Biochimica et Biophysica Acta 1798 (2010) 672–680

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Rapid screening of membrane topology of secondary transport proteins

Ramon ter Horst, Juke S. Lolkema *

Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 7 October 2009
Received in revised form 6 November 2009
Accepted 13 November 2009
Available online 22 November 2009

Keywords:
Membrane topology
Structural class
Topology prediction
MemGen
Transport protein
Reporter fusion

ABSTRACT

Limited experimental data may be very useful to discriminate between membrane topology models of membrane proteins derived from different methods. A membrane topology screening method is proposed by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method involves amplification of the encoding genes or gene fragments by PCR, rapid cloning in dedicated vectors by ligation independent cloning, and determination of the cellular disposition of the three sites using conventional techniques. The N-terminus was determined by labeling with a fluorescent probe, the central position and the C-terminus by the reporter fusion technique using alkaline phosphatase (PhoA) and green fluorescence protein (GFP) as reporters. The method was evaluated using 16 transporter proteins of known function from four different structural classes. For 13 proteins a complete set of three localizations was obtained. The experimental data was used to discriminate between membrane topology models predicted by TMHMM, a widely used predictor using the amino acid sequence as input and by MemGen that uses hydropathy profile alignment and known 3D structures or existing models. It follows that in those cases where the models from the two methods were similar, the models were consistent with the experimental data. In those cases where the models differed, the MemGen model agreed with the experimental data. Three more recent predictors, MEMSAT3, OCTOPUS and TOPCONS showed a significantly higher consistency with the experimental data than observed with TMHMM.

1. Introduction

Secondary transporters catalyze the translocation of substrates across membranes driven by (electro) chemical gradients of substrates and co-ions. They are universal to all biological cells and involved in many biological processes. They are integral membrane proteins that in most cases are encoded by a single gene. In spite of their simple architecture, their phylogenetic diversity is enormous as evidenced by the approximately 100 gene (super)families in the Electrochemical Potential-driven transporters section of the Transport Classification system [1]. Most likely, the genetic diversity is at least in part a consequence of divergent evolution and, consequently, many of the different families may represent a similar fold and translocation mechanism. Support for this view is provided by the recently reported high resolution structures of the Na^+-leucine transporter LeuT [2], the Na^+-galactose transporter vSGLT [3], the Na^+-benzyl-hydantoin transporter Mhp1 [4], the Na^+-betain transporter BetP [5] and, very recently, the arginine/agmatine exchanger AdiC [6,7] that all show the same core structure. The five transporters are from different gene families and no significant sequence similarity could be identified between the transporters. Structural similarity between transporters from these families was predicted before by the MemGen method that is based on hydropathy profile analysis of amino acid sequences of membrane proteins [8,9]. MemGen groups families of transporters with the same global fold in structural classes and the five transporters LeuT, vSGLT, Mhp1, BetP and AdiC are all found in structural class ST[2] [10]. Classification of all transporter families, and, more in general, all membrane proteins, in structural classes by a computational technique like MemGen would be a major step forward in membrane protein research.

At a much lower level of resolution, a membrane topology model describing the number of transmembrane segments (TMS) and the orientation of the protein in the membrane is an essential first step in any study of the structure/function relationships of a membrane protein. Computational analysis of the primary structure is especially amenable for membrane topology prediction due to the back and forward folding of the polypeptide chain through the hydrophobic phospholipid bilayer, and many methods have been developed and are used extensively. It has been argued that the prediction methods may be significantly improved by constraining the prediction algorithm with limited experimental data [11]. The cellular location of the C-terminus of a significant part of the membrane proteome of *Escherichia coli* [12,13] and *Saccharomyces cerevisiae* [14] was determined experimentally by fusing the reporters alkaline

Abbreviations: LIC, ligation independent cloning; GFP, green fluorescent protein; FM, fluorescein-5-maleimide; TMS, transmembrane segment

* Corresponding author. Molecular Microbiology, Biological Center, Kerklaan 30, 9751NN, Haren, The Netherlands. Tel.: +31 50 3632155; fax: +31 50 3632154.
E-mail address: j.s.lolkema@rug.nl (J.S. Lolkema).

0005-2736/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2009.11.010

phosphatase (PhoA) and green fluorescent protein (GFP) to the proteins. Subsequently, this data was used to produce improved topology models by constraining the location of the C-terminus of the proteins. Using homology, the experimentally constrained topology models for the *E. coli* and *S. cerevisiae* proteome could be assigned to tens of thousands other membrane proteins in the public databases [14,15].

In a different approach, limited experimental data may also be used to discriminate between topology models produced by different methods. MemGen structural class ST[3] contains 33 families of bacterial secondary transporters for organic and inorganic anions and Na^+/H^+ antiporters [16]. MemGen is not a secondary structure prediction method, but since the fold of the proteins in the different families in one class is identical, membrane topology for the different families may be predicted by hydropathy profile alignment using a known model of one of the families [16,17]. Topology models for the different families in ST[3] were predicted on the basis of the experimentally determined topology of CitS of *Klebsiella pneumoniae*, a member of the 2HCT family in ST[3] [18,19]. For many, if not all families the MemGen models differ from the models produced by TMHMM [20], a widely used secondary structure predictor that produces topology models using the amino acid sequence as input. The global topology analysis of the *E. coli* proteome [12] revealed the C-terminal location of 19 proteins in structural class ST[3] that were distributed over seven different families. The MemGen models for the proteins showed a C-terminal location that was in agreement with the experimentally determined location in all cases, while TMHMM predicted the C-terminus correctly only in roughly half of the cases [21]. Remarkably, constraining the C-terminus of the proteins in the TMHMM prediction [11] resulted in consensus between the MemGen and TMHMM methods for only 3 out of 19 proteins.

It follows that the experimental localization of a single site in a protein sequence is not sufficient to discriminate between different topology models. Here we present and explore a method in which the localization of three sites in the sequence are determined experimentally, the N- and C-terminus and a site in the middle of the protein. Where a single site may just detect an erroneously predicted orientation of the protein in the membrane, multiple sites are more likely to detect missing or extra transmembrane segments in the model. Moreover, selection of a site in the middle of the sequence acknowledges that many membrane proteins are two-domain proteins. A missed transmembrane segment in the first domain might be compensated for by a similar error in the second domain which would not be identified by the location of the N- and C-termini. The method was tested on a set of 16 secondary transporters from four different structural classes defined by MemGen including class ST[3], thereby representing four different structures. Experimental results were used to evaluate the membrane topology models predicted by MemGen and TMHMM and three more recently presented predictors, TOPCONS, MEMSAT3 and OCTOPUS.

2. Materials and methods

2.1. Materials

Phusion DNA polymerase was obtained from Finnzymes (Espoo, Finland). T4 ligase was obtained from New England Biolabs (Frankfurt am Main, Germany). All other enzymes were obtained from Fermentas (Burlington, Canada). Mutagenic oligonucleotides were obtained from Biolegio (Nijmegen, The Netherlands), or from Operon (Ebersberg, Germany) for ligation independent cloning. *p*-Nitrophenyl phosphate (pNPP) was obtained from Sigma (Zwijndrecht, The Netherlands), fluorescein maleimide (FM) from Invitrogen (Carslbad, USA).

