternary structure. Several techniques were used to demonstrate that the proteins are dimeric in the detergent solubilized state (48, 64, 77, 108, 126).

Extensive studies of the CitS and GltS proteins showed that even when lacking an X-ray structure it is possible to obtain good structural information about a membrane protein, by using the MemGen approach to predict the same structural fold for many other families of membrane proteins. In addition to the 2HCT and ESS families, class ST[3] contains over 30 other families of secondary transporters including the ion transporter (IT) superfamily (127). Strong support for MemGen classification was obtained by a similar organization of the core in high-resolution structures of the proteins described in 4.2 containing proteins with a "LeuT fold", all classified in ST[2] (90).

4.4 DAACS fold (ST[4])

The Sodium/Aspartate Symporter from *Pyrococcus horikoshii* (Glt_{Ph}), an archaeal homologue of the EAATs, was one of the first sodium-coupled transporters for which a 3D structure was determined (180). The Glt_{Ph} transporter belongs to the Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) Family (Table 1).

The transporter is a bowl-shaped trimer with a solvent-filled extracellular basin extending halfway across the membrane bilayer. In each Glt_{Ph} protomer the first six transmembrane helices form a distorted cylinder, which in turn encloses a compact core domain containing two reentrant helical hairpins, called HP1 and HP2. The HP1 and HP2 loops, together with flanking regions from TM7 and TM8, are structurally related and can be superimposed, even though HP1 and HP2 have no significant amino acid sequence identity. At first it was considered that an antiparallel topology of duplicated segments was observed only in the helices involved in transport but not in the rest of protein (9, 180). However, a new structure for Glt_{Ph} (131) trapped in an inward-facing conformation by the introduction of a disulfide bridge, showed that, in this state, there is an additional antiparallel topology in the N-terminal region that is not obvious in the outward-facing structure (Fig. 5D). The members of DAACS family are grouped together in structural class ST[4] in the MemGen classification.

4.5 RND fold

AcrB of *E. coli*, a proton dependent multidrug transporter belonging to the Resistance-Nodulation-cell Division (RND) superfamily was the first secondary transporter for which a 3D crystal structure was reported in 2002 (109). Other proteins from this superfamily with known structure are CusA from *E. coli*, a metal ion efflux pomp (93) and the multidrug transporter MexB from *Pseudomonas aeruginosa* (146) (Table 1). All these proteins are subunits of a larger complex that is responsible for the extrusion of toxic compounds over the cell envelope (inner and outer membrane) of Gram-negative bacteria. For example, AcrB forms together with AcrA and outer membrane pore TolC a complex that together export drugs from the cell. The transmembrane part of AcrB shows pseudo-two-fold symmetry: six N-terminal helices are symmetrically arranged with six C-terminal helices and thus form two parallel structurally homologous domains that are different from the MFS domains (Fig. 5E). Two large homologous extracellular domains are inserted in-between the first and second TMSs of the two domains (TMSs I-II and VI-VII). These two periplasmic domains extent up to the TolC pore in the outer membrane.

The AcrB is organized as a trimeric complex with a three-fold symmetry axis perpendicular to the membrane in which periplasmic parts form a central channel ending

in a funnel at the side of TolC. At the opposite side, the channel is connected to a central cavity at the level of the outer leaflet of the bilayer, which contains the substrate binding sites. Three vestibules at the interface of the AcrB protomers provide the access pathway through which the substrates diffuse in via lateral movement from the lipid bilayer (109, 145).

4.6 MATE fold

NorM from *Vibrio cholerae* represents another multi drug resistant protein with a two-domain structure (46). NorM is a member of the Multi Antimicrobial Extrusion (MATE) family (Table 1). MATE transporters are involved in a variety of important biological functions across all kingdoms of life (78, 107, 115). The members of the MATE family are the latest of MDR transporters that were structurally characterized. The NorM transporter consists of 12 TMSs organized like MFS (see section 4.1) and RND transporters in two parallel domains (TMS 1-6 and TMS 7-12), but in contrast to RND transporters they do not contain extracellular domains (Fig. 5 A and E). The two domains of NorM form an internal cavity open to the extracellular space. The topology of NorM is unique among all known transporters, the TMSs are organized in the different way then in other 12 TMSs transporters with two parallel domains.

4.7 FNT fold

The members of the Formate–Nitrite Transporter (FNT) family are predicted to transports low molecular weight acids like formate and nitrate in bacteria, archea, fungi, algae, and parasites but not in higher eukaryotes (137). The mode of energy coupling and substrate transport mechanism used by FNT proteins is unknown. It is even unclear whether they function as channels or transporters (132, 157), a confusion that is reflected in the nomenclature. The founding member of the family is named as a channel (hence the “c” in FocA), but the entire group of proteins is classified as a transporter family (hence the “T” in FNT). The best-characterized FNT member thus far is the above mentioned formate transport protein FocA (Table 1, (157)). High resolution crystal structures of FocA homologues were obtained from three sources, *E. coli* (EcFocA) (172), *Vibrio cholerae* (VcFocA) (171) and *Salmonella typhimurium* (StFocA) (96). The three obtained

structures are in general similar but also some differences were observed (see below). FocA forms a symmetric pentamer with each protomer consisting of six transmembrane segments. Each FocA protomer contains an axial passage that is roughly perpendicular to the plane of the lipid membrane. Interestingly, the pentamer assembly of FocA also contains a central pore, which is probably occupied by lipid molecules within the plasma membrane. The six transmembrane segments of each FocA pentamer subunit have the N and C terminus in the cytoplasm. Each protomer contains an internal structural repeat formed by two homologous domains (Fig. 5F). The N-terminal domain of the protomer, TM1–TM3 is structurally related to the C-terminal domain, TM4–TM6, with a quasi-two-fold axis in the plane of the lipid membrane. Despite a low sequence identity of about 8%, these two halves can be superimposed. In each domain, the middle TMS (TMS2 and TMS5) consists of two lined α-helices connected by an extended loop, which is highly conserved among the FNT family members. These two signature loops, placed roughly parallel to the plane of the membrane, are located in the axial passage of the FocA protomer.

FocA has been proposed to switch its mode of operation from a passive export channel at high external pH to a secondary active formate/H^+ importer at low pH. *Ec*FocA and *Vc*FocA structures were obtained at pH 7.5 and thus represent the high-pH form, in which FocA was suggested to function as a formate channel. The *St*FocA was crystallized at pH 4.0 and the obtained structure shows that the switch between channel and transporter involves a major rearrangement of the amino termini of individual protomers in the pentameric channel. The amino-terminal helices open or block transport in a concerted, cooperative action showing how FocA is gated in a pH-dependent way (96).

Strikingly, the structure of FocA is similar to the structure of proteins belonging to the Aquaporin family, a family of water and glycerol channels. The FocA protomer can be superimposed on that of the *Escherichia coli* water channel AqpZ (140) and the glycerol channel GlpF (40). The similarity is particularly evident in the inverted twofold symmetry, in the total number of transmembrane α-helices (six), in their topology, in the right-handed twist of the helix bundle, and even in the existence of a pore in the middle of the monomer. FocA is an integral membrane protein with no sequence homology with AQPs but they share a similar fold. Despite the overall structural similarity, FocA has

prominent features that distinguish it from AQPs. One of the most important is that the two loops that disrupt TM2 and TM5, have different configurations from those of AQPs were they adopt a reentrant loop structure.

4.8 ClC

CLC proteins form a large family of membrane proteins that transfer chloride ions across cell membranes. Present in all kingdoms of life, existing in both cytoplasmic and intracellular membranes, CLCs mediate a wide range of physiological processes (59, 98). In muscle they govern resting membrane potential, in kidney they facilitate transepithelial fluid flow, and in intracellular compartments they control pH through coupled Cl^-/H^+ exchange (154, 173, 183). Two subgroups of CLC's exist, channels and secondary active transporters (104). Even though channels and transporters catalyze different reactions, conservation of specific amino acids indicates that these functionally distinct CLC subgroups must share the same basic architecture (16, 31, 33). This remarkable finding suggests that the structural boundary separating channels and transporters is not as clear-cut as previously thought (103).

By now structures of four members of CLC family have been determined by X-ray crystallography, two closely related transporters from *Escherichia coli* (EcClC) and *Salmonella typhimurium* (StClC) (28, 29), one eukaryotic CmCLC from a thermophilic red algae *Cyanidioschyzon merolae* (36) and finally a slow ClC Cl^-/H^+ antiporter from *Cyanobacterium* (58) (Table 1).

All CLC proteins are homodimers with each subunit consisting of a transmembrane component that forms the ion pathway. The main differences between the bacterial channels of known structure and eukaryotic ClC channels are the presence of large intracellular domains at the C-terminus of most eukaryotic and some prokaryotic ClC channels (they are absent in EcClC and StClC) (32, 99). The transmembrane part of ClC contains 18 TMSs, which are remarkably tilted and variable in length. The ClC transporter protomers consist of two topologically related domains, which span the membrane in opposite orientations in an antiparallel architecture (Fig. 5G). Each homologues domain contains 8 TMSs, however the second TMS and the third TMS of each domain are short and span the membrane in one direction. Therefore, the

polypeptide chain crosses the membrane an odd number times resulting in an inverted topology. Structurally the two domains are related by a pseudo two-fold axis of symmetry in the plane of the membrane. This pseudo two-fold relationship makes it possible to bring together loops at the end of α-helices from different parts of the structure to form a selectivity filter for Cl^-.

The internal repeat pattern in ClC Cl^- transporters was not previously recognized in the amino acid sequences, but careful alignment with knowledge of the protein structure shows that the two halves are indeed weakly correlated in their sequence, particularly with respect to the position of glycine residues (29).

4.9 BASS and NhaA fold

The Na^+/H^+ antiporter NhaA of *E. coli* and the bile acid/Na^+ symporter ASBT of *Neisseria meningitides* are members of two different families that have no detectable sequence homology. Nevertheless, crystal structures of the two proteins show that they share the same fold.

NhaA is the main Na^+/H^+ antiporter of *Escherichia coli* and other enterobacteria (117) (Table 1). Na^+/H^+ antiporters are integral membrane proteins that exchange Na^+ for H^+ across the cytoplasmic membrane. The functions of these antiporters is regulation of intracellular pH, cellular Na^+ content and cell volume, which are essential processes for all living cells (116, 118, 175). NhaA is indispensable for pH and Na^+ homeostasis in *E. coli* because of its unique ability for 'sensing' the environmental signals Na^+ and H^+ and transducing them into a change in activity so that cellular homeostasis is maintained. The crystal structure of NhaA was obtained in 2005 (56) and has provided insight into the mechanism of NhaA Na^+/H^+ antiporter activity and its unique regulation by pH.

ASBT is a member of the Bile Acid/Na^+ Symporter (BASS) family (Table 1), which is a part of the Bile/Arsenite/Riboflavin Transporter (BART) superfamily (100). ASBT like other members of the BASS family catalyzes transport of bile acid across the membrane in symport with two sodium ions (174). The human ASBT is a pharmaceutical target for drugs aimed at lowering cholesterol, and several ASBT inhibitors have been developed that are effective in animal models (7). The recently reported crystal structure of ASBT

from *Neisseria meningitides* (54) showed a remarkably similar protein architecture as observed previously for the NhaA antiporter (56).

The proteins contain two homologous domains, consist of 5 TMSs each, and therefore have opposite orientation in the membrane. The NhaA antiporter contains two TMSs inserted in between the two homologous domains that are not found in the ASBT protein (Fig 5H), and that increases the number of TMSs to 12. In each domain the fourth TMS is interrupted in the middle (discontinuous) and forms a helical hairpins with the next TMS that is kinked. At the point where the two discontinuous TMSs are broken by well-conserved sequence motifs, they cross over. The unwound stretches that cross in the middle of the membrane form the binding sites for Na^+. Discontinuous transmembrane helices are a common motif in secondary active transporters (144). However, the NhaA and ASBT are the only known example in which such helices cross in this specific way.

Each repeating unit is made of an N-terminal V-motif (first two TMSs) and a core motif (next three TMSs). The core motifs from each repeat form the "core domain/bundle", whereas the two V-motifs create a "panel-like domain/bundle". The movements between these two bundles of TMSs, during transition from outward to inward–facing states, probably form a translocation pathway for substrates.

4.10 MCF fold

Membrane transport proteins found in the inner mitochondrial membrane are collectively called mitochondrial carriers. All of them are members of the same family, the Mitochondrial Carrier (MC) Family (70). The three homologous repeats of about 100 amino acids result in membrane proteins with six TMSs (Table 1). The presence of triplicated helical hairpins in mitochondrial carriers was first identified in the amino acid sequence of the ADP/ATP translocator (71, 139). The structure of the ADP/ATP translocase from bovine mitochondria crystallized in the presence of carboxyatractyloside showed the six TMSs bundle (Fig. 5I) with the helical hairpins related by a pseudo threefold axis (123). The N- and C-termini face the cytoplasmic side of the membrane and the three hairpin loops of the repeats face the matrix. The six TMSs form a compact barrel domain that shows a deep cone-shaped depression at the surface facing the intermembrane space. The ADP/ATP carrier is probably a dimer, which is consistent with

other structural, biochemical and functional data published so far (113). The ADP/ATP translocator is the most abundant and probably the most important carrier in the mitochondrion. It is an electrogenic transporter that exchanges ADP from the cytoplasm with ATP synthesized by F_1F_0-ATPase.

5. Other two-domain membrane proteins

The two-domain organization of membrane proteins is not unique to secondary transporters but is also observed in many channel proteins and also in the translocon of the protein export machinery and the transmembrane part of complex I of the respiratory chain. In this subsection some of these proteins will be briefly presented (Table 2 and Fig. 6).

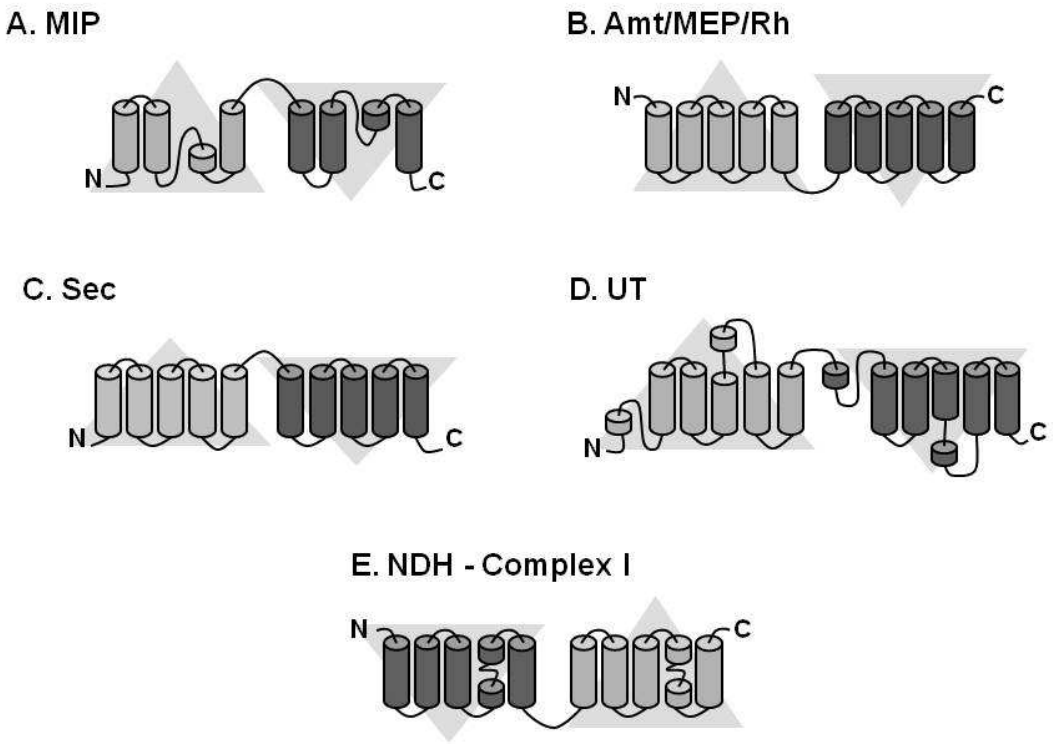


Figure 6. Membrane topology of the "core structure" of 2-domain membrane proteins. The topologies are shown for (A) MIP channels, (B) Amt channels (C) SecY translocon, (D) UT channels, and (E) antiporter like subunits (NuoM) of Complex I of respiratory chain. For further explanation see the legend to figure 5.

The Major Intrinsic Protein (MIP) Family is a large and diverse family of transmembrane channels. The members of MIP function in water, glycerol, urea, NH_3, CO_2, H_2O_2 and ion transport by energy-independent mechanisms and are found ubiquitously in all kingdoms of live organisms (120). The best studded are the water channels, called "aquaporins" and the glycerol channels (Table 2). The aquaporins have an important function in organisms to maintain water homeostasis. Several MIP channels, especially aquaporins, were crystallized and 3D structures were obtained (37, 40, 43, 45, 51-53, 81, 110, 111, 140, 156, 164). Aquaporins are tetrameric. The most important structural feature of the monomer is an internal duplication and each half of the protein consists of 3 TMSs and a reentrant loop (Fig. 6A). The two domains have inverted orientation in the membrane and as a consequence the two reentrant loops, containing important asparagines-proline-alanine (NPA) motifs, fold into the protein from opposite sides and meet in the middle of the bilayer. A similar fold was observed for the subunits of the pentameric FocA protein, but with a different organization of the discontinuous helices (see 4.6).

Table 2. A representative structure of membrane proteins containing two homologous domains or structural repeats.

	"Core" fold [1]	Family/TC#[3]	Protein	Function	PDB code[2]	Ref.
1	↓↑3+3	MIP/1.A.8	AqpO	H_2O channel	1YMG	(45)
2			GlpF1	Glycerol,O_2,NH_3, H_2O channel	1FX8	(40)
3	↓↑5+5	Amt/MEP/Rh 1.A.11	AmtB	NH_3 channel	1U7G	(66)
4			Amt1	NH_3;CH_2NH_2 channel	2B2F	(4)
5			NeRh	NH_3;CO_2 channel	3B9Y	(83)
6			RhC	NH_3 channel	3HD6	(44)
7	↓↑5+5	Sec/3.A.5	SecY	Protein-conducting channel	1RHZ	(167)
8	↓↑6+6	UT/1.A.28	dvUT	Urea channel	3K3F	(82)
9	↓↑5+5	NDH/3.D.1	NuoM	Complex I respiratory chain	3RKO	(30)

The table lists examples of membrane proteins containing two homologous domains. The proteins are grouped according to core structure, based on reported X-ray structures in the PDB database (http://www.pdb.org/pdb/home/home.do). Names of families and TC numbers correspond to the Transporter Classification Database (TCDB - http://www.tcdb.org/)

[1]) arrows represent a parallel or aniparallel orientation of domains (or structural repeats) in the membrane, numbers give the number of TMSs per domain (or structural repeats)

[2]) PDB code of first protein with specific structure in the database

[3]) abbreviations: MIP - The Major Intrinsic Protein Family
Amt/MEP/Rh - The Ammonia Channel Transporter Family
Sec - The General Secretory Pathway Family
UT - The Urea Transporter Family
NDH - The H^+ or Na^+-translocating NADH Dehydrogenase (NDH) Family

Members of Ammonia Transporter Channels (Amt) family are membrane proteins that transport NH_3, NH_4^+, CH_2NH_2 or CO_2 by energy-independent, bidirectional diffusion.

They are involved in the regulation of nitrogen metabolism. The methylammonium permeases (MEPs) from yeasts and the Rh family of proteins from animals belong to the same family and, therefore the family is referred to as Amt/MEP/Rh family (Table 2). The names still reflect the confusion about the transporter/channel status of the proteins. The crystal structures of the ammonium channel from *E. coli* (66, 182), *Archaeoglobus fulgidus* (4), *Nitrosomonas europaea* (83, 97) and a human Rh protein (44) have been determined by X-ray crystallography. Amt channels form a trimer with a separate pathway in each monomer for transport. Each monomer comprises 11 TMSs. The TMSs from 1 to 5 and from 6 to 10 are organized in two structurally homologous domains with inverted topology in the membrane (Fig. 6B). TMS 11 is not the part of structural repeat.

In bacteria, export of proteins produced in the cytoplasm through the inner membrane requires the SecYEG translocon complex (101). The translocon structures were obtained from *Methanococcus jannaschii* (167) and *Thermotoga maritima* (165, 184). The SecY channel-forming subunit has 10 TMSs that are arranged in two structurally related domains; TMS1 to TMS 5 and TMS6 to TMS10 are related by a pseudo twofold symmetry axis parallel to the membrane (Fig. 6C). The detailed arrangement of the helices in SecY is unique between other proteins with 2 times 5 TMSs. Subunits SecG and SecE are located at the periphery of the complex, perhaps to provide structural stability to SecY, which is believed to undergo fairly large structural changes during protein export.

Members of The Urea Transporter (UT) Family, which have been found in animals, fungi and bacteria (105) transport urea with channels-like mechanisms. The structure of urea channel from *Desulfovibrio vulgari* (dvUT – Table 2) was obtained by Levin and co workers (82). In the dvUT trimer, each protomer contains two homologous domains of 6 TMSs, oppositely oriented in the membrane (Fig. 6D). The first TMS of each domain is short and ends in the middle of the membrane. The loop connecting this short helix with next TMS turns sharply and exits on the same side of the membrane. Therefore, the polypeptide chain crosses the membrane an odd number of times resulting in an inverted topology. The next four TMSs from each domain span the entire membrane. The last TMSs of each domain are perpendicular to the membrane and unwind at the middle of the membrane into an extended coil to cross the membrane. Interestingly, a mammalian

urea transporter UT-A1 consists of two tandem copies of UT sequences resulting in a total of 20 TMSs (105) like was observed for the NarK, of *Paracoccus pantotrophus* (see section 4.1).

Finally, the last examples of 2-domain membrane proteins are the subunits of complex I of the respiratory chain of *E. coli* the structure of which was reported at a resolution of 3.0 Å resolution (30). The complex includes six subunits, NuoL, NuoM, NuoN, NuoA, NuoJ and NuoK, with a total of 55 TMSs. The fold of the homologous antiporter-like subunits L, M and N is novel, with two inverted structural repeats of five TMSs arranged, unusually, face-to-back. Each repeat includes a discontinuous forth TMS and forms half of a channel across the membrane (Fig. 6E).

6. Summary

The growing number of reported 3D structures of membrane proteins provides new insights into molecular organization and functionality of membrane proteins. It is clear that many of them contain two homologous domains or structural repeats, a feature that is especially widespread among secondary transporters. A few topology models of two-domain membrane proteins were identified not by crystallography, but by bioinformatics and biochemical studies, such as those described in section 4.3 for the members of 2HCT and ESS family. Another example of two-domain protein recognized by bioinformatic tools is, not mentioned above, the Ca^{2+}/Cation Antiporter (CaCA) Family (TC #2.A.19) (14).

The relatively high number of secondary transporter genes found in sequenced genomes that are grouped in a large number of families, is likely to represent a smaller number of unique structures and, consequently, transport mechanisms. It is clear that different genes families with similar functions may represent homologous 3D structures. The motif of two homologous domains present in most of the secondary transporters evolved by a gene duplication and fusion. There is an enormous variety of gene duplication both in the numbers of TMSs that are duplicated and in their relative orientation, parallel or antiparallel in the membrane. The most primitive transporters possibly consisted of multimers of identical domains, presumably pairs of TMSs as seen in subunit c of the F-type ATPases (102). It is logical to imagine that from a simpler structure of multiple

identical subunits, evolution favored a reduction in subunit number by gene fusion together with an increase in the complexity of the sequence, therefore allowing greater specificity or efficiency of a given protein transporter but also functional diversity (170). Other feature observed in the structures of secondary transporters is the discontinuity in the middle TMSs and the presence of reentrant loops.

Now that quite a number of folds of secondary transport proteins are known, the goal is to solve the structure of the proteins in each step of the transport cycle to fully understand the different transport mechanisms.

7. Outline of the thesis

The work described in this thesis focuses on the investigation of structural similarity between two families of secondary transporters and at the same time on the experimental validation of the MemGen classification system. The main transporter proteins in the studies described in this thesis are CitS of *Klebsiella pneumoniae*, a Na^+-citrate transporter that belongs to the 2-hydroxycarboxylate transporter (2HCT) family and GltS of *Escherichia coli*, a Na^+-glutamate transporter that belongs to the Glutamate Sodium Symporter (ESS) family. Members of these two families are not related in amino acid sequence but share similar hydropathy profiles, which suggest a similar folding of the proteins in the membrane. Starting from the bioinformatics data obtained from the MemGen classification system and a previous topology study of CitS we were able to predict and confirm experimentally the membrane topology of GltS and in this way show that analysis by hydropathy profiles of membrane proteins is a powerful tool to study structures of membrane proteins in the absence of X-ray structural data.

The aims of the research described in this thesis were (i) verifying experimentally that the CitS and GltS share a similar folding in the membrane, with special emphasis on two trans reentrant loops, (ii) showing that GGXG sequence motifs present in the putative reentrants loops are important for the activity of the transporters, (iii) prove that two reentrant loops in N and C domains of the transporters interact at the interface of the two domains and form a translocation pathway for substrates and co-ions, and (iv) demonstrate that genetic engineering allows the reconstruction and manipulation of evolutionary pathways of two-domain membrane proteins.

Chapter 1 reviews secondary transporters with known structure and consisting of two homologous domains or structural repeats. Additionally, the possible evolutionary pathways of these proteins are discussed.

In **chapter 2**, we describe the first detailed experimental support for the MemGen classification scheme. The known structural model of the 2HCT family was used to predict the membrane topology of the ESS family. In the model, the transporters fold into two domains containing five transmembrane segments and a reentrant loop each. The two reentrant loops enter the membrane-embedded part of the proteins from opposite sides of the membrane. The model was verified by accessibility studies of cysteine residues in single-Cys mutants of the Na^+-glutamate transporter GltS of *Escherichia coli*, a member of the ESS family. Importantly, two cysteine residues in the predicted reentrant loop entering the membrane from the cytoplasmic side were shown to be accessible for small, membrane-impermeable thiol reagents from the periplasm, as was demonstrated before for the Na^+-citrate transporter CitS of *Klebsiella pneumoniae*, a member of the 2HCT family. The data strongly suggests that GltS of the ESS family and CitS of the 2HCT family share the same fold as was predicted by comparing the averaged hydropathy profiles of the two families.

Chapter 3 reports a study of GGXG sequence motifs present in the putative reentrant loops of the CitS and GltS proteins. Experiments show that the motifs are important for the activity of the transporters. Mutation of the conserved Gly residues to Cys in the motifs of the Na^+-citrate transporter CitS in the 2HCT family and the Na^+-glutamate transporter GltS in the ESS family resulted in strongly reduced transport activity. Similarly, mutation of the variable residue "X" to Cys in the N-terminal half of GltS essentially inactivated the transporter. The corresponding mutations in the N- and C-terminal halves of CitS reduced transport activity down to 60 and 25% of wild type activity, respectively. Importantly, the X to Cys mutation (S405C) in the cytoplasmic loop in the C-terminal half of CitS rendered the protein sensitive to the bulky, membrane impermeable thiol reagent AMdiS added at the periplasmic side of the membrane,

providing further evidence that this part of the loop is positioned between the TMSs. The putative reentrant loop in the C-terminal half of the ESS family does not contain the GGXG motif, but a conserved stretch rich in Gly residues. Cysteine-scanning mutagenesis of a stretch of 18 residues in the GltS protein revealed two residues important for function. The data support, in general, the structural and mechanistic similarity between the ESS and 2HCT transporters and, more particularly, the two-domain structure of the transporters and the presence and functional importance of the reentrant loops present in each domain. It is proposed that the GGXG motifs are at the vertex of the reentrant loops.

Chapter 4 describes studies verifying the hypothesis that the reentrant loops in the N and C domains in the structural model of the 2HCT interact at the interface of the two domains. This was validated by cross-linking studies using a split transporter approach. It was shown that a CitS variant genetically split in between the two domains forms a stable complex in the membrane that is active in Na^+-coupled citrate transport and that retains its quaternary structure after purification from the membrane. In the purified complex, the two domains could be chemically cross-linked by disulfide bond formation between cysteine residues positioned in the two reentrant loops. A model is proposed in which the reentrant loops in the N and C domains are overlapping at the domain interface in the 3D structure, where they form (part of) the translocation pathway for substrate and co-ions.

Chapter 5 of this thesis deals with a plausible evolutionary pathway for two-domain membrane proteins. The proteins are believed to have evolved by gene-duplication and fusion events. By genetic manipulations we constructed a set of GltS versions corresponding to different evolutionary states: two types of gene pairs encoding domains as separated proteins forming antiparallel heterodimers, and a swapped domain GltS corresponding to a two-domain protein but with reverse order of the domains found in wild type GltS. All artificial evolutionary states were active supporting the proposed evolutionary pathway.

Chapter II

Membrane topology prediction by hydropathy profile alignment: membrane topology of the Na^+-glutamate transporter GltS

Adam Dobrowolski, Iwona Sobczak-Elbourne and Juke S. Lolkema

Published in *Biochemistry* (2007) **46**: 2326-2332

Chapter II

Abstract

Structural classification of families of membrane proteins by bioinformatics techniques has become a critical aspect of membrane protein research. We have proposed hydropathy profile alignments to identify structural homology between families of membrane proteins. Here, we demonstrate experimentally that two families of secondary transporters, the ESS and 2HCT families indeed share similar folds. Members of the two families show highly similar hydropathy profiles, but cannot be shown to be homologous by sequence similarity. A structural model was predicted for the ESS family transporters based upon an existing model of the 2HCT family transporters. In the model, the transporters fold into two domains containing 5 transmembrane segments and a reentrant or pore-loop each. The two reentrant loops enter the membrane embedded part of the proteins from opposite sides of the membrane. The model was verified by accessibility studies of cysteine residues in single-Cys mutants of the Na^+-glutamate transporter GltS of *Escherichia coli*, a member of the ESS family. Two cysteine residues in the predicted reentrant loop entering the membrane from the cytoplasmic side were shown to be accessible for small, membrane impermeable thiol reagents from the periplasmic side of the membrane as was demonstrated before for the Na^+-citrate transporter CitS of *Klebsiella pneumoniae,* a member of the 2HCT family. The data strongly suggests that GltS of the ESS family and CitS of the 2HCT family share the same fold as was predicted by comparing the averaged hydropathy profiles of the two families.

Introduction

Historically, hydropathy profiles of the amino acid sequence have played an important role in membrane protein research. In 1982, the proposal of Kyte and Doolittle for a simple method for displaying the hydropathic character of a protein resulted in the first widely used secondary structure prediction programs for membrane proteins (79). The alternating hydrophobic and hydrophilic regions in the profiles that correspond to the transmembrane α-helical segments (TMS) and connecting loop regions of the protein, respectively, provided a membrane topology model of the protein which is a basic aspect of its structure. In the past, we have explored the idea that, in addition to secondary structure information, hydropathy profiles also contain tertiary structure information in the sense that a hydropathy profile is characteristic for a specific fold of the protein in the membrane (88). Within families of membrane proteins, the hydropathy profiles of different members are strikingly similar even though sequence identity may be as low as 20-25 %. Apparently, the hydropathy profiles, like the 3D-structures of homologous proteins are much better conserved than their amino acid sequence and, therefore, they report on the global fold of the proteins in a family. Examining hydropathy profiles provides a mechanism to identify distantly related membrane proteins even when sequence identity is too low to detect homology. This has led to the MemGen classification in which families of membrane proteins are grouped in structural classes by comparing the averaged hydropathy profiles of the families (87, 88). Two classes of secondary transporters have been characterized in detail by the MemGen approach. Class ST[3] groups 33 families that together contain 2051 sequences and class ST[4] that contains 399 sequence distributed over 2 families (87, 91). The relevance of a structural classification scheme for membrane proteins is evident when considering the high sequence diversity of, for example, secondary transport proteins that results in over 100 families in the Transporter Classification (TC) system (134).

The MemGen classification system is not a membrane topology prediction method *per se*, but a major consequence of the approach is that all proteins in the different families in one class share the same fold, i.e. knowing the topology of one is knowing them all. This provides a mechanism to validate the principle that hydropathy profiles represent a certain fold. Experimental support for the method came from an analysis of the

membrane proteome of *Escherichia coli* in which the cellular location of the C-terminus of all membrane proteins was determined ('in' or 'out') (19). The C-terminal location of all 19 *E. coli* proteins found in structural class ST[3] in the MemGen classification (covering 7 different families) was correctly predicted based upon the structural model of the 2-hydroxycarboxylate transporter (2HCT) family that is also in ST[3] (84).

In this study, we present the first detailed experimental support for the MemGen classification scheme. Prominent features of the structural model of the 2HCT family are two homologous domains that are oppositely oriented in the membrane (inverted topology) and each containing a reentrant or pore-loop structure (reviewed in (149)). The model was used to predict the membrane topology of the ESS family in class ST[3], a family of Na^+ dependent glutamate transporters. No significant (local) sequence identity is detected in a BLAST search (3) between members of the two families. Importantly, the MemGen model differs from the model predicted by TMHMM (75) which in a recent survey was shown to perform well relative to a number of other membrane topology predictors (18). Here, we demonstrate by experiment that the model based on the MemGen classification is the right model.

Experimental Procedures

Bacterial strains, growth conditions and GltS constructs. *Escherichia coli* strains DH5α and ECOMUT2 (41, 163) were routinely grown in Luria Bertani Broth (LB) medium at 37^{o}C under continuous shaking at 150 rpm. When appropriate, the antibiotics ampicillin and chloramphenicol were added at final concentrations of 50 and 30 μg/ml, respectively. All genetic manipulations were done in *E. coli* DH5α while the GltS protein was expressed in *E. coli* ECOMUT2 harbouring plasmid pBAD24 derivatives coding for wild type or cysteine mutants of GltS extended with 6 histidine residues at the N-terminus (His-tag). Production of the GltS proteins was induced by addition 0.1% arabinose when the optical density of the culture measured at 660 nm (OD_{660}) reached a value of 0.6. The Cys-less version of GltS and the single-Cys mutants of GltS were constructed by PCR using the QuickChange Site-Directed Mutagenesis kit (Stratgene, La Jolla, Ca, USA). All mutants were sequenced to confirm the presence of the desired mutations (ServiceXS, Leiden, The Netherlands).

Transport assays in RSO membranes. *E. coli* ECOMUT2 cells expressing GltS variants were harvested from a 1 L culture by centrifugation at 10,000 x g for 10 min at 4 ^{o} C. Right-side-out (RSO) membrane vesicles were prepared by the osmotic lysis procedure as described (61). RSO membranes were resuspended in 50 mM KPi pH 7, rapidly frozen and stored in liquid nitrogen. Membrane protein concentration was determined by the DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA). Uptake by RSO membranes was measured by the rapid filtration method in 50 mM KPi pH 6.0 containing 70 mM NaCl at 30 ^{o}C as described (151). The membranes were energized using the K-ascorbate/phenazine methosulfate (PMS) electron donor system (73).

Partial purification of GltS derivatives by Ni^{2+}-NTA affinity chromatography. *E. coli* ECOMUT2 cells expressing single-Cys mutants of GltS were harvested from a 200 mL culture by centrifugation at 10,000 x g for 10 min at 4 ^{o} C. The cells were washed once with 50 mM KPi pH 7.0 buffer and, subsequently, resuspended in 2 ml of the same buffer. The cells were broken by a Soniprep 150 sonicator operated at amplitude of 8 μm by 9 cycles consisting of 15 sec ON and 45 sec OFF. Cell debris and unbroken cells were removed by centrifugation at 9,000 rpm for 10 min. Membranes were collected by ultracentrifugation for 25 min at 80,000 rpm at 4^{o}C in a Beckman TLA 100.4 rotor and washed once with 50 mM KPi pH 7.0. The His-tagged GltS derivatives were partially purified from the cytoplasmic membranes using Ni^{2+}-NTA affinity chromatography as follows. The membranes (4 mg/ml) were solubilized in 50 mM KPi pH 8, 400 mM NaCl, 20% glycerol and 1% Triton X-100 followed by incubation on ice for 30 min. Undissolved material was removed by ultracentrifugation at 80,000 rpm for 25 min at 4^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (50 μl bed volume per 5 mg protein), equilibrated in 50 mM potassium phosphate pH 8.0, 600 mM KCl, 10% glycerol, 0.1% Triton X-100, 10 mM imidazole and incubated overnight at 4 ^{o}C under continuous shaking. Subsequently, the column material was pelleted by pulse centrifugation and the supernatant was removed. The resin was washed with 10 volumes of equilibration buffer containing 300 mM KCl and 40 mM imidazole. The protein was

eluted with half a bed volume of the washing buffer but containing 150 mM imidazole. The eluted fraction was stored at –20 ^{o}C until use.

