Phylogenetic and conservation analyses of MFS transporters

Poonam Vishwakarma^{1,2} · Atanu Banerjee¹ · Ritu Pasrija² · Rajendra Prasad³ · Andrew M. Lynn¹

Received: 3 May 2018 / Accepted: 11 October 2018 / Published online: 27 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Major facilitator superfamily is one of the largest superfamily of secondary transporters present across the kingdom of life. Considering the physiological and clinical importance of MFS proteins, we attempted to explore the phylogenetic and structural aspects of the superfamily. To achieve the objectives, we performed global sequence-based analyses of MFS proteins encompassing multiple taxa. Notably, phylogenetic analysis of MFS proteins resulted in the clustering of MFS proteins based on their function, rather than lineage of the respective organisms. Additionally, we employed information theoretic measures, Relative entropy (RE) and Cumulative relative entropy (CRE) to decipher fold-specific and function-specific residues, respectively, in the MFS proteins. The residues with high RE score when mapped on to the 3D-structure of MFS transporter LacY, were found to be distributed throughout the tertiary structure of the protein. On the other hand, CRE calculation was employed to contrast two subfamilies Drug H⁺ antiporter 1 and 3 (DHA1 and DHA3). The particular analysis unveiled certain differentially conserved residues in DHA1 as compared to DHA3 highlighting family-specific importance of them. Remarkably, a number of high scoring CRE residues have already established functional roles, for instance, the arginine residue present in TMH4. Altogether, the current study apart from providing an insight into the functional clustering of MFS proteins also identifies residues with established or plausible roles in the transport mechanism. Thus, the study lays a platform for future structure–function studies of these proteins.

Keywords Major facilitator superfamily · Sequence conservation · Phylogeny · Relative entropy · Cumulative relative entropy

Introduction

A significant proportion of the genome of living organisms encode for proteins involved in transport processes across cell membranes (Shi 2013; Yan 2015). These transport proteins mediate movement of a wide variety of molecules ranging from ions and drugs to peptides (Yan 2015; Diallinas 2016). Broadly, they are necessary for maintenance

of cellular growth and homeostasis, cellular response, metabolism and defense from xenobiotics (Shi 2013; Diallinas 2016; Redhu et al. 2016). Among these proteins, there exist channels and transporters (Shi 2013; Yan 2015). While channels serve to move substrates down their electrochemical gradients, the transporters have the ability to extrude substrates against their electrochemical gradients (Shi 2013). Mechanistically, a transporter has to undergo a sequence of conformational changes to extrude its substrate out. In contrast, channels are not required to undergo any prominent conformational transitions during the transport process (Shi 2013). Thus, while the former process requires energy source (active process), the latter is a passive process and is not dependent on any energy source. Transporters are classified into primary active transporters, secondary active transporters and facilitators. While the primary active transporters depend on energy from other chemical processes, such as ATP-hydrolysis to catalyze the uphill movement of their substrates, the secondary active transporters rely on electrochemical potentials of a ion or solute which is

✉ Rajendra Prasad
rp47jnu@gmail.com

✉ Andrew M. Lynn
andrew@jnu.ac.in

¹ School of Computational and Integrative Sciences,
Jawaharlal Nehru University, New Delhi, India

² Department of Biochemistry, Maharshi Dayanand University,
Rohtak, Haryana, India

³ Amity Institute of Integrative Sciences and Health, Amity
Institute of Biotechnology, Amity University Haryana,
Gurgaon, India

co-transported. Moreover, the direction could either be same (symporter) or opposite (antiporter) (Lee et al. 2016). Contrarily, the facilitators or uniporters are transporters that mediate the facilitative diffusion of specific substrates along their concentration gradients (Shi 2013). Secondary active transporters and facilitators are often together referred to as secondary transporters (Yan 2015).

Among secondary transporters, major facilitator superfamily (MFS) represents the largest and most ubiquitously found group of transporter proteins (Saier et al. 2014; Finn et al. 2016). They are known for their wide array of substrates which range from nucleosides, ions, peptides, sugars to neurotransmitters and drugs (Kasho et al. 2006; Gaur et al. 2008; Lee et al. 2016). Currently, the Transport Classification database (TCDB) classifies the superfamily into 87 families (<http://www.tcdb.org/superfamily.php?id=3>) (Table 1). However, apart from the primary MFS group (TC# 2.A.1) which includes the 87 families, there exist 9 other families (Table 2) which are distant members of the MFS.