2.2. Bacterial strains and growth conditions

Escherichia coli strain SF100 (*recA Δlac ΔompT*) [22] harboring the indicated pLIC vector (see below) was routinely grown in Luria Broth medium at 37 °C, with ampicillin added at a final concentration of 50 μg/ml. Overnight cultures were diluted 30-fold in 3 ml of fresh medium and when the optical density measured at 660 nm (OD_{660}) reached a value between 0.6 and 0.8, arabinose was added at a final concentration of 0.002–0.05 % (w/v) to induce protein production from the plasmids. Following growth for another 1.5–2 h, cells were harvested by centrifugation in a table top centrifuge operated at 4 °C. Cells were resuspended in the indicated buffer and kept on ice until use.

2.3. Construction of LIC vectors and ligation independent cloning

All genetic manipulations were done using standard techniques. A *Kpn*I/*Xba*I fragment of pPHO7 [23] containing the *phoA* gene encoding alkaline phosphatase without its signal sequence was ligated downstream of the arabinose promoter in the commercial pBAD24 vector (Invitrogen) [24] using the same two restriction enzymes. The resulting vector was digested with *Nco*I and *Kpn*I and a synthetic double stranded piece of DNA (the LIC cassette) with *Nco*I and *Kpn*I compatible overhang at the 5′ and 3′ ends, respectively, was inserted. The double stranded DNA fragment was prepared by hybridizing two oligonucleotides. The nucleotide sequence of the sense strand was 5′-C *ATG* GGT *CAT CAT CAC CAT CAC CA***T TTA AA**T AGT GGT GTG GTA C- 3′ in which the initiation codon was underlined, the sequence coding for a His_6-tag set in italics and a *Swa*I restriction site in bold. The resulting vector, termed pLIC1 was used to construct alkaline phosphatase fusion proteins (see below). Vector pLIC2 which is used for GFP fusions was constructed by replacing the *phoA* gene in pLIC1 by the *gfp* gene encoding green fluorescence protein. A *Kpn*I/*Xba*I fragment of pLIC1 was replaced by a fragment taken from vector pBADcLIC_GFP [25] digested with *Kpn*I and *Xba*I. Similarly, the *phoA* fragment was replaced by a double stop codon resulting in vector pLIC3 which is used to produce His-tagged proteins. The second codon of the cassette in pLIC3, GGT (Gly) was mutated into TGT (Cys) by site directed mutagenesis rendering vector pLIC4 which is used for N-terminal localizations.

Ligation Independent Cloning (LIC) was done essentially as described [25]. The LIC vectors pLIC1-4 were linearized by *Swa*I digestion. Transporter genes were amplified using forward and backward primers containing 5′ flanking regions corresponding to the nucleotide sequences upstream and downstream of the *Swa*I site in the LIC cassette as follows: 5′-C*ATG*GGTCATCATCACCATCACCATTG……-3′ and 5′-TACCACACCACTATTTTG……-3′, respectively. For each individual gene one forward and two reverse primers were used to amplify full length and half length transporter genes (see also Fig. 1). Single-stranded overhangs of the PCR products and vectors were generated using T4 DNA polymerase in the presence of dGTP (vector) and dCTP (PCR product). The complementary overhangs of PCR product and vector annealed upon mixing, after which the resulting heteroduplexes were transformed to *E. coli*.

2.4. GFP and PhoA assays

2.4.1. PhoA assay

Cells from 0.5 ml of culture of *E. coli* SF100 cells harboring a pLIC1 vector carrying the indicated insert were washed once, and subsequently, resuspended in 1 M Tris–HCl pH 8.0 buffer after which the OD_{660} was measured. Following incubation for 5 min at 37 °C, 200 μl of a 1.5 mM *p*-nitrophenyl phosphate (pNPP) solution was added and the suspension was incubated until a yellow color developed. Then, the cells were spun down and the absorption was read at 420 and 550 nm using a Hitachi U-1100 spectrophotometer. The phosphatase activity was calculated in Miller units [26].

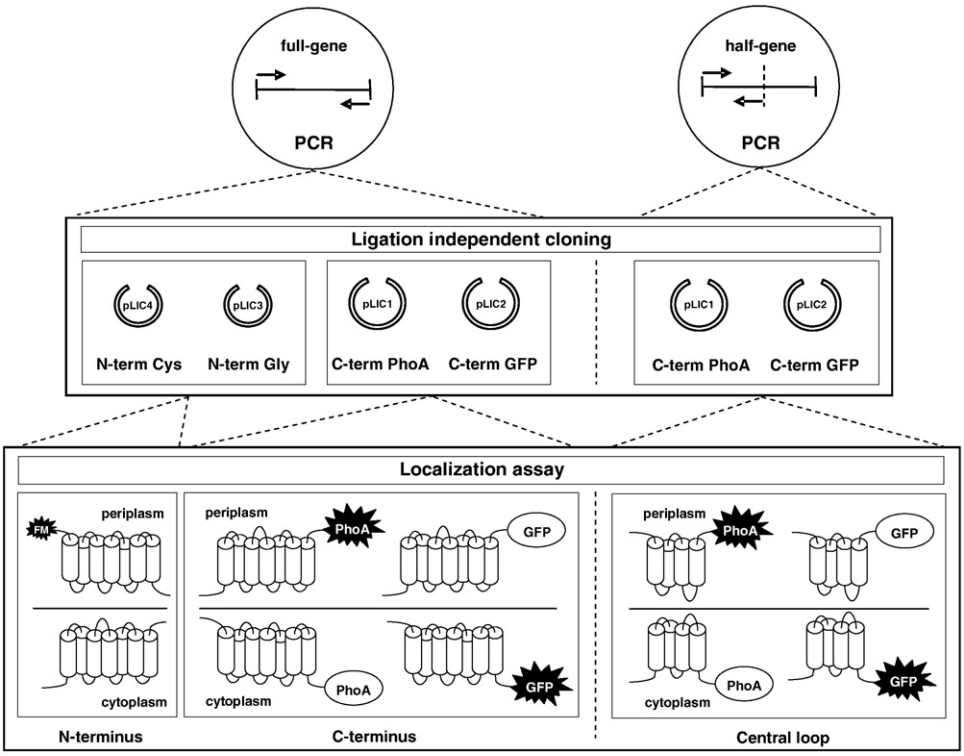

Fig. 1. Schematic representation of the membrane topology screening method. The method consists of three steps: (i) amplifying of the full-length gene and half-gene by PCR, (ii) cloning into a set of dedicated vectors by ligation independent cloning, and (iii) the localization assays. The dedicated vectors introduce a Cys (pLIC4) or Gly (pLIC3) residue at the N-terminus, or fuse PhoA (pLIC1) or GFP (pLIC2) at the C-terminus of the protein encoded by the insert. The N-terminal localization of the full-length protein is determined by the accessibility of the introduced cysteine residue at the external face of the membrane using FM, a fluorescent sulfhydryl reagent. The C-terminal localizations of the full-length and half-proteins are reported by PhoA (periplasmic) and GFP (cytoplasmic) activity. For further details, see the text.

2.4.2. GFP assay

Cells of *E. coli* SF100 harboring a pLIC2 vector carrying the indicated insert were washed once and, subsequently, resuspended in 50 mM Tris–HCl pH 8.0, 200 mM NaCl, and 15 mM EDTA to a final OD_{660} of 0.2. Following incubation for 30 min at room temperature, 150 µl cell suspension was transferred into a precision cell (Hellma, Quartz SUPRSIL) and GFP fluorescence emission intensity at 508 nm was measured at an excitation wavelength of 470 nm using an AMINCO Bowman Series 2 Luminescence Spectrometer. Background signals caused by light scattering were estimated from cells harboring the pLIC3 vector without insert [27].