Treatment with thiol reagents. Stock solutions of MTSES, MTSET, NEM and AMdiS were prepared freshly in water. The treatment of the different reagents was stopped by addition of an equal concentration of dithiothreitol (DTT) in case of NEM and AMdiS, and L-cysteine in case of the MTS-reagents. The presence of DTT or L-cysteine did not affect the uptake rate in control experiments.

RSO membranes at a concentration of 1 mg/ml were treated for the indicated times and at the indicated temperatures with the thiol reagents in 50 mM KPi pH 7.0. Following treatment, RSO membranes were diluted twice into 50 mM KPi pH 5.0 containing 140 mM NaCl. The pH of the resulting suspension was 6.0 and the suspension was immediately used for uptake measurements.

E. coli ECOMUT2 cells expressing single-Cys mutants of GltS from a 400 mL culture were washed and resuspended in 4 mL 50 mM KPi pH 7.0. Half of the cells were sonicated (see above) in the presence of 1 mM AMdiS, followed by incubation for 30 min at 30 ^{o}C and quenching with 1 mM DTT. The other half was treated with 1 mM AMdiS for the same time and at the same temperature, quenched with DTT, and then sonicated. Control experiments were done in an identical manner but omitting the AMdiS treatment. The GltS derivatives were partially purified using Ni-NTA affinity chromatography (see above) and treated with 0.1 mM FM for 5 min at 20 ^{o}C. The reaction was stopped with 0.5 mM DTT. Samples of 25 µl volume were mixed with SDS sample buffer and run on a 12% SDS-PAGE gel. Fluorescence of proteins labeled with FM was visualized on a Lumi-Imager F1 imager (Roche Diagnostic GmbH, Mannheim, Germany) by irradiation with UV light using a 520 nm filter. All samples containing FM were kept out of bright light until the gel was exposed. After exposure, the gel was stained with Coomassie Brilliant Blue (CBB) to compare the GltS protein levels. Quantification was done using the LumiAnalyst 3.1 software package supplied by Roche Diagnostic (Roche Diagnostic GmbH, Mannheim, Germany).

Computational methods. A structural class in the MemGen classification contains a subset of the entries in the NCBI protein database (http://www.ncbi.nlm.nih.gov/entrez/) that are stored locally in the MemGen database (http://molmic35.biol.rug.nl/). Building of a structural class has been described before in detail (87, 91) and involves a combination of BLAST searches (3), multiple sequences alignments (162) and hydropathy profile alignments (88). The [st324]ESS family in class ST[3] contains 96 entries, 76 of which represent unique proteins. A set of 27 sequences shows no pair wise sequence identity higher than 60% with any of the other sequences in the set, which prevents a bias towards very similar sequences. This set was used to construct the multiple sequence alignment and the family profile in Figure 1. The sequences are listed in the 'multiple sequence alignment' section in the MemGen database.

Materials. The methanethiosulfonate (MTS) derivatives MTSET and MTSES were purchased from Anatrace Inc. (Ohio, USA). NEM was purchased from Sigma-Aldrich BV (Zwijndrecht, The Netherlands), AMdiS and FM, were purchased from Molecular Probes Europe BV (Leiden, The Netherlands).

Results

Structural model of GltS of Escherichia coli by MemGen. The Na^+-glutamate transporter GltS of *E. coli* is a member of the [st324]ESS family in the MemGen classification scheme. The ESS family and the 2HCT family ([st326]2HCT) are found both in structural class ST[3] of the classification (87). A detailed structural model of the transporters in the latter family is available (149). Optimal alignment of the averaged hydropathy profiles of the two families resulted in a similarity score (88) of 0.8, indicating very similar profiles (Fig. 1). The alignment allows for a projection of the 2HCT structural model on the ESS family. The transporters of the 2HCT family contain 11 transmembrane segments (TMS) with the N-terminus in the cytoplasm. Two times five segments (TMSs II-VI and TMSs VII-XI) form two homologous domains that have opposite orientations in the membrane. The two domains are connected by a large hydrophilic loop. In between the 4^{th} and 5^{th} TMS of each domain the connecting loop

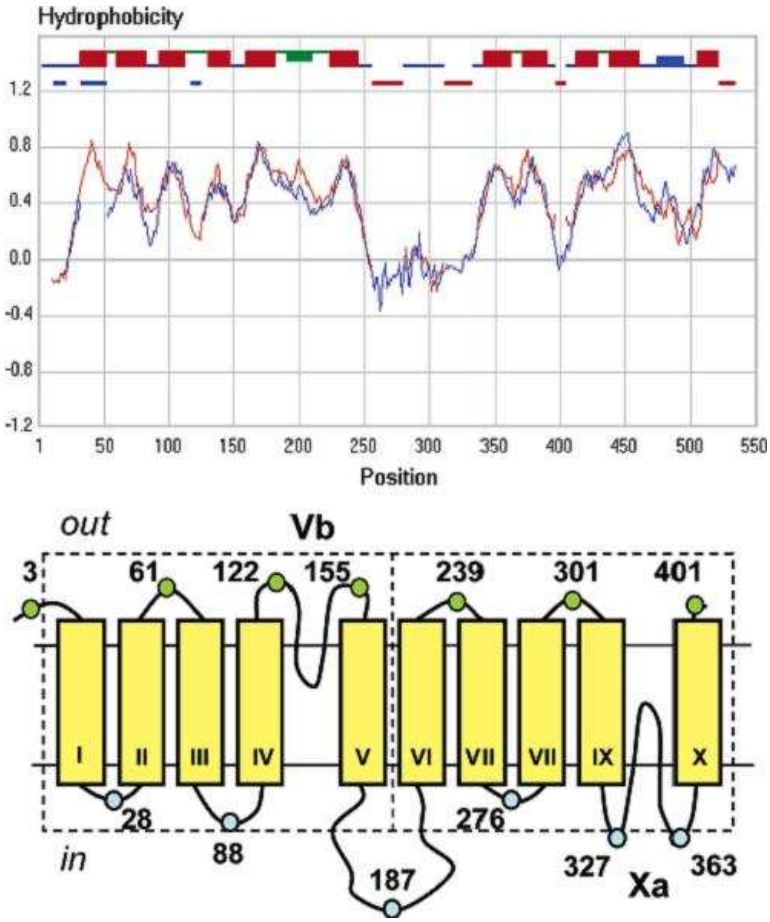


Figure 1. Topology model of the ESS family by MemGen. (top) Hydropathy profile alignment of the family profiles of the 2HCT family (red) and the ESS family (blue). The profiles of the 2HCT and ESS families represent 23 and 27 members, respectively, with pair wise sequence identities between 20 and 60%. SDS values (88) of the profiles were 0.114 and 0.117, respectively. The alignment resulted in a similarity score (88) of 0.8. The membrane topology model of the 2HCT family was indicated in the upper part. Trans membrane segments (red boxes), cytoplasmic loops (blue lines), periplasmic loops (green lines). Thickened parts of loop regions indicate the positions of reentrant loops. Horizontal blue and red lines indicate positions of gaps introduced by the algorithm in the alignment in the blue and red profiles, respectively. (bottom) Membrane topology model of the ESS family based on the profile alignment. Dashed boxes indicate two homologous domains with inverted orientation in the membrane. Vb and Xa correspond to reentrant loop structures. Circles in the loop regions indicate positions for which the location was determined. The numbers correspond to the residues in the GltS protein.

folds back between the TMSs forming a so-called reentrant or pore-loop. The reentrant loop in the N-terminal domain (loop Vb) enters the membrane embedded part from the periplasmic side, the one in the C-terminal domain (loop Xa) from the cytoplasmic side of the membrane. The two loops are believed to contact each other in the three dimensional structure where they would form the translocation path. The N-terminal TMS is not part of the two domain structure and seems to form a separate domain by

itself. The alignment of the two family hydropathy profiles shows that this N-terminal transmembrane segment is absent in the ESS transporters. The resulting model for GltS of *E. coli* consists of two domains containing five TMSs and one reentrant loop structure each (Fig. 1). Both the N- and C-termini are located in the periplasm. The model corresponds to the core structure of the ST[3] proteins (92).

Cloning of the gltS gene of E. coli and construction of set of single-Cys mutants. The *gltS* gene of *E. coli* was cloned into vector pBAD24, yielding pBADHNGltS that codes for the GltS protein extended with 6 histidine residues at the N-terminus under control of the arabinose promoter. A Cys-less version of the GltS protein was constructed by mutating the four native cysteine residues one by one into serine residues. Transport activity of GltS with L-[^{14}C]-glutamate as the substrate was measured in right-side-out (RSO) membrane vesicles prepared from *E.coli* ECOMUT2 cells which contain a copy of the *gltS* gene on the chromosome (163). Thus, the background activity was estimated in RSO membranes containing the Na^+-citrate transporter CitS of *Klebsiella pneumonia* from the 2HCT family that was produced from exactly the same expression system (pBADCitS) and which does not transport L-glutamate (5). RSO vesicles containing recombinant GltS showed at least a five times higher uptake activity in comparison to RSO vesicles containing only endogenous GltS (Fig. 2, ● and ■). RSO vesicles containing the Cys-less vesrsion of the GltS protein showed a comparable level of uptake activity as the wild type GltS protein. (Fig. 2, ▲). A set of 15 single-Cys mutants was constructed, using site-directed mutagenesis and the Cys-less version of the *gltS* gene as a template, to probe the membrane topology of the protein. The positions of the mutations were selected in putative loop regions. Uptake activities of the single-Cys mutants measured in RSO membrane vesicles were comparable to the wild type protein, indicating that the mutations did not significantly affect the folding of the proteins in the membrane (not shown). The positions of the cysteine residues introduced in the protein were indicated in Fig. 1 and correspond to the following mutants: H3C, S28C, N61C, R88C, S117C, D122C, S145C, T155C, S187C, A239C S276C, S301C, R327C, R363C and G401C.

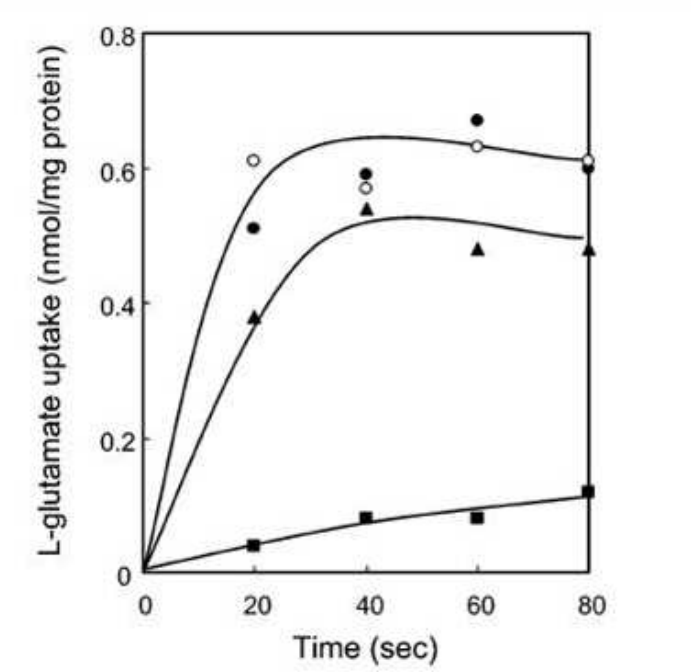


Figure 2. Glutamate uptake activity in RSO membranes. L-[^{14}C]-glutamate uptake was measured in RSO membrane vesicles expressing CitS of *K. pneumoniae* (■), GltS of *E. coli* before (●) and after (○) treatment with 1 mM NEM for 10 min at room temperature and the Cys-less version of GltS (▲).

Accessibility of the cysteine residues in the single-Cys mutants of GltS from either side of the membrane. Accessibility of the cysteine residues introduced into the GltS protein from the water phase was shown by sonicating cells expressing the mutants in the presence and absence of 4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMdiS), a membrane impermeable, negatively charged maleimide derivative. Following sonication and quenching of the excess of the AMdiS reagent, the GltS single-Cys mutants were purified from the membrane by Ni-NTA affinity chromatography. To identify labeling with AMdiS in the first step, purified proteins were treated with the fluorescent thiol reagent fluorescein-5-maleimide (FM). Labeling with FM was detected by fluorescence imaging of the gel after SDS-PAGE. The upper rows (FM) in Fig. 3 show the fluorescence image of the gel, and the bottom rows (CBB) show the same part of the gel after staining with Coomassie Brilliant Blue. All mutants that were not treated with AMdiS showed a clear fluorescent band when treated with FM in detergent solution. In contrast, all mutants, except S117C and S145C, were not labeled with FM after treatment with AMdiS during the sonication step, demonstrating the accessibility of the

thiol group to AMdiS when the mutants reside in the membrane. The protein bands in the Coomassie Brilliant Blue stained image show that the lack of fluorescence was not due to loss of the protein. Mutants S117C and S145C were only partly labeled with AMdiS in the first step, which suggests a restricted accessibility. These two mutants that reside both in the Vb region of the sequence were not considered further in this study.

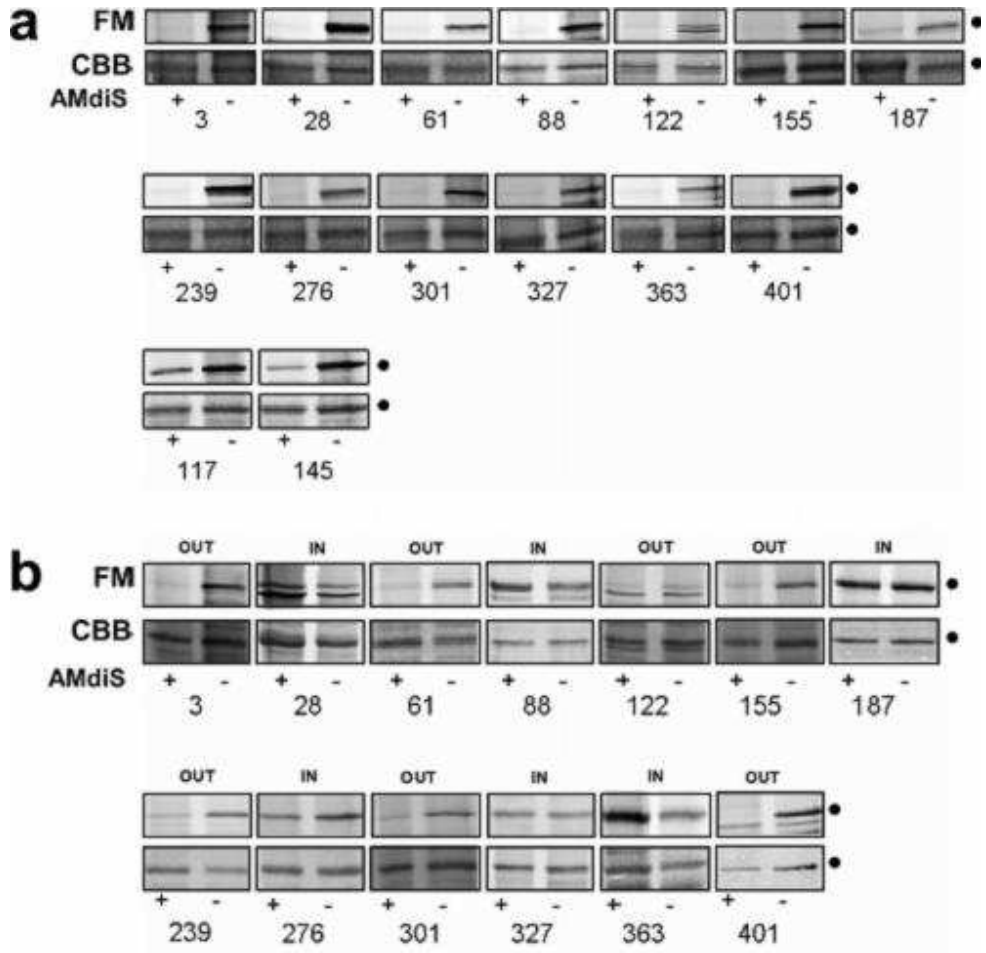


Figure 3. Accessibility of cysteine residues in single-Cys mutants of GltS by AMdiS. Single-Cys mutants were treated with (+) or without (-) 1 mM AMdiS for 30 min followed by the addition of 1 mM DTT. (a) AMdiS was present during the breaking of the cells by sonication and 30 min there after. (b) The AMdiS treatment and quenching with DTT was done before breaking the cells. Following purification, the GltS proteins were treated with 0.1 mM FM for 5 min. Top rows labeled 'FM' show the fluorescence image of the gel and those labeled 'CBB' show the gel after Coomassie Brilliant Blue staining. Numbers indicate the position of cysteine residues in the corresponding single-Cys mutants. Dots mark the position of the GltS protein.

Accessibility of the cysteine residues in single-Cys mutants of GltS from the periplasmic side of the membrane. The sidedness of the cysteine residues introduced into the GltS protein was determined by treating whole cells with AMdiS before sonication. The AMdiS reagent cannot permeate the membrane, and therefore, can reach only residues located at the periplasmic side of the membrane. The labeling by AMdiS was identified as above by treating purified proteins with FM followed by fluorescence imaging. The results for mutants H3C, N61C, D122C, T155C, A239C, S301C and G401C were very similar as in the previous experiment. Treatment with AMdiS inhibited subsequent labeling of the purified protein by FM (Fig. 3b, lanes +), indicating that AMdiS reacted at these sites in whole cells, indicating that these residues are located in the periplasm. Opposite results were obtained for mutants S28C, R88C, S187C, S276C, R327C and R363C. The ratio of fluorescence intensity in AMdiS treated (+ lanes) and untreated (- lanes) samples was similar as the protein intensities of the bands in the Coomassie Brilliant Blue stained gel, indicating that AMdiS did not react with these sites in whole cells. Since the experiment above demonstrated that these sites are accessible from the water phase, it is concluded that these residues are exposed to the cytoplasm. The two groups of mutants are presented as green and blue circles in the model in Fig. 1. The results support the MemGen topology model. The Vb region in between TMSs IV and V and the Xa region in between TMSs IX and X are clearly located in the periplasm and cytoplasm, respectively.

Accessibility of cytoplasmic loop Xa from the periplasm. Treatment of wild type GltS, which contains four cysteine residues, three of which are located in transmembrane segments and one in the Xa region, with the membrane permeable thiol reagent *N*-ethylmaleimide (NEM) did not affect glutamate uptake by RSO membranes (Fig. 2, ● and ○). As expected, the activity of the Cys-less version of GltS was insensitive to the same treatment (not shown). More surprisingly, none of the single-Cys mutants was affected by NEM treatment. Apparently, labeling of the Cys residues did not affect the transport activity. It was shown before that labeling of two endogenous cysteine residues located in reentrant loop region Xa of CitS of *Klebsiella pneumoniae* in the 2HCT family resulted in decreased uptake activity, which emphasized the importance of this part of the

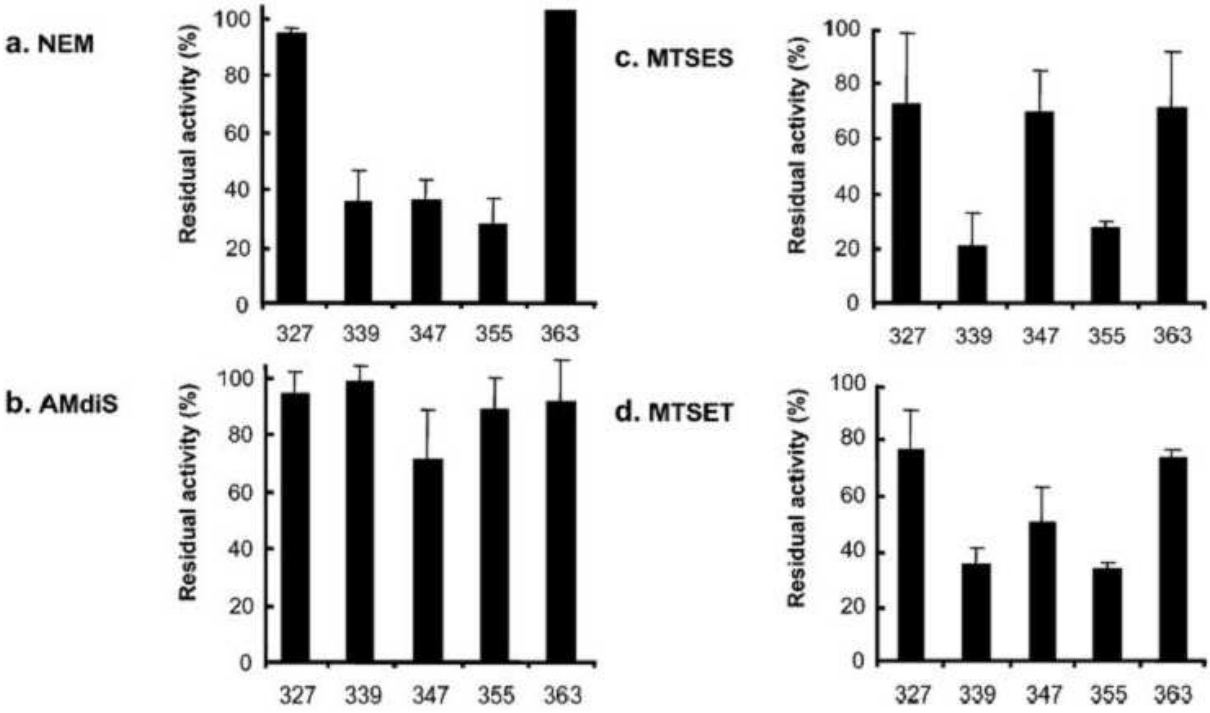


Figure 4. Inactivation of single-Cys mutants of GltS by various thiol reagents. RSO membrane vesicles containing single-Cys mutants R327C, A339C, L347C, A355C and R363C were treated at room temperature with 1 mM NEM for 10 min (a), 0.25 mM AMdiS for 20 min (b), 10 mM MTSES for 20 min (c), and 1 mM MTSET for 10 min (d). Residual uptake activity was plotted as the percentage of the rate obtained with untreated membranes. Bars represent the average and standard deviation obtained from at least three independent measurements.

sequence in catalysis (150), and, ultimately, resulted in the identification of the reentrant loop structure (151). Three additional single-Cys mutants were constructed in GltS yielding, a total of five single-Cys mutants in the Xa region of the GltS protein: R327C, A339C, L347C, A355C and R363C. Uptake activity measured in RSO membranes revealed a comparable level of activity as wild type GltS. Treatment of the RSO vesicles with NEM, which is non-polar and membrane permeable, reduced uptake activity for the middle 3 single-Cys mutants (A339C, L347C, A355C) down to 30-40%, while the activity of the two flanking mutants (R327C and R363C) was not affected (Fig. 4a). About 10-20% of the uptake activity in RSO vesicles is due to the chromosomal copy of wild type *gltS* gene, which is not sensitive to the thiol reagent (Fig. 2). Increasing the NEM concentration and time of treatment did not result in lower activity of the single-Cys mutants suggesting that the labeled molecules still have residual activity. The same observation was made for the endogenous cysteines in the Xa region of CitS of *Klebsiella*

pneumoniae (150, 151). Treatment of RSO membranes containing the 5 single-Cys mutants with membrane impermeable AMdiS did not significantly decrease the uptake activity, which is in line with a cytoplasmic location of the Cys residues (Fig. 4b).

The methanethiosulfonate (MTS) derivatives 2-(trimethylammonium)ethyl methanethiosulfonate bromide (MTSET) and sodium (2-sulfonatoethyl) methanethiosulfonate (MTSES) represent small thiol reactive reagents that react with cysteine residues in proteins to form mixed disulfides (2). MTSET and MTSES differ in the charge of the groups attached to the reactive MTS moiety, MTSET is positively charged and MTSES is negatively charged which makes them membrane impermeable. Treatment of RSO membranes with these reagents has been shown before to compromise the energy generating system in the membranes to a certain level (150). Accordingly, treatment of RSO membranes containing wild type and Cys-less GltS with MTSET and MTSES resulted in a decrease of 20-40% of the uptake activity (not shown). A similar reduction in uptake activity was observed for mutants R327C, L347C, R363C in case of MTSES and for mutants R327C and R363C in case of MTSET. However, treatment of mutants A339C and A355C with MTSES or MTSET resulted in a significantly lower residual activity (Fig. 4c,d). The activities were down to the level observed after treatment with NEM. Treatment of mutant L347C with MTSET resulted in slightly less inactivation while the same mutant was not affected by MTSES. The results demonstrate that residues in the Xa region in the cytoplasmic loop between TMS IX and X are accessible from the periplasmic side of the membrane for small, but not for more bulky membrane impermeable reagents. The same observation was made for the endogenous Cys residues in the Xa region of CitS of *K. pneumoniae* in the 2HCT family (151).

Discussion

The glutamate transporter GltS of *E. coli* is the only characterized member of the ESS family of secondary transporters. The *gltS* gene product was shown be a L-glutamate transporter that also has affinity, albeit much lower, for D-glutamate, α-methylglutamate and homocysteate (21, 22, 62, 163). Glutamate transport was observed only in the presence of Na^+, irrespective of pH, suggesting an obligatory coupling of L-glutamate and Na^+ translocation. GltS is one of three L-glutamate transporters coded on the genome

of *E. coli* K12. The [st324]ESS family in the MemGen classification consists of 76 unique members, all from bacterial origin, mostly from the γ subdivision of the Proteobacteria. The sequences in the family show a remarkably narrow size distribution with most of the sequences consisting of 400-410 residues. Since no significant sequence similarity could be detected with members of other families, the ESS family is not in any of the known superfamilies in the Transporter Classification (TC) system (2.A.27 ESS) (134). In the MemGen classification, the ESS family is in structural class ST[3] together with 32 other families of membrane proteins.

The structural model for the members of the ESS family presented in Figure 1 is based on a different approach than models produced by predictors like TMHMM (75). While the latter rely on general features of membrane proteins, the MemGen method is basically homology modeling, porting information between families in the same structural class. Therefore, the MemGen method relies on the correct classification of families in structural classes. The MemGen method allows for the incorporation of structural details like domain structure and reentrant loop structures, while TMHMM is restricted to loops and transmembrane segments. In a recent analysis in which the performance of transmembrane helix predictors was tested on a database of high resolution structures of membrane proteins (18), Hidden Markov Model based approaches were shown to perform well for proteins consisting of transmembrane helices and loops, but less so when the protein contained structural elements like reentrant loops. For most sequences in the ESS family, TMHMM predicts the putative reentrant loop in the second domain (Xa, Fig 1) to be a transmembrane segment, while this is the case for about half of the putative reentrant loops in the first domain (Vb). This results in models with 9-12 TMS for the different members. TMHMM predicts 9 transmembrane segments for GltS of *E. coli*, 3 in the N-terminal half and 6 in the C-terminal half. The experimental data presented here was according the 10 TMS model predicted by the MemGen method. Cysteine residues placed in the N- and C-termini (H3C and G401C) of the GltS protein of *E. coli* were both localized in the periplasm by the accessibility of the bulky, membrane impermeable thiol reagent AMdiS. The same was observed for five other cysteine residues in mutants N61C, D122C, T155C, A239C and S301C, suggesting the presence of four periplasmic loops. The periplasmic location of both cysteine residues at positions

122 and 155 in region Vb indicates that the region is not transmembrane. Six cysteine residues in single-Cys mutants S28C, R88C, S187C, S276C, R327C and R363C were not accessible for AMdiS in whole cells while they were accessible from the water phase which indicates a cytoplasmic location and 5 cytoplasmic loops. Again, the cytoplasmic location of the cysteine residues in R327C and R363C shows that the Xa region is not transmembrane.

Reentrant loop structures are commonly found in channel proteins like the well studied K^+ channels (27) and aquaporins (110) where they function as selectivity filters. The recently reported crystal structure of a glutamate transporter homologue of the archaeon *Pyrococcus horikoshi* revealed that they may be essential features in secondary transporters as well (180). The reentrant loop structure in the Xa region of the transporters of the 2HCT family is based mostly on experimental studies of the Na^+-citrate transporter CitS of *K. pneumoniae* (150, 151) and, to a lesser extent, the citrate/malate transporter CimH of *Bacillus subtilis* (76). Identification followed from the accessibility of sites in the cytoplasmic loop by small water-soluble thiol reagents from the periplasmic side of the membrane. The access pathway is believed to be the translocation pathway of substrate and co-ions through the protein (149, 151). Using similar criteria, the present study identifies the reentrant loop structure in the corresponding Xa region in the GltS transporter of the ESS family. Labeling of cysteine residues in the Xa region in three single-Cys mutants of GltS, A339C, L347C and A355C with membrane permeable NEM resulted in reduced transport activity. The same sites were not accessible to membrane impermeable AMdiS when added at the periplasmic side of the membrane in RSO membranes, but two of the three, A339C and A355C were clearly accessible for the small membrane impermeable MTSES and MTSET reagents in the same experimental system. The third mutant L347C that was inactivated by NEM was clearly less reactive with MTSES and MTSET, which suggest that not all positions of the Xa region are equally accessible to all reagents. No experimental evidence for the reentrant loop in the N-terminal domain is available for any of the proteins in the 2HCT family or any other family in structural class ST[3]. The reentrant loop is based on weak homology between the N- and C-terminal halves of the proteins in ST[3] detected by bioinformatics tools (92). The reentrant loop regions in both domains are characterized

by the well-conserved sequence motif GGXG which is believed to be at the vertex of the loops (149). The motif is present in the putative reentrant loop sequence in the N-terminal domain of GltS as GGHG at positions 136-139.

The folding of the GltS protein in the membrane in 10 TMS and the identification of the reentrant loop structure in between TMS IX and X strongly indicates a similar fold for the members of the ESS and 2HCT families. Apart from yielding a detailed structural model for the members of the ESS transporter family, the experimental data presented in this study supports the MemGen classification method, which is based on hydropathy profile alignment.

Acknowledgments

This work was supported by a grant from the Dutch Organization for Scientific Research (NWO-ALW).

Chapter III

Functional importance of GGXG sequence motifs in putative reentrant loops of 2HCT and ESS transport proteins

Adam Dobrowolski and Juke S. Lolkema

Published in *Biochemistry* (2009) **48**: 7448-7456

Chapter III

Abstract

Members of the 2HCT and ESS families of secondary transporters are unrelated in amino acid sequence but are believed to share the same fold. Structural models show two homologous domains containing a reentrant or pore-loop each. Here we show that GGXG sequence motifs present in the putative reentrant loop structures are crucial for the activity of the transporters. Mutation of the conserved Gly residues to Cys in the motifs of the Na^+-citrate transporter CitS in the 2HCT family and the Na^+-glutamate transporter GltS in the ESS family resulted in strongly reduced transport activity. Similarly, mutation of the variable residue 'X' to Cys in the N-terminal half of GltS essentially inactivated the transporter, while the corresponding mutations in the N- and C-terminal halves of CitS reduced transport activity down to 75 and 40 % of wild type, respectively. Residual activity of any of the mutants could be further reduced by treatment with the membrane permeable thiol reagent N-ethylmaleimide. The Cys residue in CitS mutant S405C ('X') was accessible to the bulky, membrane impermeable thiol reagent 4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMdiS) from the opposite side of the membrane, providing further evidence for the existence of the reentrant loop structure. The putative reentrant loop structure in the C-terminal halve of the ESS family does not contain the GGXG motif, but rather a conserved stretch rich in Gly residues. Cysteine-scanning mutagenesis of the stretch of 18 residues revealed that mutant N356C was completely inactivated by treatment with NEM, while mutant P351C appeared to be the counterpart of mutant S405C of CitS; the mutant was inactivated by AMdiS added at the opposite side of the membrane. The data supports, in general, the structural and mechanistic similarity between the ESS and 2HCT transporters, and, more particularly, the two domain structure of the transporters and the presence and functional importance of the reentrant loops present in each domain. It is proposed that the GGXG motifs are at the vertex of the reentrant loops.

Introduction

The 2HCT (2-HydroxyCarboxylate Transporter; TC 2.A.24, (134)) and ESS (Glutamate:Na^+ Symporter; TC 2.A.27) families represent families of ion-driven transporter proteins that are exclusively found in the bacterial domain. Members of the two families do not share any significant amino acid sequence similarity but the hydropathy profiles of the sequences are very similar. For this reason, the two families are in the same structural class (ST[3]) in the MemGen classification system that we have introduced to identify membrane proteins sharing the same fold (87-89, 91). The MemGen classification system groups membrane proteins in structural classes based on hydropathy profile analysis. The hydropathy profile of the amino acid sequence of the membrane protein is taken to be characteristic for the folding of the protein in the membrane. Recently, strong support for the MemGen classification was obtained by the similar organization of the core in the high resolution structures of members of the NSS (Neurotransmitter Sodium Symporter, TC 2A.22, (179)), SSS (Sodium Solute Symporter, TC 2A.21, (34)) and NCS1 (Nucleobase Cation Symport 1, TC 2A.39, (176)) transporter families. While the members of the NSS, SSS and SCN1 families do not share sequence similarity, the families are all found in the same structural class (ST[2]) in the MemGen classification (89, 90).

Experimental support for the same fold of the proteins in the 2HCT and ESS families was obtained by demonstrating a similar membrane topology for two transporters from the two families (26). The well-established membrane topology model of the 2HCT family, mostly based on studies of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* (reviewed in reference (149)), was used to predict the membrane topology of the Na^+-glutamate transporter GltS of *Escherichia coli* (22, 62), a member of the ESS family. The model was verified by accessibility studies of cysteine residues introduced into the GltS protein (26). Though membrane topology of a protein represents a low structural resolution, the result is not trivial and does validate the MemGen classification because of specific structural features of these proteins. Secondary structure predictors like TMHMM (75) predict different models for both CitS and GltS that were inconsistent with the experimental data (26).

The structural model of the transporters in the ESS and 2HCT families, and in all 33 families of structural class ST[3] in the MemGen classification, consist of two domains each containing 5 transmembrane segments (Fig. 1A). The total number of transmembrane segments is variable between different families. For instance, the 2HCT proteins have an additional segment at the N-terminus, which is missing in the ESS proteins. In the model, the two domains of 5 TMSs each share a similar fold but have opposite orientations in the membrane (84, 92, 149), a structural motif seen more frequently in membrane proteins (inverted topology; (34, 56, 66, 110, 167, 176, 179). The loops between the 4^{th} and 5^{th} transmembrane segment in each domain are believed to form so-called pore loops or reentrant loops, which fold back in between the transmembrane segments from opposite sides of the membrane (*trans* reentrant loops (84, 92)). The reentrant loops in the N- and C-terminal domains are believed to be in close vicinity in the 3D structure and to form the translocation pathway for substrate and co-ions.

The putative reentrant loop regions, termed Vb in the N-terminal domain and Xa in the C-terminal domain, are well conserved within the families throughout structural class ST[3] and contain a remarkable high fraction of residues with small side chains like Gly, Ala and Ser (149). In almost all families of ST[3] small stretches of highly conserved residues are found in the putative reentrant loop regions Vb and Xa. Sequence analysis of the 138 members of the ESS family and the 74 members of the 2HCT family showed that in the Vb regions of both families and the Xa region of the 2HCT family these conserved stretches contain a GGXG sequence motif in which X is a less conserved residue (see Fig. 1B). The same motif cannot be found in the Xa region of the ESS family but also here, 3 conserved Gly residues are found in a stretch of 8 residues.

Here, we present a mutational study of the sequence motifs GGXG found in the putative reentrant loops of the CitS and GltS proteins to determine the relevance of the motifs for the transport function of the proteins. The functional relevance of the Xa region in GltS in which the motif is not found is addressed by cysteine scanning mutagenesis. It follows that the motifs play an important role in the transport mechanism catalyzed by both transporters and further evidence for the existence of the reentrant loops is obtained. The

corresponding properties of CitS and GltS further support a similar core structure and mechanism for the two transport proteins.