Irrespective of the differences in the primary protein sequence, all MFS proteins adopt a common protein-fold known as the MFS fold (Shi 2013; Zhang et al. 2015). The MFS fold is made up of four structural repeats (represented by four distinct colors in Fig. 1), each comprising of three consecutive transmembrane helices (TMHs). While the N-terminal domain includes TMHs 1–3 and 4–6, the C-terminal domain contains TMHs 7–9 and 10–12 (Shi 2013). Interestingly, the two repeats within each domain are related to each other by 180° rotation around an imaginary axis that lies parallel to the membrane plane. Further, the N- and C-terminal domains are themselves related by a pseudo-twofold symmetry axis perpendicular to the phospholipid bilayer (Shi 2013; Zhang et al. 2015) (Fig. 1). Functionally, the first helix from each structural repeat (TMHs 1, 4, 7 and 10), contributes to the substrate-binding pocket. In contrast, the second (TMHs 2, 5, 8 and 11) and third (TMHs 3, 6, 9 and 12) helices of the structural repeats take part in inter-domain conformational changes and interactions with the lipid bilayer, respectively (Zhang et al. 2015). Even though several structures of bacterial MFS transporters are known, for instance, lactose:H⁺ symporter LacY (Abramson et al. 2003), the multidrug transporter EmrD (Yin et al. 2006), the glycerol-3-phosphate:phosphate antiporter GlpT (Huang et al. 2003), xylose:H⁺ symporter XylE (Sun et al. 2012) etc (Shi 2013), eukaryotic MFS transporters with known structures are relatively fewer in numbers. Three eukaryotic MFS transporters whose structures have recently been resolved include the *Piriformospora indica* phosphate transporter PiPT (Pedersen et al. 2013), plant nitrate transporter NRT1.1 (Sun et al. 2014) and human glucose transporter GLUT1 (Deng et al. 2014).

With respect to the mechanism of transport, the alternating access mechanism seems to be the most plausible one

(Jardetzky 1966; Forrest et al. 2011; Kaback et al. 2011; Yan 2015). This model envisions the protein to undergo at least two conformational switches to complete the transport process. While one is an outward facing conformation, the other is inward facing, which facilitates alternating access to the substrate binding pocket from either extracellular side or the cytoplasmic side of the membrane (Yan 2013). Broadly, according to this model, the catalytic cycle does not involve any significant movement of the substrate binding site relative to the bilayer, however, sequential conformational changes of the transporter lead to a sequential exposure of the binding site, i.e., alternate accessibility of this site to one side of the membrane bilayer or the other (Forrest et al. 2011). Alterations in MFS transporter functioning have been often found to be associated with various human diseases such as Amyotrophic lateral sclerosis (ALS), De Vivo syndrome, schizophrenia and Alzheimer's disease (Yan 2013; Deng et al. 2014). Furthermore, in case of many pathogens such as *Candida albicans*, these proteins have been found to be responsible for multidrug resistance (MDR) further necessitating the need for an in-depth study (Sanglard et al. 1995; Redhu et al. 2016). Recent advancements in structure-determination methods have allowed deduction of essential molecular details such as the conserved MFS fold and various conformational stages in the transport mechanism (Abramson et al. 2003; Kaback et al. 2011; Yan 2013; Jiang et al. 2016; Kumar et al. 2018). However, there exists a lacuna in overall understanding of the superfamily in-terms of phylogeny and conserved functional residues. Keeping in view the medical implications, MFS superfamily representatives from the TCDB database were analyzed. In the present study, we performed extensive computational analysis to retrieve essential details regarding phylogenetic relationships of various MFS members. Additionally, to facilitate further structure–function studies on MFS transporters, we employed information theoretic measures to identify important functional residues in the MFS members. Since, certain residues we identified as important have already been found to be crucial for functioning of some of the MFS proteins, our predictions appear valid. Altogether, the results provide an essential framework for structural and functional studies on these critical proteins.