2.4.3. Evaluation of data

Mean values and standard deviations were calculated from at least three independent measurements. PhoA and GFP activities of the cells were normalized by the mean PhoA activity of all positive PhoA fusions (>100 Miller units) and all positive GFP fusions (>0.5 emission units), respectively. The logarithm of the ratio of the normalized PhoA and GFP activities was calculated for each full-length and half protein to obtain a measure for the cellular localization of the fusion point. Values >2.5 or <-2.5 were arbitrarily set to 2.5 and -2.5, respectively.

2.5. Labeling studies

E. coli SF100 cells harboring pLIC4 or pLIC3 vectors carrying the indicated insert were washed once and, subsequently, resuspended in ice-cold 50 mM potassium phosphate buffer pH 7.0. Cells from 200 ml of culture were resuspended in 1 ml of buffer. A 5 mM solution of fluorescein 5-maleimide (FM) was freshly prepared by diluting a 50 mM solution in DMSO with 50 mM potassium phosphate buffer pH 7.0. FM was added to the cell suspension to a final concentration of 100 µM. After incubation for 20 min at room temperature, excess FM was quenched by adding 2 mM dithiothreitol (DTT). Cells were washed three times with 50 mM potassium phosphate buffer pH 7.0, and disrupted by sonication. Following removal of debris, membranes were collected from the supernatant by centrifugation for 25 min at 80,000 rpm in a Beckman TLA100.2 rotor at 4 °C. The pellet was solubilized in 1 ml 50 mM potassium phosphate buffer pH 8.0 containing 400 mM NaCl, 20% glycerol, 10 mM imidazole, 1 mM DTT and 2% Triton X-100 and left on ice for 30 min after which undissolved material was removed by centrifugation at 90,000 rpm for 25 min at 4 °C. The supernatant was mixed with Ni^{2+}-NTA resin (25 µl bed volume) equilibrated in 50 mM potassium phosphate buffer pH 8.0 containing 600 mM NaCl, 20% glycerol, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100, and incubated overnight at 4 °C under continuous shaking. The resin was pelleted by brief centrifugation in a table top centrifuge and the supernatant was removed. The resin was washed with 0.5 ml of equilibration buffer containing 300 mM NaCl and 20–40 mM imidazole. The protein was eluted with 25 µl of the same buffer, but at a pH of 7.0 and containing 200 mM imidazole. Subsequently, samples were loaded onto a 12% sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE). Following electrophoresis, in-gel fluorescence was recorded using a Fujifilm LAS-4000 luminescent image analyzer, and the gel was stained with Coomassie Brilliant Blue (CBB).

3. Results

3.1. Strategy

To discriminate between different topology models of membrane proteins, the cellular disposition of three positions in each of the polypeptide chains was determined by a combination of conventional techniques. The locations of the C-termini of full-length proteins and of half-proteins, truncated at a central loop, were determined using reporter gene fusion techniques. Green fluorescent protein (GFP) and

alkaline phosphatase (PhoA), devoid of its signal sequence, were used as positive reporters of a cytoplasmic and periplasmic localization, respectively. GFP folds into an active conformation when localized in the cytoplasm, but not in the periplasm [28]. In contrast, PhoA matures into the active state only when residing in the periplasm [29]. The third position, the N-terminus, was determined by the accessibility of an engineered cysteine residue using a membrane impermeable fluorescent sulfhydryl reagent.

The procedure as outlined in Fig. 1 was set up using rapid, ligation independent cloning (LIC) [25] and a set of dedicated cloning vectors (see Materials and methods) to allow for the screening of many proteins. A single PCR product encoding the full-length protein was cloned into a set of four cloning vectors resulting in the fusions with PhoA and GFP at the C-terminus of the protein and the introduction of a Cys and Gly residue at position 2 at the N-terminus. The construct with the Gly residue was used as a control in the labeling studies. The Cys and Gly residues were placed upstream of a His_6-tag that was present in all constructs to allow for maximum accessibility of the Cys residue. A second PCR product encoding the N-terminal half of the protein (half-protein) was cloned into the two vectors resulting in PhoA and GFP fusion proteins. Whenever possible, the protein was truncated in between the (putative) N-terminal and C-terminal domain.

Localization assays were done with resting cells using standard assays for alkaline phosphatase activity and fluorescence measurements for GFP content. Cells expressing the appropriate construct were treated with the fluorescent membrane-impermeable sulfhydryl reagent, fluorescein maleimide (FM) followed by processing of the cells to determine N-terminal location by the labeling efficiency.

3.2. Selection of transporter proteins and topology models

A set of 16 secondary transport proteins of known function were selected to evaluate the procedure outlined in Fig. 1. All transporters were from bacterial origin, originating from the phyla proteobacteria and firmicutes (Table 1). They were selected from 4 different structural classes in the MemGen classification and, therefore, putatively represent four different structures. The transporters from one structural class were mostly selected from different gene families (Transport Classification; TC system. [1]), thereby warranting high amino acid sequence diversity. In general, no significant sequence similarity can be detected between transporters from different families in one structural class. Secondary structure predictor TMHMM [20] predicts a topology model for each single sequence. In contrast, MemGen needs a template structure to model membrane topologies of other proteins in the same structural class by hydropathy profile alignment [8,9,10,16,17].

The three transporters selected from MemGen structural class ST [1], KgtP, LmrP, and MelB are from different gene families in the same superfamily: the Major Facilitator Superfamily (MFS) in the TC system [1]. The MemGen topology model of the ST[1] transporters is based on the high resolution structures of the LacY and GlpT transporters [30,31] and shows two (homologous) domains containing six TMSs each with both termini in the cytoplasm (Fig. 2). Two of the three exchangers selected from MemGen structural class ST[2] are from different families in the APC superfamily and a third (TyrP) from yet another family, HAAAP. High resolution structures showing the same fold of the core of the proteins are available for transporters from five unrelated families in ST[2] (see Introduction). Hydropathy profile alignment reveals topology models for the selected transporters with two (homologous) domains containing five TMSs each, that are oriented in opposite directions in the membrane (inverted topology). The transporters are predicted to have two or three additional TMSs at the C-terminus (Fig. 2). Seven transporters were selected from three different families in structural class ST[3]. The CitMHS family is found in the IT (Ion Transporters) superfamily. No high resolution structure is available for a transporter in class ST[3]. The topology model is based on extensive studies of the membrane topology of CitS of *K. pneumoniae* (2HCT family) and GltS of *E. coli* (ESS family) that both were included in the set as references [17–19]. The model reveals two (homologous) domains containing five TMSs each resulting in inverted topology. In between the fourth and fifth TMS in each domain the loop region folds back in between the TMSs (reentrant or pore-loop). The 2HCT family proteins have an additional TMS at the N-terminus. Three transporters were selected from ST[4]. ST[4] consists of a single family (DAACS). GltT shares 38% and 19% sequence identities with DctA and SstT, respectively, while the identity between the latter two is 19%. The high resolution structure of the Na^+-aspartate transporter Glt_{Ph} [32], was used to determine the membrane topology by MemGen. The models reveal eight TMS with two reentrant loops in the C-terminal half of the proteins.

3.3. C-terminal location of full-length and half-proteins

The alkaline phosphatase activity of cells expressing PhoA fusions of the full-length and half-proteins of each of the transporters in Table 1 was plotted against the GFP fluorescence of cells expressing the corresponding GFP fusion (Fig. 3A,B). The data points formed two

Table 1
Transporter proteins used in this study.