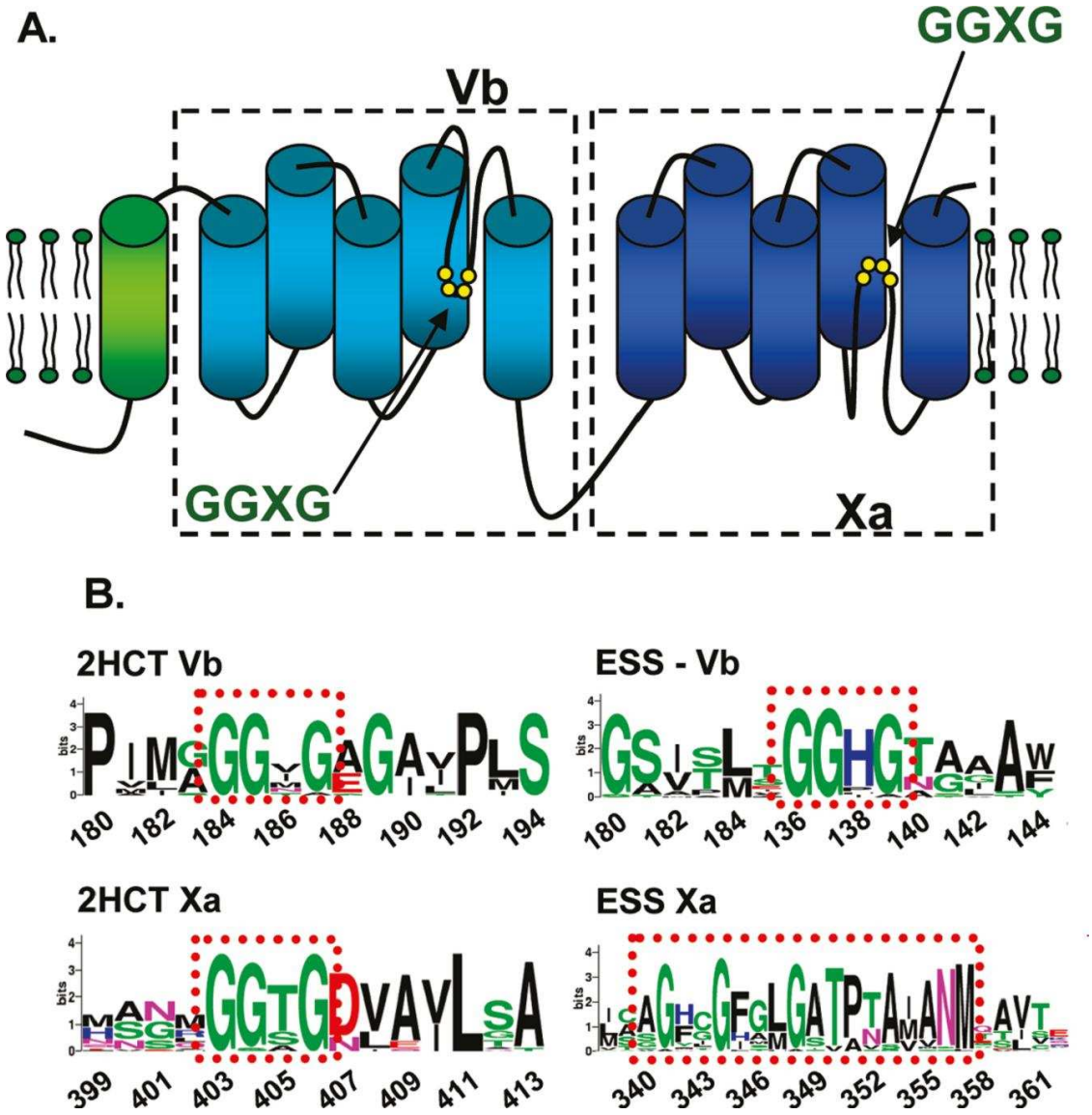


Figure 1. (A) Structural model for the transporters of the 2HCT and ESS families. Two homologous domains containing 5 TMS each with an inverted topology in the membrane were indicated in *dashed boxes*. Each domain contains a reentrant loop structure entering the membrane embedded part of the protein from the periplasmic and cytoplasmic side of the membrane, respectively (Vb and Xa). Members of the 2HCT family have an additional TMS at the N-terminus that is not present in members of the ESS family. (B) Sequence logos of regions Vb and Xa showing the sequence motifs GGXG in the 2HCT (left) and ESS (right) families. Position numbers correspond to the residue numbers in the CitS (2HCT) and GltS (ESS) sequences. Residues mutated into cysteines were indicated in the red dotted boxes. The logos were generated using WebLogo, version 2.8.1 (http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi).

Experimantal procedures

Bacterial strains, growth conditions and GltS and CitS constructs. *Escherichia coli* strain DH5α was routinely grown in Luria Bertani Broth (LB) medium at 37 ^{o}C under continuous shaking at 150 rpm. Ampicillin was used at a final concentration of 50 μg/ml. The GltS and CitS proteins were expressed in *E. coli* DH5α cells harboring plasmid pBAD24 (Invitrogen) derivatives coding for wild type or cysteine mutants of GltS (26) and CitS (150) extended with 6 additional histidine residues at the N-terminus (his-tag). In case of CitS variants a sequence encoding an enterokinase cleavage site was present in between the his-tag and the *citS* gene. Expression of genes cloned in pBAD24 is under control of the arabinose promoter. Production of the GltS and CitS proteins was induced by addition 0.01% arabinose when the optical density of the culture measured at 660 nm (OD_{660}) reached a value of 0.6. The cysteine mutants of GltS and CitS were constructed by PCR using the QuickChange Site-Directed Mutagenesis kit (Stratgene, La Jolla, Ca, USA). All mutants were sequenced to confirm the presence of the desired mutations (ServiceXS, Leiden, The Netherlands).

Transport assays in RSO membranes. *E. coli* DH5α cells expressing CitS or GltS variants were harvested from a 1 L culture by centrifugation at 10,000 x g for 10 min at 4 ^{o}C. Right-side-out (RSO) membrane vesicles were prepared by the osmotic lysis procedure as described (61). RSO membranes were resuspended in 50 mM KPi pH 7, rapidly frozen and stored in liquid nitrogen. Membrane protein concentration was determined by the DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA).

Uptake by RSO membranes was measured by the rapid filtration method. The membranes were energized using the K-ascorbate/phenazine methosulfate (PMS) electron donor system (73). Membranes were diluted to a final concentration of 0.5 mg/ml into 50 mM KPi pH 6.0 containing 70 mM Na^{+}, in a total volume of 100 μl at 30 ^{o}C. Under a constant flow of water-saturated air, and while stirring magnetically, 10 mM K-ascorbate and 100 μM PMS (final concentrations) were added and the proton motive force was allowed to develop for 2 min. Then, L-[^{14}C]-glutamate or [^{14}C]-citrate was added at a final concentration of 1.9 μM and 4.4 μM, respectively. Uptake was stopped by the addition of 2 ml of ice-cold 0.1 M LiCl, followed by immediate filtration over

cellulose nitrate filters (0.45 μm, pore size). The filters were washed once with 2 ml of a 0.1 M LiCl solution and assayed for radioactivity. The background was estimated by adding the radiolabeled substrate to the vesicles suspension after the addition of 2 ml of ice-cold LiCl, immediately followed by filtering.

Partial purification of GltS and CitS derivatives by Ni^{2+}-NTA affinity chromatography. *E. coli* DH5α cells expressing CitS or GltS variants were harvested from a 200 ml culture by centrifugation at 10,000 x g for 10 min at 4 ^{o}C. Cells were washed with 50 mM KPi buffer pH 7 and resuspended in 2 ml of the same buffer and, subsequently, broken by a Soniprep 150 sonicator operated at an amplitude of 8 μm by 9 cycles consisting of 15 sec ON and 45 sec OFF. Cell debris and unbroken cells were removed by centrifugation at 9,000 rpm for 5 min. Membranes were collected by ultracentrifugation for 25 min at 80,000 rpm at 4^{o}C in a Beckman TLA 100.4 rotor and washed once with 50 mM KPi pH 7.0.

His-tagged GltS and CitS derivatives were partially purified from the cytoplasmic membranes or RSO membranes prepared as described above using Ni^{2+}-NTA affinity chromatography as follows. Membranes (4 mg/ml) were solubilized in 50 mM KPi pH 8, 400 mM NaCl, 20% glycerol and 1% Triton X-100 followed by incubation for 30 min at 4 ^{o}C under continuous shaking. Undissolved material was removed by ultracentrifugation at 80,000 rpm for 25 min at 4 ^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (50 μl bed volume per 5 mg protein), equilibrated in 50 mM potassium phosphate pH 8.0, 600 mM NaCl, 10% glycerol, 0.1% Triton X-100, 10 mM imidazole and incubated overnight at 4 ^{o}C under continuous shaking. Subsequently, the column material was pelleted by pulse centrifugation and the supernatant was removed. The resin was washed with 10 volumes of equilibration buffer containing 300 mM NaCl and 40 mM imidazole. The protein was eluted with half a bed volume of the washing buffer but containing 150 mM imidazole. The eluted fraction was stored at –20 ^{o}C until use.

Treatment of RSO membrane vesicles with thiol reagents. Stock solutions of the thiol reagents N-ethylmaleimide (NEM) and 4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMdiS) were prepared freshly in water. The treatment of the reagents

was stopped by addition of an equal concentration of dithiothreitol (DTT). The presence of DTT did not affect the initial rate of uptake in transport assays. RSO membranes at a concentration of 1 mg/ml were treated for the indicated times and at the indicated temperatures with the thiol reagents in 50 mM KPi pH 7.0. Following treatment, RSO membranes were diluted twice into 50 mM KPi pH 5.0 containing 140 mM NaCl. The pH of the resulting suspension was 6.0 and the suspension was immediately used for uptake measurements.

Materials. NEM was purchased from Sigma-Aldrich BV (Zwijndrecht, The Netherlands), AMdiS was purchased from Molecular Probes Europe BV (Leiden, The Netherlands). L-[^{14}C]-glutamate and [^{14}C]-citrate were obtained from Amersham Pharmacia, Roosendaal, The Netherlands.

Results

Sequence motif GGNG in region Vb of CitS. Each of the amino acid residues in sequence motif 184-GGNG-187 found in the Vb region of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* in the 2HCT family was substituted with a cysteine residue. The 4 mutants, G184C, G185C, N186C and G187C were tested for their ability to accumulate [^{14}C]-citrate in right-side-out (RSO) membrane vesicles prepared from *E. coli* DH5α cells expressing the mutants. Citrate uptake was measured in the presence of a proton motive force (pmf) that was generated using the artificial electron donor system K-ascorbate/PMS (Fig. 2A). RSO membrane vesicles prepared from cells not expressing CitS lack citrate uptake activity because of the absence of a citrate transport system in the *E. coli* membrane. Membranes containing mutants G184C and G187C showed a similar uptake activity that was about 10-15 % of the activity of membranes containing wild type CitS, while mutant G185C resulted in complete lack of activity (Fig. 2A). Mutation of the non-conserved position in the motif, N186 to Cys resulted in an uptake activity of approximately 75% of wild type CitS. Protein levels of the mutants in the membranes were estimated by small-scale purifications making use of the N-terminally fused His-tag and Ni-NTA affinity chromatography (Fig. 2B). All mutants showed similar expression levels as observed for wild type CitS indicating that the lower transport activity of the

membranes containing the mutants was a consequence of the mutation in CitS rather than lack of production or degradation of the proteins. It follows that the conserved Gly residues in the GGXG motif at positions 184 and 187, and especially Gly185 appear to be critical for the activity of the protein while mutation of the non-conserved N186 to Cys had only a marginally effect on the specific activity of CitS.

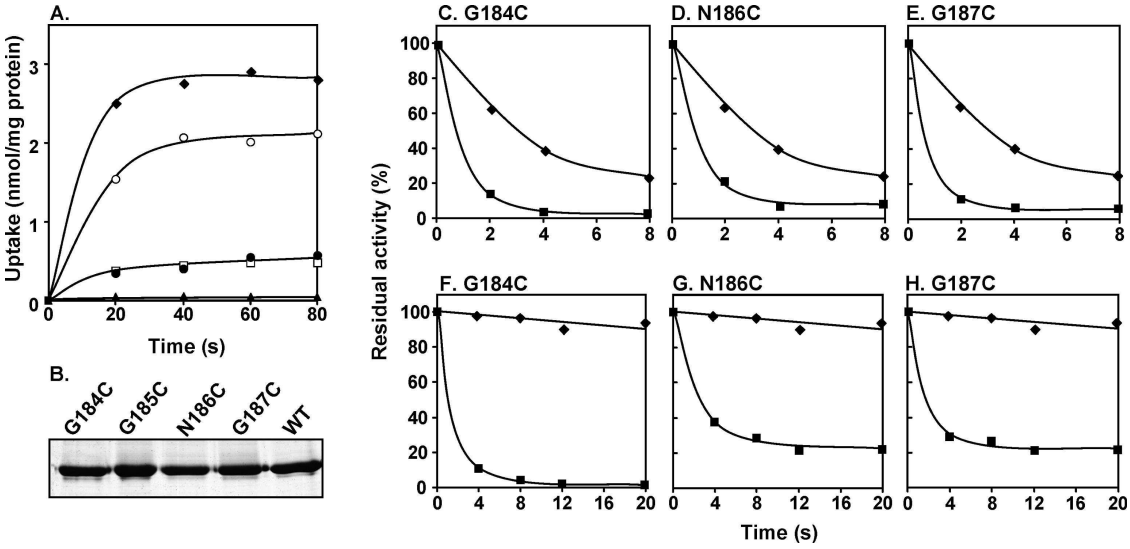


Figure 2. Sequence motif GGNG in Vb of CitS of *K. pneumoniae*. (A) [^{14}C]-citrate uptake in RSO membrane vesicles containing CitS (♦) and the CitS mutants G184C (●), G185C (▲), N186C (○) and G187C (□). (B) SDS-PAGE of partial purified CitS and the G184C, G185C, N186C and G187C mutants purified from the RSO membranes used in the uptake assays shown in panel A. (C-H) Residual activity after treating RSO membranes containing CitS (♦) and the mutants (■) G184C (C,F), N186C (D,G) and G187C (E,H) with 1 mM NEM (C,D,E) or 0.25 mM AMdiS (F,G,H) for the indicated time. Initial rates were expressed as the percentage of the initial rate catalyzed by untreated membranes.

N-ethylmaleimide (NEM) is a small, membrane permeable thiol reagent. Treatment of the RSO membrane vesicles containing wild type CitS, which contains 5 cysteine residues, with 1 mM NEM results in a slow inactivation of the protein down to 10%-20% of wild type activity with a half-time of inactivation of about 4 min ((150); Fig. 2C,D,E). In contrast, the three mutations in the GGXG motif that resulted in CitS proteins with residual activity (G184C, N186C and G187C) rendered the proteins highly sensitive to NEM with inactivation half-times shorter than 1 min. Moreover, treatment with NEM resulted in lack of significant residual activity (Fig. 2C,D,E).

The site of reaction of wild type CitS with NEM are two cysteine residues in cytoplasmic region Xa (150). Consequently, wild type CitS is not inactivated by AMdiS, a maleimide derivative containing a bulky, negatively charged group that, in contrast to NEM, is membrane impermeable and cannot reach these sites (Fig. 2F, G, H). In contrast, the three active cysteine mutants in the GGNG motif, G184C, N186C and G187C, were rapidly inactivated by AMdiS, with half-times ranging from 1 to 3 min, demonstrating that the introduced cysteine residues were readily accessible from the periplasmic side of the membrane (Fig. 2F,G,H). Residual activity of the membranes containing mutants N186C and G187C was around 20 % after treatment with AMdiS, while membranes containing mutant G184C did not show significant activity anymore.

Mutants G184C, N186C and G187C were constructed in a Cys-less version of CitS (151) to reveal the site of reaction of the mutants with the thiol reagents. Unfortunately, uptake activity by these mutants in RSO membrane vesicles was to low to measure transport in a reliable way. The low activity correlated with a low level of the CitS protein in the membrane (not shown). As an alternative, corresponding mutants with Ala replacements were constructed in the wild-type background. Mutants G184A and N186A showed the same residual activity after treatment with NEM as observed for the wild-type, while both mutants were insensitive to treatment with AMdiS, strongly indicating that in the G184C and N186C mutants, the introduced Cys residues were modified by the thiol reagents (Table 1). Similarly, mutant G187A was insensitive to AMdiS showing that the Cys residue at position 187 in the G187C mutant was modified by AMdiS and responsible for the inactivation of the transporter. Remarkably, inactivation of the G187A mutant by NEM was the same as observed for the G187C mutant, suggesting that the Gly to Ala mutation increased the reactivity of the endogenous Cys residues with NEM. In this respect, the G187A mutant in the Vb region of CitS behaves like the C398S mutant described before (150). Cys398 is situated in the Xa region of CitS.

Sequence motif GGHG in region Vb of GltS. In the Na^+-glutamate transporter GltS of *Escherichia coli* the GGXG motif in the periplasmic Vb region is represented by residues 136-GGHG-139. In a similar approach as a described for CitS above, each position was

Table 1. Residual activity after treating RSO membrane vesicles containing CitS wild type (WT) and cysteine (Cys) and alanine (Ala) mutants with NEM and AMdiS[a].

Position	Inhibitor	Residual activity[b] (%)	
		Cys	Ala
G184	NEM	12 ± 3	62 ± 6
	AMdiS	2 ± 3	95 ± 5
N186	NEM	21 ± 4	55 ± 4
	AMdiS	22 ± 6	95 ± 2
G187	NEM	11 ± 3	5 ± 8
	AMdiS	21 ± 4	99 ± 5
S405	NEM	42 ± 5	68 ± 5
	AMdiS	5 ± 5	97 ± 3
WT	NEM	63 ± 6	
	AMdiS	96 ± 3	

[a] RSO membranes were treated with 1 mM NEM for 2 min and 0.25 mM AMdiS for 12 min. [b] The indicated values give the residual uptake activity in RSO membranes as the percentage of an untreated sample. The average and standard deviation of 2-3 independent measurements were reported.

mutated to a Cys residue and activities of the mutants were evaluated in RSO membranes. Membrane vesicles prepared from *E. coli* DH5α cells contain a basal level of glutamate transport activity due to endogenous glutamate transporters encoded on the chromosome (21). The background activity was estimated in membrane vesicles containing the citrate transporter CitS produced from the same expression system (Fig. 3A). CitS does not transport L-glutamate. RSO membranes containing plasmid-encoded GltS showed an activity that was approximately 5 times higher than the background activity (Fig. 3A). RSO membrane vesicles containing mutants G136C, G137C and H138C of GltS showed an uptake activity similar to the background level (Fig. 3A). Mutant G139C repeatedly revealed an activity slightly above background, but potential changes in the background activity do not allow for a firm conclusion about the significance of this residual activity. Partial purification of the proteins showed that the lack of activity was not due to a lack of production of the proteins (Fig. 3B). It follows that mutations in the sequence motif in region Vb of GltS are deleterious for the transport activity of the protein.

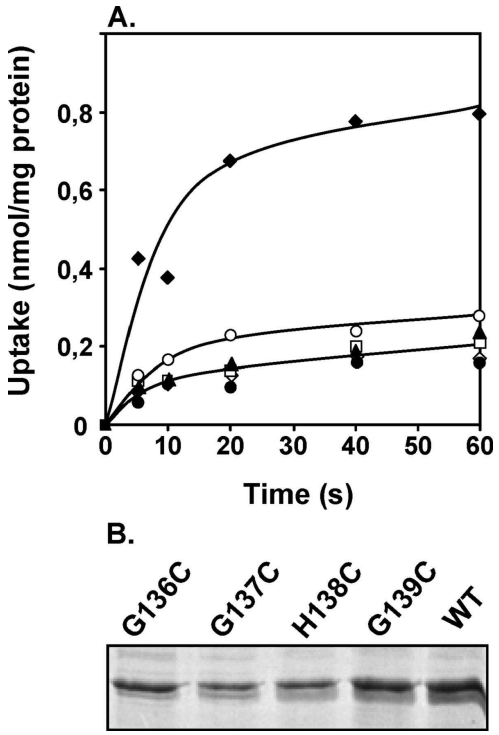


Figure 3. Sequence motif GGHG in Vb of GltS of *E. coli*. (A) L-[^{14}C]-glutamate uptake in RSO membrane vesicles containing GltS (♦) and the GltS mutants G136C (□), G137C (▲), H138C (●) and G139C (○), and CitS as control (▼). (B) SDS-PAGE of partial purified GltS and the mutants G136C, G137C, H138C and G139C purified from the RSO membranes used in the uptake assays shown in panel A.

Sequence motif GGSG motif in region Xa of CitS. The GGXG sequence motif in cytoplasmic region Xa of CitS is represented by residues 403-GGSG-406. Similar as observed for the motif in the Vb region of CitS, the least conserved residue in the motive was the least sensitive to mutation to a Cys residue. Membranes containing mutant S405C showed a relative uptake activity of approximately 40 % of wild type CitS, while mutation of the 3 Gly residues to Cys (G403C, G404C, G406C) resulted in membranes with no significant citrate uptake activity (Fig. 4A). Expression levels of the mutants were not significantly different from the wild type level (Fig. 4B).

Evidence for the folding of the cytoplasmic Xa region in between the transmembrane segments as a reentrant loop structure followed from the accessibility of two endogenous Cys residues, Cys398 and Cys414 for the small, charged thiol reagent 2-(trimethylammonium) ethyl methanethiosulfonate (MTSET) from the periplasmic site of the membrane (150). Both sites could not be reached by the more bulky charged reagent AMdiS, suggesting a restricted access pathway from the periplasm. In motif mutant

S405C the introduced Cys residue is in between Cys398 and Cys414. Treatment of RSO membranes containing S405C with membrane permeable NEM resulted in a higher rate of inactivation than observed for the wild type with no residual activity indicating that position 405 site is accessible from the water phase (Fig. 4C). Remarkably, the transport activity of the S405C mutant was also effectively inhibited by treatment with AMdiS under conditions where the wild-type is insensitive to the reagent (Fig. 4D).

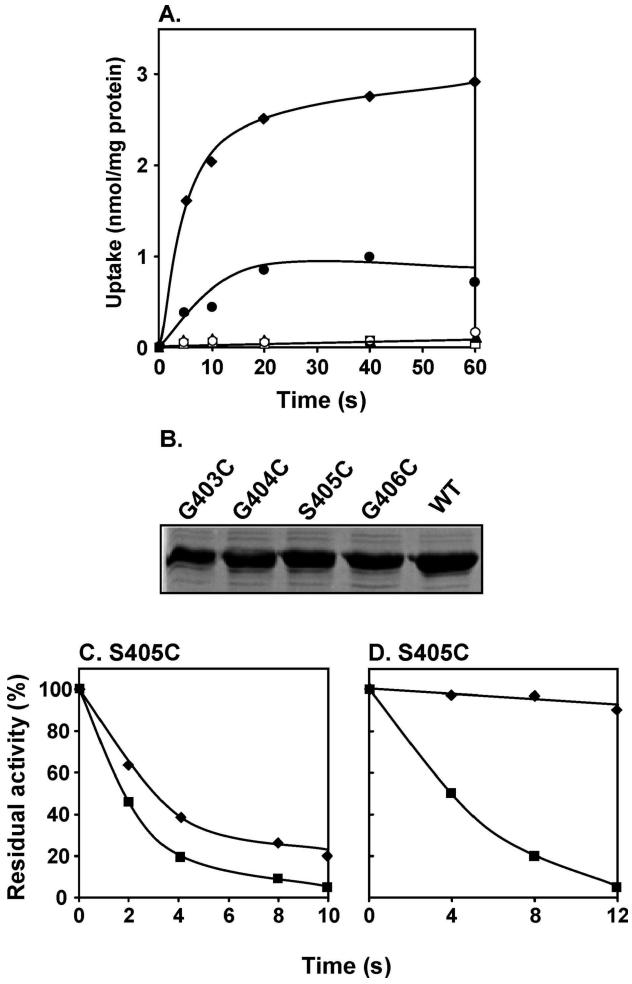


Figure 4. Sequence motif GGSG in Xa of CitS of *K. pneumoniae.* (A) [^{14}C]-citrate uptake in RSO membrane vesicles containing CitS (♦) and the CitS mutants G403C (□), G404C (▲), S405C (●) and G406C (○). (B) SDS-PAGE of partial purified CitS and the mutants G403C, G404C, S405C and G406C purified from the RSO membranes used in the uptake assays shown in panel A. (C,D) Residual activity after treating CitS (♦) and the mutant G405C (■) with 1 mM NEM (C) or 0.25 mM AMdiS (D) for the indicated time. Initial rates were expressed as the percentage of the initial rate catalyzed by untreated membranes.

The half-time of inactivation was approximately 4-5 min with essentially no residual activity. Apparently, the Cys residue at position 405 is well exposed to the periplasmic side of the membrane. Mutant S405C was also rapidly inactivated by treatment with the small, positively charged thiol reagent MTSET (not shown). In a control experiment, S405 was replaced by Ala. The S405A mutant was inactivated by NEM at the same rate as observed for the wild-type and no inactivation was observed with AMdiS (Table 1).

Cysteine scanning mutagenesis of the Xa region of GltS. A set of 17 cysteine mutants of the GltS protein was constructed having mutations in cytoplasmic region Xa between TMS IX and X. Together with Cys343 in wild type GltS, the mutants cover a stretch of 18 residues from A340 to M357, which represents the most conserved region in the C-terminal half of the ESS family (Fig. 1B). Partial purification following expression in *E. coli* DH5α showed that the mutations did not significantly affect the levels of the mutant proteins in the membrane (Fig. 5A, top panel). As described above for the Cys mutants in region Vb of GltS, the mutants were tested for their ability to accumulate glutamate in right-side-out (RSO) membrane vesicles (Fig. 5A). The mutations were remarkably well tolerated by the transporter. Twelve of the 17 mutants showed glutamate uptake activities not significantly different from the wild type activity. The GltS proteins mutated at the two adjacent positions G341 and H342 showed a significantly decreased activity of 30-40% of wild type activity. Similarly, a second cluster of 4 mutants A349C, P351C, T352C and A353C showed an activity of 30-60% of wild type.

The wild type GltS protein contains 4 cysteine residues, three in transmembrane segments (TMS IV, V and VI) and one in region Xa at position 343. Nevertheless, it was shown before that the activity of wild type GltS in RSO membranes is not sensitive to treatment with the membrane permeable thiol reagent NEM, nor to the membrane impermeable, bulky reagent AMdiS ((26); see also Fig. 5B, C343). In the group of 12 Xa mutants showing comparable specific activity as observed for the wild type, only the activity of N356C was dramatically affected by treatment with NEM. The mutant was essentially inactivated by NEM. In contrast, treatment with the membrane impermeable reagent AMdiS did not affect the activity suggesting that the residue at position 356 is not accessible to the latter in RSO membranes (Fig. 5B). In the pair of mutants G341C and

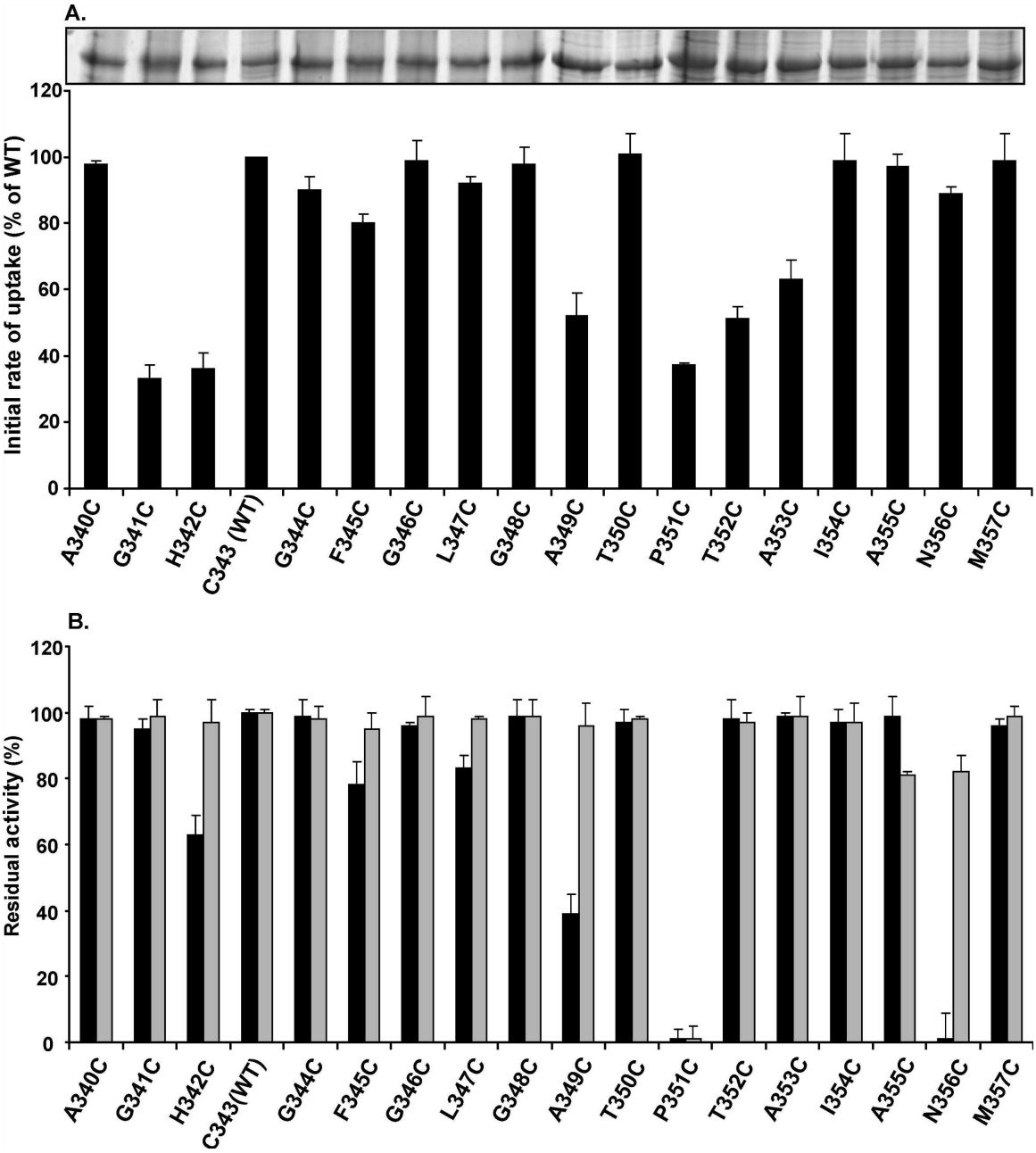


Figure 5. Cys-scanning mutagenesis of the Xa region of GltS of *E. coli*. (A) Relative L-[^{14}C]-glutamate uptake activity in RSO membrane vesicles containing GltS mutants with Cys residues at positions 340 through 357. Initial rates of glutamate uptake were expressed as percentage of the rate measured in membrane vesicles containing wild type GltS (C343). Uptake rates were corrected for the rate observed in control membranes that contained CitS produced from the same expression system (see also Figure 3). Expression levels of wild type GltS and the mutants after partial purification form cells were indicated at the top. (B) Residual activity of glutamate uptake in RSO membrane vesicles after treatment with 1 mM NEM (black bars) and 0.25 mM AMdiS (grey bars) for 10 min. Residual activity represents the initial rates as percentage of the initial rate catalyzed by untreated membranes.

H342C, that both showed reduced transport activity relative to wild type, the former appeared to be unaffected by both NEM and AMdiS, while the latter was marginally sensitive to NEM resulting in 60 % residual activity and not sensitive to AmdiS. Mutants T352C and A353C in the cluster of 4 mutants with reduced activity between positions 349 and 353 were also not affected by the thiol reagent, while the activitiy of mutant A349C was reduced to 40 % by NEM, but not sensitive to AMdiS. Clearly, the most interesting mutant in this cluster was P351C. P351C was completely inactivated by NEM and, most importantly, also by AmdiS. The result indicates that the proline residue at position 351 in cytoplasmic loop Xa of GltS is accessible from the periplasmic side of the membrane. In this respect, P351 in GltS behaves like residue S405 in the Xa region of CitS.

Two types of control experiments demonstrated that the Cys residues at positions 351 and 356 in GltS mutants P351C and N356C, respectively, were the actual target sites for the thiol reagents. One, mutants containing P351C and N356C in the Cys-less background of GltS (26) showed similar residual activities after treatment with NEM and AMdiS as observed in the wild type background. Two, replacement of P351 and N356 by Ala in the wild-type background, rendered the mutant transporters insensitive to treatment by NEM and AMdiS (Table 2).

Table 2. Residual activity after treating RSO membranes containing GltS cysteine (Cys) and alanine (Ala) mutants in wild type and Cys-less background with NEM and AMdiS[a].

Position	Inhibitor[a]	Residual activity[b] (%)		
		Cys/WT	Cys/Cys-less	Ala/WT
P351	NEM	2 ± 4	3 ± 4	100 ± 4
	AMdiS	2 ± 6	4 ± 4	96 ± 2
N356	NEM	1 ± 8	1± 5	99 ± 6
	AMdiS	82 ± 5	87± 7	96 ± 4

[a] RSO membranes were treated with 1 mM NEM and 0.25 mM AMdiS for 10 min. [b] The indicated values give the residual uptake activity in RSO membranes corrected for the background activity and as the percentage of an untreated sample. The average and standard deviation of 2-3 independent measurements were reported.

Discussion

The putative reentrant loop structures in the N- and C- terminal halves of the Na^+-citrate transporter of *Klebsiella pneumoniae* CitS and in the N-terminal half of the Na^+-glutamate transporter of *Escherichia coli* GltS contain a GGXG sequence motif. The present study demonstrates that these motifs are important for the function of the transporters. Mutant transporters in which the conserved Gly residues of the motifs were replaced by Cys residues were completely inactive or showed a severely reduced transport activity. CitS mutants with the variable positions ('X') mutated to Cys revealed a lower, but significant residual activity, while the corresponding mutant of GltS was inactive. The residual activity appears to correlate with the level of conservation at the 'X' position as the His residue in GltS is, in fact, highly conserved in the family (see sequence logo's in Fig. 1). Others have shown that mutation of the variable residue in the motif GGNG in the N-terminal half of CitS to Val reduced the affinity for citrate by one order of magnitude (63). Further evidence for a relevant role of the motifs in the properly functioning of the transporters was obtained by showing that the residual activity of the Cys mutants could be further reduced by treatment with thiol reagents.

Structural models for the CitS and GltS proteins consist of two homologous domains with opposite orientations in the membrane (Fig. 1A) (149). Prominent in the models are two reentrant loops that fold back between the trans membrane segments from opposite sides of the membrane. These are the regions in the sequence that contain the GGXG motifs. The present results give support to the models in two important aspects: (i) CitS and GltS proteins share a similar fold, and (ii) the proteins have a two-domain structure.

The CitS and GltS proteins belong to different transporter families (2HCT and ESS, respectively) and do not share any sequence similarity. The same structural model for both CitS and GltS is based on the highly similar family hydropathy profiles of the transporters in the 2HCT and ESS families (88). In the MemGen classification system the 2HCT and ESS families are in structural class ST[3] together with 31 other families of transport proteins ((87); http://molmic35.biol.rug.nl/memgen/mgweb.dll). The hypothesis is that the proteins from all families in one structural class have the same fold. The membrane topology of the CitS and GltS proteins in the models was recently confirmed experimentally by Cys accessibility studies (26). The functional importance of the GGXG

motifs in corresponding parts of the CitS and GltS sequences reported here provide further evidence for a similar structure and mechanism of the two transporters.

The two domain structure in the CitS and GltS models is based on sequence analysis of all protein families in structural class ST[3] in the MemGen classification reported before (92). A low, but significant sequence identity was identified when the N-terminal halves of the proteins were compared to the C-terminal halves suggesting sequence homology between the two halves that, consequently, would fold into two domains sharing a similar fold. The putative reentrant loop in the N-terminal domain in the model is a copy of the reentrant loop in the C-terminal domain that has been demonstrated experimentally in the CitS and GltS proteins (see below). Mutation of the three Gly residues to Cys in both GGXG motifs of CitS inactivate or strongly reduce the activity, while mutation of the variable residue reduces the activity down to 40-75 % in the motifs in the N- and C-terminal halves. The 'functional symmetry' strongly supports the 'structural symmetry' of the two domains and provides, for the first time, strong experimental support for the reentrant loop structure in the N-terminal domain.