Methods

Data source

All the sequences of Transporter classification database (TCDB) were downloaded from the database download link (<http://www.tcdb.org/download.php>). A total of 855 MFS sequences were obtained from these TCDB downloaded sequences. Apart from the sequences, we also used 24 MFS

Table 1 MFS (2.A.1) family members

S. no.	MFS_ID	MFS family
1.	2.A.1.1	The Sugar Porter (SP) Family
2.	2.A.1.2	The Drug:H + Antiporter-1 (12 Spanner) (DHA1) Family
3.	2.A.1.3	The Drug:H + Antiporter-2 (14 Spanner) (DHA2) Family
4.	2.A.1.4	The Organophosphate:Pi Antiporter (OPA) Family
5.	2.A.1.5	The Oligosaccharide:H + Symporter (OHS) Family
6.	2.A.1.6	The Metabolite:H + Symporter (MHS) Family
7.	2.A.1.7	The Fucose: H + Symporter (FHS) Family
8.	2.A.1.8	The Nitrate/Nitrite Porter (NNP) family
9.	2.A.1.9	The Phosphate: H + Symporter (PHS) Family
10.	2.A.1.10	The Nucleoside: H + Symporter (NHS) Family
11.	2.A.1.11	The Oxalate:Formate Antiporter (OFA) Family
12.	2.A.1.12	The Sialate:H + Symporter (SHS) Family
13.	2.A.1.13	The Monocarboxylate Transporter (MCT) Family
14.	2.A.1.14	The Anion:Cation Symporter (ACS) Family
15.	2.A.1.15	The Aromatic Acid:H + Symporter (AAHS) Family
16.	2.A.1.16	The Siderophore-Iron Transporter (SIT) Family
17.	2.A.1.17	The Cyanate Porter (CP) Family
18.	2.A.1.18	The Polyol Porter (PP) Family
19.	2.A.1.19	The Organic Cation Transporter (OCT) Family
20.	2.A.1.20	The Sugar Efflux Transporter (SET) Family
21.	2.A.1.21	The Drug:H + Antiporter-3 (12 Spanner) (DHA3) Family
22.	2.A.1.22	The Vesicular Neurotransmitter Transporter (VNT) Family
23.	2.A.1.23	The Conjugated Bile Salt Transporter (BST) Family
24.	2.A.1.24	The Vacuolar Basic Amino Acid Transporter (VBAAT) Family
25.	2.A.1.25	The Peptide/Acetyl-Coenzyme A/Drug Transporter (PAT) Family
26.	2.A.1.26	The Drug:H + Antiporter-4 (DHA4) Family Family
27.	2.A.1.27	The Phenyl Propionate Permease (PPP) Family
28.	2.A.1.28	The Feline Leukemia Virus Subgroup C Receptor (FLVCR)/Heme Importer Family
29.	2.A.1.29	The Potential Heme Import (HemeI) Family
30.	2.A.1.30	The Putative Abietane Diterpenoid Transporter (ADT) Family
31.	2.A.1.31	The Nickel Resistance (Nre) Family
32.	2.A.1.32	The Putative Aromatic Compound/Drug Exporter (ACDE) Family
33.	2.A.1.33	The Putative YqgE Transporter (YqgE) Family
34.	2.A.1.34	The Sensor Kinase-MFS Fusion (SK-MFS) Family
35.	2.A.1.35	The Fosmidomycin Resistance (Fsr) Family
36.	2.A.1.36	The Acriflavine-sensitivity (YnfM) Family
37.	2.A.1.37	The Unknown Major Facilitator-4 (UMF4) Family
38.	2.A.1.38	The Enterobactin (Siderophore) Exporter (EntS) Family
39.	2.A.1.39	The Vibrioferrin (Siderophore) Exporter (PrsC) Family
40.	2.A.1.40	Major Facilitator Superfamily Domain-containing Protein Family
41.	2.A.1.41	The Putative Bacteriochlorophyll Delivery (BCD) Family
42.	2.A.1.42	The Lysophospholipid Transporter (LpIT) Family
43.	2.A.1.43	The Putative Magnetosome Permease (PMP) Family
44.	2.A.1.44	The L-Amino Acid Transporter-3 (LAT3) Family
45.	2.A.1.45	The 2,4-diacetylphloroglucinol (PHL) Exporter (PHL-E) Family
46.	2.A.1.46	The Uncharacterized Major Facilitator-5 (UMF5) Family
47.	2.A.1.47	The Unknown Major Facilitator-6 (UMF6) Family
48.	2.A.1.48	The Vacuolar Basic Amino Acid Transporter (V-BAAT) Family
49.	2.A.1.49	The Endosomal Spinster (Spinster) Family
50.	2.A.1.50	The Proton Coupled Folate Transporter/Heme Carrier Protein (PCFT/HCP) Family
51.	2.A.1.51	The Unknown Major Facilitator 7 (UMF7) Family