MemGen class	Transport protein	TC superfamily/family	Organism	Lineage[a]	Function	Half-protein[b]	Reference
ST[1]	KgtP	MFS/MHS	*Escherichia coli*	B-P-c	α-ketoglutarate:H^+ symporter	S222	42
	LmrP	MFS/DHA	*Lactococcus lactis*	B-F-l	multidrug:H^+ antiporter	P193	43
	MelB	MFS/GPH	*Escherichia coli*	B-P-c	melibiose:Na^+ symporter	S203	44
ST[2]	AguD	APC/APA	*Lactobacillus brevis*	B-F-l	agmatine/putrescine exchanger	S180/M216[c]	45
	HdcP	APC/GGA	*Lactobacillus hilgardi*	B-F-l	histidine/histamine exchanger	A195/A238[c]	46
	TyrP	HAAAP	*Lactobacillus brevis*	B-F-l	tyrosine/tyramine exchanger	L183/P222[c]	47
ST[3]	CimH	2HCT	*Bacillus subtilis*	B-F-b	*L*-malate/citrate:H^+ symporter	E261	48
	CitH	IT/CitMHS	*Bacillus subtilis*	B-F-b	Ca^{2+}-citrate:H^+ symporter	R202	49
	CitM	IT/CitMHS	*Bacillus subtilis*	B-F-b	Mg^{2+}-Citrate:H^+ symporter	E208	49
	CitP	2HCT	*Leuconostoc mesenteroides*	B-F-l	citrate/lactate exchanger	V269	50
	CitS	2HCT	*Klebsiella pneumonia*	B-P-c	citrate:Na^+ symporter	S251	50
	GltS	ESS	*Escherichia coli*	B-P-c	glutamate:Na^+ symporter	Q197	51
	MleP	2HCT	*Lactococcus lactis*	B-P-c	malate/lactate exchanger	E240	50
ST[4]	DctA	DAACS	*Escherichia coli*	B-P-c	C4-dicarboxylate transporter	S271	52
	GltT	DAACS	*Bacillus stearothermophilus*	B-F-b	glutamate:H^+ symporter	T272	52
	SSTT	DAACS	*Escherichia coli*	B-P-c	serine/threonine:Na^+ symporter	A269	52

[a] B-P-c: Bacteria, Proteobacteria, gamma subdivision; B-F-l: Bacteria, Firmicutes, lactobacillales; B-F-b: Bacteria, Firmicutes, bacillales.
[b] Last residue of wild-type sequence in N-terminal half-protein.
[c] Five TMS/six TMS half-proteins, respectively.

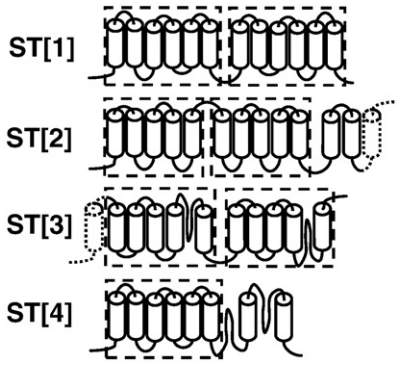

Fig. 2. Membrane topology models of structural classes ST[1] to [4] by MemGen. See text for explanation. Boxed parts represent domains. Transmembrane segments indicated by thin lines are not present in all of the proteins in one class listed in Table 1.

well separated groups falling along the two axes. The two groups correspond to a periplasmic and cytoplasmic localization of the fusion point. In general, there was a good correlation between the two reporters: proteins resulting in a high PhoA activity showed a low GFP activity and *vice versa*. The 'negative' reporter signal observed for alkaline phosphatase was clearly lower than observed for GFP, suggesting a significant background of fluorescence intensity in the GFP assay. The dashed lines tentatively limit the regions in the graph where the two groups of points are found. The results for two full-length transporters were inconclusive. Signals obtained for GltS of *E. coli* were insignificant (Fig. 3A, arrow a), suggesting lack of expression of the fusion proteins, and MleP of *Lactococcus lactis* (arrow b) showed a relatively small, but significant positive signal with both reporters. Both these transporter are from structural class ST[3].

The range of values of positive PhoA activities and GFP fluorescence intensities was considerable, and, mostly likely, represents the different levels of expression of the fusion proteins. Under the assumption that expression levels are mainly determined by the membrane protein and less so by the reporter protein, expression levels may be eliminated from the data by taking the ratio of the PhoA activity and the GFP fluorescence intensity after normalization of the two signals (see Section 2.4). Fig. 3C and D shows the logarithms of the normalized ratio of the PhoA and GFP signals for each of the full-length proteins (C) and half-proteins (D). Then, a positive bar corresponds to a periplasmic localization, a negative bar to a cytoplasmic localization. The length of the bars indicates the significance of the localization and would be independent of the expression levels. Threshold values for positive localizations were indicated by the dashed lines that correspond to the dashed lines in Fig. 3A and B.

The C-termini of the full-length transporters of structural classes ST[1] and ST[4] are all found in the cytoplasm, while ST[3] transporters have their C-termini in the periplasm. As mentioned above, the results for GltS and MleP in ST[3] were inconclusive. In ST[2], the AguD and TyrP proteins end in the cytoplasm, while HdcP ends in the periplasm. The C-termini of the half proteins of ST[1], ST[3] and ST[4] are all found in the cytoplasm, while the N-terminal half-proteins of the ST[2] transporters end in the periplasm.