Reentrant loops in membrane proteins are identified by accessibility studies of sites (usually Cys residue) in loop regions. Accessibility of the Cys residue from the opposite side of the membrane by water soluble, membrane impermeable thiol reagents is taken as evidence that the loop folds between the transmembrane segments exposing the residue more or less to the opposite side of the membrane. In case of the CitS protein, two endogenous Cys residues, Cys398 and Cys414, in the putative reentrant loop in the C-terminal domain were shown to be accessible for the membrane impermeable thiol reagent MTSET from the periplasmic side of the membrane (150). The reactivities of the sites could be modified by the presence of the substrate citrate and the co-ion Na^+, providing further evidence for an important functional role in catalysis. In GltS, two engineered Cys residues at positions 339 and 355 in the putative reentrant loop region in the C-terminal domain were accessible to MTSET from the periplasm as well (26). For both transporters, the Cys residues were not accessible for the more bulky reagent AMdiS suggesting a size restriction in the access pathway from the periplasm which would be compatible with a funnel-like pore structure with the target sites deep down in the narrow part (151). The characteristics of the putative reentrant loops in the C-terminal domains

of CitS and GltS differed in this respect from those of a reentrant loop in the glutamate transporter GltT of *Bacillus stearothermophilus*. GltT, a glutamate transporter not be confused with GltS, has a completely different structure, but, similar to the CitS/GltS model, two reentrant loops enter the core of the protein from opposite sites of the membrane. The reentrant loops were confirmed by the crystal structure of the homologous transporter Glt_{Ph} of the archaeon *Pyrococcus horikoshi* (180). Accessibility studies of Cys residues engineered in the loop entering from the cytoplasmic side showed that three consecutive positions were accessible for the bulky AMdiS reagent from the periplasmic side (148). The present results now demonstrate that a similar situation exists for the CitS and GltS proteins. Residue S405 at the variable position in the GGXG motif of CitS and P351 of GltS were both accessible for AMdiS added at the periplasmic side of the membrane. This suggests that these positions are closest to the periplasm and at the vertex of the reentrant loop. It was noted before that the putative reentrant loops in the ST[3] structural class contain an unusual high fraction of residues with small side chains which was interpreted as indicative of a compact packing of the loops in between the helices with a strong bending at the vertex allowed by the Gly residues in the GGXG motif (149). The accessibility of S405 in the GGXG motif of CitS by AMdiS is in line with the latter view. In GltS, which lacks the GGXG motif in the C-terminal domain, the P351 residue is flanked by Ala and polar Thr residues (AT<u>P</u>TA) which to some extent resembles the vertex of the reentrant loop in GltT of *B. stearothermophilus* which contains a Glu residue flanked by Ala and polar Ser and Thr residues (TASS<u>E</u>T). Please, note that 9 out of the 18 residues that form the putative reentrant loop in the C-terminal domain of GltS have small side chains (A, G or C; Fig. 1B).

By analogy to the transporter Glt_{Ph} of the archaeon *Pyrococcus horikoshi* (180), it is believed that the reentrant loops in the N- and C-terminal domains of the ST[3] transporters are in close vicinity in the 3D structure of the proteins where they would form the translocation path for substrates and co-ions (149). Turnover would follow an alternating access mechanism by which the translocation pore would open up to either side of the membrane in an alternating manner (151). It is easy to envision how changes to the reentrant loops that make up the translocation path could interfere with this process thereby explaining the high sensitivity of the GGXG motifs to mutations and chemical

modifications. Possibly, in the conformational state with the pathway opened up to the cytoplasm, alkylation of Cys residue at position 356 in GltS mutant N356C that is 5 positions away from P351 at the vertex blocks the cytoplasmic access pathway to the pore thereby inactivating the transporter (Fig. 5B). Also, giving the dynamic nature of the protein in the alternate access model, the reactivity of sites on the reentrant loops with thiol reagents may be sensitive to changes in the equilibrium between the inward and outward conformations. Three conditions have been shown to affect the sensitivity of CitS to thiol reagents in a similar way, two mutations, C398S in the Xa region (150) and G187A in the Vb region (this study), and the presence of the proton motive force (150). The three conditions do not affect the (in)sensitivity of CitS to AMdiS which is diagnostic for the periplasmic access pathway, but increase the sensitivity to membrane permeable NEM. Possibly, the Cys residues in the reentrant loop in the Xa region (Cys398 and/or Cys414) that are the target sites for the reagents have become more accessible in these conditions by a shift of the equilibrium towards the inward conformation.

Acknowledgments

This work was supported by a grant from the Dutch Organization for Scientific Research (NWO-ALW).

Chapter IV

Cross-linking of *trans* reentrant loops in the Na^+-citrate transporter CitS of *Klebsiella pneumoniae*

Adam Dobrowolski, Fabrizia Fusetti and Juke S. Lolkema

Published in *Biochemistry* (2010) **49**:4509-4515

Abstract

The membrane topology model of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* shows a core of two homologous domains with opposite orientation in the membrane and each containing a so called reentrant loop. A split version of CitS was constructed to study domain interactions and proximity relationships of the putative reentrant loops. Split CitS retained 50% transport activity of the wild-type version in membrane vesicles. Unspecific cross-linking of the purified complex with glutaraldehyde revealed a tetrameric complex with 2 N and 2 C domains corresponding to dimeric CitS. The separately expressed domains were not detected in the membrane. Strong interaction between the two domains followed from successful purification by Ni^{2+}-NTA chromatography of the whole complex when only one domain was His-tagged. Different kinetic states of the transporter did not seem to affect the interaction significantly. Successful disulfide cross-linking was obtained between single cysteine residues introduced in the highly conserved GGNG sequence motif at the vertex of the reentrant loop in the N domain and either of two endogenous cysteine residues at the base of the reentrant loop in the C domain. The disulfide bond was formed within one subunit in the dimer. It is concluded that the reentrant loops in the N and C domains are overlapping at the domain interface in the 3D structure and form (part of) the translocation pathway for substrate and co-ions. Sites at the interface involved in conformational changes pertinent to turnover are distinct from sites involved in the binding between the domains.

Introduction

The 2-hydroxycarboxylate (2HCT; TC 2.A.24, (134)) transporter family is a family of secondary transporters found exclusively in bacteria. Members of the family transport substrates that contain the 2-hydroxycarboxylate motif, as in citrate, lactate and malate (149). Like in other families of secondary transporters, the members of the 2HCT family represent different modes of energy coupling. The transporters are either H^+ or Na^+ symporters or they catalyze exchange between two substrates. More or less characterized members of the family include the Na^+ symporters CitS of *Klebsiella pneumoniae* and MaeN of *Bacillus subtilis*, the H^+ symporters CimH of *Bacillus subtilis* and MalP of *Streptococcus bovis*, and the citrate/acetate exchanger CitW of *Klebsiella pneumoniae* and the malate/lactate exchanger MleP and citrate/lactate exchanger CitP both found in lactic acid bacteria.

The 2HCT family is found in class ST[3] in the MemGen system that classifies membrane proteins in structural classes based on hydropathy profile analysis (87-89, 91). Proteins in the different families in a structural class in the MemGen system are not related in sequence but are believed to share the same global folding and, most likely, the same mechanism. Support for the approach was recently obtained for different families in classes ST[2] (90, 160) and ST[3] (25, 26). Class ST[3] contains 32 families of secondary transporters including the ion transporter (IT) superfamily (127). No high resolution 3D structure is available for any of the >10,000 transporters found in the public databases that belong to structural class ST[3]. Therefore, information on the structure and mechanism of 2HCT transporters bears upon an important group of transporter families for which an X-ray structure is still elusive.

By far the best studied transporter in the 2HCT family is the Na^+-citrate symporter CitS of *Klebsiella pneumoniae.* Studies on mostly CitS have resulted in a detailed structural model for the transporters in the family (149, 151, 168). The proteins exist as dimers in the membrane (108). The monomeric subunit consists of 11 transmembrane segments (TMSs) with the N and C termini at the cytoplasmic and periplasmic side of the membrane, respectively (see Fig. 1A). The N-terminal TMS is not part of the core

structure and absent in most other families of structural class ST[3]. The core structure is formed by two homologous domains of 5 TMSs each share a similar fold but have

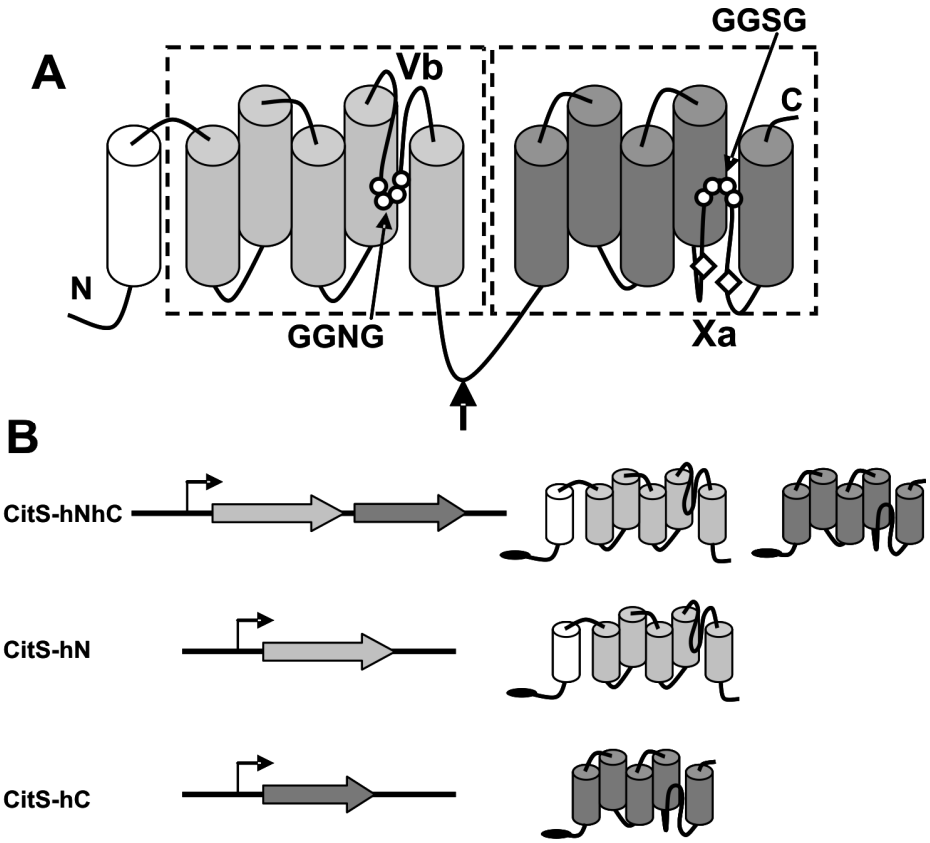


Figure 1. Schematic representation of (A) the structural model of CitS of *Klebsiella pneumoniae* and (B) the genetic constructs used to express split CitS and the separate domains. (A) Dashed boxes represent the N and C domains containing five TMSs (cylinders) that together form the core structure of CitS. Two reentrant loops termed Vb and Xa enter the membrane-embedded part of the protein from the periplasmic and cytoplasmic side of membrane, respectively. Sequence motifs GGNG and GGSG at the vertex of the reentrant loops were indicated by open dots. Two endogenous cysteine residues in reentrant loop Xa were indicated by open diamonds. The two homologous domains are connected by a long cytoplasmic loop. The arrow indicates the site where CitS was split yielding split CitS. (B) Operon structures are shown at the left, the encoded proteins at the right. Closed ovals at the N-termini of the proteins represent His-tags.

opposite orientation in the membrane (84, 92). Inverted topology of homologous domains is a structural motif observed in several other non-related transporter families, i.e. aquaporin (110), the Na^+/H^+ antiporter NhaA (56), the ammonium transporter AmtB (66), and the Na^+-leucine transporter LeuT (179). The two domains are believed to originate from an ancient duplication of a gene encoding an odd number of TMSs (85). The loops between the 4^{th} and 5^{th} transmembrane segment in each domain in the 2HCT family

model form so-called pore loops or reentrant loops, which fold back in between the transmembrane segments from opposite sides of the membrane (*trans* reentrant loops). Embedded in a different structural context, a similar pair of reentrant loops is observed in the structures of aquaporins (110) and the glutamate transporter homologue of the archaeon *Pyrococcus horikoshi* (180). The vertexes of the reentrant loops of the 2HCT transporters are formed by GGXG sequence motifs that were shown to be crucial for transport activity (25). Aspects of the model are supported by membrane topology studies, cysteine accessibility studies, mutagenesis studies and kinetic analyses that were recently reviewed in (149) or presented in subsequent papers (25, 26, 108).

An appealing hypothesis (149) positions the reentrant loops in the N and C domains in the structural model of the 2HCT transporters at the interface of the two domains where they would constitute the translocation pathway for substrate and co-ions. In this paper this possibility is addressed by cross-linking studies using a split permease approach (8). It is shown that a CitS variant genetically split in between the two domains forms a stable complex in the membrane that is active in Na^+-coupled citrate transport. Cross-linking of the two domains by disulfide bond formation demonstrates close vicinity of the reentrant loops in the two domains in the three dimensional structure of CitS.

Experimental Procedures

Bacterial strains, growth conditions and expression of CitS derivatives. *Escherichia coli* strain DH5α was routinely grown in Luria Bertani Broth (LB) medium at 37 ^oC under continuous shaking at 150 rpm. Ampicillin was used at a final concentration of 50 μg/ml. CitS and split CitS versions were expressed in *E. coli* DH5α cells harboring plasmid pBAD24 (Invitrogen) derivatives (150). The proteins were extended with 6 additional histidine residues at the N-terminus (His-tag) and enterokinase cleavage site unless otherwise stated. Expression of genes cloned in pBAD24 is under control of the arabinose promoter. Production of the proteins was induced by addition 0.01% arabinose when the optical density of the culture measured at 660 nm (OD_{660}) reached a value of 0.6.

Construction of split citS genes. All genetic manipulations were done in *E.coli* DH5α cells. Plasmid pCitS-hNC was constructed by PCR insertion mutagenesis. Synthetic primer 5'-GGGGAACTGGTGCGTAAA**TAA**GCTAGCA**GGAGG**AATTCACC**ATG**G CCTCGTTCAAAGTGG -3' was designed to insert a sequence containing a termination codon, a ribosomal biding site (RBS) and a start codon (marked in bold) into the *citS* gene in plasmid pBAD-CitS (150) between the codons encoding K249 and A250 in the CitS protein (underlined bases correspond to sequences in the *citS* gene). Plasmid pCitS-hNC encodes a protein containing the N-terminal 249 amino acid residues of the CitS protein (N domain + TMSI) extended with a N-terminal His-tag and enterokinase cleavage site and a protein consisting of the C terminal 198 amino acid residues of CitS (C domain). Plasmid pCitS-hC was constructed by deleting a *NcoI-NcoI* fragment from pCitShNC which results in a single gene encoding the C domain extended with His-tag and enterokinase cleavage site at the N-terminus. The pCitS-hN vector was constructed by first PCR insertion of sequence containing *Sma*I restriction site between stop codon after N domain and RBS site before C domain in pCitS-hNC vector, and subsequently deleting *Sma*I/*Msc*I fragment. Plasmid pCitS-hNhC was constructed by ligating a *NheI* fragment excised from pCitS-hNC into pCitS-hC digested with the same restriction enzyme. Plasmid pCitS-NhC was constructed by ligating a *Bpu*1102I/*Sal*I fragment from pCitS-hNhC into a pBAD24 vector containing the *citS* gene without His-tag and eneterokinase site (unpublished). Plasmid pSSSCC-hNhC was constructed by ligating *Bpu*1102I/*Stu*I fragment from pCitSCysless-hNhC, made in similar steps as pCitS-hNhC but with the Cys-less version of the *citS* gene, vector into similarly digested pCitS-hNhC. Plasmid pSSSCC-183-hNhC and the other mutants with Cys inserted in GGGNG motif of Vb region and two native cysteines in Xa region were constructed by cloning *Bpu*1102I/*Spe*I fragments from pSSSCC-hChN construct into similarly digested pCitS-G183C (25). All plasmids were sequenced and shown to contain the desired constructs (ServiceXS, Leiden, The Netherlands).

Transport assays in RSO membranes. *E. coli* DH5α cells expressing CitS variants were harvested from a 1 L culture by centrifugation at 10,000 x g for 10 min at 4 ^{o}C. Right-side-out (RSO) membrane vesicles were prepared by the osmotic lysis procedure as

described (61). RSO membranes were resuspended in 50 mM KPi pH 7, rapidly frozen and stored in liquid nitrogen. Membrane protein concentration was determined by the DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA).

Uptake by RSO membranes was measured by the rapid filtration method. The membranes were energized using the K-ascorbate/phenazine methosulfate (PMS) electron donor system (73). Membranes were diluted to a final concentration of 0.5 mg/ml into 50 mM KPi pH 6.0 containing 70 mM Na^+, in a total volume of 100 µl at 30 ^oC. Under a constant flow of water-saturated air, and while stirring magnetically, 10 mM K-ascorbate and 100 µM PMS (final concentrations) were added and the proton motive force was allowed to develop for 2 min. Then, [1,5-^{14}C]-citrate was added at a final concentration of 4.4 µM. Uptake was stopped by the addition of 2 ml of ice-cold 0.1 M LiCl, followed by immediate filtration over cellulose nitrate filters (0.45 µm, pore size). The filters were washed once with 2 ml of a 0.1 M LiCl solution and assayed for radioactivity. The background was estimated by adding the radiolabelled substrate to the vesicles suspension after the addition of 2 ml of ice-cold LiCl, immediately followed by filtering.

Partial purification by Ni^{2+}-NTA affinity chromatography. *E. coli* DH5α cells expressing CitS variants were harvested from a 200 ml culture by centrifugation at 10,000 x g for 10 min at 4 ^oC. Cells were washed with 50 mM KPi buffer pH 7.0 and resuspended in 2 ml of the same buffer and, subsequently, broken by a Soniprep 150 sonicator operated at an amplitude of 8 µm by 9 cycles consisting of 15 sec ON and 45 sec OFF. Cell debris and unbroken cells were removed by centrifugation at 9,000 rpm for 5 min. Membranes were collected by ultracentrifugation for 25 min at 80,000 rpm at 4 ^oC in a Beckman TLA 100.4 rotor and washed once with 50 mM KPi pH 7.0. The membranes (4 mg/ml) were solubilized in 50 mM KPi pH 8.0, 400 mM NaCl, 20% glycerol and 1% Triton X-100 followed by incubation for 30 min at 4 ^oC under continuous shaking. Undissolved material was removed by ultracentrifugation at 80,000 rpm for 25 min at 4 ^oC. The supernatant was mixed with Ni^{2+}-NTA resin (50 µl bed volume per 5 mg protein), equilibrated in 50 mM KPi pH 8.0, 600 mM NaCl, 10% glycerol, 0.1% Triton X-100, 10 mM imidazole and incubated overnight at 4 ^oC under

continuous shaking. Subsequently, the column material was pelleted by pulse centrifugation and the supernatant was removed. The resin was washed with 10 volumes of equilibration buffer containing 300 mM NaCl and 40 mM imidazole. The protein was eluted with half a bed volume of the washing buffer but containing 150 mM imidazole. The eluted fraction was stored at –20 ^{o}C until use.

Labeling and cross-linking studies. Aliquots of purified proteins were gently thawed on ice and treated with 0.1 mM fluorescein-5-maleimide (FM) for 5 min at room temperature in the dark. The treatment was stopped with 1 mM DTT. Samples were treated with 2.5 mM glutaraldehyde (GA, Sigma) at room temperature for 20 min. The treatment was quenched with 100 mM Tris-HCl pH 7.4 after which the samples were left at room temperature for 10 min. In control experiments 0.1% SDS was added to the samples, before GA treatment. Samples were treated with 2 mM sodium tetrathionate (NaTT, Sigma) at 37 °C for 30 min. In control experiments, 5 mM DTT was added to the samples following NaTT treatment. Following treatment, samples were mixed with SDS sample buffer and run on a 12% SDS-PAGE gel. Fluorescent labelling of proteins labeled with FM was visualized on a Fujifilm LAS-4000 luminescent image analyzer (Fuji).

Mass Spectrometry Analysis. Partially purified split CitS variants treated with sodium tetrathionate (NaTT) were separated by SDS-PAGE, using a 12% gel, and stained with Coomassie Brilliant Blue. Selected bands were cut from the gel. The pieces of gel were fragmented and distained in 50 mM ammonium bicarbonate with 50% acetonitrile, Reduction and alkylation of cysteine residues were achieved via incubation with 100 mM dithiothreitol, followed by iodoacetamide treatment. Acetonitriledehydrated pieces of the gel were reswollen via addition of 10 μL of a 10 ng/μL trypsin solution and incubated overnight at 37 °C. Tryptic peptides were extracted twice with 30 μL of 60% acetonitrile in 1% trifluoroacetic acid (TFA) in water and vacuum-dried. Dried peptides were resuspended in 0.1% TFA and separated on a C18 capillary column (C18 PepMap 300, 75 μm x 150 mm, 3 μm particle size, LC-Packing, Amsterdam, The Netherlands) mounted on an UltiMate 3000 nanoflow liquid chromatography system (LCPacking). Aqueous solutions of 0.05% TFA (A) and 80% acetonitrile with 0.05%TFA (B) were

used for elution. Agradient from 4 to 40%B over 50 min was used at a flow rate of 300 nL/min. Column effluent was mixed in a 1:4 (v/v) ratio with a solution of 2.3 mg/mL R-cyano-4-hydroxycinnamic acid (LaserBio Laboratories, Sophia-Antipolis, France) in a 60% ACN/0.07% TFA mixture. Fractions of 12 s were spotted on a blank MALDI target with a Probot MALDI spotter system (Dionex). Mass spectrometric analysis was conducted with a MALDI-TOF/TOF 4800 Proteomics Analyzer (Applied Biosystems) in the range of *m/z* 600-4000, in positive ion mode. Peptides with signal-to-noise levels of >50 were selected for MS/MS fragmentation. Matching of the MSMS spectra to the CitS sequences was performed with Mascot, version 2.1 (Matrix Science, London, U.K.).

Results

***Split CitS is an active transporter*.** The CitS protein was genetically split in two nonoverlapping polypeptides corresponding to the N and C terminal domains (Fig. 1). Plasmid pCitS-hNhC contains an artificial operon of two genes, the first one coding for the first 249 amino acids of CitS (TMSI plus the N domain), the second for the last 198 amino acids (C domain). The operon is under control of the arabinose promoter. The two halves of the *citS* gene were cloned separately yielding plasmids pCitS-hN and pCitS-hC. In all cases, the encoded polypeptides were extended with six histidine residues (His-tag) and an enterokinase site at the N-terminus.

Split CitS-hNhC and the two domains CitS-hN and CitS-hC were tested for the ability to accumulate [1,5-^{14}C]-citrate in right-side-out (RSO) membrane vesicles prepared from *E.coli* DH5α cells harboring the appropriate plasmid. Citrate uptake was measured in the presence of a proton motive force (pmf) that was generated using the artificial electron donor system ascorbate/PMS (Fig. 2A). RSO membrane vesicles prepared from cells harboring the empty vector lack citrate uptake activity because of the absence of a citrate transport system in the *E. coli* membrane. RSO membranes prepared from cells co-expressing the two domains retained about 50% uptake activity of membranes containing the wild-type version of CitS. In contrast, membranes prepared form cells expressing the isolated domains did not show any citrate uptake activity.

N-terminally His-tagged CitS purified by Ni^{2+}-NTA affinity chromatography migrates as a single band with apparent molecular mass of 38 kDa on SDS-PAGE (Fig. 2B, left

panel) (150, 151). The co-expressed N and C domains with His-tags at the N-termini were purified using the same protocol. On SDS-PAGE two bands showed up with apparent molecular masses of 19 and 25 kDa, strongly suggesting that they correspond to the C domain (198 residues) and N domain (249 residues), respectively. The bands were identified by labeling the purified proteins before SDS-PAGE with fluorescein-5-maleimide (FM), followed by fluorescence imaging of the gel. The CitS protein contains 5 cysteine residues, all of them located in the C domain. The image showed the labeling of full-length CitS and the 19 kDa band (Fig. 2B, right panel). Therefore, the latter corresponds to the C domain. Production of one domain in the absence of the other from constructs pCitS-hN and pCitS-hC was not detected. Apparently, the domains are not stably inserted into the membrane as separate entities.

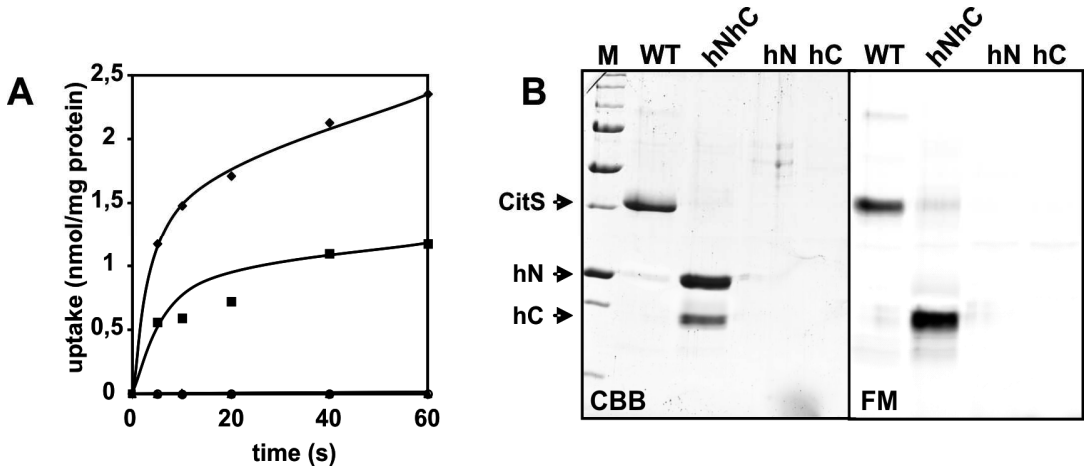


Figure 2. Activity and expression of split CitS. (A) [1,5-^{14}C]-citrate uptake by RSO membrane vesicles containing CitS (♦), split CitS-hNhC (■) and the separately expressed domains of CitS, CitS-hN (▲) and CitS-hC (○). (B) SDS-PAGE of purified CitS, split CitS-hNhC, CitS-hN and CitS-hC after Coomassie Brilliant Blue staining (left panel-CBB) and fluorescence imaging of the same gel (right panel-FM). Samples were treated with fluorescein-5-maleimide before running the gel. Arrows point at CitS and the domains hN and hC. Lane M, molecular mass standards. From top to bottom the bands correspond to 200, 150, 100, 75, 50, 37, 25, 20 and 15 kDa. The same marker was used in all figures.

CitS was shown before to be a dimeric protein by electron microscopy, BN-PAGE and by single-molecule fluorescence spectroscopy (64, 108). In agreement, treatment of purified CitS with the unspecific cross-linker glutaraldehyde followed by analysis by SDS-PAGE resulted in complete disappearance of the 38 kDa band and, at the same time, a new, somewhat fuzzy, band appeared running at approximately double the mass (Fig. 3). A small fraction of the protein did not enter the gel to any significant extent, suggesting

some aggregation in the protein preparation. Treatment with glutaraldehyde in the presence of SDS prevented cross-linking of the protein showing that cross-linking was the result of complex formation rather than random collisions. The same fuzziness of the band suggests that it is due to random labeling of CitS molecules with glutaraldehyde. Treatment of purified split CitS-hNhC with glutaraldehyde resulted in a similarly sized complex as observed with wild type CitS suggesting a complex consisting of 2 N and 2 C domains like in dimeric CitS (Fig. 3). Again, treatment with glutaraldehyde in the presence of SDS did not result in cross-linking.

It follows that split CitS, expressed from an artificial operon encoding the two domains as separate proteins forms a complex of two N domains and two C domains that is active in pmf-driven citrate uptake. The separately expressed domains of CitS are not stably produced, suggesting that the N and C domains stabilize one another.

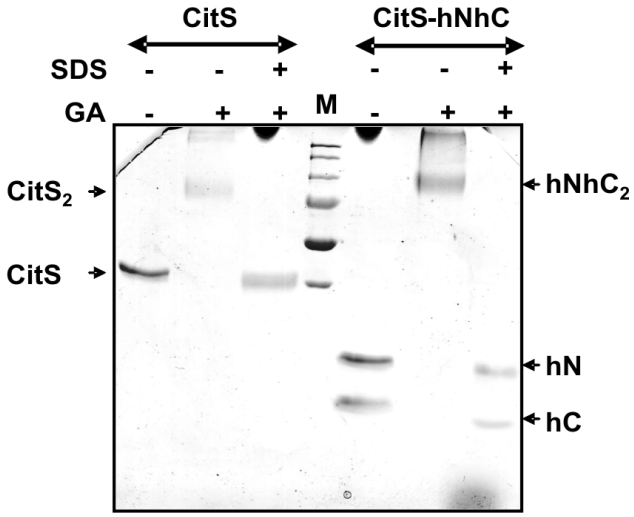


Figure 3. Glutaraldehyde cross-linking of CitS and split CitS-hNhC. Purified CitS and split CitS-hNhC were treated with glutaraldehyde (GA) at a concentration of 2.5 mM for 20 min in the presence and absence of 0.1% SDS as indicated at the top followed by SDS-PAGE. Left arrows point at monomeric CitS and dimeric CitS ($CitS_2$). Right arrows point at the domains hN and hC and the tetrameric complex of two N domains and two C domains $hNhC_2$.

Domain interactions in CitS. Two additional vectors encoding the two domains in one operon were constructed to study the interaction between the N and C domains. CitS-hNC represents split CitS with only the N domain His-tagged, while CitS-NhC has only the C domain His-tagged. RSO membrane vesicles prepared from cells expressing the latter showed citrate uptake activity similar to the activity observed for membranes containing split CitS with a His-tag at both domains (CitS-hNhC), i.e. about 50% of wild

type CitS activity (Fig. 2A and 4A). In contrast, with the His-tag only at the N domain (CitS-hNC) the activity was approximately three times lower.

For both combinations, the untagged domain was co-purified with the His-tagged domain by Ni^{2+}-NTA affinity chromatography indicating high affinity between the two domains (Fig. 4B, left panel). The N domain of CitS without His-tag migrated on SDS-PAGE at an apparent molecular mass of about 22 kDa, while untagged C domain migrated at 17 kDa. The C domains were identified by FM labeling (Fig. 4B, right panel). The expression level of CitS-hNC was clearly lower than observed for CitS-NhC, which is in line with the citrate uptake experiments. Apparently, tagging of the C domain with the His-tag and enterokinase site at the N terminus results in a more stable integration into the membrane. Cross-linking of both variants hNC and NhC by glutaraldehyde resulted in a band of around 80 kDa indicating that the presence or absence of His-tags does not interfere with complex formation (data not shown).

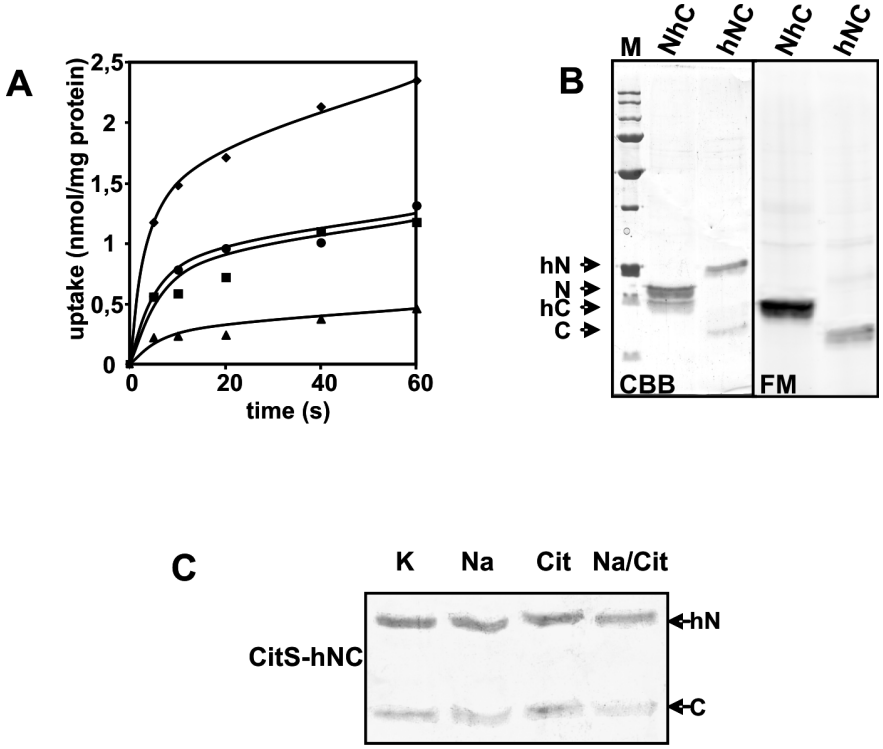


Figure 4. Domain interactions in CitS. (A) [1,5-^{14}C]-citrate uptake by RSO membrane vesicles containing CitS (♦), split CitS-hNhC (■), split CitS-hNC (▲) and split CitS-NhC (●). (B) SDS-PAGE of CitS-NhC (NhC) and CitS-hNC (hNC) after Coomassie Brilliant Blue staining (left panel-CBB) and fluorescence imaging of the gel (right panel-FM). Samples were treated with fluorescein-5-maleimide. Arrows indicate the position of the N and C domains with (hN, hC) or without (N, C) a His-tag at the N-terminus. (C) SDS-PAGE of CitS-hNC purified in 50 mM KPi pH 7 in the presence of 200 mM KCl (K), 200 mM NaCl (Na), 10 mM citrate (Cit), and 10 mM citrate plus 200 mM NaCl. Arrows point at the hN and C domains.

The interface of the N and C domains has been suggested to form the substrate translocation pathway (149). The interaction between the two domains was investigated in different catalytic states of the transporter by binding the CitS-hNC and CitS-NhC complexes to Ni^{2+}-NTA resin and washing the columns with buffer containing 10 mM of citrate or 200 mM of Na^{+}, or both. SDS-PAGE analysis after elution of the His-tagged domain revealed that neither substrate or co-ion did affect co-purification of the untagged domain (Fig. 4C), suggesting that the binding affinity between the two domains is the same in the free and bound conformations of the transporter.

Subunit interaction in dimeric CitS. In the model of the structure of CitS each of the two domains contains one reentrant-loop that together have been hypothesized to form the translocation pathway through the proteins at the interface of the two domains (Fig. 1A). The functional complex formed by CitS-hNhC allows for analysis of proximity relationships between the two domains by formation of disulfide cross-links between cysteine residues located in the two domains. Cross-linking of cysteine residues present in the two reentrant loops may be readily detected by SDS-PAGE.

Wild type CitS contains five cysteine residues, all in the C-terminal domain (Fig. 2B, right panel). Wild type CitS purified in the absence of reducing agents revealed in addition to the bulk of the protein running at the monomer position, a minor band running at the position of a dimer (Fig. 5A). Treatment of the protein with the oxidant sodium tetrathionate (NaTT) significantly increased the intensity of the dimer band and a new band appeared with an apparent molecular mass of approximately a tetrameric complex. At the same time the intensity of the monomeric band decreased. The multimeric species were not observed in the Cys-less version of the protein indicating that they were the consequence of the formation of disulfide cross-links between cysteine residues present in the C domains of the two subunits. Two of the five cysteines (Cys-398 and Cys-414) are in the putative reentrant loop (Fig. 1A, open diamonds) and important for activity of the transporter, while the other three (Cys-278, Cys-317 and Cys-347) do not seem to play a role in catalysis (150, 151). Treatment of a CitS mutant in which the former two cysteine residues were mutated to Ser (CCCSS) with NaTT showed the same behavior as

wild type CitS (Fig. 5A; compare lanes 6&7 with lanes 3&4). In contrast, treatment of the complementary mutant with the latter three cysteines mutated to Ser (SSSCC) behaved like the Cys-less mutant. It follows, that the three non-essential Cys residues are involved in the cross-linking of dimeric CitS by NaTT.

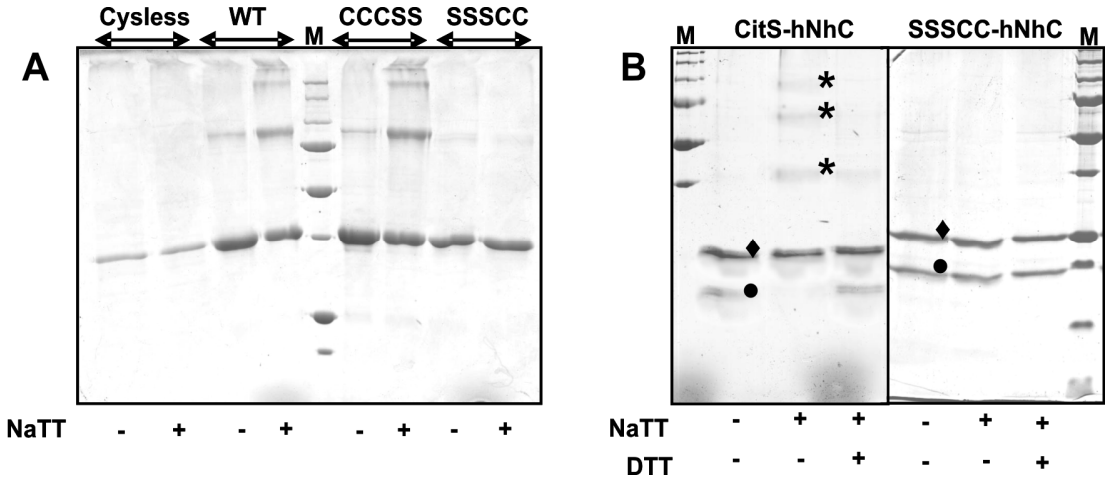


Figure 5. Disulfide cross-linking of CitS (A) and split CitS (B) derivatives. (A) Purified CitS (A, WT), Cys-less CitS (A, Cysless), Cys mutant CCCSS (A, CCCSS), Cys mutant SSSCC (A, SSSCC), split CitS-hNhC (B, CitS-hNhC) , and split Cys mutant SSSCC-hNhC (B, SSSCC-hNhC) were treated with NaTT in the presence and absence of DTT as indicated at the bottom followed by SDS-PAGE. * cross-linked products, ♦ N domain hN, ● C domain hC.

