Biochimica et Biophysica Acta 1818 (2012) 72–81

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

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

Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity

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 10 August 2011
Received in revised form 22 September 2011
Accepted 23 September 2011
Available online 29 September 2011

Keywords:
Membrane topology
Structural classification
ST[3]
Topology prediction
MemGen
Reporter fusion

ABSTRACT

The MemGen structural classification of membrane proteins groups families of proteins by hydropathy profile alignment. Class ST[3] of the MemGen classification contains 32 families of transporter proteins including the IT superfamily. Transporters from 19 different families in class ST[3] were evaluated by the TopScreen experimental topology screening method to verify the structural classification by MemGen. TopScreen involves the determination of the cellular disposition of three sites in the polypeptide chain of the proteins which allows for discrimination between different topology models. For nearly all transporters at least one of the predicted localizations is different in the models produced by MemGen and predictor TMHMM. Comparison to the experimental data showed that in all cases the prediction by MemGen was correct. It is concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other families in class ST[3]. The core structure of the model consists of two homologous domains, each containing 5 transmembrane segments, which have an opposite orientation in the membrane. A reentrant loop is present in between the 4th and 5th segments in each domain. Nearly all of the identified and experimentally confirmed structural variations involve additions of transmembrane segments at the boundaries of the core model, at the N- and C-termini or in between the two domains. Most remarkable is a domain swap in two subfamilies of the [st312] NHAC family that results in an inverted orientation of the proteins in the membrane.

1. Introduction

Secondary transporters are integral membrane proteins commonly encoded by a single gene. They consist of a bundle of α-helices that are more or less perpendicular to the plane of the membrane. Despite this 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 (TC system) [1]. Most likely, the genetic diversity is a consequence of divergent evolution and many different families may represent a similar fold and translocation mechanism. In recent years high resolution X-ray structures have been presented of transporters from different gene families that showed the same core structure while no significant sequence similarity could be identified between the transporters. The families include the Neurotransmitter: Sodium Symporter (2.A.22 NSS) family (structure of LeuT) [2], the Solute:Sodium Symporter (2.A.21 SSS) family (vSGLT) [3], the Nucleobase:Cation Symporter-1 (2.A.39 NCS1) family (Mhp1) [4], the Betaine/Carnitine/Choline Transporter (2.A.15 BCCT) family (BetP) [5] and the Amino Acid-Polyamine-Organocation (2.A.3 APC) family (AdiC, ApcT) [6,7,8]. Structural similarity between transporters of the NSS, SSS, NCS1 and APC families was predicted by the MemGen classification system that identifies distant evolutionary relationships by hydropathy profile alignment [9,10,11]. MemGen groups families of transporters with the same global fold in structural classes. So far, 4 structural classes were defined termed ST[1], ST[2], ST[3] and ST[4]. The NSS, SSS, NCS1 and APC families are found in class ST[2]. In addition to ST[2], high resolution structures are available for class ST[1] (LacY, GlpT) [12,13] and class ST[4] (Glt_{Ph}) [14]. Class ST[3] in the MemGen classification groups 32 protein families and includes the IT superfamily [15] in the TC system [1,16]. Functional characterized members of the 32 families are secondary transporters of inorganic and organic anions and Na^+/H^+ antiporters. Most families contain exclusively transporters from prokaryotic origin, but f. i. the Divalent Anion:Na^+ Symporter (2.A.47 DASS) family (also known as SLC13) contains eukaryotic transporters for organic di- and tricarboxylate Krebs cycle intermediates as well as dicarboxylate amino acids, and inorganic sulfate and phosphate ions. No high resolution structure is available of any of the transporters in class ST [3]. Detailed studies of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* and the Na^+-glutamate transporter GltS of *Escherichia coli*, members of the [st326]2HCT and [st324]ESS families in class ST [3], respectively, have resulted in a model for the common core

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

* Corresponding author at: Molecular Microbiology, Center for Life Sciences, University of Groningen, Nijenborgh 7, 9747AG Groningen, 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 © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2011.09.021

structure of the transporters. The core consists of two homologous domains with opposite orientation in the membrane and containing 5 TMSs and a re-entrant (or pore) loop each [17]. Members of the two families share no significant sequence similarity, but the data strongly suggested that CitS and GltS share the same fold which confirmed their assignment to the same structural class [18].