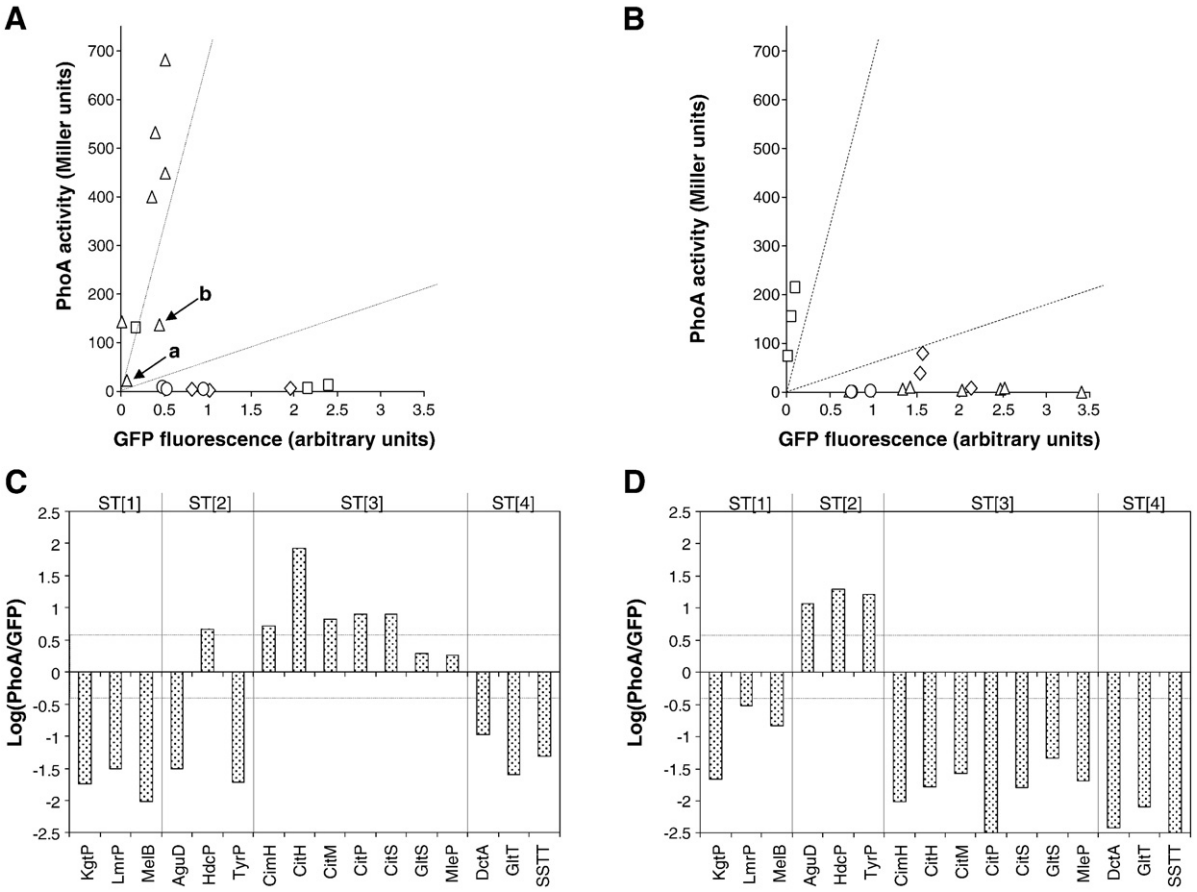

Fig. 3. C-terminal localization of full-length and half-proteins. (A, B) PhoA activity plotted against the GFP fluorescence intensities for full length proteins (A) and half-proteins (B) of the transporters listed in Table 1. ◊, ST[1] proteins, □, ST[2] proteins △, ST[3] proteins, ○, ST[4] proteins. Dotted lines indicate the limits for a periplasmic and cytoplasmic localization. (C, D) Logarithm of the ratio of the normalized PhoA and GFP signals for each of the full-length proteins (C) and half-proteins (D). Threshold values for positive localizations were indicated by the dotted lines which correspond to the dotted lines in panels A and B.

3.4. Domain structure of ST[2] transporters

TMHMM predicts 12 TMSs for most of the APC superfamily transporters in structural class ST[2], while MemGen models the same proteins using the available high resolution structures of transporters in ST[2] which results in two domains of five TMSs each plus two or three additional TMSs (see above). The half-proteins of the ST[2] transporters in the studies above were based on the latter prediction i.e. the half-proteins consisted of the predicted first domains (five TMSs). On the other hand, the TMHMM topology might suggest a two times six TMSs structure like the transporters of the MFS in class ST[1]. Since the choice of the truncation point may affect the stability of the fusion protein and even erroneously report the localization [33], the fusions of the half-proteins of the three ST[2] transporters were also constructed based on this assumption, i.e. consisting of the first six TMSs. The results are presented in Fig. 4. In case of the AguD and TyrP half-proteins, the C-terminus of the 6th TMS is in the cytoplasm which is consistent with the result from the five TMS half-proteins. However, the results for the HdcP transporter were found to be inconclusive, both a high PhoA and GFP signal were observed. It follows that the five TMS half-proteins give a more consistent result than the six TMS half-proteins.

3.5. N-terminal location of full-length proteins

For each of the transporters in Table 1 two constructs were made to determine the location of the N-terminus, one variant with a Cys and one with a Gly residue at position 2, immediately upstream of the His-tag. The cytoplasmic localization of the N-terminus of CitS of *K. pneumoniae* has been well documented [34–36]. Treatment of resting cells expressing the Gly variant of CitS with 0.1, 0.33 and 1.5 mM of fluorescence maleimide for 20 min showed significant labeling of the protein only at the highest FM concentration (Fig. 5A). Apparently, the five endogenous Cys residues of CitS were only poorly accessible for FM added to the extra-cellular medium. The same experiment with the Cys variant showed dramatically more labeling at the highest concentration of FM indicating that under these conditions the introduced Cys residue in the cytoplasmic N-terminus was accessible to the reagent. It also shows that successful labeling of a single Cys residue results in a much stronger signal than obtained with Coomassie Brilliant Blue staining of the protein. Labeling of the Cys variant was considerably less at a concentration of 0.33 mM FM and could only be marginally detected at a concentration of 0.1 mM. It follows that treatment of cells with 0.1 mM of FM for 20 min does not result in significant labeling of Cys residues located at the cytoplasmic face of the membrane. The N-terminus of GltS of *E. coli* is located in the periplasm [17]. Treatment of cells expressing the Gly variant with 0.1 mM FM for 20 min does not reveal any of the endogenous Cys residues in the fluorescence image of the gel (Fig. 5B). In contrast to what was observed with CitS, the Cys variant was clearly labeled after the same treatment, showing that the introduced Cys residue was readily accessible for the reagent, consistent with a periplasmic location. Treatment of cells expressing the two variants of the transporters with 0.1 mM FM for 20 min and at room temperature was used to discriminate between a cytoplasmic and periplasmic location of the N-terminus of the proteins.

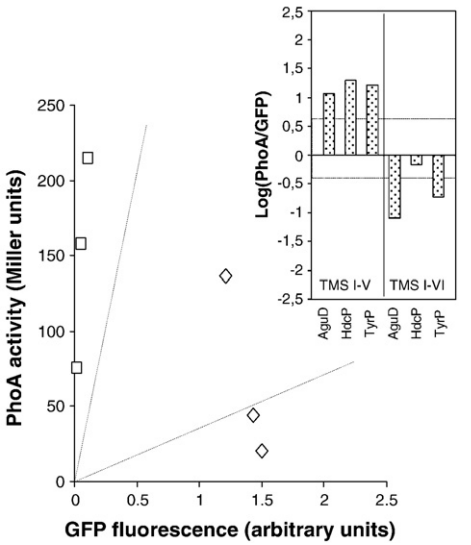

Fig. 4. C-terminal localization of half-proteins in structural class ST[2]. PhoA activity plotted against the GFP fluorescence intensities for ST[2] proteins truncated after TMS V (□) and TMS VI (◊). Dotted lines correspond to the corresponding lines in Fig. 3A, B. Insert. Logarithm of the ratio of the normalized PhoA and GFP signals for each of the fusion proteins.

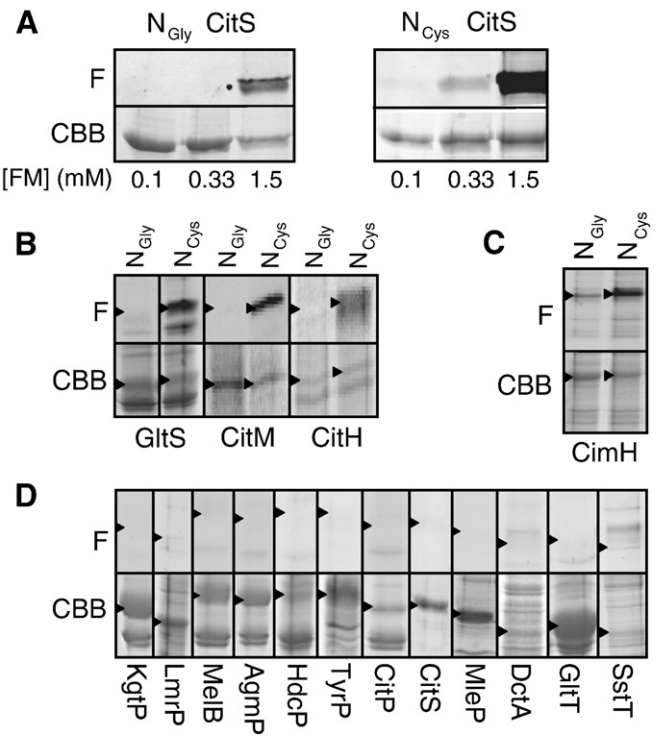

Fig. 5. N-terminal localization of full-length proteins. *E. coli* SF100 cells expressing the Gly or Cys variant of the full-length proteins as indicated were treated with the indicated concentration of FM (A) or 0.1 mM FM (B–D) for 20 min. Following partial purification, the proteins were analyzed by SDS-PAGE (see Materials and methods). F, fluorescence image of the part of the gel containing the protein; CBB, same part of the gel after staining with Coomassie Brilliant Blue. Arrow heads indicate the positions of the partially purified proteins.