In agreement, treatment of split CitS-hNhC with NaTT under the same conditions resulted in the disappearance of the C domain and at the same time the appearance of bands with apparent molecular masses corresponding to the di-, tri and tetrameric C domain (Fig. 5B). The N domain was not affected by the oxidant. Treatment of the oxidized protein complex with the reducing agent DTT showed that the disulfide formation was fully reversible.

A split CitS version of the SSSCC mutant was constructed for cross-linking studies between the N and C domain. RSO membrane vesicles containing SSSCC-hNhC showed about half of the citrate uptake activity that was observed for CitS-hNhC which largely correlated with a lower level of expression (Fig. 6A). Unspecific cross-linking of the purified SSSCC-hNhC complex with glutaraldehyde revealed the high molecular weight band on SDS-PAGE corresponding to dimeric CitS (not shown). Most importantly,

treatment of purified SSSCC-hNhC with NaTT did not result in the formation of cross-links between the C domains (Fig. 5B).

Cross-linking of the reentrant loops. A set of 5 single-Cys mutants of the N domain of the split CitS version SSSCC-hNhC was constructed. Each of the amino acid residues in the conserved sequence motif 184-GGNG-187 believed to form the vertex of the reentrant loop (Fig. 1A; (25)) and the glycine residue at position 183 was substituted with a cysteine residue. Then, each mutant results in a SSSCC-hNhC complex with a single cysteine in the putative reentrant loop in the N domain and two cysteines, Cys-398 and Cys-414, in the reentrant loop in the C domain. The G183C mutation did not affect the activity of the complex to a significant extent (Fig. 6A). The N186C variant showed about 50 % of the citrate uptake activity observed with the SSSCC-hNhC complex. Membranes containing mutants G184C and G187C exhibited a low but significant activity of less than 10%, while the G185C mutation resulted in complete lack of activity. These results correlate well with the activities of the same mutations in wild type CitS that were reported before (25).

The split SSSCC mutants were purified under reducing conditions and analysis by SDS-PAGE revealed the 19 and 25 kDa bands corresponding to the N and C domains, respectively (Fig. 6B, C; closed circle and diamond). Treatment of the samples with NaTT resulted in a decrease of the intensity of the two bands (Fig. 6B,C). The decrease in intensity was stronger for the N domain than for the C domain. Mainly, two new bands showed up running at molecular masses of about 45 kDa and 48 kDa (arrow head and square, respectively). Analysis by mass spectrometry identified peptides originating from both the N domain and C domain in the lower 45 kDa band (arrow head), while the upper 48 kDa band contained peptides of the N domain only. A third minor new band observed with some of the mutants and running at a apparent molecular mass of about 40 kDa was identified as the *E. coli* cAMP receptor protein (CRP) which apparently was present as an impurity in the preparations. It is concluded that the 48 kDa band corresponds to two cross-linked N domains and the 45 kDa band to cross-linked N and C domains of CitS. Cross-linking efficiency between two N domains appears to be more or less the same for all mutants while the efficiency of cross-linking between the N and C domain was higher

for the G183 and G184 mutants than observed for the G185, N186 and G187 mutants. Repeating the purification under non-reducing conditions resulted in a small fraction of the domains running as the cross-linked products. Apparently, the disulfide bonds forms spontaneously between the cysteines, but this spontaneous cross-linking was less efficient (data not shown).

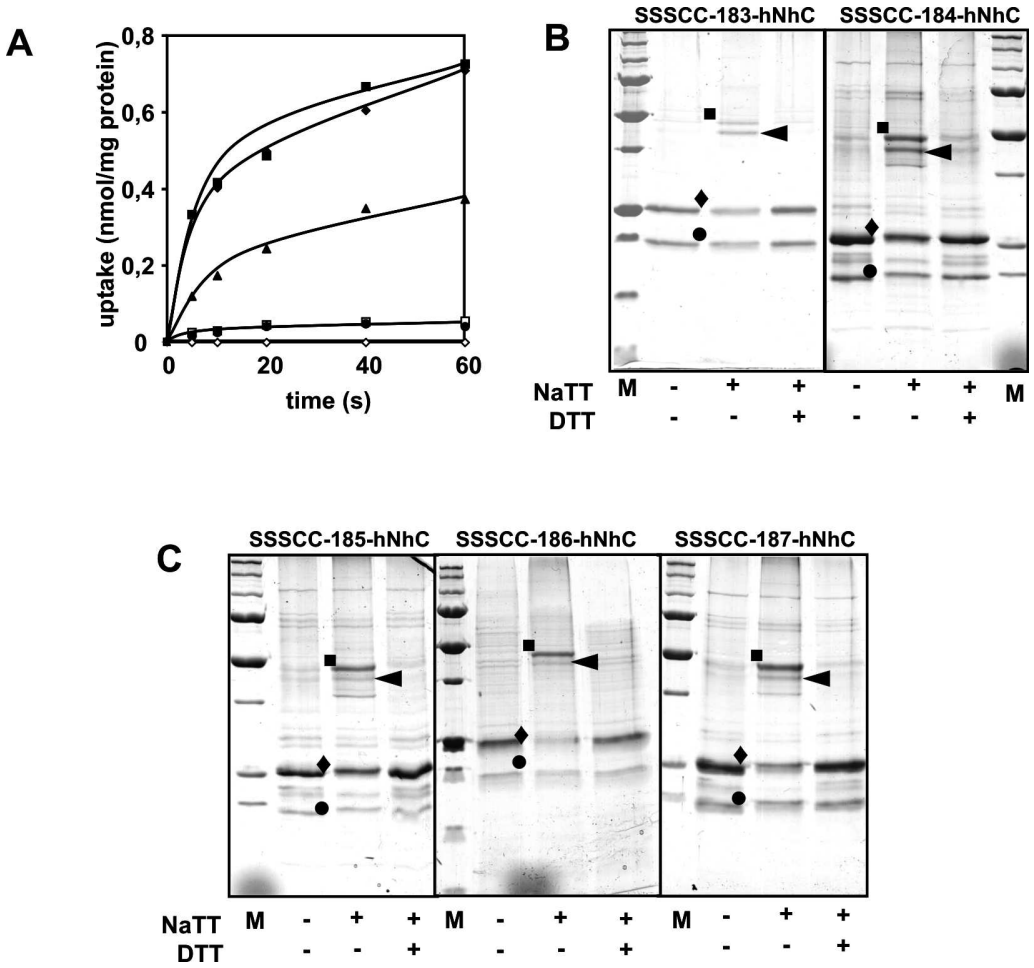


Figure 6. Disulfide cross-linking of the reentrant loops in the N and C domains. (A) [1,5-^{14}C]-citrate uptake by RSO membrane vesicles containing SSSCC-hNhC (■), SSSCC-183-hNhC (♦), SSSCC-184-hNhC (□), SSSCC-185-hNhC (◊), SSSCC-186-hNhC (▲) and SSSCC-187-hNhC (●). Numbers refer to the position in the reentrant loop in the N domain that was mutated to a cysteine residue. (B,C) Split SSSCC variants SSSCC-183-hNhC, SSSCC-184-hNhC, SSSCC-185-hNhC, SSSCC-186-hNhC and SSSCC-187-hNhC were treated with NaTT and with and without consecutive treatment with DTT as indicated at the bottom followed by SDS-PAGE. Proteins were purified under reducing condition. ■ cross-linked product of two N domains, ◄ cross-linked product of a N and C domain, ♦ N domain hN, ● C domain hC.

Since CitS is dimeric, the cross-link between the N and C domain in the mutant split SSSCC complexes may be formed between the domains of one subunit or between the domains of different subunits. Obviously, the cross-link between the N domains is between different subunits. Mutants of 'full length' CitS were constructed containing the G183C and N186C mutants in the SSSCC background. No cross-linking was detected between the subunits upon treatment of the purified proteins with NaTT demonstrating that the cross-link between the N and C domain is in one subunit and the cross-link between the two N domains is only observed in the split CitS constructs (data not shown). The results support a close proximity of reentrant loops Vb and Xa (Fig. 1A) in the N and C domain of one subunit of dimeric CitS.

Discussion

The structural model of the CitS transporter of *Klebsiella pneumoniae* shows two homologous domains, each containing 5 TMSs, with inverted topology in the membrane (92, 149). Prominent in the model are two reentrant loops connecting the fourth and fifth TMSs in each domain that fold back in between the TMSs from opposite sides of the membrane (*trans* reentrant loops). The reentrant loop regions are highly conserved in the 2HCT family and contain a sequence motif GGXG that was shown to be essential for activity of the transporter (25). By analogy to other two-domain transporter structures (1, 55, 66, 147, 179), it was proposed that the translocation site would be formed at the interface of the two domains and, consequently, the two reentrant loops would be positioned at the interface. In this study we demonstrate that the reentrant loops are in close vicinity in the 3D structure. Cysteine residues engineered in the reentrant loop in the N domain (VB, Fig. 1A) were shown to form a disulfide bond with either of two endogenous cysteine residues in the reentrant loop in the C domain (XA, Fig. 1A, open diamonds). The disulfide bond cross-linked the N and C domains in a split version of CitS, but not the subunits in 'full-length' CitS indicating that in dimeric CitS (108) the reentrant loops of the same subunit are in close contact. In the N domain, the cysteine residues involved in the cross-links were positioned at the 184-GGNG-187 sequence motif that is believed to be at the vertex of the reentrant loop (25). They cross-link to cysteine residues Cys398 and/or Cys414 positioned at the base of the reentrant loop in the

C domain where the sequence motif is 403-GGSG-406. Apparently, the two loops overlap one another at the domain interface suggesting that they may constitute a major part of the wall of the translocation pore (Fig. 7). Accessibility of a Cys residue substituting for Ser405 in sequence motif 403-GGSG-406 by a bulky, membrane impermeable thiol reagent 4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMdiS) from the periplasmic side of the membrane demonstrated previously (25) is the best evidence for the reentrant loop in the C domain to stick through the membrane all the way. The accessibility of Cys residues at position 398 and 414 from the periplasm was shown to be restricted to small thiol reagents like the [2-(trimethylammonium)ethyl] methanethiosulfonate bromide (MTSET, (150)) suggesting that they are located deeper in a more narrow part of the pore from a periplasmic perspective. The disulfide cross-links to the latter demonstrated here represent the first experimental evidence that the putative reentrant loop in the N domain may actually enter the membrane embedded part of the protein as depicted in Figure 7. At consecutive positions 183 through 187, highest cross-linking efficiency was observed with Cys residues at positions 183 and 184, suggesting that movement of the tip of the loop is constrained by the conformation of the protein.

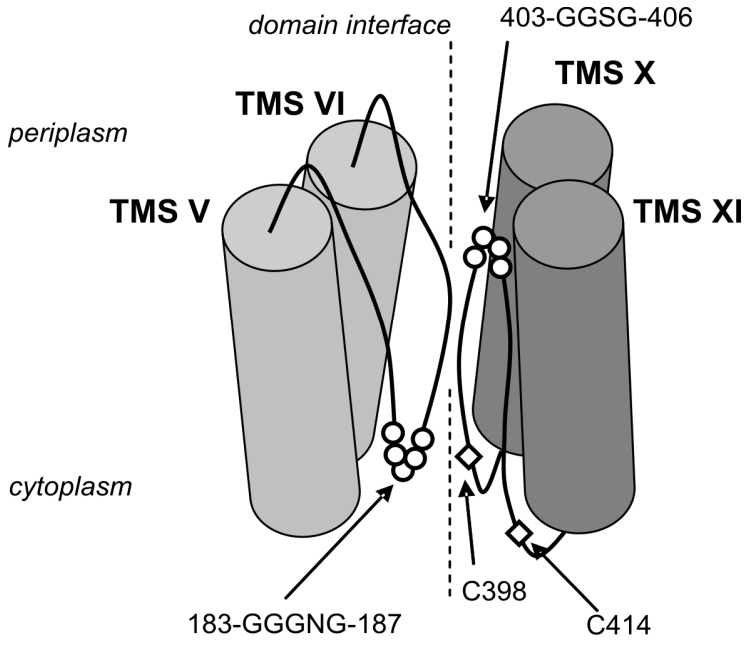


Figure 7. Structural model of the domain interface of CitS of *K. pneumoniae*. Shown are TMSs V and VI in the N domain and TMSs X and XI in the C domain together with the connecting reentrant loops. See text for further explanation.

The disulfide bond formed between the putative reentrant loops of CitS upon oxidation, was identified by cross-linking the two domains in a split version of CitS. Split CitS was expressed from an artificial operon encoding two non-overlapping polypeptides corresponding to TMSI plus the N domain and the C domain, respectively. Polytopic membrane proteins are cotranslationally inserted in the membrane by the Sec machinery. A single ribosome sitting on the SecYEG channel translates the sequence on the messenger and pushes the nascent polypeptide into the translocation channel. The transmembrane segments move laterally into the phospholipid bilayer one by one and probably start assembling into the native conformation of the protein as they leave the insertion machinery. Therefore, assembly of the two domains of CitS expressed from a single *citS* gene is a sequential and coupled event. In addition, the single gene warrants a 1:1 stoichiometry of the two domains. In contrast, in case of the artificial two-gene operon encoding split CitS, translation and insertion is parallel and not coupled. N and C domains are synthesized and inserted into the membrane by different ribosome/SecYEG complexes in an uncoordinated manner with no guarantee for a 1:1 stoichiometry. Once in the membrane they assemble into a conformation independent of the other domain, after which they have to find each other to allow the formation of the native complex. Nevertheless, the different steps proceed successfully since stable insertion of the two domains in the membrane was observed and significant Na^+-citrate transport activity detected (Fig. 2A). The same result was obtained before for a split version of the lactose transportrer LacY (8, 178). Two observations suggest that the interaction between the two domains are crucial for the stability of the CitS protein, (i) individually expressed N and C domains could not be detected in the membrane (also observed for split LacY; Fig. 2), and (ii) only a single complex containing 2 N domains and 2 C domains corresponding to dimeric CitS was detected in the membrane (Fig. 3). Titration of purified split CitS with increasing concentrations of the non-specific cross-linker glutaraldehyde resulted in two intermediate bands on SDS-PAGE in the lower concentration domain (data not shown). The bands disappeared at higher concentrations of glutaraldehyde to yield the single cross-linked product shown in Figure 3. It follows that only the complex corresponding to the original CitS dimer is stable in the membrane and any N or C domain produced in excess is being taken care of by the proteolytic system of the cells. Apparently, during

production, time allows for a significant part of the domain proteins to find their counterpart before being degraded. The importance of the domain interaction is also evident from the strength of the interaction. Binding of one domain to Ni^{2+}-NTA affinity resin through an engineered His-tag resulted in co-purification of the other domain (Fig. 4). The co-purified domain could not be eluted by extensive washing under different conditions unless non-physiological conditions were used (not shown). The high affinity between the domains was not affected by the different catalytic states of the complex (Fig. 4C), while, at the same time, it is proposed that translocation takes place at the interface of the two domains (Fig. 7). Apparently, the sites at the interface involved in the conformational changes associated with turnover are different from the sites that keep the two domains together.

The split CitS complex consists of 2 N and 2 C domains in an arrangement as in dimeric CitS. In addition to cross-linking of the N and C domains within one subunit, the same split CitS variants resulted in cross-linking of two N domains. The cross-link must be the result of disulfide formation between the single cysteine residues in the reentrant loops in the N domains and correspond to a inter subunit cross-link.

Nevertheless, cross-linking of purified CitS to a dimer was not observed when the single Cys183 and Cys186 variants of 'full-length' CitS were treated with NaTT under the same conditions, suggesting that cross-linking of two N-domains is an artifact of the split versions of CitS. The conclusion is supported by the lack of specificity of cross-linking efficiency for the cysteine residues at the different positions in the reentrant loop. One possible scenario might be that a fraction of the reentrant loops in the N domain are not properly inserted in the membrane embedded part of the complex as a consequence of the alternative biogenesis route for split CitS (see above). The unstructured loop outside of the translocation pore and with a length of up to the thickness of the membrane may easily reach a similar structure in the other subunit and form a disulfide bond upon oxidation. Possibly, proper insertion of the reentrant loop in the N domain requires the coordinated insertion of the C domain, which is the normal situation when CitS is synthesized from a single gene. When translated from two different genes, the time delay in the interaction between N and C domain and/or the different folding states of the two proteins may prevent proper insertion of part of the loops during assembly of the

complex. The lack of overlap between the sites that keep the two domains together and the sites involved in translocation (the reentrant loops) noted above would be consistent with such a proposal. In addition, the assisted insertion of TMS VIII of CitS by downstream TMS IX that was reported before (169) provides a precedent for a mechanism by which proper insertion of the reentrant loop in the N domain would depend on C-terminal sequences in CitS.

Acknowledgment. We thank Wim Huibers for help with preparing the samples for mass spectrometry. This work was supported by a grant from the Dutch Organization for Scientific Research (NWO-ALW).

Chapter V

Evolution of antiparallel two-domain membrane proteins. Swapping domains in the glutamate transporter GltS

Adam Dobrowolski and Juke S. Lolkema

Published in *Biochemistry* (2010) **49**: 5972-5974

Chapter V

Abstract

X-ray crystallography has revealed that many integral membrane proteins consist of two domains with a similar fold but opposite (antiparallel) orientation in the membrane. The proteins are believed to have evolved by gene-duplication and fusion events from a 'dual topology' ancestral membrane protein, which adapted both orientations in the membrane and formed antiparallel homodimers. We used a Na^+-glutamate transporter GltS of *Escherichia coli*, an antiparallel two-domain protein to study evolutionary pathways of membrane proteins. A set of evolutionary states of the GltS was engineered. The two half genes encoding the two domains were placed in a single operon in both orders ($GltS^{split}$) and the split genes were fused in the reverse order compared to the original protein ($GltS^{swap}$). The transporter halves were produced and shown to be active in Na^+-coupled glutamate transport. $GltS^{swap}$ was equally active as the original transporter provided that the domains were connected by a linker of the same size connecting them in the original transporter.

Introduction

Crystallography studies have shown that many membrane proteins consist of two domains that share a similar fold. The two domains have the same (parallel) or opposite (antiparallel) orientation in the membrane corresponding to even or odd numbers of transmembrane segments (TMS) per domain. Representative examples of proteins with a parallel domain organization are LacY (1), GlpT (55), EmrD (181) and FucP (20) in the Major Facilitator Superfamily (MFS), AcrB (109) and the ADP/ATP carrier (123). The antiparallel domain organization is observed more frequently in the 3D structures of membrane proteins, and examples of this organization are the aquaporins (40, 110), AmtB (66), SecY (167), LeuT (179), the H^+/Cl^- exchanger CLC (28), NhaA (56) FocA (172), vSGLT (34). All of these proteins contain two easily recognizable domains or, at least, structural elements, that have the same fold and that are oppositely oriented in the membrane. In addition, biochemical evidence has been presented indicating this structural organization to be more widespread (25, 26, 92, 133, 149). Although the domain structure is clearly recognizable in the high-resolution crystal structures, homology was in most cases not obvious from the amino acid sequences of the two domains probably because they have diverged too far.

A plausible model for the evolution of two-domain membrane proteins with an internal repeat involves duplication of a primordial gene followed by fusion and, thus, resulting in a single gene encoding a protein with two homologous domains (11, 124, 125, 128, 129) (Fig. 1). Alternatively, the duplication and fusion steps proceed in a single step ((85), see below). As mentioned above, the two domains have the same (parallel) or opposite (antiparallel) orientation in the membrane. To account for the antiparallel orientation of the two domains, the ancestral membrane protein is hypothesized to be 'dual topology', i.e. it inserts with random orientation into the membrane. Following duplication, the two dual topology proteins would adopt fixed, but opposite orientations by genetic drift resulting in the introduction of positively charged amino acid residues in cytoplasmic loops (positive-inside rule) (47). The first gene on the chromosome may either encode one or the other orientation; at the protein level this has no consequences for the antiparallel heterodimer that is formed. However, in the fused state this results in two

different proteins with the N-terminus either inside or outside the cell (Fig. 1). Therefore, the evolutionary pathway has two outcomes representing proteins that differ in the order of the two domains in the primary structure. In the course of evolution, one of these outcomes will be selected at random or because of a selective advantage.

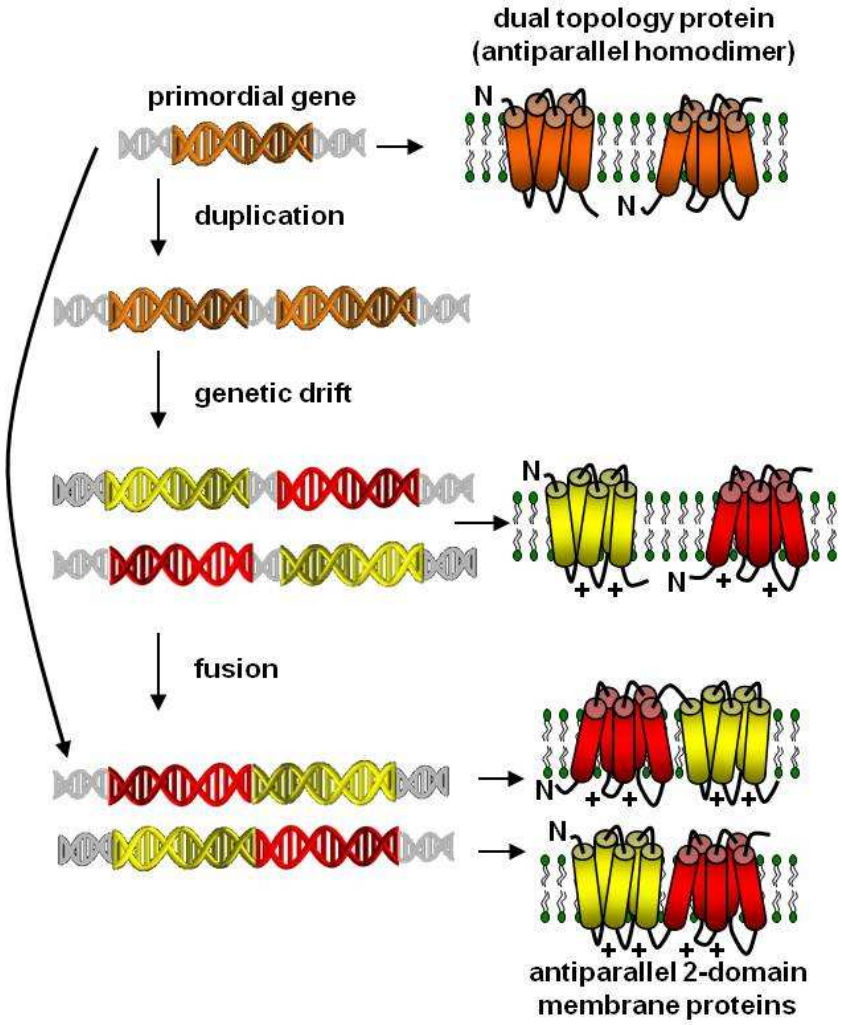


Figure 1. Model for the evolution of antiparallel 2-domain membrane proteins. The left column shows a genetic states. Genes are color-coded as follows: orange, gene encoding a dual topology protein; yellow, gene encoding a protein with a fixed orientation in the membrane; red, same as yellow but with the opposite orientation. The right column shows encoded proteins embedded in the membrane. Cylinders represent transmembrane segments. N, N-terminus. '+' indicates the positions of positively charged amino acid residues in the protein sequence.

Experimental support for the proposed evolutionary pathway comes from studies of members of the small multidrug-resistance (SMR) transporter family. The best-studied SMR protein is EmrE from *E. coli*, an inner-membrane drug efflux pump with four TMSs, which is coded by a single gene and forms an antiparallel homodimer (128, 129, 166). It should be noted that the dual topology character of EmrE and the antiparallel orientation of subunits are still under debate and data has been presented that favors a parallel homodimer (141, 142, 152, 153). In the same SMR family, the EbrA and EbrB proteins from *B. subtilis* are encoded in an operon and the gene products form a heterodimer with antiparrallel oriented subunits providing strong evidence for the antiparallel orientation to be true in general (67). By genetic manipulation, it was shown that these proteins could be mutated back to a dual topology protein (68).

The model presenting the antiparrallel two-domain proteins as the direct result of an in-gene duplication event rather than a sequential mechanism, in which fusions evolve from a pair of genes, was suggested by a bioinformatics study of the DUF606 family (85). The DUF606 family is especially rich in evolutionary states proposed in the evolutionary pathway: single genes that would code for dual topology homodimeric proteins, paired genes coding for homologues proteins with fixed but opposite orientation in the membrane, that would form heterodimers, and fused genes that encode antiparallel two domain fusion proteins.

GltS of *Escherichia coli* transports glutamate in symport with Na^+ ions (22, 62) and is a member of the bacterial ESS family (Glutamate:Na^+ symporter; TC 2.A.27; (134)). The transporter is active as a homodimer (108). Previous studies on this transporter have resulted in a detailed structural model showing that GltS is an antiparallel two-domain protein (25, 26). GltS contains two homologues domains of 5 TMSs each. The loops between the 4^{th} and 5^{th} TMS in each domain form so called reentrant loops or pore-loops, which fold back in between the TMS from opposite sides of membrane (*trans* reentrant loops, Fig. 2A).

In this study, we used the GltS protein to study the evolution of membrane proteins consisting of two antiparallel domains by reconstructing the proposed evolutionary intermediates and showing that they produce active transporter proteins and finally to redirect the pathway to the alternative outcome. The gene encoding GltS protein was split

into two genes encoding the two domains after which the genes were placed in both orders in artificial operons ($GltS^{split}$). Next, a set of genes encoding swapped GltS proteins ($GltS^{swap}$) was constructed by fusing the two half genes in the reverse order. Presumably, $GltS^{swap}$ has the N-terminus of the first domain in the cytoplasm (Fig. 2B). The study shows that the intermediates in the evolutionary pathway can be reconstructed successfully.

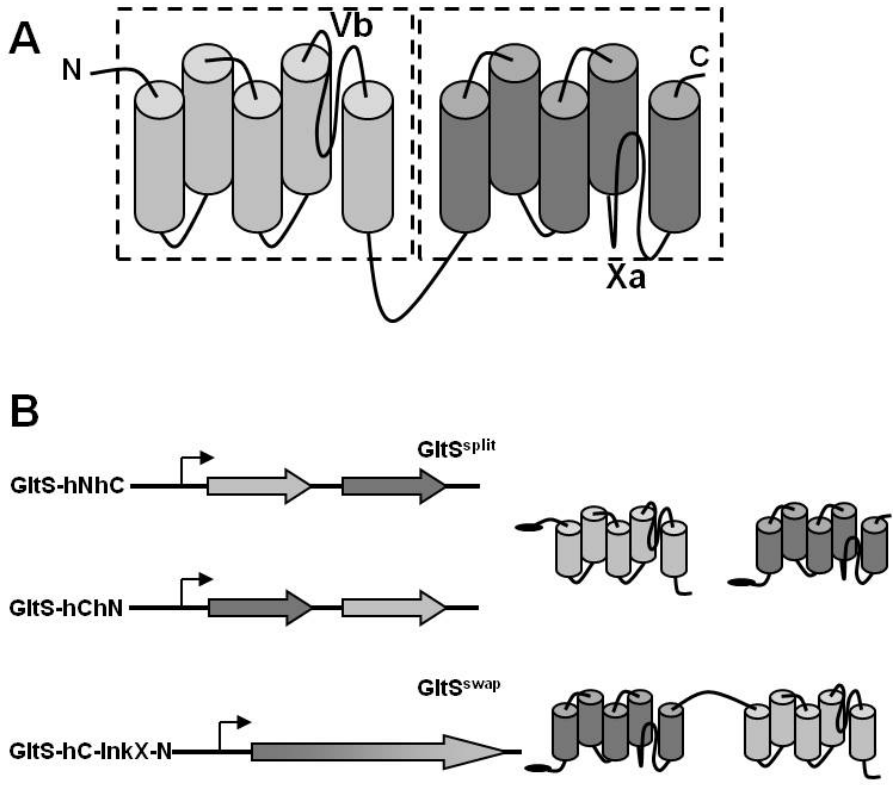


Figure 2. Schematic representation of (A) the structural model of GltS of *Escherichia coli* and (B) reconstructed evolutionary states of GltS used in this study. (A) Dashed boxes represent the N and C domains containing five TMSs (cylinders). Two reentrant loops termed Vb and Xa enter the membrane-embedded part of the protein from the periplasmic and cytoplasmic side of membrane, respectively. The two homologous domains are connected by a long cytoplasmic loop. (B) The left column shows genetic constructs. The N and C domain encoding part of the genes were indicated in two different gray colors. Upper, $gltS^{split}$ operons containing the genes encoding the N and C domain in both orders, bottom $gltS^{swap}$ gene with swapped N and C domains. The right column shows a topology models of encoded $GltS^{split}$ and $GltS^{swap}$ proteins. Cylinders represent transmembrane segments and closed ovals at the N-termini of the proteins represent His-tags. The periplasm is above, the cytoplasm below the models.

Experimental Procedures

Bacterial strains, growth conditions and expression of CitS derivatives. *Escherichia coli* strain DH5α was routinely grown in Luria Bertani Broth (LB) medium at 37 ^{o}C under continuous shaking at 150 rpm. Ampicillin was used at a final concentration of 50 μg/ml. GltS and GltS versions were expressed in *E. coli* DH5α cells harboring plasmid pBAD24 (Invitrogen) derivatives (150). The proteins were extended with 6 additional histidine residues at the N-terminus (His-tag). Expression of genes cloned in pBAD24 is under control of the arabinose promoter. Production of the proteins was induced by addition 0.01% arabinose when the optical density of the culture measured at 660 nm (OD_{660}) reached a value of 0.6.

Construction of gltS^{split} and gltS^{swap} carrying plasmids. Construction of the plasmids encoding the GltS variants used in this study was based on plasmid pBADHN-GltS described before (26). All genetic manipulations were done in *E. coli* DH5α cells. Plasmid pGltS-hNC was constructed by PCR insertion mutagenesis. Synthetic primers GltS_RBS_F and GltS_RBS_R (Table 1) were designed to insert a sequence containing a termination codon, a ribosomal biding site (RBS) and a start codon into the *gltS* gene in plasmid pBADHN-GltS between the codons encoding T201 and A202 in GltS. Plasmid pGltS-hNC encodes a protein containing the N-terminal 201 amino acid residues of the GltS protein (N domain) extended with a N-terminal His-tag and a protein consisting of the C terminal 200 amino acid residues of GltS with an additional Met residue on N-terminus (C domain). Plasmid pGltS-hC was constructed by deleting a *Nco*I-*Nco*I fragment from pGltS-hNC which results in a single gene encoding the C domain extended with His-tag at the N-terminus. Plasmid pGltS-hNhC was constructed by ligating a *EcoR*I- *EcoR*I fragment excised from pGltS-hNC into pGltS-hC digested with the same restriction enzyme. Plasmid pGltS-hNhC encodes a protein containing the N-terminal 201 amino acid residues of the GltS protein (N domain) extended with a N-terminal His-tag and a protein consisting of the C terminal 200 amino acid residues of GltS (C domain) and an additional Met residue at N-terminus, and, also, extended with a N-terminal His-tag (GltS^{split}). Plasmid pGltS-hChN was constructed by ligating a DNA

fragment containing the N domain of GltS, produced by PCR with the use of primers GltS_hN_F and GltS_N_R and pGltS-hNC as template, and digested with *Hpa*I/*Sph*I restriction enzymes into pGltS-hC digested with the same enzymes. Plasmid pGltS-hChN encodes the N and C domain as in pGltS-hNhC but in the reverse order ($GltS^{split}$). Plasmid pGltS-hC2N was constructed by ligating a DNA fragment, produced by PCR with the use of primers GltS_N_F and GltS_N_R and pGltS-hNC as template, and digested with *Stu*I/*Sph*I into pGltS-hC digested with *Hpa*I/*Sph*I restriction enzymes. Plasmid pGltS-hC2N encodes a single protein containing 404 amino acid residues (a Met residue followed by the C domain (200 residues) linked with 2 Ser residues linked to the N domain (201 residues)) extended with a N-terminal His-tag ($GltS^{swap}$). Plasmid pGltS-hC12N was constructed by ligating a DNA fragment, produced by PCR with the use of primers GltS_Nlnk12_F and GltS_N_R and pGltS-hNC as template, and digested with *Stu*I/*Sph*I into pGltS-hC digested with *Hpa*I/*Sph*I restriction enzymes. Plasmid pGltS-hC12N encodes a single protein containing 414 amino acid residues (like pGltS-hC2N, but with 12 residues in the linker) extended with a N-terminal His-tag ($GltS^{swap}$). Plasmid pGltS-hC19N was constructed by ligating a DNA fragment, produced by PCR with the use of primers GltS_Nlnk19_F and GltS_N_R and pGltS-hC12N as template, and digested with *Stu*I/*Sph*I into pGltS-hC digested with *Hpa*I/*Sph*I restriction enzymes. Plasmid pGltS-hC19N encodes a single protein containing the 421 amino acid (like pGltS-hC2N, but with 19 residues in the linker) extended with a N-terminal His-tag. Plasmids pGltS-hNC and pGltS-hC were not used during this study to produce proteins. All plasmids were sequenced and shown to contain the designed inserts (ServiceXS, Leiden, The Netherlands).

Table 1. Sequences of the primers used in this study

Primer name	Primer sequence[a]	Introduced restriction sites
GltS_RBS_F	5'-*CCAGGAAGTCCCGACG*TAA**GCTAGC**AGGAG**GAATTC**A**CC**A**TGG***CGTTTGAAAAGCCGG*[b,c]	*NheI,NcoI,EcoRI*
GltS_RBS_R	5'-*CCGGCTTTTCAAACG***C**CAT**GG**T**GAATTC**CCTCCTG**CTAGC**TTA*CGTCGGGACTTCCTGG*[b,c]	*NheI,NcoI, EcoRI*
GltS_hN_F	5'-CGCG**GTTAAC**GCTAGCAGGAGGAATTC	*HpaI*
GltS_N_R	5'-CGCGC**GCATGC**TTCCTCCTGCTAGCTTAC	*SphI*
GltS_N_F	5'-CGCG**AGGCCT**CA*ATGGTTCATCTCGATAC*[c]	*StuI*
GltS_Nlnk12_F	5'-CGCG**AGGCCT**CAGGATCCGGGAGCGGTTCGGGAAGTGGCTCA*ATGGTTCATCTCGATAC*[c]	*StuI*
GltS_N_lnk19_F	5'CGCG**AGGCCT**CAGGGTCAGGCTCGGGATCAGGTGGATCCGGGAGCGGTTC	*StuI*

[a] indicated in bold are introduced restriction sites, [b] indicated by underline are introduced termination codon, a ribosomal biding site (RBS) and a start codon, [c] indicated in italic are bases, which correspond to sequences in the *gltS* gene.

Transport assays in RSO membranes. *E. coli* DH5α cells expressing GltS variants were harvested from a 1 L culture by centrifugation at 10,000 x g for 10 min at 4 ^{o}C. Right-side-out (RSO) membrane vesicles were prepared by the osmotic lysis procedure as described (61). RSO membranes were resuspended in 50 mM KPi pH 7, rapidly frozen and stored in liquid nitrogen. Membrane protein concentration was determined by the DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA).