To verify the structural classification by MemGen, we have developed the TopScreen membrane topology screening method that experimentally discriminates between different topology models [19]. TopScreen involves determination of the cellular disposition of three positions of the polypeptide chain of a membrane protein using a combination of conventional techniques and was used before to confirm the distribution of a total of 16 secondary transporters over the 4 MemGen structural classes ST[1], ST[2], ST[3] and ST[4]. Here, the TopScreen approach was used to confirm the assignment of 19 transporter families to class ST[3] thereby strongly indicating that the [st326]2HCT/[st324]ESS structural core model applies to all the families in ST[3] including the families of the IT superfamily. Structural variations adopted by different (sub)families in class ST[3] are discussed.

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), and fluorescein maleimide (FM) was obtained from Invitrogen (Carlsbad, United States).

2.2. Bacterial strains and growth conditions

E. coli strain SF100 (*recA Δlac ΔompT*) [20] 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. Ligation independent cloning

Ligation independent cloning (LIC) was done as described [19,22]. Briefly, a synthetic double stranded piece of DNA (the LIC cassette) was inserted downstream of the arabinose promoter in the commercial pBAD24 vector (Invitrogen) [21]. The genes encoding alkaline phosphatase and green fluorescence protein (GFP) were inserted at the 3′-end of the LIC cassette in frame with the initiation codon yielding pLIC1 or pLIC2, respectively. The two vectors were used to produce PhoA and GFP fusion proteins, respectively. Vector pLIC3 which is used to produce His-tagged proteins contains a double stop codon inserted at the same site. 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. 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.

2.4. GFP and PhoA assays

Alkaline phosphatase activity of *E. coli* SF100 cells harboring a pLIC1 vector carrying the indicated insert was measured in Miller units [23] as described before [19]. GFP fluorescence emission intensity at 508 nm was measured at an excitation wavelength of 470 nm using an AMINCO Bowman Series 2 Luminescence Spectrometer. Cells of *E. coli* SF100 harboring a pLIC2 vector carrying the indicated insert were transferred into a precision cell (Hellma, Quartz SUPRSIL). For further details see [19,24].

2.5. 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. A positive value corresponds to a periplasmic localization, a negative value to a cytoplasmic localization. The length of the bar indicates the significance of the localization and would be independent of the expression levels [19]. Values >2.5 or <−2.5 were arbitrarily set to 2.5 and −2.5, respectively.

2.6. 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. Following treatment of the cells with fluorescein maleimide and, subsequently, Ni^{2+}-NTA affinity purification of the His-tagged transporters from the membranes, the 25 μl samples were run on a 12% sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE). In-gel fluorescence was recorded using a Fujifilm LAS-4000 luminescent image analyzer, and the gel was stained with Coomassie Brilliant Blue (CBB). Experimental details were as described before [19].

3. Results

3.1. TopScreen and selection of ST[3] transporters

The TopScreen approach involves the determination of the cellular disposition of three positions of a membrane protein, the N- and C-termini and a position in the central loop that is measured by the localization of the C-terminus of the half-protein [19]. The locations of the C-termini of full-length proteins and half-proteins are determined using complementary reporter fusion techniques. Alkaline phosphatase fused at the C-terminus results in an active enzyme when located in the periplasm, while fusion of GFP results in high fluorescence only when located in the cytoplasm. The N-termini are determined by the accessibility of an introduced cysteine residue by the fluorescent sulfhydryl reagent fluorescein maleimide.

Table 1 lists a total of 38 proteins from structural class ST[3] of the MemGen classification (http://molmic35.biol.rug.nl/memgen/mgweb.dll) that have been evaluated by the TopScreen method. Most of the proteins have been functionally characterized as secondary transporters. All of them are from bacterial origin, originating from the phyla proteobacteria and firmicutes. The proteins are distributed over 19 different families. In general no significant sequence similarities can be detected between transporters from different families. Transporters in the [st301]CITMHS, [st324]ESS and [st326]2HCT families were evaluated before [19] and were included for completeness throughout this study. Detailed topology models are available for the [st324]GLTS and [st326]2HCT families. The IT superfamily [15] from