The CitM and CitH transporters of *Bacillus subtilis* in structural class ST[3] gave the same results as observed for GltS. No labeling was observed with the Gly variant, while the Cys variant showed labeling (Fig. 5B). It follows that these transporters expose their N-termini at the exterior of the cell surface. The amount of CitH protein obtained after the partial purification was very small indicating low expression which is consistent with previous observations [37,38]. In addition, the more sensitive fluorescent signal shows that the CitH protein runs as a diffuse band on SDS-PAGE.

CimH of *B. subtilis* in structural class ST[3] was the only protein that showed labeling with FM in the Gly variant at the conditions used (Fig. 5C). CimH contains two endogenous Cys residues. With the Cys variant, fluorescence intensity increased somewhat, but the ratio of fluorescence intensity over protein stain was lower than for GltS, CitM and CitH above. A firm assignment of the N-terminal location of CimH could not be made. All other transporters showed no labeling in the Gly (not shown) or Cys variant (Fig. 5D) while the proteins could be

Table 2
Comparison[a] of experimentally determined (Exp) and predicted localizations[b] by MemGen, TMHMM and $TMHMM_{fix}$[c].

Class	Protein	N-terminus				Central loop				C-terminus			
		Exp	MemGen	TMHMM	$TMHMM_{fix}$	Exp	MemGen	TMHMM	$TMHMM_{fix}$	Exp	MemGen	TMHMM	$TMHMM_{fix}$[d]
ST[1]	KgtP	i	i	i	i	i	i	i	i	i	i	i	i
	LmrP	i	i	i	i	i	i	i	i	i	i	i	i
	MelB	i	i	i	i	i	i	i	i	i	i	i	i
ST[2]	AguD	i	i	i	i	o	o	o	o	i	i	i	i
	HdcP	i	i	i	i	o	o	o	o	o	o	o	o
	TyrP	i	i	i	i	o	o	o	o	i	i	i	i
ST[3]	CitH	o	o	i	i	i	i	i	i	o	o	o	o
	CitM	o	o	i	i	i	i	i	i	o	o	o	o
	CitP	i	i	i	i	i	i	o	o	o	o	o	o
	CitS	i	i	o	o	i	i	i	i	o	o	o	o
ST[4]	DctA	i	i	i	i	i	i	i	i	i	i	i	i
	GltT	i	i	o	o	i	i	i	i	i	i	i	i
	SSTT	i	i	i	i	i	i	i	i	i	i	i	i

[a] Grayed areas indicate predicted localizations at variance with the experimental data.
[b] i, cytoplasmic (in); o, external (out).
[c] TMHMM with constraint C-terminal localization.
[d] Copied from column C-terminus Exp.

identified by CBB staining. It was concluded that the N-termini of the proteins were situated in the cytoplasm.

Summarizing, the transporters of structural classes ST[1], ST[2] and ST[4] have cytoplasmic N-termini. For ST[3] the results are more diverse. The 2HCT transporters (CitS, CitP and MleP) have cytoplasmic N-termini as well, while the N-termini of the ESS transporter (GltS) and the CitMHS transporters (CitM and CitH) were found in the periplasm. No firm conclusion could be drawn for CimH, a 2HCT transporter in ST[3].

4. Discussion

4.1. Overall performance

Limited experimental data may be very useful to improve membrane topology predictions by computational methods [11,13] and to discriminate between different models derived from different methods. Here, a method is presented by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method involves rapid cloning followed by conventional localization assays. The procedure was set up to allow for rapid screening of many proteins in a funnel type of approach by which proteins that do not result in positive localizations at all three positions are dropped from the screen. In this way, when a number of proteins from one family are included, an improved topology model for the family will still be obtained, even if individual proteins drop out of the scheme. A total of 16 secondary transporter proteins taken from four different structural classes were used to test the method. A positive result was obtained for 15 of the N-termini, all of the central positions, and 14 of the C-termini. The localization of the N-terminus of CimH could not be obtained mainly because of background labeling of endogenous Cys residues. The localization of the C-terminus of GltS was not obtained, due to low expression of the fusion proteins and of MleP because the results were ambiguous. It follows that a full complement of three localizations was obtained for 13 out 16 proteins, giving a drop out rate of approximately 20%. Of course, incomplete sets of localizations may still be used to check for consistency with different models.

4.2. The central loop

The use of reporter fusions to determine the membrane topology of integral membrane proteins has been widely used and, also, widely criticized. The criticism mainly focuses on the use of truncated N-terminal fragments to determine the localization of internal positions in the sequence, as used here for the central loop and does not apply when the localization of the C-terminus is determined [12,13,14]). The assumption is that the truncate itself is stable and has the same topology as in the full-length protein, i.e. the topology is not dependent on missing C-terminal parts. While in many cases the assumption appears to be valid, in specific cases erroneous assignments have been reported [30]. The internal position in the present method was selected in the middle of the protein, thereby acknowledging that many transporter proteins consist of two (homologous) domains containing the same number of TMSs, and of approximately equal length. It is assumed that the N-terminal domain in the fusion protein is a stable entity by itself and therefore less prone to erroneous results. With no further experimental information the truncation site can only be chosen based upon the number of predicted TMSs. In case of MemGen, the site may be selected based on the domain structure found in the template structure. The difference between these two cases is nicely demonstrated for the three transporters AguD, HdcP and TyrP from structural class ST[2] (Fig. 4). The corresponding parts of the transporters contains 12 TMSs (HdcP has one additional TMS at the C-terminus). Therefore, the proteins would be truncated in the loop between TMS VI and VII. However, the model by MemGen was based on the high resolution structures of the LeuT and vSGLT proteins (see Introduction; [10]). (In the mean time, the structural similarity between the families has been confirmed by the structure of the AdiC transporter, a homologue of HdcP [6,7].) These transporters consist out of two times five TMSs plus two additional segments at the C-terminus (see Fig. 2) and the site of truncation was chosen between TMSs V and VI. The fusion proteins showed a more consistent result in the latter case, giving a high PhoA activity in combination with a low GFP signal. Truncation of the HdcP protein after TMS VI, which includes the first TMS of the second domain, resulted in both a high PhoA activity and GFP fluorescence, suggesting a scrambled C-terminus. Apparently, truncation between the two domains results in a more stable situation. In general, the fusions of the truncated proteins in this test behaved quite well, even better than the fusions of the full-length proteins. All 16 half-proteins resulted in well-defined localizations (Fig. 3), 13 in the cytoplasm and 3 at the external face of the membrane. The two full-length fusion proteins for which no localization was obtained both had their C-terminus predicted in the exterior.