Uptake by RSO membranes was measured by the rapid filtration method. The membranes were energized using the K-ascorbate/phenazine methosulfate (PMS) electron donor system (73). Membranes were diluted to a final concentration of 0.5 mg/ml into 50 mM KPi pH 6.0 containing 70 mM Na^{+}, in a total volume of 100 µl at 30 ^{o}C. Under a constant flow of water-saturated air, and while stirring magnetically, 10 mM K-ascorbate and 100 µM PMS (final concentrations) were added and the proton motive force was allowed to develop for 2 min. Then, L-[^{14}C]-glutamate (Amersham Pharmacia, Roosendaal, The Netherlands) was added at a final concentration of 1.9 µM. Uptake was stopped by the addition of 2 ml of ice-cold 0.1 M LiCl, followed by immediate filtration over cellulose nitrate filters (0.45 µm, pore size). The filters were washed once with 2 ml of a 0.1 M LiCl solution and assayed for radioactivity. The background was estimated by adding the radiolabelled substrate to the vesicles suspension after the addition of 2 ml of ice-cold LiCl, immediately followed by filtering.

Partial purification by Ni^{2+}-NTA affinity chromatography. *E. coli* DH5α cells expressing GltS variants were harvested from a 200 ml culture by centrifugation at 10,000 x g for 10 min at 4 ^{o}C. Cells were washed with 50 mM KPi buffer pH 7 and resuspended in 2 ml of the same buffer and, subsequently, broken by a Soniprep 150 sonicator operated at an amplitude of 8 µm by 9 cycles consisting of 15 sec ON and 45 sec OFF. Cell debris and unbroken cells were removed by centrifugation at 9,000 rpm for 5 min. Membranes were collected by ultracentrifugation for 25 min at 80,000 rpm at 4 ^{o}C in a Beckman TLA 100.4 rotor and washed once with 50 mM KPi pH 7.0. The membranes (4 mg/ml) were solubilized in 50 mM KPi pH 8, 400 mM NaCl, 20% glycerol and 1% Triton X-100 followed by incubation for 30 min at 4 ^{o}C under

continuous shaking. Undissolved material was removed by ultracentrifugation at 80,000 rpm for 25 min at 4 ^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (50 µl bed volume per 5 mg protein), equilibrated in 50 mM KPi pH 8.0, 600 mM NaCl, 10% glycerol, 0.1% Triton X-100, 10 mM imidazole and incubated overnight at 4 ^{o}C under continuous shaking. Subsequently, the column material was pelleted by pulse centrifugation and the supernatant was removed. The resin was washed with 10 volumes of equilibration buffer containing 300 mM NaCl and 40 mM imidazole. The protein was eluted with half a bed volume of the washing buffer but containing 150 mM imidazole. The eluted fraction was stored at –20 ^{o}C until use.

Cross-linking studies. Aliquots of purified proteins were gently thawed on ice treated with 2.5 mM glutaraldehyde (GA, Sigma) at room temperature for 20 min. The treatment was quenched with 100 mM Tris-HCl pH 7.4 after which the samples were left at room temperature for 10 min. In control experiments 0.1% SDS was added to the samples, before GA treatment. Following treatment, samples were mixed with SDS sample buffer and run on a 12% SDS-PAGE.

Mass spectrometry analysis. Partially purified GltS variants were separated by SDS-PAGE using a 12 % gel and stained by Coomassie Brilliant Blue. Selected bands were cut from the gel. The pieces of gel were fragmented in smaller pieces, distained in 50 mM ammonium bicarbonate in 40% ethanol, dehydrated by a three times repeated treatment with 100 µl acetonitril, and dried completely using a SpeedVac centrifuge. The pieces of gel were reswollen by adding 20 µl of a 10 ng/µl trypsin solution and the samples were incubated overnight at 37°C. The peptides were extracted from the fluid by shaking for 20 min with 30 µl of a mixture of 60% acetonitril and 1% trifluoroacetic acid (TFA) in water. The extracted peptides were dried in a SpeedVac centrifuge and dissolved in 10 µl of 0.1% TFA in water. Aliquots of 0.75 µl of the peptide suspension were spotted on the MALDI target and mixed on the target in a 1:1 ratio with the matrix solution consisting of 10 mg/ml α-cyano-4-hydroxycinnamic acid (dissolved in 70% acetonitril and 0.1% TFA). The spots were allowed to dry completely before the MALDI-TOF experiment was performed on the Applied Biosystems 4700 Proteomics Analyzer.

Results

***Split GltS is an active transporter*.** The GltS protein was genetically split in two non-overlapping polypeptides corresponding to the N and C terminal domains (Fig. 2B). Plasmid pGltS-hNhC contains an artificial operon of two genes, the first one coding for the first 201 amino acids of GltS (N domain), the second for the last 200 amino acids of GltS (C domain) plus an additional Met residue at the N-terminus (start codon). The operon is under control of the arabinose promoter. In both cases, the encoded polypeptides were extended with six histidine residues (His-tag) at the N-terminus.

$GltS^{split}$ (GltS-hNhC) was tested for the ability to accumulate L-[^{14}C]-glutamate in right-side-out (RSO) membrane vesicles prepared from cells harboring the pGltS-hNhC plasmid. Glutamate uptake was measured in the presence of a proton motive force (pmf) that was generated using the artificial electron donor system ascorbate/PMS. Membrane vesicles prepared from *Escherichia coli* DH5α cells contain a basal level of glutamate transport activity due to endogenous glutamate transporters encoded on the chromosome (21). The background activity was estimated in membrane vesicles containing the citrate transporter CitS produced from the same expression system (Fig. 3B, close circles). CitS does not transport L-glutamate. RSO membranes containing wild type GltS showed an activity that was approximately 5 times higher than the background activity (Fig. 3B, diamonds). RSO membranes prepared from cells co-expressing the two domains of GltS retained about 50% uptake activity of membranes containing the wild-type version of GltS (Fig. 3B, open circles).

N-terminally His-tagged GltS purified by Ni^{2+}-NTA affinity chromatography migrates as a single band with apparent molecular mass of 35 kDa on SDS-PAGE (Fig. 3A, right panel) (26). The co-expressed N and C domains with His-tags at the N-termini were purified using the same protocol. On SDS-PAGE two bands showed up close to each other with apparent molecular masses around 17-18 kDa, strongly suggesting that they correspond to the N domain (201 residues) and C domain (200 residues) (Fig. 3A, left panel). Mass spectrometry analysis of the bands showed that the upper band corresponded to the N domain and the lower band to C domain. It follows that split GltS, expressed from an artificial operon encoding the two domains as separate proteins forms an active complex.

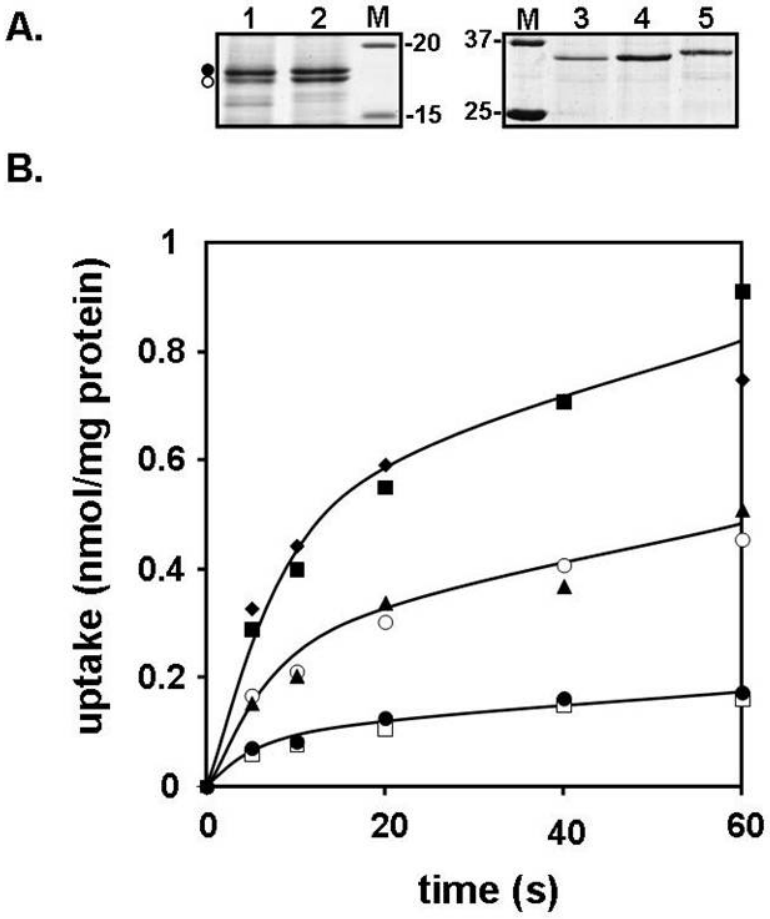


Figure 3. Expression (A) and activity (B) of reconstructed evolutionary states of GltS. (A) SDS-PAGE of purified GltS^{split} (left) produced from plasmids with the two half-genes in both orders (lanes 1 and 2) and the full-length GltS variants (right): wild-type (lane 3), swapped with the 12-residue linker (lane 4) and swapped with the 19-residue linker (lane 5). Black and white circles indicate the positions of the N and C domain, respectively. Lanes marked M show molecular mass standards with the masses in kDa as indicated. The gel was stained with Coomassie Brilliant Blue. (B) Glutamate uptake activity of GltS variants in RSO membranes. L-[^{14}C]-glutamate uptake was measured in RSO membrane vesicles containing wild type GltS (♦), the two versions of GltS^{split} (○,▲), the two versions of GltS^{swap} with the 12 (□) and the 19 (■) linker, and control membranes (●).

Almost identical results were obtained when the order of the two genes encoding the two domains of GltS in the artificial operon was reversed. In plasmid pGltS-hChN the first gene in the operon encodes the polypeptide corresponding to the C domain, while the second gene encodes the N domain (Fig. 2B). Membrane vesicles prepared from cells harboring plasmid pGltS-hChN showed glutamate uptake activity similar as observed for membrane vesicles prepared from cells harboring plasmid pGltS-hNhC, i.e. about 50% of

wild type GltS activity (Fig. 3B, triangles). The co-expressed, His-tagged domains of GltS^{split} (plasmid pGltS-hChN) showed a similar behavior during SDS-PAGE like domains expressed from pGltS-hNhC (Fig. 3A, left panel). It follows that the order of the genes in the operon does not affect the assembly or activity of the GltS complex in membrane.

Domain swap in full length GltS. Three vectors were constructed encoding in a single gene versions of GltS with swapped domains that differed in the length of the linker between the C-terminus of the C domain and the N-terminus of the N-domain (Fig. 2B). GltS-hC2N represents a swapped GltS protein with two Ser residues between the two domains. In GltS-hC12N the linker consists of 12-residues (SSGSGSGSGSGS) and in GltS-hC19N the linker is 19-residues long (SSGSGSGSGSGSGSGGSGS). All versions of the GltS^{swap} versions were extended with a N-terminal His-tag.

As mentioned above, N-terminally His-tagged wild type GltS purified by Ni^{2+}-NTA affinity chromatography migrates as a single band with apparent molecular mass of 35 kDa on SDS-PAGE (Fig. 3A, right panel). Following purification using the same protocol, no protein product of the expected size was observed for GltS^{swap} with the shortest linker of two Ser residues (not shown). In agreement, membrane vesicles containing GltS^{swap} with the 2-residue linker showed glutamate uptake activity corresponding to the background level, indicating that this version of GltS^{swap} is not active (not shown). Apparently, the protein is not stably assembled in the membrane. In contrast, the versions with 12- and 19-residue long linkers resulted in bands on SDS-PAGE with apparent molecular masses slightly larger than wild type GltS, in line with the additional mass introduced by the linkers (Fig. 3A, right panel). However, only the version with the longest linker of 19 residues was active in glutamate uptake activity. The activity was comparable to the activity observed for wild type GltS (Fig. 3B, closed squares). The vesicles containing GltS^{swap} with a 12-residue linker showed activity at the background level (Fig. 3B, open squares).

Dimer structure of swapped GltS. GltS was shown before to be a dimeric protein by electron microscopy and BN-PAGE (108). In contrast, it was not possible to demonstrate

the dimeric structure of purified wild type GltS by treatement with an unspecific cross-linker like glutaraldehyde as it was demonstrated for other transporter proteins (77). A possible explanation for this apparent discrepancy would be that the GltS dimer does not contain suitable reactive sites that would be productive in cross-linking. Surprisingly, treatment of the purified GltS^{swap} with 12- and 19-residue linkers with glutaraldehyde followed by analysis by SDS-PAGE resulted in complete disappearance of the 35 and 36 kDa bands, respectively, and, at the same time, new, somewhat fuzzy, bands appeared running at approximately double the mass (Fig. 4A and B). A small fraction of the protein did not enter the gel to any significant extent, suggesting some aggregation in the protein preparation. Treatment with glutaraldehyde in the presence of SDS prevented cross-linking of the proteins showing that cross-linking was the result of complex formation rather than random collisions. The same fuzziness of the band suggests that it is due to random labeling of versions of GltS protein molecules with glutaraldehyde. The results suggest that the introduced linkers are the target sites for cross-linking in the swapped GltS proteins.

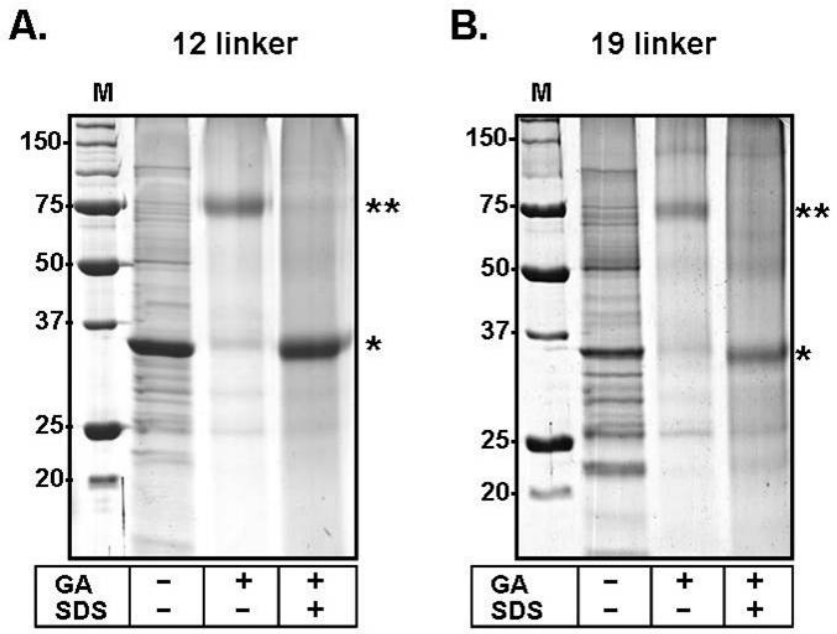


Figure 4. Dimeric state of GltS^{swap}. Purified GltS^{swap} variants with 12 (A) and 19 (B) residue long linkers were treated with glutaraldehyde (GA) at a concentration of 2.5 mM for 20 min in the presence and absence of 0.1% SDS as indicated, followed by SDS-PAGE. *, monomeric GltS, **, dimeric GltS. Lane M – molecular mass standards as indicated on the left in kDa.

Discussion

The structural model of the GltS transporter of *E. coli* shows two homologous domains, each containing five TMSs with inverted topology in the membrane (26). The motif of two antiparallel domains appears to be widespread in structures of many membrane proteins. The proposed model of evolution of these proteins involves duplication of a primordial gene followed by fusion and, thus, resulting in a single gene encoding a protein with two homologous domains (124, 129) (Fig. 1). In case of the antiparallel orientation of the two domains, the ancestral membrane protein is hypothesized to be 'dual topology', which means, the protein inserts with random orientation into the membrane. In this study a set of *gltS* genetic constructs was engineered that correspond to different evolutionary states in the pathway presented in Figure 1 with the aim to redirect the pathway to the alternative outcome. First, the protein was taken back one step in evolution by splitting the *gltS* gene in half and constructing two artificial operons containing the genes encoding the N and C domains in both orders ($GltS^{split}$; Fig. 2B). Next, a set of genes encoding swapped GltS proteins ($GltS^{swap}$) was constructed by fusing the two half genes in the reverse order (Fig. 2B).

The $GltS^{split}$, expressed from an artificial operon encoding the two domains as separate proteins is stably expressed, correctly assembled into the membrane and active in glutamate transport. The order of the two half genes in the artificial operons did not seem to affect the expression and activity levels. Three versions of $GltS^{swap}$ were constructed that differed in the length of the linker that connects the C-terminus of the C domain and the N-terminus of the N-domain. The linkers consisted of 2, 12 or 19 residues. The experiments show that the version of $GltS^{swap}$ with the shortest linker was not stably assembled in the membrane. The two remaining $GltS^{swap}$ versions were produced and inserted into the membrane but only the version with the 19-residue linker was active in glutamate transport. The activity was comparable to the activity observed for wild type GltS. Additionally, based on the previous experiments suggesting that GltS is a dimeric protein (108) cross-linking experiments were performed. Both expressed versions of $GltS^{swap}$ formed a dimeric complex after treatment with glutaraldehyde. The results indicate that even though $GltS^{swap}$ with the 12-residue linker was not active in Na^+-coupled glutamate transport, the protein still formed a dimeric complex.

Apparently, there is a threshold distance between the C-terminal of one domain and the N-terminus of the other domain above which stable assembly in the membrane occurs. A linker of 2 residues resulted in complete degradation of the polypeptide chain suggesting conformational stress and misfolding. Additional 12 residues in the linker would be above the threshold and results in a stable dimeric complex in the membrane (Fig. 4). However, the complex is not active. Activity was observed when the linker was increased up to 19 residues. The latter suggests that turnover of the complex involves inter subunit movement and that the 12-residue linker version is locked in one catalytic state of the protein.

GltS of *E. coli* is a member of the ESS family, that is believed to share the antiparallel two-domain structure with 32 other families in structural class ST[3] of the MemGen classification (87, 90, 91). Sequence analysis showed that in all but one family, [st312]NhaC (TC 2.A.35), the N-terminus of the N domain is at the external face of the membrane and the N-terminus of the C domain in the cytoplasm (87). The high frequency of one particular order of the domains suggests an evolutionary advantage for this organization or, alternatively, the families originate all from a single ancient duplication event before diversification into the different gene families occurred. Arguing against a single duplication event, an analysis of the DUF606 family, a family of membrane proteins in which different evolutionary states depicted in Figure 1 are found in nowadays members, demonstrated that duplication events within one family are quite frequent (85). In this study we demonstrate that changing the order of the domains in an antiparallel two-domain membrane protein that is at the end of its evolution with respect to the orientation in the membrane does not significantly affect its biogenesis and function. $GltS^{swap}$ is produced by the cells, inserted in the membrane, and equally active in ion-coupled solute transport. The only condition appears to be that the linker connecting the two domains is of about the same length as the linker in the original protein. Evolution towards the same domain order in almost all families sharing this structural organization must be driven by less pronounced features that nevertheless are important at an evolutionary time scale. Possibly, the cellular disposition of the connecting loop in the cytoplasm or exposed to the hazardous external medium may have played a role. At any rate, the study demonstrates that genetic engineering allows for the

reconstruction of evolutionary pathways and that evolutionary pathways can be manipulated.

Acknowledgment

We thank Fabrizia Fusetti of the Netherlands Proteomics Centre/Membrane Enzymology group of the University of Groningen for analysing samples by mass spectrometry. Research was funded by the Dutch Organization for Scientific Research (NWO-ALW).

Chapter VI

Summary

Chapter VI

All cells of living organisms are surrounded by a membrane that encloses the content of the cell (the cytoplasm). The membranes not only have a very important role in sustaining contact with the outside of the cell, but also in protecting the cell against the stressful environment surrounding it. The plasma membranes have to be a resistant to harmful compounds and, at the same time, they have to enable uptake of different substrates, like nutrition factors, and extrusion of waste products. Biomembranes consist of a phospholipid bilayer and a variety of proteins, called membrane proteins. The membrane proteins serve multiple functions, most of them are involved in transport and signaling, or they are key components in energy transduction, such as converting the chemical energy in ATP into electrochemical energy, or, in reverse, in ATP synthesis.

Membrane proteins involved in transport catalyze the physical process of movement of substances from one side of the membrane to the other. Important for the survival of the cell is uptake of required compounds. Membrane proteins participating in the transport of solutes and ions across the membrane may be divided in channels and transporters. Channels or pores form holes in the membrane that may or may not be gated, which facilitate the diffusion of substrates without requiring energy. Transporters are enzymes that couple the translocation of the solute to conformational changes of the transporter protein. The latter show a great variety of transport mechanisms, structures and specificities and they differ in the source of energy required for the transport process. Based on the energy requirement for the transport process they are divided in primary transporters using chemical energy, such as the free energy released in the hydrolysis of ATP, and secondary transporters using energy stored in electrochemical gradients of solutes and/or ions.

Secondary transporters, which are the main subject of this thesis, are one of the largest functional categories of membrane transport proteins. They use the free energy stored in ion or solute gradients to drive the transport of a solute across the cytoplasmic or internal membranes of biological cells. Accumulation of the solute at one side of the membrane is achieved by coupling the translocation of the solute to the translocation of one or more ions (H^+ or Na^+) that move down their own gradients, the proton motive force and/or Na^+ -ion motive force, respectively. Secondary transporters are commonly classified in three groups based on their mode of energy coupling: (i) uniporters catalyze the translocation

of a single solute across the membrane, (ii) symporters couple the translocation of a solute to the translocation of a co-ion(s) in the same direction, and (iii) antiporters couple the translocation of a solute and a co-ion(s) in opposite directions. The different modes of energy coupling enable transporters to play an important role in different aspects of the physiology of the cell.

Secondary transporters are typical integral membrane proteins that fold as a bundle of hydrophobic α-helices, which are oriented more or less perpendicular to the membrane. At the two sides of the membrane, the transmembrane segments (TMSs) are connected by hydrophilic loops of various lengths. Membrane proteins traverse the hydrophobic membrane in a zig-zag fashion. The hydropathy profile of the amino acid sequence of a membrane protein shows the distribution of these hydrophobic and hydrophilic residues over the sequence, which is believed to represent the folding of the protein in the membrane. The hydropathy profiles, like the 3D structures of homologous proteins are much better conserved than their amino acid sequence and, therefore, they report on the global fold of the proteins in a family. Knowledge about the structure of membrane proteins is an important source of information in molecular biology, which allowed understanding the function of transporters and providing insight into the molecular mechanism by which they work. Unfortunately, obtaining a high-resolution structure of membrane proteins in general is problematic because of their hydrophobic character, which makes it difficult to obtain good 3D crystals for X-ray diffraction. For this reason, bioinformatics and biochemical structural study of membrane proteins are so important.

Examining hydropathy profiles provides a mechanism to identify distantly related membrane proteins even when sequence identity is too low to detect homology. This has led to the MemGen classification in which families of membrane proteins are grouped in structural classes by comparing the average hydropathy profiles of the families. The MemGen classification system is not a membrane topology prediction method *per se*, but a major consequence of the approach is that all proteins in the different families in one class share the same fold, i.e., knowing the topology of one, is knowing them all.

The work described in this thesis focuses on the investigation of structural similarity between two families of secondary transporters and at the same time on the experimental validation of the MemGen classification system. Experiments were performed on two

transport proteins, CitS and GltS. The CitS protein of *Klebsiella pneumoniae* is a sodium dependent citrate transporter that belongs to the 2-hydroxycarboxylate transporter (2HCT) family. The GltS transporter of *Escherichia coli* is a sodium dependent glutamate transporter that belongs to the Glutamate Sodium Symporter (ESS) family. Both proteins transport the substrate in symport with two sodium ions. These two proteins, and all members of their families, are not related in amino acid sequence but share similar hydropathy profiles (Fig. 1A) and are found in the same structural class ST[3] of the MemGen classification system. Since the MemGen approach states that all proteins in one structural class share a similar fold, the well-established membrane topology model of the 2HCT family, mostly based on studies of the Na^+-citrate transporter CitS, was used to predict the membrane topology of the members of the ESS family.

The model was verified by accessibility studies of cysteine residues in single-Cys mutants of the GltS protein (**chapter 2** of this thesis). The structural model of the transporters shows a core of two homologous domains consisting of five TMSs each that are connected by a large cytoplasmic loop region (Fig. 1B). The CitS protein and all members of the 2HCT family have an additional TMS at the N-terminal end of the core structure, placing the N-terminus in the cytoplasm. Members of the ESS family including GltS do not have this additional segment and their structure corresponds to the core structure that has the N-terminus in the periplasm. Because of the odd number of helices in the two domains that form the core structure, they have opposite orientations in the membrane. In between the 4^{th} and 5^{th} TMS in each domain, the connecting loop folds back in between the TMSs to form a so-called 'reentrant loop'. The reentrant loop in the N-terminal domain enters the membrane from the periplasmic side, the one in the C-terminal domain from the cytoplasm. Importantly, it was shown experimentally that two cysteine residues in the predicted reentrant loop of GltS entering the membrane from the cytoplasmic side are accessible for small, membrane-impermeable thiol reagents from the periplasm, as was demonstrated before for the CitS.

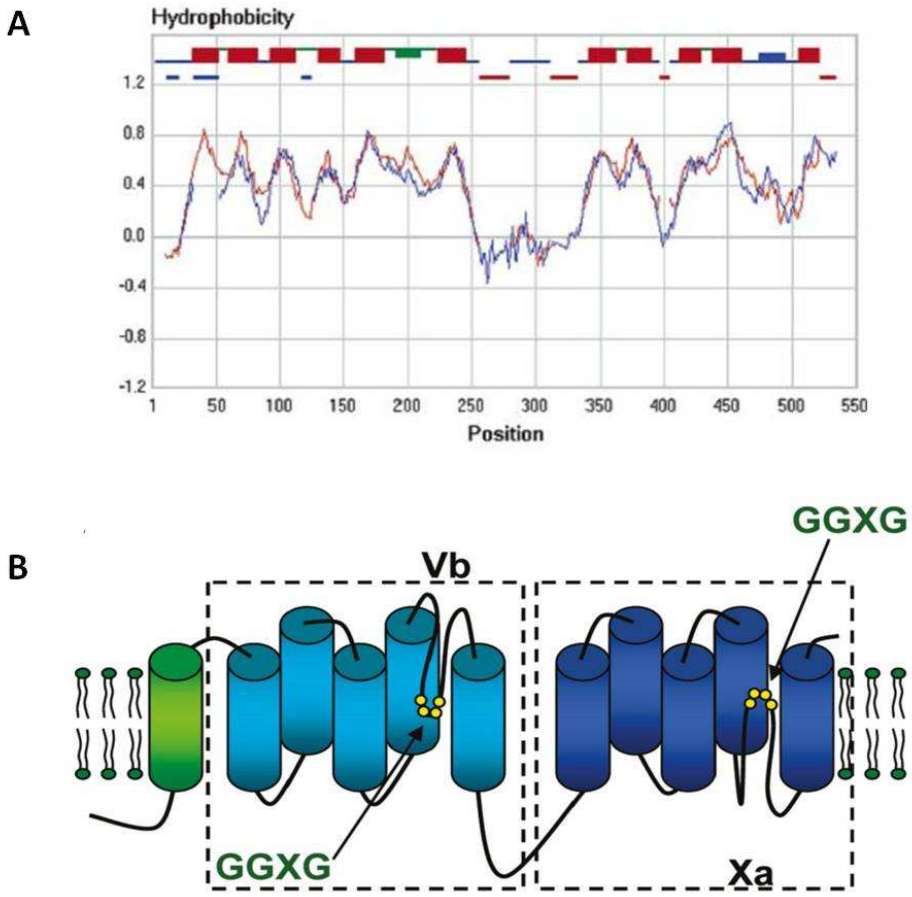


Figure 1.(A) Hydropathy profile alignment of the family profiles of the 2HCT family (red) and the ESS family (blue). The membrane topology model of the 2HCT family was indicated in the upper part. Trans membrane segments (red boxes), cytoplasmic loops (blue lines), periplasmic loops (green lines). Thickened parts of loop regions indicate the positions of reentrant loops. Horizontal blue and red lines indicate positions of gaps introduced by the algorithm in the alignment in the blue and red profiles, respectively.
(B) Structural model for the transporters of the 2HCT and ESS families. Two homologous domains containing 5 TMSs (cylinders) each with an inverted topology in the membrane were indicated in dashed boxes. Each domain contains a reentrant loop structure entering the membrane embedded part of the protein from the periplasmic and cytoplasmic side of the membrane, respectively (Vb and Xa). Members of the 2HCT family have an additional TMS at the N-terminus that is not present in members of the ESS family. GGXG sequence motifs present in reentrants loops are indicated.

Chapter 3 reports on the sequence motifs GGXG present in the reentrant loops of the transporters of both the 2HCT and ESS families. The motifs were shown to be at the vertex of the reentrant loops and were demonstrated to be crucial for the activity of the CitS and GltS proteins (Fig. 2). Importantly, the X to Cys mutation (S405C) in the cytoplasmic loop in the C-terminal half of CitS rendered the protein sensitive to the bulky, membrane impermeable thiol reagent AMdiS added at the periplasmic side of the

membrane, providing further evidence that this part of the loop is positioned between the TMSs. Similar conclusions were obtained from cysteine-scanning mutagenesis of a stretch of 18 residues in the reentrant loop in the C-domain of GltS.

Evidence presented in **chapter 4** suggested that in the 3D structure, the reentrant loops in the N and C domains are in close vicinity and overlapping at the interface of the two domains (Fig. 2). This was validated by cross-linking studies using a split transporter approach. It is believed that the two reentrant loops form (part of) the translocation pore and that translocation proceeds through an alternate access mechanism involving movement of the two domains relative to one another.

Membrane proteins, like GltS and CitS that consist of two homologous domains are believed to have evolved by gene-duplication and fusion events. In **chapter 5** of this thesis reconstruction and manipulation of the proposed evolutionary pathway by genetic engineering are described. A set of GltS versions corresponding to different evolutionary

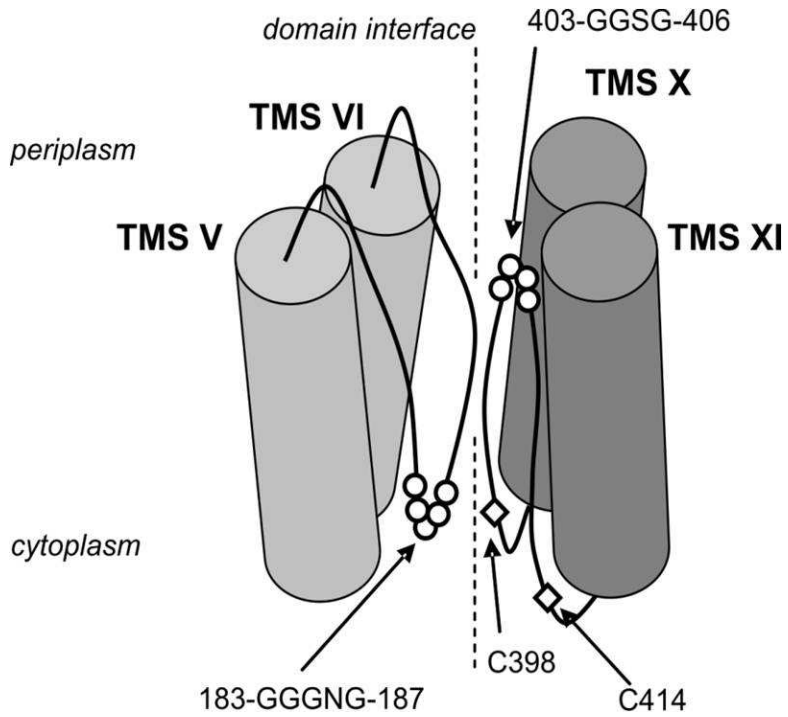


Figure 2. Structural model of the domain interface of CitS of *K. pneumoniae*. Shown are TMSs V and VI in the N domain and TMSs X and XI in the C domain together with the connecting reentrant loops. Two native cysteine residues in the Xa region are represented as open diamonds, and conserved GGXG motifs are represented as a open circles.

states was constructed. The set consisted of two types of gene pairs encoding domains as separated proteins forming antiparallel heterodimers, and a gene with swapped 3' and 5'halves of the *gltS* gene encoding GltS protein but with reverse order of the domains found in wild type GltS. All artificial evolutionary states were active supporting the proposed evolutionary pathway.

The main object of the research described in this thesis was to obtain structural data of the GltS and CitS proteins and at the same time find experimental validation of the MemGen classification system. The data support, in general, the structural and mechanistic similarity between the ESS and 2HCT transporters and, more particularly, the two-domain structure of the transporters and the presence and functional importance of the reentrant loops present in each domain of this type of secondary transporters. Starting from the bioinformatic data obtained from MemGen and a previous topology study of CitS we were able to predict and confirm experimentally the membrane topology of GltS and in this way show that the analysis by hydropathy profiles of families of membrane proteins is a powerful tool to study structures of membrane proteins in the absence of X-ray structural data. In addition to the 2HCT and ESS families, class ST[3] of MemGen contains over 30 other families of secondary transporters including the ion transporter (IT) superfamily, which shows the usefulness of MemGen. Additionally, the experimental data showed that reentrant loops are important structural feature of secondary transporters.

Now that a structure at the low resolution level is available for the GltS, CitS and for all proteins in ST[3], the next step will be to obtaining high-resolution structural data. Next, the goal will be to solve the structure of the transport proteins in each step of the transport cycle to fully understand the transport mechanism. All this information will be helpful for a better understanding of the function and importance of the secondary transporter proteins for living organisms.

Chapter VII

Samenvatting

Chapter VII

Alle cellen van organismen zijn omgeven door een membraan die de inhoud van de cel, het cytoplasma, omsluit. Het membraan heeft niet alleen een belangrijke rol in het contact met de omgeving, maar ook in het beschermen van de cel tegen potentiële gevaren van buiten. Het moet schadelijke stoffen buiten houden, maar ook de opname van allerlei stoffen, zoals bijvoorbeeld nutriënten, en het uitscheiden van afvalproducten mogelijk maken. Biomembranen bestaan uit een dubbele laag fosfolipiden en een assortiment van eiwitten, de zogenaamde membraaneiwitten. Deze verzorgen allerlei functies. De meeste zijn betrokken bij transport en het overbrengen van signalen uit de omgeving; andere zijn belangrijke componenten in energiehuishouding van de cel, zoals bij het omzetten van de chemische energie van ATP in electrochemische energie, of, omgekeerd, van electrochemische energie in ATP.

Membraaneiwitten die nodig zijn voor transport katalyseren het overbrengen van stoffen van één kant van het membraan naar de andere kant (translocatie). Ze kunnen ingedeeld worden in “kanalen” en “transporters”. Kanalen, ook wel poriën genoemd, vormen openingen in het membraan die al dan niet afgesloten kunnen worden en faciliteren diffusie van substraten zonder dat daar energie bij verbruikt wordt. Transporters werken als enzymen die verplaatsing van een verbinding koppelen aan conformationele veranderingen van het transporteiwit. Transporters vertonen een grote diversiteit aan transportmechanismen, structuren en substraatvoorkeuren en ook verschillen ze onderling met betrekking tot het soort energie dat gebruikt wordt in het transportproces. Aan de hand van de bovengenoemde eigenschappen zijn ze ingedeeld in primaire transporters, die chemische energie zoals die van ATP gebruiken, en secundaire transporters die de energie van elektrochemische gradiënten van verbindingen en/of ionen gebruiken.

Secundaire transporters, het onderwerp van dit proefschrift, vormen een van de grootste categorieën van membraan-transporteiwitten. Ze gebruiken de vrije energie die opgeslagen ligt in de gradiënten van ionen of verbindingen, als drijvende kracht voor het transport van een verbinding over het membraan. Accumulatie van een verbinding aan een kant van het membraan is mogelijk doordat het transport gekoppeld wordt aan de translocatie van een of meerdere ionen die met hun elektrochemische gradiënt meegaan. Deze ionen zijn meestal protonen (H^+) of natrium ionen (Na^+) en hun elektrochemische gradiënten worden respectievelijk proton motive force of Na^+-ion motive force genoemd.

Secundaire transporters worden doorgaans ingedeeld in drie groepen, gebaseerd op de wijze van energiekoppeling: (i) “uniporters” katalyseren de translocatie van slechts één verbinding over het membraan, (ii) “symporters” koppelen de translocatie van een verbinding aan de translocatie van één of meerdere “co-ionen” in gelijke richting, en (iii) “antiporters” koppelen de translocatie van een verbinding aan de translocatie van een andere verbinding of co-ion in tegengestelde richting. Door deze verschillende manieren van energiekoppeling kunnen transporters een belangrijke rol spelen in verschillende aspecten van de fysiologie van de cel.

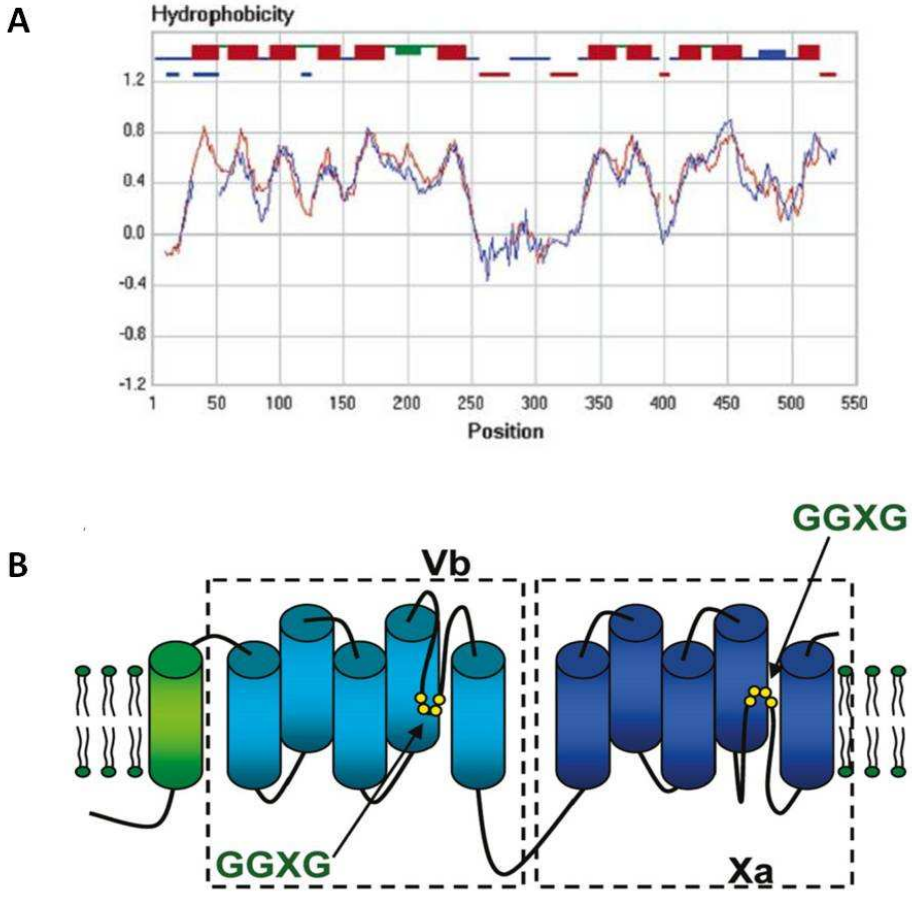
Secundaire transporters zijn eiwitten die opgevouwen zijn als een bundel hydrofobe α-helices die, ingebed in het membraan, min of meer loodrecht op het membraan staan. Aan beide zijden van het membraan zijn deze helices, ook wel transmembraansegmenten (TMS) genoemd, aan elkaar verbonden door hydrofiele lussen of “loops” van variabele lengte. Membraaneiwitten doorsteken zo het membraan op een zigzag manier. Het profiel van hydrofobiciteit van de aminozuursequentie van een membraaneiwit laat de verdeling van hydrofiele en hydrofobe aminozuren over de sequentie zien. Er wordt aangenomen dat deze de vouwing van het eiwit in het membraan weerspiegelt. De hydrofobiciteitsprofielen zijn, evenals 3D structuren, binnen een groep homologe eiwitten veel beter geconserveerd dan de aminozuursequenties en geven de globale vouwing van eiwitten binnen een familie weer. Kennis van de structuur van membraaneiwitten is een belangrijke bron van informatie in de moleculaire biologie en heeft het mogelijk gemaakt de functie van transporters en hun transportmechanismen beter te begrijpen. Helaas is het verkrijgen van structuren van membraaneiwitten met hoge resolutie over het algemeen problematisch, omdat het hydrofobe karakter het verkrijgen van goede kristallen voor röntgen diffractie bemoeilijkt. Om deze reden is bioinformatisch en biochemisch onderzoek aan membraaneiwitten zo belangrijk.

Het vergelijken van hydrofobiciteitsprofielen maakt het mogelijk om overeenkomsten in structuur te identificeren tussen membraaneiwitten die ver verwant zijn, zelfs als deze zo weinig overeenkomen in aminozuursequentie dat op basis daarvan geen homologie te ontdekken is. Dit heeft geleid tot de MemGen classificatie, waarin families van membraaneiwitten gegroepeerd zijn in structurele klassen door de gemiddelde hydrofobiciteitsprofielen van de families te vergelijken. Het MemGen classificatie

systeem is niet een methode voor het voorspellen van membraantopologie *per se*, maar een belangrijke consequentie van de benadering is dat alle eiwitten van verschillende families binnen één klasse dezelfde vouwing hebben. Dat betekent dat als je de topologie van één weet, je ook die van alle andere weet.

Het werk dat in deze studie beschreven staat, richt zich op de overeenkomsten in structuur tussen twee families van secundaire transporters en daarnaast op de experimentele validatie van het MemGen classificatie systeem. Twee transporteiwitten werden gebruikt in deze studie: CitS en GltS. Het CitS eiwit van *Klebsiella pneumoniae* is een natrium afhankelijke citraat transporter die behoort tot de 2-hydroxycarboxylaat transporter (2HCT) familie. De GltS transporter van *Escherichia coli* is een natrium afhankelijke glutamaat transporter die behoort tot de Glutamate Sodium Symporter (ESS) familie. Beide eiwitten transporteren het substraat en, in symport, twee natrium ionen. Deze eiwitten, en hun families, vertonen geen overeenkomst in aminozuursequentie, maar hebben wel hetzelfde hydrofobiciteitsprofiel (Fig. 1A) en zitten in dezelfde structurele klasse van het MemGen classificatiesysteem. Omdat volgens het MemGen systeem alle eiwitten binnen een structurele klasse een overeenkomstige vouwing hebben, is het gevestigde topologiemodel van de 2HCT familie, gebaseerd op onderzoek aan met name de Na^+-citraat transporter CitS, gebruikt om de membraantopologie van de leden van de ESS familie te voorspellen.

Het model werd geverifieerd met experimenten waarin de toegankelijkheid van cysteïne residuen in GltS mutanten, die slechts één cysteïne hebben, getest werd (**hoofdstuk 2** van dit proefschrift). Het structurele model van de transporters bestaat uit een kernstructuur van twee homologe domeinen die elk uit vijf TMS'en bestaan en die verbonden zijn door een grote cytoplasmatische "loop" (Fig. 1B). Het CitS eiwit en alle andere leden van de 2HCT familie hebben een additionele TMS aan het N-terminale uiteinde van de kernstructuur, waardoor de N-terminus in het cytoplasma uitkomt. Leden van de ESS familie, inclusief GltS, hebben dit segment niet en hun structuur komt overeen met de kernstructuur, waarvan de N-terminus in het periplasma uitkomt. Omdat het aantal helices van elk domein in de kernstructuur oneven is, hebben de domeinen een tegengestelde oriëntatie ten opzichte van elkaar in het membraan. Tussen het vierde en vijfde TMS van elk domein vouwt de verbindende loop terug, in plaats van het

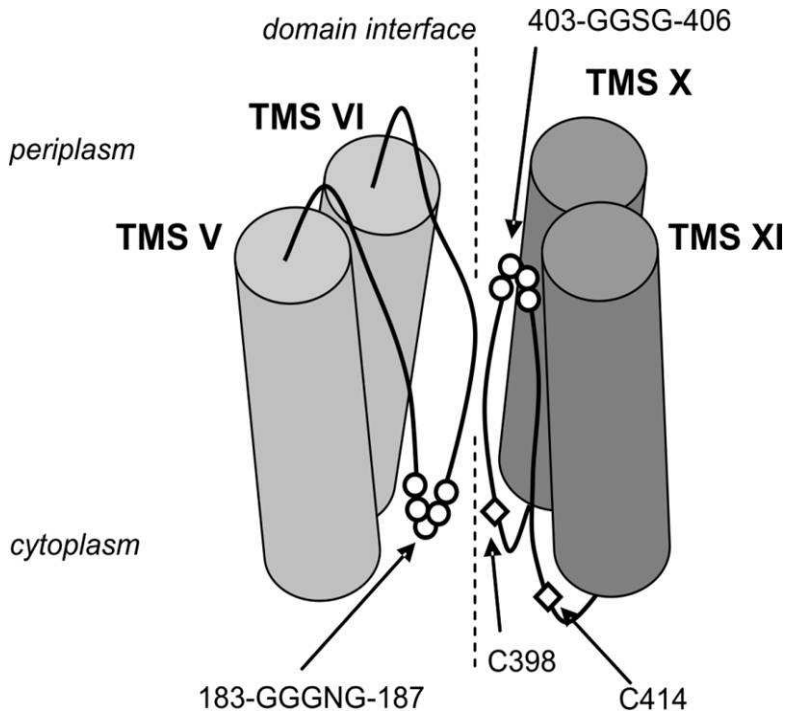


Figuur 1. (A) Alignment van de gemiddelde hydrofobiciteitsprofielen van de 2HCT familie (rood) en de ESS familie (blauw). Het membraantopologiemodel van de 2HCT familie staat aangegeven boven de profielen, met de transmembraansegmenten als rode rechthoeken verbonden door de cytoplasmatische en periplasmatische loops, weergegeven als respectievelijk blauwe en groene lijnen. De verdikte delen van de loops geven de positie van de reentrant loops aan. De losse horizontale blauwe en rode lijnen geven de posities van ”gaps” aan die voortkomen uit het algoritme van de alignment van respectievelijk de blauwe en rode profielen. (B) Structureel model van de transporters van de 2HCT en ESS families. De twee homologe domeinen met elk 5 TMS’en (cilinders) en hun tegengestelde oriëntatie zijn afgebeeld binnen de gestippelde vierkanten. Beide domeinen hebben een reentrant loop die het membraan-ingebedde deel van het eiwit binnengaan aan, respectievelijk, de periplasmatische of cytoplasmatische zijde van het membraan (Vb en Xa). Leden van de 2HCT familie hebben een additionele TMS aan de N-terminus die niet aanwezig is in de ESS familie. De GGXG sequentie-motieven in de reentrant loops zijn aangegeven.

membraan in zijn geheel te doorkruisen (Fig. 1B) en vormt daarmee een zogenaamde “reentrant loop”. De reentrant loop in het N-terminale domein gaat het membraan binnen aan de periplasmatische zijde en die van het C-terminale domein aan de cytoplasmatische. Een belangrijk experimenteel resultaat toonde aan dat de twee cysteïne residuen in de voorspelde reentrant loop van GltS (degene die het membraan aan de cytoplasmatische zijde binnengaat) vanuit het periplasma bereikbaar zijn voor kleine thiol reagentia die niet door het membraan kunnen diffunderen, zoals eerder was aangetoond voor CitS.

Hoofdstuk 3 handelt over het GGXG sequentie-motief dat aanwezig is in de reentrant loops van de transporters van zowel de 2HCT als de ESS familie. Er is aangetoond het motief op het uiteinde van de reentrant loops ligt en dat het cruciaal is voor de activiteit van CitS en GltS (Fig. 2). Een belangrijke vinding was dat de mutatie van de serine naar een cysteïne op positie 405, oftewel X in het GGXG motief van de reentrant loop van het C-terminale domein, het eiwit gevoelig maakte voor van buiten toegevoegd AMdiS, een relatief groot en membraan impermeabel thiol reagens. Dit verschafte extra bewijs dat dit deel van de loop gepositioneerd is tussen de TMS’en. Overeenkomstige conclusies konden worden getrokken uit “cysteine scanning” mutagenese van een reeks van 18 residuen in de reentrant loop in het C-domein van GltS.

Resultaten beschreven in **hoofdstuk 4** suggereerden dat in de 3D-structuur, de reentrant loops van het N- en C-domein dicht bij elkaar liggen en overlappen daar waar de twee domeinen tegen elkaar aan liggen (Fig. 2). Dit werd bevestigd door cross-linking experimenten waarbij een opgesplitste transporter gebruikt is. Er wordt aangenomen dat de twee reentrant loops (een deel van) de translocatie porie vormen en dat translocatie zich voltrekt volgens een mechanisme van alternerende toegang waarbij de twee domeinen zich ten opzichte van elkaar bewegen.



Figuur 2. Model van het snijvlak van de twee domeinen van CitS van *K. pneumoiae*. Afgebeeld zijn TMS V en VI in het N-domein en TMS X en XI in het C-domein, samen met de verbindende reentrant loops. Twee oorspronkelijke cysteïnes in de Xa loop zijn afgebeeld als ruiten; de residuen van de geconserveerde GGXG motieven als cirkels.

Er wordt aangenomen dat membraaneiwitten die, zoals CitS en GltS, uit twee domeinen bestaan, geëvolueerd zijn door genduplicaties en -fusies. In **hoofdstuk 5** van dit proefschrift worden de reconstructie en manipulatie van de voorgestelde evolutionaire route door middel van genetische recombinatie beschreven. Een set van GltS mutanten die corresponderen met de verschillende evolutionaire tussenvormen werd geconstrueerd. De set bestond uit twee types genenparen die coderen voor de domeinen als gescheiden eiwitten die antiparallelle heterodimeren vormen, en een gen waarin de 3’- en 5’-helften van het *gltS* gen van positie gewisseld zijn en dat daarmee codeert voor een GltS eiwit waarbij de domeinen in de omgekeerde volgorde achter elkaar gezet zijn. Alle artificiële evolutionaire tussenvormen waren actief en ondersteunden daarmee de voorgestelde evolutionaire route.

Het hoofddoel van het onderzoek dat in dit proefschrift beschreven staat was om structuur data van het GltS en CitS eiwit te verkrijgen en tegelijkertijd experimentele

validatie van het MemGen classificatie systeem. De data ondersteunen in algemene zin de structurele en mechanistische overeenkomsten tussen de ESS en 2HCT transporters en, meer in het bijzonder, de twee-domein structuur van de transporters, de aanwezigheid van de reentrant loops - die in elk domein van dit type secundaire transporters aanwezig zijn – en hun belangrijke rol in de functie van het eiwit. Beginnend met bioinformatische data verkregen met MemGen en een eerder onderzoek aan de topologie van CitS, konden we een membraantopologie van GltS voorspellen en die vervolgens experimenteel bevestigen. Hiermee lieten we zien dat de analyse van hydrofobiciteitsprofielen van families van membraaneiwitten een effectieve methode is om de structuur van membraaneiwitten te onderzoeken als een kristalstructuur niet voorhanden is.

Nu een structuur van lage resolutie van GltS, CitS en daarmee alle andere eiwitten in ST[3], bekend is, is de volgende stap het verkrijgen van hoge-resolutie data van de structuur. Vervolgens zal het doel zijn om de structuur van de transporters in elke afzonderlijke stap van de transportcyclus op te lossen om het transportmechanisme volledig te kunnen begrijpen. Al deze informatie zal helpen bij het beter begrijpen van de functie en belang van secundaire transporters in levende organismen.

(translated by Hein Trip)

Chapter VIII

Streszczenie

Chapter VIII

Wszystkie komórki żywych organizmów otoczone są przez błonę cytoplazmatyczną, która utrzymuje zawartość komórki (cytoplazmę). Błony komórkowe pełnią nie tylko niezwykle ważną role w podtrzymywaniu kontaktów ze światem zewnętrznym, ale jednocześnie ochraniają komórkę przed szkodliwym środowiskiem zewnętrznym. Błony aby spełniać swą rolę ochronna, musza być nieprzepuszczalną barierą dla substancji szkodliwych dla żywej komórki, ale jednocześnie, muszą umożliwiać pobieranie różnorodnych substratów, takich jak czynniki odżywcze oraz pozwalać na usuwanie niekorzystnych produktów metabolizmu. Biomembrany składają się z dwu-warstwy lipidowej oraz białek, nazywanych białkami błonowymi. Białka błonowe pełnią różnorodne funkcje, większość z nich zaangażowana jest w transport i przekazywanie sygnałów, lub są one głównymi komponentami w szlakach przekazywania energii, takich jak przekształcanie energii chemicznej zawartej w ATP w elektrochemiczną, lub przeciwnie, zaangażowane są w syntezę ATP.

Białka błonowe zaangażowane w transport katalizują fizyczny proces przenoszenia transportowanej substancji z jednej strony błony na drugą. Niezwykle ważne dla przeżycia komórki jest pobieranie niezbędnych do życia czynników. Białka błonowe biorące udział w transporcie można podzielić na kanały i transportery. Kanały lub pory tworzą w błonie komórkowej „dziury", które mogą posiadać system bramkujący pozwalający na swobodną dyfuzje związków bez nakładu energii. Transportery są enzymami które podczas przenoszenia substratów zmieniają swoją konformację. Transportery są bardzo zróżnicowaną grupą białek, wykazującą różnorodne mechanizmy transportu, struktury oraz różnią się sposobami pozyskiwania energii potrzebnej do przeprowadzenia procesu przenoszenia substancji. Bazując na sposobie pozyskiwania energii do transportu, wśród transporterów można wyróżnić (i) pierwotne transportery aktywne, wykorzystujące energię chemiczną, wydzieloną na przykład podczas hydrolizy cząsteczki ATP oraz (ii) wtórne transportery aktywne wykorzystujące energię zmagazynowaną w elektrochemicznym gradiencie jonów.

Wtórne transportery, które są głównym przedmiotem tej rozprawy, są jedną z największych grup funkcjonalnych transportujących białek błonowych. Używają one wolnej energii zakumulowanej w gradiencie elektrochemicznym jonów lub innych związków do transportu innych substancji przez błony cytoplazmatyczne lub wewnętrzne

błony komórek. Akumulacja danego związku po jednej ze stron błony jest uzyskiwana poprzez sprzężony transport tego związku wraz z transportem jednego lub większej ilości jonów (H^+ lub Na^+), które są przenoszone zgodnie z gradientem ich stężenia, tzw. siła motoryczna jonów wodorowych lub siła motoryczna jonów sodowych. Wtórne transportery są powszechnie dzielone na trzy typy: (i) uniportery katalizują transport jednej substancji przez błonę, (ii) symportery sprzęgają transport danej substancji z transportem jonu/jonów w tym samym kierunku, oraz (iii) antyportery transportują substancje w przeciwnych kierunkach. Wszystkie typy transportu pozwalają wtórnym transporterom odgrywać ważną role w rożnych aspektach fizjologii komórki.

Wtórne transportery są typowymi białkami błonowymi. Ze względu na to że białka te wbudowują się w hydrofobiczna błonę komórkową, część ich łańcucha polipeptydowego tworzy charakterystyczne spirale, α-helisy, które przecinają błonę mniej lub bardziej prostopadle. Hydrofobiczne α-helisy, nazywane często segmentami transbłonowymi (z ang. transmebrane segments, TMSs), są połączone przez hydrofilowe pętle znajdujące się na zewnątrz błony. W ten sposób białka błonowe przecinają hydrofobową błonę komórkową w sinusoidalny sposób. Profile hydrofobowości sekwencji aminokwasów budujących białko błonowe pokazują transbłonowe obszary hydrofobowe (α-helisy) oraz hydrofilowe obszary odpowiadające łączącym je pętlą co jednocześnie daje informacje na temat budowy (fałdowania) danego białka w błonie. Profile hydrofobowości, podobnie jak trójwymiarowe struktury homologicznych białek są znacznie lepiej zachowywane w ewolucji niż ich sekwencja aminokwasowa, dlatego też zawarta jest w nich informacja na temat struktury w całej rodzinie białek. Znajomość struktury białek błonowych jest ważnym źródłem informacji w biologii molekularnej ponieważ umożliwia zrozumienie samego procesu transportu oraz dostarcza informacji na temat jego mechanizmu i sposobu pracy transporterów. Niestety, otrzymywanie struktur białek błonowych w wysokiej rozdzielczości jest bardzo problematyczne i trudne do osiągnięcia z powodu ich właściwości hydrofobowych, które powodują problemy z otrzymaniem dobrej jakości kryształów do dyfrakcji rentgenowskiej. Z tego powodu ważne w analizie białek błonowych stały się narzędzia bioinformatyczne i biochemiczne.

Badanie profili hydrofobowości pozwala na zidentyfikowanie podobnej budowy białek błonowych nawet wówczas gdy białka te różnią się na poziomie sekwencji

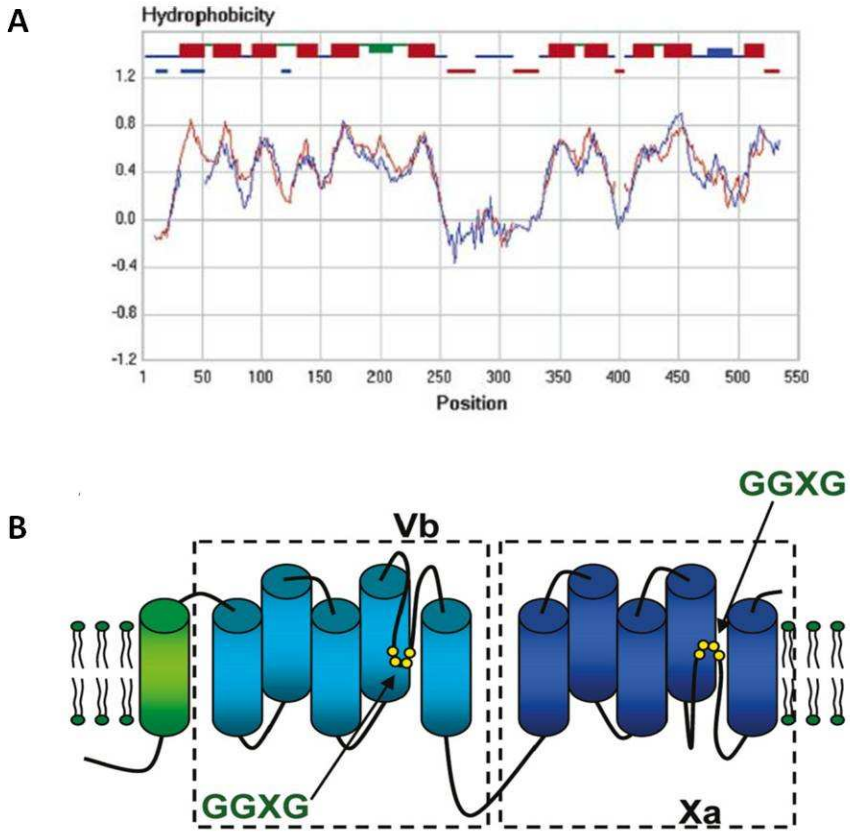
aminokwasowej na tyle, że trudno wykryć homologie pomiędzy nimi. Powyższy tok działania doprowadził do powstanie bazy danych MemGen w której system klasyfikacji polega na grupowaniu rodzin białek błonowych w oparciu o podobieństwa uśrednionych dla rodzin profili hydrofobowości. System klasyfikacji białek MemGen nie jest metodą przewidywania topologii (fałdowania) białek *per se*, ale główną konsekwencją tego sposobu grupowanie białek jest to że wszystkie białka z różnych rodzin znajdujące się w tej samej klasie są podobnie sfałdowane w błonie komórkowej. Z tego wynika następna konkluzja, iż znając budowę jednego białka z danej klasy, znamy budowę wszytych białek znajdujących się w tej samej klasie.

Praca opisana w niniejszej rozprawie skupiona jest na poszukiwaniu podobieństwa w budowie białek znajdujących się w dwóch rodzinach wtórnych transporterów oraz jednocześnie na eksperymentalnym potwierdzeniu prawdziwości systemu klasyfikacji bazy MemGen. Prace były prowadzone na dwóch transporterach, CitS oraz GltS. Białko CitS jest zależnym od jonów sodu transporterem cytrynianu z bakterii *Klebsiella pneumoniae*, należącym do rodziny białek 2HCT (ang. 2-hydroxyccarboxylate). Natomiast białko GltS z bakterii *Eschericha coli* transportuje w symporcie z dwoma jonami sodowymi glutaminian i należy do rodziny biaek ESS (ang. glutamate-E sodium symporter). Białka te, oraz członkowie ich rodzin, nie są podobne do siebie na poziomie sekwencji aminokwasowej ale mają podobne profile hydofobowści (Rys. 1A) i znajdują się w tej samej klasie strukturalnej ST[3] w systemie klasyfikacji bazy MemGen. Ponieważ uważa się, iż wszystkie białka znajdujące się w tej samej klasie strukturalnej mają podobną budowę, użyliśmy dobrze znanego modelu budowy białek z rodziny 2HCT, bazującego głównie na badaniach przeprowadzonych na białku CitS, do zbudowania modelu fałdowania białek z rodziny ESS.

Model ten został zweryfikowany poprzez studia dostępności cystein wprowadzonych do mutantów GltS zawierających tylko jeden taki aminokwas (ang. single-Cys) co zostało opisane w **rozdziale 2** tej rozprawy. Zaproponowany model strukturalny tych transporterów zakłada iż białka te zawierają główną cześć (rdzeń) zbudowaną z dwóch homologicznych domen zawierających po pięć segmentów transbłonowych (TMS), które są połączone za pomocą długiej cytoplazmatycznej pętli (Rys. 1B). Białko CitS, oraz wszystkie inne białka z rodziny 2HCT mają dodatkowy transbłonowy segment na N-

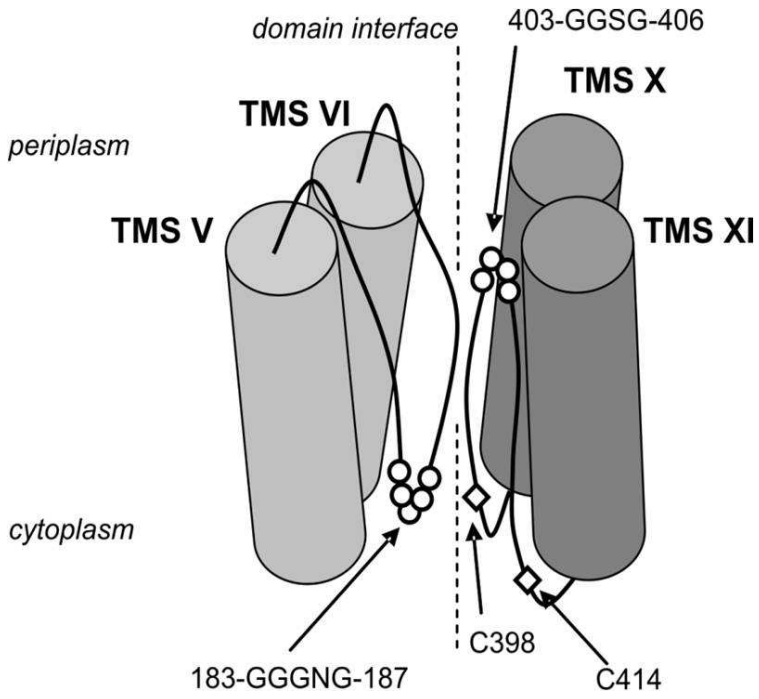
terminalnym końcu głównej struktury (rdzeniu) białka. Białka należące do rodziny ESS nie posiadają dodatkowego segmentu transbłonowego i dlatego ich budowa jest równoznaczna z budową głównej części (rdzenia) modelu. Ponieważ każda z dwóch homologicznych domen budujących rdzeń białka zbudowana jest z nieparzystej liczby segmentów transbłonowych posiadają tzw. przeciwną orientacje w błonie. Pętla łącząca 4 i 5 segment transbłonowy w każdej z domen, fałduje się tworząc nietypową strukturę tzw. powtórnej pętli (ang. reentrant loop) lub pętli kanałowej (pore loop). Pętla ta zawraca i zawija się w błonie wchodząc pomiędzy α-helisy, ale nigdy nie przebija się całkowicie na drugą stronę błony. Pętle znajdujące się w N-domenie białka wchodzą do błony z peryplazmatycznej strony, natomiast pętle znajdujące się w C-domenie wchodzą w błonę z cytoplazmatycznej strony. Należy podkreślić iż zostało eksperymentalnie udowodnione że dwie cysteiny znajdujące się w powrotnej pętli białka GltS wchodzącej do błony z cytoplazmatycznej strony (C-domena) są dostępne dla małych ale nieprzechodzących swobodnie przez błonę reagentów z przeciwnej, peryplazmatycznej strony. Udowadnia to obecnośc pętli kanałowych w białku GltS, wykazanych w ten sam sposób w białku CitS.

Badania opisane w **rozdziale 3** skupione sa na obecności i właściwościach motywu GGXG sekwencji aminokwasowej zlokalizowanej w poprzednio wspomnianych pętlach. Motyw ten znajduje sie w sekwencji białek w rodzinach 2HCT oraz ESS. Udowodnione zostało że motyw ten znajduje sie na wierzchołku pętli i jest kluczowy dla aktywności białek CitS oraz GltS (Rys. 2). Należy zaznaczyć, że mutacja wprowadzająca aminokwas cysteinę w miejsce aminokwasu X z motywu w pętli cytoplazmatycznej w białku CitS (tzw. mutant S405C) wykazała że aminokwas ten jest dostępny dla dużego, nieprzchodzącego przez błony reagentu AMdiS dodanego z peryplazmatycznej - przeciwnej strony błony. Eksperyment ten był kolejnym dowodem na to że białko CitS posiada pętle kanałową która wchodzi pomiędzy segmenty błonowe zawija się i zawraca. Podobne wyniki uzyskano dla badań nad pętlą cytoplazmatyczną w białku GltS, tutaj jednak badania dotyczyły wprowadzenia mutacji w 18 aminokwasowym odcinku pętli.



Rysunek 1. (A) Alignment (dopasowanie) profili hydrofobowości rodzin 2HTC (czerwony) i ESS (niebieski). W górnej części znajduje się model fałdowania (topologii) w błonie białek z rodziny 2HCT. Segmenty transbłonowe (czerwone prostokąty), pętle cytoplazmatyczne (linie niebieskie), pętle peryplazmatyczne (linie zielone). Pogrubione pętle oznaczają pętle powrotne (ang. reentrat loop). Poziome czerwone i niebieskie linie oznaczają miejsca wystąpienia przerw w alignmencie odpowiednio; profili czerwonych i niebieskich. (B) Model strukturalny transporterów z rodzin 2HTC i ESS. Dwie homologiczne domeny zbudowane z 5 segmentów transbłonowych (walce), każda z odwróconą orientacja w błonie, zaznaczone są prostokątami z przerywanej linii. Każda domena zawiera pętle powrotną wchodzącą w błonę z peryplazmatycznej (Vb) lub cytoplazmatycznej (Xa) strony. Białka należące do rodziny 2HCT mają dodatkowy segment transbłonowy na N-terminalnym końcu łańcucha polipeptydowego, którego brak jest u białek z rodziny ESS. Zaznaczone zostały motywy GGXG sekwencji aminokwasów w pętlach powrotnych.

Badania przedstawione w **rozdziale 4** sugerują iż w strukturze przestrzennej pętle powrotne z N (peryplazmatyczna) i C domen (cytoplazmatyczna) opisywanych białek znajdują sie blisko i nachodzą na siebie w miejscu styku dwóch domen budujacych te białka (Rys. 2). Zbadane zostało to za pomocą eksperymetów zwanych „cross-loinking" z użyciem rozdzielonych domen transporterów (uzyskanych drogą genetycznej modyfikacji). Uważa się że pętle powrotne tworzą szlak transportu substancji w białku i poprzez zmianę konformacji dwóch domen względem siebie pozwalają na transport w tzw. mechanizmie alternatywnego dostępu (ang. alternate access model).



Rysunek 2. Model strukturalny powierzchni pomiędzy domenami w białku CitS z *K. pneumoniae*. Pokazane są segmenty transbłonowe (TMS) numer V i VI z N-domeny oraz segmenty transbłonowe numer X i XI z C-domeny wraz z łączącymi je pętlami powrotnymi. Dwie naturalnie występujące cysteiny w regionie Xa zaznaczone są jako romby, natomiast motywy sekwencji GGXG zostały oznaczone jako koła.

Uważa się że białka błonowe, wśród nich CitS oraz GltS, zbudowane z dwóch homologicznych domen wyewoluowały na drodze duplikacji i późniejszej fuzji genów pierwotnych. W **rozdziale 5** rozprawy opisana została próba rekonstrukcji i manipulacji proponowanej drogi ewolucji. Skonstruowano, drogą manipulacji genetycznych, zestaw wersji białka GltS odpowiadających różnym etapą ewolucji. Zestaw składał się z dwóch typów par genów kodujących domeny jako osobne białka, w różnej kolejności w operonie, formujące przeciwstawne hetrodimery, oraz gen z przestawioną kolejnością domen w porównaniu do oryginalnego genu *gltS*. Wszystkie sztucznie utworzone białka korespondujące do różnych etapów ewolucji dwu-domenowych białek były aktywne przez co potwierdzały zaproponowany szlak ewolucyjny.

Głównym celem badań opisanych w tej rozprawie było zdobycie danych na temat budowy białek GltS i CitS oraz jednoczesne eksperymentalne potwierdzenie prawidłowości systemu klasyfikacji bazy MemGen. Dane uzyskane w czasie badań potwierdziły podobieństwo w strukturze oraz w mechanizmie transportu pomiędzy transporterami z rodzin ESS i 2HTC. Poznane oraz udowodnione zostały nowe szczegóły na temat budowy dwu-domenowej tych białek oraz obecność, i kluczowym znaczeniu, pętli powrotnych w każdej z domen tych wtórnych transporterów. Rozpoczynając prace na danych bioinformatycznych uzyskanych za pomocą bazy danych MemGen oraz danych z wcześniejszych badań nad topologią białka CitS mogliśmy zbudować oraz udowodnić eksperymentalnie model budowy transportera GltS i tym samym wykazać że analiza profilów hydrofobowości rodzin białek błonowych jest potężnym i skutecznym narzędziem do studiowania struktur białek błonowych, zwłaszcza przy braku danych strukturalnych uzyskanych drogą krystalografii rentgenowskiej. Klasa strukturalna ST[3] z bazy MemGen, oprócz rodzin 2HCT i ESS, zawiera ponad 30 innych rodzin wtórnych transporterów włączając w to superrodzinę IT (ang. ion transporter), co wskazuje na dużą użyteczność MemGen. Dodatkowo uzyskane wyniki pokazały, że pętle powrotne zawijające sie w błonie stanowią niezwykle ważny element budowy i działania we wtórnych transporterach.

Obecnie, gdy znane są ogólne zarysy budowy białek GltS i CitS, oraz tym samym wszystkich białek w ST[3], następnym etapem będzie uzyskanie dokładnych danych na temat struktury tych transporterów, np. za pomocą krystalografii rentgenowskiej. Kolejnym krokiem będzie uzyskanie danych strukturalnych dla białek transportowych w każdym z etapów cyklu transportu, co umożliwi lepsze zrozumienie mechanizmu transportu. Wszystkie powyższe informacje pozwolą na lepsze i dokładniejsze zrozumienie funkcji i działania wtórnych transporterów w żywych organizmach.

Abbreviations

TMS – transmembrane segment

pmf – proton motive force

RSO – right-side-out

PMS – phenazine methosulfate

NEM – *N*-ethylmaleimide

FM – fluorescein-5-maleimide

AMdiS – 4-acetamido-4’-maleimidylstilbene-2,2’-disulfonic acid

MTSET – [2-(trimethylammonium)ethyl] methanethiosulfonate bromide

MTSES – sodium (2-sulfonatoethyl) methanethiosulfonate

DTT – dithiothreitol

NaTT – sodium tetrathionate

GA – glutaraldehyde

References

1. **Abramson, J., I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata.** 2003. Structure and mechanism of the lactose permease of Escherichia coli. Science **301:**610-5.

2. **Akabas, M. H., D. A. Stauffer, M. Xu, and A. Karlin.** 1992. Acetylcholine receptor channel structure probed in cysteine-substitution mutants. Science **258:**307-10.

3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.

4. **Andrade, S. L., A. Dickmanns, R. Ficner, and O. Einsle.** 2005. Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus. Proc Natl Acad Sci U S A **102:**14994-9.

5. **Bandell, M., V. Ansanay, N. Rachidi, S. Dequin, and J. S. Lolkema.** 1997. Membrane potential-generating malate (MleP) and citrate (CitP) transporters of lactic acid bacteria are homologous proteins. Substrate specificity of the 2-hydroxycarboxylate transporter family. J Biol Chem **272:**18140-6.

6. **Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne.** 2000. The Protein Data Bank. Nucleic Acids Res **28:**235-42.

7. **Bhat, B. G., S. R. Rapp, J. A. Beaudry, N. Napawan, D. N. Butteiger, K. A. Hall, C. L. Null, Y. Luo, and B. T. Keller.** 2003. Inhibition of ileal bile acid transport and reduced atherosclerosis in apoE-/- mice by SC-435. J Lipid Res **44:**1614-21.