Table 3
Performance of different predictors on the data set.

Predictor	ST[1]	ST[2]	ST[3]	ST[4]	All (%)
MemGen	3/3[a]	3/3	4/4	3/3	100
TMHMM	3/3	3/3	0/4	2/3	62
MEMSAT3	3/3	3/3	4/4	3/3	100
OCTOPUS	3/3	3/3	4/4	2/3	92
TOPCONS	3/3	3/3	2/4	3/3	85

[a] Number of consistent topology models/number of transporters. Only those transporters were included for which a location was obtained for all three sites in the proteins.

4.3. The N-terminus

Determining the localization of the N-termini of the proteins is the most elaborate and cumbersome part of the screen. Following treatment of the cells with the fluorescent probe FM, identification of labeling of the engineered Cys residue at the N-terminus requires partial purification of the membrane protein. For this purpose, a His_6-tag was engineered at the N-terminus which allows for a rapid, small scale partial purification of the proteins by affinity chromatography. Adverse affects of the presence of the His-tag at the N-terminus on the expression of the proteins may increase the drop out rate of proteins from the screen. An additional disadvantage of the labeling approach is that the absence of labeling is a negative result and it cannot be excluded that the introduced Cys residue is not accessible at the exterior of the cell rather than having a cytoplasmic location. To minimize this possibility, the Cys residue was introduced in front of the His_6-tag at the N-terminus. Then, the six histidine residues form a hydrophilic linker between the Cys residue and the hydrophobic membrane protein promoting its localization in the water phase. A further problem may arise from endogenous Cys residues that are accessible from the exterior of the cell as observed for the CimH protein in this test (Fig. 5C). In general, bacteria tend to not expose Cys residues to the external surface of the cell to prevent damage to the proteins. The problem will be minimized by analyzing the amino acid sequence of the proteins in a family for the presence and position of endogenous Cys residues before a selection of proteins to be included in the screen is made.

All 16 proteins in this test of 16 transporters were identified in the N_{cys} and N_{gly} variants by SDS-PAGE indicating no expression problems caused by the N-terminal His-tags. For 15 of the 16 transporters, no labeling of endogenous Cys residues was observed in the N_{gly} variant. Nevertheless, an attempt to identify labeling of the N_{cys} variant by SDS-PAGE of isolated cytoplasmic membranes, thereby skipping the purification step, failed because of too much background labeling.

4.4. Evaluation of predictors

The topology models produced by MemGen and TMHMM were evaluated for those proteins for which the complete set of three experimentally determined localizations was obtained (Table 2). The models for the proteins in structural classes ST[1] and ST[2] were similar for the two methods and agreed well with the experimental data. In contrast, for each of the proteins in class ST[3], one of the predicted localizations is different in the models produced by MemGen and TMHMM. Comparison to the experimental data shows that in all cases the prediction by MemGen is correct. Similarly, the experimental data is consistent with the MemGen model in case of the GltT protein in class ST[4] and not with the TMHMM model. Prediction for the other two proteins in class ST[4] was the same and in agreement with the experimental data. Summarizing, MemGen models agree with the experimental data in all cases (100%), while the TMHMM models agree in 8 out of 13 cases (62%) Constraining the C-terminal localization in the algorithm ($TMHMM_{fix}$; [11]) in none of the cases changed the prediction of the other two sites and, therefore, no improvement of the erroneously predicted models by TMHMM was obtained. It appears that the success rate of TMHMM is strongly dependent on the structural class with correct predictions in classes ST[1] and ST[2] and erroneous predictions in classes ST[3] and ST[4]. The low success rate for proteins in ST[3] is in agreement with observations made before (see Introduction; [21]). The transporters in class ST[3] are believed to contain two so-called pore-loops or reentrant loops [17,19,21]. Such loops are also prominent features of the transporters of class ST[4] [32] but are absent in proteins from ST [1] and ST[2]. TMHMM does not predict pore-loops and easily confuses them with transmembrane segments.

The TMHMM server was released a decade ago. During the last couple of years a number of new topology predictors for membrane proteins have been presented among which MEMSAT3, a neural network predictor that makes use of evolutionary information [39], OCTOPUS, a Hidden Markov model predictor that includes reentrant-, membrane dip- and transmembrane hairpin regions [40] and TOPCONS, a consensus predictor [41], These new methods score considerably better with our dataset giving 100%, 92% and 85% consistency between models and experimental data for MEMSAT3, OCTOPUS and TOPCONS, respectively (Table 3). Clearly, there has been considerable progress in the development of membrane topology prediction methods. It should be mentioned though, that models consistent with the experimental data are not necessarily the same models. In many cases, the models predicted by these predictors are still different from the MemGen models. The topology screen that is proposed here positively identifies models that are wrong. There may still be several models that are consistent with the three experimentally determined localizations, in which case more experiments are required to positively identify the correct model.

Acknowledgements

This work was supported by a grant from The Netherlands Organization for Scientific Research (NWO). The authors appreciate the help of Eric Geertsema of the Membrane Enzymology Group of the University of Groningen for his help with setting up the LIC system.

References

[1] M.H. Saier, A functional-phylogenetic classification system for transmembrane solute transporters, Microbiol. Mol. Rev. 64 (2000) 354–411.
[2] A. Yamashita, S.K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters, Nature 437 (2005) 215–223.
[3] S. Faham, A. Watanabe, G. Mercado Bessemer, D. Cascio, A. Szpecht, B.A. Hirayama, E.M. Wright, J. Abramson, The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport, Science 321 (2008) 810–814.
[4] S. Weyand, 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, A.D. Cameron, Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter, Science 322 (2008) 709–713.
[5] S. Ressl, A.C. Terwisscha van Scheltinga, C. Vonrhein, V. Ott, C. Ziegler, Molecular basis of transport and regulation in the Na+/betaine symporter BetP, Nature 458 (2009) 47–52.
[6] Y. Fang, H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, C. Miller, Structure of a prokaryotic virtual proton pump at 3.2 A resolution, Nature 460 (2009) 1040–1043.
[7] Xiang Gao, Lu Feiran, Lijun Zhou, Shangyu Dang, Linfeng Sun, Xiaochun Li, Jiawei Wang, Yigong Shi, Structure and mechanism of an amino acid antiporter, Science 324 (2009) 1565–1568.
[8] J.S. Lolkema, D.J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, Mol. Membr. Biol. 15 (1998) 33–42.
[9] J.S. Lolkema, D.J. Slotboom, Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins, FEMS Microbiol. Rev. 22 (1998) 305–322.
[10] J.S. Lolkema, D.J. Slotboom, The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-Leucine transporters, Mol. Membr. Biol. 25 (2008) 567–570.
[11] K. Melén, A. Krogh, G. von Heijne, Reliability measures for membrane protein topology prediction algorithms, J. Mol. Biol. 327 (2003) 735–744.

[12] D.O. Daley, M. Rapp, E. Granseth, K. Melén, D. Drew, G. von Heijne, Global topology analysis of the *Escherichia coli* inner membrane proteome, Science 308 (2005) 1321–1323.
[13] M. Rapp, D. Drew, D.O. Daley, J. Nilsson, T. Carvalho, K. Melén, J.W. De Gier, G. Von Heijne, Experimentally based topology models for *E. coli* inner membrane proteins, Protein Sci. 13 (2004) 937–945.
[14] H. Kim, K. Melén, M. Osterberg, G. von Heijne, A global topology map of the *Saccharomyces cerevisiae* membrane proteome, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 11142–11147.
[15] E. Granseth, D.O. Daley, M. Rapp, K. Melén, G. von Heijne, Experimentally constrained topology models for 51,208 bacterial inner membrane proteins, J. Mol. Biol. 352 (2005) 489–494.
[16] J.S. Lolkema, D.J. Slotboom, Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis, J. Mol. Biol. 327 (2003) 901–909.
[17] A.J. Dobrowolski, I. Sobczak, J.S. Lolkema, Experimental validation of membrane topology prediction by hydropathy profile alignment: membrane topology of the Na^+-glutamate transporter of *Escherichia coli*, Biochemistry 46 (2007) 2326–2332.
[18] I. Sobczak, J.S. Lolkema, The 2-hydroxycarboxylate transporter (2HCT) family: physiology, structure and mechanism, Microbiol. Mol. Biol. Rev. 69 (2005) 665–695.
[19] A. Dobrowolski, J.S. Lolkema, Functional importance of GGXG sequence motifs in putative reentrant loops of 2HCT and ESS transport proteins, Biochemistry 48 (2009) 7448–7456.
[20] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[21] J.S. Lolkema, Domain structure and pore-loops in the 2-hydroxycarboxylate transporter family, J. Mol. Microbiol. Biotechnol. 11 (2006) 318–325.
[22] F. Baneyx, G. Georgiou, In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT, J. Bacteriol. 172 (1990) 491–494.
[23] C. Gutierrez, J.C. Devedjian, A plasmid facilitating *in vitro* construction of phoA gene fusions in *Escherichia coli*, Nucleic Acids Res. 17 (1989) 3999.
[24] L.M. Guzman, D. Belin, M.J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter, J. Bacteriol. 177 (1995) 4121–4130.
[25] E.R. Geertsma, B. Poolman, High-throughput cloning and expression in recalcitrant bacteria, Nature Methods 4 (2007) 705–707.
[26] C. Manoil, Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions, Methods Cell Biol. 34 (1991) 61–75.
[27] D. Drew, D. Sjöstrand, J. Nilsson, T. Urbig, C.N. Chin, J.W. de Gier, G. von Heijne, Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 2690–2695.
[28] B.J. Feilmeier, G. Iseminger, D. Schroeder, H. Webber, G.J. Phillips, Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*, J. Bacteriol. 182 (14) (2000) 4068–4076.
[29] C. Manoil, J. Beckwith, A genetic approach to analyzing membrane protein topology, Science 233 (1986) 1403–1408.
[30] J. Abramson, I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*, Science 301 (2003) 610–615.
[31] Y. Huang, M.J. Lemieux, J. Song, M. Auer, D.N. Wang, Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*, Science 301 (2003) 616–620.
[32] D. Yernool, O. Boudker, Y. Jin, E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, Nature 431 (2004) 811–818.
[33] M. van Geest, J.S. Lolkema, Membrane topology and insertion of membrane proteins: search for topogenic signals, Microbiol. Mol. Biol. Rev. 64 (2000) 13–33.
[34] M. van Geest, J.S. Lolkema, Membrane topology of the sodium ion-dependent citrate carrier of *Klebsiella pneumoniae*. Evidence for a new structural class of secondary transporters, J. Biol. Chem. 271 (1996) 25582–25589.
[35] M. van Geest, J.S. Lolkema, Membrane topology of the Na(+)/citrate transporter CitS of *Klebsiella pneumoniae* by insertion mutagenesis, Biochim. Biophys. Acta 1466 (2000) 328–338.
[36] I. Sobczak, J.S. Lolkema, Accessibility of cysteine residues in a cytoplasmic loop of CitS of *Klebsiella pneumoniae* is controlled by the catalytic state of the transporter, Biochemistry 42 (2003) 9789–9796.
[37] A. Boorsma, M.E. van der Rest, J.S. Lolkema, W.N. Konings, Secondary transporters for citrate and the Mg(2+)-citrate complex in *Bacillus subtilis* are homologous proteins, J. Bacteriol. 178 (1996) 6216–6222.
[38] V.S. Blancato, C. Magni, J.S. Lolkema, Functional characterization and Me ion specificity of a Ca-citrate transporter from *Enterococcus faecalis*, FEBS J. 273 (2006) 5121–5130.
[39] D.T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information, Bioinformatics 23 (2007) 538–544.
[40] H. Viklund, A. Elofsson, Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar, Bioinformatics 24 (2008) 1662–1668.
[41] A. Bernsel, H. Viklund, A. Hennerdal, A. Elofsson, TOPCONS: consensus prediction of membrane protein topology, Nucleic Acids Res. 37 (Web Server issue) (2009) W465–W468.
[42] W. Seol, A.J. Shatkin, Membrane topology model of *Escherichia coli* alpha-ketoglutarate permease by phoA fusion analysis, J. Bacteriol. 175 (1993) 565–567.
[43] G.J. Poelarends, P. Mazurkiewicz, W.N. Konings, Multidrug transporters and antibiotic resistance in *Lactococcus lactis*, Biochim. Biophys. Acta 1555 (2002) 1–7.
[44] T. Pourcher, E. Bibi, H.R. Kaback, G. Leblanc, Membrane topology of the melibiose permease of *Escherichia coli* studied by melB-phoA fusion analysis, Biochemistry 35 (1996) 4161–4168.
[45] P.M. Lucas, V.S. Blancato, O. Claisse, C. Magni, J.S. Lolkema, A. Lonvaud-Funel, Agmatine deiminase pathway genes in *Lactobacillus brevis* are linked to the tyrosine decarboxylation operon in a putative acid resistance locus, Microbiology 153 (2007) 2221–2230.
[46] P.M. Lucas, W.A.M. Wolken, O. Claisse, J.S. Lolkema, A. Lonvaud-Funel, Histamine-producing pathway encoded on an unstable plasmid in *Lactobacillus hilgardii* 0006, Appl. Environ. Microbiol. 71 (2005) 1417–1424.
[47] W.A.M. Wolken, P.M. Lucas, A. Lonvaud-Funel, J.S. Lolkema, The mechanism of the tyrosine transporter TyrP supports a proton motive tyrosine decarboxylation pathway in *Lactobacillus brevis*, J. Bacteriol. 188 (2006) 2198–2206.
[48] B.P. Krom, R. Aardema, J.S. Lolkema, *Bacillus subtilis* YxkJ is a secondary transporter of the 2-hydroxycarboxylate transporter family that transports L-malate and citrate, J. Bacteriol. 183 (2001) 5862–5869.
[49] B.P. Krom, J.B. Warner, W.N. Konings, J.S. Lolkema, Complementary metal ion specificity of the metal-citrate transporters CitM and CitH of *Bacillus subtilis*, J. Bacteriol. 182 (2000) 6374–6381.
[50] I. Sobczak, J.S. Lolkema, The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism, Microbiol. Mol. Biol. Rev. 69 (2005) 665–695.
[51] A. Dobrowolski, I. Sobczak-Elbourne, J.S. Lolkema, Membrane topology prediction by hydropathy profile alignment: membrane topology of the Na(+)-glutamate transporter GltS, Biochemistry 46 (2007) 2326–2332.
[52] D.-J. Slotboom, W.N. Konings, J.S. Lolkema, Structural features of the glutamate transporter family, Microbiol. Mol. Biol. Rev. 63 (1999) 293–307.