8. **Bibi, E., and H. R. Kaback.** 1990. In vivo expression of the lacY gene in two segments leads to functional lac permease. Proc Natl Acad Sci U S A **87:**4325-9.

9. **Boudker, O., R. M. Ryan, D. Yernool, K. Shimamoto, and E. Gouaux.** 2007. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. Nature **445:**387-93.

10. **Boudker, O., and G. Verdon.** 2010. Structural perspectives on secondary active transporters. Trends Pharmacol Sci **31:**418-26.

11. **Bowie, J. U.** 2006. Flip-flopping membrane proteins. Nat Struct Mol Biol **13:**94-6.

12. **Brett, C. L., M. Donowitz, and R. Rao.** 2005. Evolutionary origins of eukaryotic sodium/proton exchangers. Am J Physiol Cell Physiol **288:**C223-39.

13. **Busch, W., and M. H. Saier, Jr.** 2004. The IUBMB-endorsed transporter classification system. Mol Biotechnol **27:**253-62.

14. **Cai, X., and J. Lytton.** 2004. The cation/Ca(2+) exchanger superfamily: phylogenetic analysis and structural implications. Mol Biol Evol **21:**1692-703.

15. **Chang, A. B., R. Lin, W. Keith Studley, C. V. Tran, and M. H. Saier, Jr.** 2004. Phylogeny as a guide to structure and function of membrane transport proteins. Mol Membr Biol **21:**171-81.

16. **Chen, M. F., and T. Y. Chen.** 2003. Side-chain charge effects and conductance determinants in the pore of ClC-0 chloride channels. J Gen Physiol **122:**133-45.

17. **Chen, Y. J., O. Pornillos, S. Lieu, C. Ma, A. P. Chen, and G. Chang.** 2007. X-ray structure of EmrE supports dual topology model. Proc Natl Acad Sci U S A **104:**18999-9004.

18. **Cuthbertson, J. M., D. A. Doyle, and M. S. Sansom.** 2005. Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng Des Sel **18:**295-308.

19. **Daley, D. O., M. Rapp, E. Granseth, K. Melen, D. Drew, and G. von Heijne.** 2005. Global topology analysis of the Escherichia coli inner membrane proteome. Science **308:**1321-3.

20. **Dang, S., L. Sun, Y. Huang, F. Lu, Y. Liu, H. Gong, J. Wang, and N. Yan.** 2010. Structure of a fucose transporter in an outward-open conformation. Nature **467:**734-8.

21. **Deguchi, Y., I. Yamato, and Y. Anraku.** 1989. Molecular cloning of gltS and gltP, which encode glutamate carriers of Escherichia coli B. J Bacteriol **171:**1314-9.

22. **Deguchi, Y., I. Yamato, and Y. Anraku.** 1990. Nucleotide sequence of gltS, the Na+/glutamate symport carrier gene of Escherichia coli B. J Biol Chem **265:**21704-8.

23. **Dobrowolski, A., F. Fusetti, and J. S. Lolkema.** 2010. Cross-linking of trans reentrant loops in the Na(+)-citrate transporter CitS of Klebsiella pneumoniae. Biochemistry **49:**4509-15.

24. **Dobrowolski, A., and J. S. Lolkema.** 2010. Evolution of antiparallel two-domain membrane proteins. Swapping domains in the glutamate transporter GltS. Biochemistry **49:**5972-4.

25. **Dobrowolski, A., and J. S. Lolkema.** 2009. Functional importance of GGXG sequence motifs in putative reentrant loops of 2HCT and ESS transport proteins. Biochemistry **48:**7448-56.

26. **Dobrowolski, A., I. Sobczak-Elbourne, and J. S. Lolkema.** 2007. Membrane topology prediction by hydropathy profile alignment: membrane topology of the Na(+)-glutamate transporter GltS. Biochemistry **46:**2326-32.

27. **Doyle, D. A., J. Morais Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon.** 1998. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science **280:**69-77.

28. **Dutzler, R., E. B. Campbell, M. Cadene, B. T. Chait, and R. MacKinnon.** 2002. X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. Nature **415:**287-94.

29. **Dutzler, R., E. B. Campbell, and R. MacKinnon.** 2003. Gating the selectivity filter in ClC chloride channels. Science **300:**108-12.

30. **Efremov, R. G., and L. A. Sazanov.** 2011. Structure of the membrane domain of respiratory complex I. Nature **476:**414-20.

31. **Engh, A. M., and M. Maduke.** 2005. Cysteine accessibility in ClC-0 supports conservation of the ClC intracellular vestibule. J Gen Physiol **125:**601-17.

32. **Estevez, R., and T. J. Jentsch.** 2002. CLC chloride channels: correlating structure with function. Curr Opin Struct Biol **12:**531-9.

33. **Estevez, R., B. C. Schroeder, A. Accardi, T. J. Jentsch, and M. Pusch.** 2003. Conservation of chloride channel structure revealed by an inhibitor binding site in ClC-1. Neuron **38:**47-59.

34. **Faham, S., A. Watanabe, G. M. Besserer, D. Cascio, A. Specht, B. A. Hirayama, E. M. Wright, and J. Abramson.** 2008. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. Science **321:**810-4.

35. **Fang, Y., H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, and C. Miller.** 2009. Structure of a prokaryotic virtual proton pump at 3.2 A resolution. Nature **460:**1040-3.

36. **Feng, L., E. B. Campbell, Y. Hsiung, and R. MacKinnon.** 2010. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. Science **330:**635-41.

37. **Fischer, G., U. Kosinska-Eriksson, C. Aponte-Santamaria, M. Palmgren, C. Geijer, K. Hedfalk, S. Hohmann, B. L. de Groot, R. Neutze, and K. Lindkvist-Petersson.** 2009. Crystal structure of a yeast aquaporin at 1.15 angstrom reveals a novel gating mechanism. PLoS Biol **7:**e1000130.

38. **Forrest, L. R., R. Kramer, and C. Ziegler.** 2011. The structural basis of secondary active transport mechanisms. Biochim Biophys Acta **1807:**167-88.

39. **Forrest, L. R., and G. Rudnick.** 2009. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. Physiology (Bethesda) **24:**377-86.

40. **Fu, D., A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud.** 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. Science **290:**481-6.

41. **Gaillard, I., D. J. Slotboom, J. Knol, J. S. Lolkema, and W. N. Konings.** 1996. Purification and reconstitution of the glutamate carrier GltT of the thermophilic bacterium Bacillus stearothermophilus. Biochemistry **35:**6150-6.

42. **Gao, X., L. Zhou, X. Jiao, F. Lu, C. Yan, X. Zeng, J. Wang, and Y. Shi.** 2010. Mechanism of substrate recognition and transport by an amino acid antiporter. Nature **463:**828-32.

43. **Gonen, T., P. Sliz, J. Kistler, Y. Cheng, and T. Walz.** 2004. Aquaporin-0 membrane junctions reveal the structure of a closed water pore. Nature **429:**193-7.

44. **Gruswitz, F., S. Chaudhary, J. D. Ho, A. Schlessinger, B. Pezeshki, C. M. Ho, A. Sali, C. M. Westhoff, and R. M. Stroud.** 2010. Function of human Rh based on structure of RhCG at 2.1 A. Proc Natl Acad Sci U S A **107:**9638-43.

45. **Harries, W. E., D. Akhavan, L. J. Miercke, S. Khademi, and R. M. Stroud.** 2004. The channel architecture of aquaporin 0 at a 2.2-A resolution. Proc Natl Acad Sci U S A **101:**14045-50.

46. **He, X., P. Szewczyk, A. Karyakin, M. Evin, W. X. Hong, Q. Zhang, and G. Chang.** 2010. Structure of a cation-bound multidrug and toxic compound extrusion transporter. Nature **467:**991-4.

47. **Heijne, G.** 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J **5:**3021-7.

48. **Heuberger, E. H., L. M. Veenhoff, R. H. Duurkens, R. H. Friesen, and B. Poolman.** 2002. Oligomeric state of membrane transport proteins analyzed with blue native electrophoresis and analytical ultracentrifugation. J Mol Biol **317:**591-600.

49. **Hirai, T., J. A. Heymann, D. Shi, R. Sarker, P. C. Maloney, and S. Subramaniam.** 2002. Three-dimensional structure of a bacterial oxalate transporter. Nat Struct Biol **9:**597-600.

50. **Hirai, T., and S. Subramaniam.** 2004. Structure and transport mechanism of the bacterial oxalate transporter OxlT. Biophys J **87:**3600-7.

51. **Hiroaki, Y., K. Tani, A. Kamegawa, N. Gyobu, K. Nishikawa, H. Suzuki, T. Walz, S. Sasaki, K. Mitsuoka, K. Kimura, A. Mizoguchi, and Y. Fujiyoshi.** 2006. Implications of the aquaporin-4 structure on array formation and cell adhesion. J Mol Biol **355:**628-39.

52. **Ho, J. D., R. Yeh, A. Sandstrom, I. Chorny, W. E. Harries, R. A. Robbins, L. J. Miercke, and R. M. Stroud.** 2009. Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. Proc Natl Acad Sci U S A **106:**7437-42.

53. **Horsefield, R., K. Norden, M. Fellert, A. Backmark, S. Tornroth-Horsefield, A. C. Terwisscha van Scheltinga, J. Kvassman, P. Kjellbom, U. Johanson, and R. Neutze.** 2008. High-resolution x-ray structure of human aquaporin 5. Proc Natl Acad Sci U S A **105:**13327-32.

54. **Hu, N. J., S. Iwata, A. D. Cameron, and D. Drew.** 2011. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. Nature **478:**408-11.

55. **Huang, Y., M. J. Lemieux, J. Song, M. Auer, and D. N. Wang.** 2003. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. Science **301:**616-20.

56. **Hunte, C., E. Screpanti, M. Venturi, A. Rimon, E. Padan, and H. Michel.** 2005. Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature **435:**1197-202.

57. **Jardetzky, O.** 1966. Simple allosteric model for membrane pumps. Nature **211:**969-70.

58. **Jayaram, H., J. L. Robertson, F. Wu, C. Williams, and C. Miller.** 2011. Structure of a slow CLC Cl/H+ antiporter from a cyanobacterium. Biochemistry **50:**788-94.

59. **Jentsch, T. J.** 2008. CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. Crit Rev Biochem Mol Biol **43:**3-36.

60. **Jin, J., A. A. Guffanti, C. Beck, and T. A. Krulwich.** 2001. Twelve-transmembrane-segment (TMS) version (DeltaTMS VII-VIII) of the 14-TMS Tet(L) antibiotic resistance protein retains monovalent cation transport modes but lacks tetracycline efflux capacity. J Bacteriol **183:**2667-71.

61. **Kaback, H. R.** 1974. Transport in isolated bacterial membrane vesicles. Methods Enzymol **31:**698-709.

62. **Kalman, M., D. R. Gentry, and M. Cashel.** 1991. Characterization of the Escherichia coli K12 gltS glutamate permease gene. Mol Gen Genet **225:**379-86.

63. **Kastner, C. N., P. Dimroth, and K. M. Pos.** 2000. The Na+-dependent citrate carrier of Klebsiella pneumoniae: high-level expression and site-directed mutagenesis of asparagine-185 and glutamate-194. Arch Microbiol **174:**67-73.

64. **Kastner, C. N., M. Prummer, B. Sick, A. Renn, U. P. Wild, and P. Dimroth.** 2003. The citrate carrier CitS probed by single-molecule fluorescence spectroscopy. Biophys J **84:**1651-9.

65. **Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp.** 2005. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Res **33:**D334-7.

66. **Khademi, S., J. O'Connell, 3rd, J. Remis, Y. Robles-Colmenares, L. J. Miercke, and R. M. Stroud.** 2004. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. Science **305:**1587-94.

67. **Kikukawa, T., S. Miyauchi, T. Araiso, N. Kamo, and T. Nara.** 2007. Anti-parallel membrane topology of two components of EbrAB, a multidrug transporter. Biochem Biophys Res Commun **358:**1071-5.

68. **Kikukawa, T., T. Nara, T. Araiso, S. Miyauchi, and N. Kamo.** 2006. Two-component bacterial multidrug transporter, EbrAB: Mutations making each component solely functional. Biochim Biophys Acta **1758:**673-9.

69. **Klabunde, T., and G. Hessler.** 2002. Drug design strategies for targeting G-protein-coupled receptors. Chembiochem **3:**928-44.

70. **Klingenberg, M.** 2008. The ADP and ATP transport in mitochondria and its carrier. Biochim Biophys Acta **1778:**1978-2021.

71. **Klingenberg, M.** 1981. Membrane protein oligomeric structure and transport function. Nature **290:**449-54.

72. **Kolbusz, M. A., R. ter Horst, D. J. Slotboom, and J. S. Lolkema.** 2010. Orientation of small multidrug resistance transporter subunits in the membrane: correlation with the positive-inside rule. J Mol Biol **402:**127-38.

73. **Konings, W. N., E. M. Barnes, Jr., and H. R. Kaback.** 1971. Mechanisms of active transport in isolated membrane vesicles. 2. The coupling of reduced phenazine methosulfate to the concentrative uptake of beta-galactosides and amino acids. J Biol Chem **246:**5857-61.

74. **Krishnamurthy, H., C. L. Piscitelli, and E. Gouaux.** 2009. Unlocking the molecular secrets of sodium-coupled transporters. Nature **459:**347-55.

75. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.

76. **Krom, B. P., and J. S. Lolkema.** 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:**467-74.

77. **Krupnik, T., A. Dobrowolski, and J. S. Lolkema.** 2011. Cross-linking of dimeric CitS and GltS transport proteins. Mol Membr Biol **28:**243-53.

78. **Kuroda, T., and T. Tsuchiya.** 2009. Multidrug efflux transporters in the MATE family. Biochim Biophys Acta **1794:**763-8.

79. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol **157:**105-32.

80. **Law, C. J., P. C. Maloney, and D. N. Wang.** 2008. Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol **62:**289-305.

81. **Lee, J. K., D. Kozono, J. Remis, Y. Kitagawa, P. Agre, and R. M. Stroud.** 2005. Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A. Proc Natl Acad Sci U S A **102:**18932-7.

82. **Levin, E. J., M. Quick, and M. Zhou.** 2009. Crystal structure of a bacterial homologue of the kidney urea transporter. Nature **462:**757-61.

83. **Li, X., S. Jayachandran, H. H. Nguyen, and M. K. Chan.** 2007. Structure of the Nitrosomonas europaea Rh protein. Proc Natl Acad Sci U S A **104:**19279-84.

84. **Lolkema, J. S.** 2006. Domain structure and pore loops in the 2-hydroxycarboxylate transporter family. J Mol Microbiol Biotechnol **11:**318-25.

85. **Lolkema, J. S., A. Dobrowolski, and D. J. Slotboom.** 2008. Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. J Mol Biol **378:**596-606.

86. **Lolkema, J. S., B. Poolman, and W. N. Konings.** 1996. Chapter 11 Secondary transporters and metabolic energy generation in bacteria, p. 229-260, Handbook of Biological Physics, vol. Volume 2. North-Holland.

87. **Lolkema, J. S., and D. J. Slotboom.** 2003. Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis. J Mol Biol **327:**901-9.

88. **Lolkema, J. S., and D. J. Slotboom.** 1998. Estimation of structural similarity of membrane proteins by hydropathy profile alignment. Mol Membr Biol **15:**33-42.

89. **Lolkema, J. S., and D. J. Slotboom.** 1998. Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins. FEMS Microbiol Rev **22:**305-22.

90. **Lolkema, J. S., and D. J. Slotboom.** 2008. The major amino acid transporter superfamily has a similar core structure as Na+-galactose and Na+-leucine transporters. Mol Membr Biol **25:**567-70.

91. **Lolkema, J. S., and D. J. Slotboom.** 2005. Sequence and hydropathy profile analysis of two classes of secondary transporters. Mol Membr Biol **22:**177-89.

92. **Lolkema, J. S., I. Sobczak, and D. J. Slotboom.** 2005. Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops. FEBS J **272:**2334-44.

93. **Long, F., C. C. Su, M. T. Zimmermann, S. E. Boyken, K. R. Rajashankar, R. L. Jernigan, and E. W. Yu.** 2010. Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. Nature **467:**484-8.

94. **Lu, F., S. Li, Y. Jiang, J. Jiang, H. Fan, G. Lu, D. Deng, S. Dang, X. Zhang, J. Wang, and N. Yan.** 2011. Structure and mechanism of the uracil transporter UraA. Nature **472:**243-6.

95. **Lu, M., and D. Fu.** 2007. Structure of the zinc transporter YiiP. Science **317:**1746-8.

96. **Lu, W., J. Du, T. Wacker, E. Gerbig-Smentek, S. L. Andrade, and O. Einsle.** 2011. pH-dependent gating in a FocA formate channel. Science **332:**352-4.

97. **Lupo, D., X. D. Li, A. Durand, T. Tomizaki, B. Cherif-Zahar, G. Matassi, M. Merrick, and F. K. Winkler.** 2007. The 1.3-A resolution structure of Nitrosomonas europaea Rh50 and mechanistic implications for NH3 transport by Rhesus family proteins. Proc Natl Acad Sci U S A **104:**19303-8.

98. **Maduke, M., C. Miller, and J. A. Mindell.** 2000. A decade of CLC chloride channels: structure, mechanism, and many unsettled questions. Annu Rev Biophys Biomol Struct **29:**411-38.

99. **Maduke, M., C. Williams, and C. Miller.** 1998. Formation of CLC-0 chloride channels from separated transmembrane and cytoplasmic domains. Biochemistry **37:**1315-21.

100. **Mansour, N. M., M. Sawhney, D. G. Tamang, C. Vogl, and M. H. Saier, Jr.** 2007. The bile/arsenite/riboflavin transporter (BART) superfamily. FEBS J **274:**612-29.

101. **Manting, E. H., and A. J. Driessen.** 2000. Escherichia coli translocase: the unravelling of a molecular machine. Mol Microbiol **37:**226-38.

102. **Meier, T., P. Polzer, K. Diederichs, W. Welte, and P. Dimroth.** 2005. Structure of the rotor ring of F-Type Na+-ATPase from Ilyobacter tartaricus. Science **308:**659-62.

103. **Miller, C.** 2006. ClC chloride channels viewed through a transporter lens. Nature **440:**484-9.

104. **Miller, C., and W. Nguitragool.** 2009. A provisional transport mechanism for a chloride channel-type Cl-/H+ exchanger. Philos Trans R Soc Lond B Biol Sci **364:**175-80.

105. **Minocha, R., K. Studley, and M. H. Saier, Jr.** 2003. The urea transporter (UT) family: bioinformatic analyses leading to structural, functional, and evolutionary predictions. Receptors Channels **9:**345-52.

106. **Mitchell, P.** 1957. A general theory of membrane transport from studies of bacteria. Nature **180:**134-6.

107. **Moriyama, Y., M. Hiasa, T. Matsumoto, and H. Omote.** 2008. Multidrug and toxic compound extrusion (MATE)-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics. Xenobiotica **38:**1107-18.

108. **Moscicka, K. B., T. Krupnik, E. J. Boekema, and J. S. Lolkema.** 2009. Projection structure by single-particle electron microscopy of secondary transport proteins GltT, CitS, and GltS. Biochemistry **48:**6618-23.

109. **Murakami, S., R. Nakashima, E. Yamashita, and A. Yamaguchi.** 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. Nature **419:**587-93.

110. **Murata, K., K. Mitsuoka, T. Hirai, T. Walz, P. Agre, J. B. Heymann, A. Engel, and Y. Fujiyoshi.** 2000. Structural determinants of water permeation through aquaporin-1. Nature **407:**599-605.

111. **Newby, Z. E., J. O'Connell, 3rd, Y. Robles-Colmenares, S. Khademi, L. J. Miercke, and R. M. Stroud.** 2008. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite Plasmodium falciparum. Nat Struct Mol Biol **15:**619-25.

112. **Newstead, S., D. Drew, A. D. Cameron, V. L. Postis, X. Xia, P. W. Fowler, J. C. Ingram, E. P. Carpenter, M. S. Sansom, M. J. McPherson, S. A. Baldwin, and S. Iwata.** 2011. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. EMBO J **30:**417-26.

113. **Nury, H., C. Dahout-Gonzalez, V. Trezeguet, G. Lauquin, G. Brandolin, and E. Pebay-Peyroula.** 2005. Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers. FEBS Lett **579:**6031-6.

114. **Oberai, A., Y. Ihm, S. Kim, and J. U. Bowie.** 2006. A limited universe of membrane protein families and folds. Protein Sci **15:**1723-34.

115. **Omote, H., M. Hiasa, T. Matsumoto, M. Otsuka, and Y. Moriyama.** 2006. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. Trends Pharmacol Sci **27:**587-93.

116. **Orlowski, J., and S. Grinstein.** 2004. Diversity of the mammalian sodium/proton exchanger SLC9 gene family. Pflugers Arch **447:**549-65.

117. **Padan, E., T. Tzubery, K. Herz, L. Kozachkov, A. Rimon, and L. Galili.** 2004. NhaA of Escherichia coli, as a model of a pH-regulated Na+/H+antiporter. Biochim Biophys Acta **1658:**2-13.

118. **Padan, E., M. Venturi, Y. Gerchman, and N. Dover.** 2001. Na(+)/H(+) antiporters. Biochim Biophys Acta **1505:**144-57.

119. **Pao, S. S., I. T. Paulsen, and M. H. Saier, Jr.** 1998. Major facilitator superfamily. Microbiol Mol Biol Rev **62:**1-34.

120. **Park, J. H., and M. H. Saier, Jr.** 1996. Phylogenetic characterization of the MIP family of transmembrane channel proteins. J Membr Biol **153:**171-80.

121. **Patlak, C.** 1957. Contributions to the theory of active transport: II. The gate type non-carrier mechanism and generalizations concerning tracer flow, efficiency, and measurement of energy expenditure. Bulletin of Mathematical Biology **19:**209-235.

122. **Paulsen, I. T., M. K. Sliwinski, and M. H. Saier, Jr.** 1998. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J Mol Biol **277:**573-92.

123. **Pebay-Peyroula, E., C. Dahout-Gonzalez, R. Kahn, V. Trezeguet, G. J. Lauquin, and G. Brandolin.** 2003. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature **426:**39-44.

124. **Poolman, B., E. R. Geertsma, and D. J. Slotboom.** 2007. Biochemistry. A missing link in membrane protein evolution. Science **315:**1229-31.

125. **Pornillos, O., and G. Chang.** 2006. Inverted repeat domains in membrane proteins. FEBS Lett **580:**358-62.

126. **Pos, K. M., M. Bott, and P. Dimroth.** 1994. Purification of two active fusion proteins of the Na(+)-dependent citrate carrier of Klebsiella pneumoniae. FEBS Lett **347:**37-41.

127. **Prakash, S., G. Cooper, S. Singhi, and M. H. Saier, Jr.** 2003. The ion transporter superfamily. Biochim Biophys Acta **1618:**79-92.

128. **Rapp, M., E. Granseth, S. Seppala, and G. von Heijne.** 2006. Identification and evolution of dual-topology membrane proteins. Nat Struct Mol Biol **13:**112-6.

129. **Rapp, M., S. Seppala, E. Granseth, and G. von Heijne.** 2007. Emulating membrane protein evolution by rational design. Science **315:**1282-4.

130. **Ressl, S., A. C. Terwisscha van Scheltinga, C. Vonrhein, V. Ott, and C. Ziegler.** 2009. Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. Nature **458:**47-52.

131. **Reyes, N., C. Ginter, and O. Boudker.** 2009. Transport mechanism of a bacterial homologue of glutamate transporters. Nature **462:**880-5.

132. **Rossmann, R., G. Sawers, and A. Bock.** 1991. Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. Mol Microbiol **5:**2807-14.

133. **Saaf, A., L. Baars, and G. von Heijne.** 2001. The internal repeats in the Na+/Ca2+ exchanger-related Escherichia coli protein YrbG have opposite membrane topologies. J Biol Chem **276:**18905-7.

134. **Saier, M. H., Jr.** 2000. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev **64:**354-411.

135. **Saier, M. H., Jr.** 2003. Tracing pathways of transport protein evolution. Mol Microbiol **48:**1145-56.

136. **Saier, M. H., Jr., J. T. Beatty, A. Goffeau, K. T. Harley, W. H. Heijne, S. C. Huang, D. L. Jack, P. S. Jahn, K. Lew, J. Liu, S. S. Pao, I. T. Paulsen, T. T. Tseng, and P. S. Virk.** 1999. The major facilitator superfamily. J Mol Microbiol Biotechnol **1:**257-79.

137. **Saier, M. H., Jr., B. H. Eng, S. Fard, J. Garg, D. A. Haggerty, W. J. Hutchinson, D. L. Jack, E. C. Lai, H. J. Liu, D. P. Nusinew, A. M. Omar, S. S. Pao, I. T. Paulsen, J. A. Quan, M. Sliwinski, T. T. Tseng, S. Wachi, and G. B. Young.** 1999. Phylogenetic characterization of novel transport protein families revealed by genome analyses. Biochim Biophys Acta **1422:**1-56.

138. **Saier, M. H., Jr., C. V. Tran, and R. D. Barabote.** 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res **34:**D181-6.

139. **Saraste, M., and J. E. Walker.** 1982. Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. FEBS Lett **144:**250-4.

140. **Savage, D. F., P. F. Egea, Y. Robles-Colmenares, J. D. O'Connell, 3rd, and R. M. Stroud.** 2003. Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z. PLoS Biol **1:**E72.

141. **Schuldiner, S.** 2009. EmrE, a model for studying evolution and mechanism of ion-coupled transporters. Biochim Biophys Acta **1794:**748-62.

142. **Schuldiner, S.** 2007. When biochemistry meets structural biology: the cautionary tale of EmrE. Trends Biochem Sci **32:**252-8.

143. **Schulze, S., S. Koster, U. Geldmacher, A. C. Terwisscha van Scheltinga, and W. Kuhlbrandt.** 2010. Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT. Nature **467:**233-6.

144. **Screpanti, E., and C. Hunte.** 2007. Discontinuous membrane helices in transport proteins and their correlation with function. J Struct Biol **159:**261-7.

145. **Seeger, M. A., A. Schiefner, T. Eicher, F. Verrey, K. Diederichs, and K. M. Pos.** 2006. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. Science **313:**1295-8.

146. **Sennhauser, G., M. A. Bukowska, C. Briand, and M. G. Grutter.** 2009. Crystal structure of the multidrug exporter MexB from Pseudomonas aeruginosa. J Mol Biol **389:**134-45.

147. **Shaffer, P. L., A. Goehring, A. Shankaranarayanan, and E. Gouaux.** 2009. Structure and mechanism of a Na+-independent amino acid transporter. Science **325:**1010-4.

148. **Slotboom, D. J., I. Sobczak, W. N. Konings, and J. S. Lolkema.** 1999. A conserved serine-rich stretch in the glutamate transporter family forms a substrate-sensitive reentrant loop. Proc Natl Acad Sci U S A **96:**14282-7.

149. **Sobczak, I., and J. S. Lolkema.** 2005. The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism. Microbiol Mol Biol Rev **69:**665-95.

150. **Sobczak, I., and J. S. Lolkema.** 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:**9789-96.

151. **Sobczak, I., and J. S. Lolkema.** 2004. Alternating access and a pore-loop structure in the Na+-citrate transporter CitS of Klebsiella pneumoniae. J Biol Chem **279:**31113-20.

152. **Soskine, M., S. Mark, N. Tayer, R. Mizrachi, and S. Schuldiner.** 2006. On parallel and antiparallel topology of a homodimeric multidrug transporter. J Biol Chem **281:**36205-12.

153. **Steiner-Mordoch, S., M. Soskine, D. Solomon, D. Rotem, A. Gold, M. Yechieli, Y. Adam, and S. Schuldiner.** 2008. Parallel topology of genetically fused EmrE homodimers. EMBO J **27:**17-26.

154. **Steinmeyer, K., B. Schwappach, M. Bens, A. Vandewalle, and T. J. Jentsch.** 1995. Cloning and functional expression of rat CLC-5, a chloride channel related to kidney disease. J Biol Chem **270:**31172-7.

155. **Stevens, T. J., and I. T. Arkin.** 2000. Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins **39:**417-20.

156. **Sui, H., B. G. Han, J. K. Lee, P. Walian, and B. K. Jap.** 2001. Structural basis of water-specific transport through the AQP1 water channel. Nature **414:**872-8.

157. **Suppmann, B., and G. Sawers.** 1994. Isolation and characterization of hypophosphite--resistant mutants of Escherichia coli: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. Mol Microbiol **11:**965-82.

158. **Tang, L., L. Bai, W. H. Wang, and T. Jiang.** 2010. Crystal structure of the carnitine transporter and insights into the antiport mechanism. Nat Struct Mol Biol **17:**492-6.

159. **Ter Horst, R., and J. S. Lolkema.** 2012. Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity. Biochim Biophys Acta **1818:**72-81.

160. **Ter Horst, R., and J. S. Lolkema.** 2009. Rapid screening of membrane topology of secondary transport proteins. Biochim Biophys Acta **1798:**672-80.

161. **Theobald, D. L., and C. Miller.** 2010. Membrane transport proteins: surprises in structural sameness. Nat Struct Mol Biol **17:**2-3.

162. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 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:**4673-80.

163. **Tolner, B., T. Ubbink-Kok, B. Poolman, and W. N. Konings.** 1995. Cation-selectivity of the L-glutamate transporters of Escherichia coli, Bacillus stearothermophilus and Bacillus caldotenax: dependence on the environment in which the proteins are expressed. Mol Microbiol **18:**123-33.

164. **Tornroth-Horsefield, S., Y. Wang, K. Hedfalk, U. Johanson, M. Karlsson, E. Tajkhorshid, R. Neutze, and P. Kjellbom.** 2006. Structural mechanism of plant aquaporin gating. Nature **439:**688-94.

165. **Tsukazaki, T., H. Mori, S. Fukai, R. Ishitani, T. Mori, N. Dohmae, A. Perederina, Y. Sugita, D. G. Vassylyev, K. Ito, and O. Nureki.** 2008. Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature **455:**988-91.

166. **Ubarretxena-Belandia, I., J. M. Baldwin, S. Schuldiner, and C. G. Tate.** 2003. Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. EMBO J **22:**6175-81.

167. **Van den Berg, B., W. M. Clemons, Jr., I. Collinson, Y. Modis, E. Hartmann, S. C. Harrison, and T. A. Rapoport.** 2004. X-ray structure of a protein-conducting channel. Nature **427:**36-44.

168. **van Geest, M., and J. S. Lolkema.** 2000. Membrane topology of the Na(+)/citrate transporter CitS of Klebsiella pneumoniae by insertion mutagenesis. Biochim Biophys Acta **1466:**328-38.

169. **van Geest, M., and J. S. Lolkema.** 1999. Transmembrane segment (TMS) VIII of the Na(+)/Citrate transporter CitS requires downstream TMS IX for insertion in the Escherichia coli membrane. J Biol Chem **274:**29705-11.

170. **Vinothkumar, K. R., and R. Henderson.** 2010. Structures of membrane proteins. Q Rev Biophys **43:**65-158.

171. **Waight, A. B., J. Love, and D. N. Wang.** 2010. Structure and mechanism of a pentameric formate channel. Nat Struct Mol Biol **17:**31-7.

172. **Wang, Y., Y. Huang, J. Wang, C. Cheng, W. Huang, P. Lu, Y. N. Xu, P. Wang, N. Yan, and Y. Shi.** 2009. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. Nature **462:**467-72.

173. **Weinert, S., S. Jabs, C. Supanchart, M. Schweizer, N. Gimber, M. Richter, J. Rademann, T. Stauber, U. Kornak, and T. J. Jentsch.** 2010. Lysosomal pathology and osteopetrosis upon loss of H+-driven lysosomal Cl- accumulation. Science **328:**1401-3.

174. **Weinman, S. A., M. W. Carruth, and P. A. Dawson.** 1998. Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. J Biol Chem **273:**34691-5.

175. **West, I. C., and P. Mitchell.** 1974. Proton/sodium ion antiport in Escherichia coli. Biochem J **144:**87-90.

176. **Weyand, S., T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E. P. Carpenter, N. G. Rutherford, J. M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S. G. Patching, R. J. Hope, H. T. Norbertczak, P. C. Roach, S. Iwata, P. J. Henderson, and A. D. Cameron.** 2008. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. Science **322:**709-13.

177. **Wood, N. J., T. Alizadeh, D. J. Richardson, S. J. Ferguson, and J. W. Moir.** 2002. Two domains of a dual-function NarK protein are required for nitrate uptake, the first step of denitrification in Paracoccus pantotrophus. Mol Microbiol **44:**157-70.

178. **Wu, J., and H. R. Kaback.** 1996. A general method for determining helix packing in membrane proteins in situ: helices I and II are close to helix VII in the lactose permease of Escherichia coli. Proc Natl Acad Sci U S A **93:**14498-502.

179. **Yamashita, A., S. K. Singh, T. Kawate, Y. Jin, and E. Gouaux.** 2005. Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature **437:**215-23.

180. **Yernool, D., O. Boudker, Y. Jin, and E. Gouaux.** 2004. Structure of a glutamate transporter homologue from Pyrococcus horikoshii. Nature **431:**811-8.

181. **Yin, Y., X. He, P. Szewczyk, T. Nguyen, and G. Chang.** 2006. Structure of the multidrug transporter EmrD from Escherichia coli. Science **312:**741-4.

182. **Zheng, L., D. Kostrewa, S. Berneche, F. K. Winkler, and X. D. Li.** 2004. The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli. Proc Natl Acad Sci U S A **101:**17090-5.

183. **Zifarelli, G., and M. Pusch.** 2007. CLC chloride channels and transporters: a biophysical and physiological perspective. Rev Physiol Biochem Pharmacol **158:**23-76.

184. **Zimmer, J., Y. Nam, and T. A. Rapoport.** 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature **455:**936-43.

List of publications

Dobrowolski, A., Sobczak-Elbourne, I. and Lolkema, J. (2007) Membrane topology prediction by hydropathy profile alignment: membrane topolgy of the Na^+-glutamate transporter GltS. *Biochemistry* 46(9):2326-32

Lolkema, J., **Dobrowolski, A**. and Slotboom, DJ. (2008) Evolution of antiparallel two-domain membrane proteins: Tracing multiple gene duplication events in the DUF606 family. *J. Mol. Biol.* 378(3):596–606

Dobrowolski, A., and Lolkema, J. (2009) Functional importance of GGXG sequence motifs in putative reentrant loops of 2HCT and ESS transport proteins. *Biochemistry* 48(31):7448-56

Dobrowolski, A., Fusetti, F. and Lolkema, J. (2010) Cross-linking of *trans* reentrant loops in the Na^+-citrate transporter CitS of *Klebsiella pneumoniae*. *Biochemistry* 49(21):4509-15

Dobrowolski, A., and Lolkema, J. (2010) Evolution of antiparallel two-domain membrane proteins. Swapping domains in the glutamate transporter GltS. *Biochemistry* 49(29): 5972-74

Krupnik, T., **Dobrowolski, A.** and Lolkema JS. (2011) Cross-linking of dimeric CitS and GltS transport proteins. *Mol Membr Biol.* 28(5):243-53

Acknowledgements

Finally, after a long, long… long time, I did it. I finished my thesis.
Now is the time to acknowledge some people.
First of all I would like to thank Juke, my supervisor and copromoter. Juke, thank you for proposing me a PhD position in your lab, help and advice during the project and finally thank you for a great patience and support during the writing of this thesis.
Thank you once more.
Next, I would like to thank Arnold my promoter. Arnold it was a pleasure and honor to be a part of your team.

I would like to thank all the MolMic members for their cooperation and unique atmosphere at work. Especially I would like to mention all my lab mates, Magda with whom I shared a bench, Ramon my paranimf, Tomek and Hein (thank you for translating my summary into Dutch).
It was a pleasure to work with all MolMiccers. I wish all of you good luck and lots of success with your future exploits in life.

Chciałbym również podziękować wszystkim tym którzy pozostali w Polsce. W pierwszej kolejności moim rodzicom (to Wam dedykowana jest ta rozprawa) i siostrze, oraz wszystkim przyjaciołom i znajomym którzy pomimo odległości pamiętali o mnie.

Na końcu szczególnie chciałbym podziękować Aleksandrze. Dziękuje za to, że byłaś ze mną tu w Groningen, dziękuję za to że wspierałaś mnie podczas pisania doktoratu już w Polsce, mam nadzieję że pozostaniemy razem gdziekolwiek byśmy się nie udali.

Dziekuję.

Adam