Table 1 (continued)

S. no.	MFS_ID	MFS family
52.	2.A.1.52	The Glycerophosphodiester Uptake (GlpU) Family
53.	2.A.1.53	The Proteobacterial Intrapagosomal Amino Acid Transporter (Pht) Family
54.	2.A.1.54	The Unknown (Archaeal/Bacterial) Major Facilitator-9 (UMF9) Family
55.	2.A.1.55	Unknown Major Facilitator-8 (UMF8) Family
56.	2.A.1.56	The 1,3-Dihydroxybenzene/Drug Transporter (DHB-T) Family
57.	2.A.1.57	The Ferripyochelin Transporter (FptX) Family
58.	2.A.1.58	The N-acetylglucosamine Transporter (NAG-T) Family
59.	2.A.1.59	Unidentified Major Facilitator-10 (UMF10) Family
60.	2.A.1.60	The Rhizopine-related MocC (MocC) Family
61.	2.A.1.61	The Microcin C51 Immunity Protein (MccC) Family
62.	2.A.1.62	The Unidentified Major Facilitator-11 (UMF11) Family
63.	2.A.1.63	The Unidentified Major Facilitator-12 (UMF12) Family
64.	2.A.1.64	The Unidentified Major Facilitator-13 (UMF13) Family
65.	2.A.1.65	The Unidentified Major Facilitator-14 (UMF14) Family
66.	2.A.1.66	The Unidentified Major Facilitator-15 (UMF15) Family
67.	2.A.1.67	The Unidentified Major Facilitator-16 (UMF16) Family
68.	2.A.1.68	The Glucose Transporter (GT) Family
69.	2.A.1.69	Unidentified Major Facilitator-17 (UMF17) Family
70.	2.A.1.70	Unidentified Major Facilitator-18 (UMF18) Family
71.	2.A.1.71	The Valanimycin-resistance (Val-R) Family
72.	2.A.1.72	The Uncharacterized Major Facilitator-20 (UMF20) Family
73.	2.A.1.73	The Unidentified Major Facilitator-21 (UMF21) Family
74.	2.A.1.74	The Uncharacterized Major Facilitator-22 (UMF22) Family
75.	2.A.1.75	The Unidentified Major Facilitator-23 (UMF23) Family
76.	2.A.1.76	The Uncharacterized Major Facilitator 24 Family
77.	2.A.1.77	Uncharacterized Major Facilitator-25 (UMF25) Family
78.	2.A.1.78	The Uncharacterized Major Facilitator-26 (UMF26) Family
79.	2.A.1.79	The Uncharacterized Major Facilitator-27 (UMF27) Family
80.	2.A.1.80	The Uncharacterized Major Facilitator-28 (UMF28) Family
81.	2.A.1.81	The Copper Uptake Porter (Cu-UP)
82.	2.A.1.82	The Plant Copper Uptake Porter (PI-Cu-UP)
83.	2.A.1.83	The 1-arseno-3-phosphoglycerate exporter (APGE) Family
84.	2.A.1.84	The 1-arseno-3-phosphoglycerate exporter (APGE) Family
85.	2.A.1.85	The Uncharacterized Major Facilitator-29 (UMF29) Family
86.	2.A.1.86	The Uncharacterized Major Facilitator-30 (UMF30) Family
87.	2.A.1.87	The Uncharacterized Major Facilitator-31 (UMF31) Family

The families which exclusively include transporter proteins with 12 Transmembrane helices (TMHs) are highlighted in bold

structures from PDB (<https://www.rcsb.org/>) which exhibit a 12 TMH topology. Table 3 includes the PDB IDs of the proteins and their respective details.

Sequence and structure alignment

Multiple alignments of membrane protein sequences is a challenging task mainly due to high conservation of hydrophobic residues at the transmembrane regions by default, and hence, requires methodology which is distinct from that used for soluble proteins. For the alignment, we developed

an in-house strategy similar to that used by the membrane-specific alignment program PRALINE (Pirovano et al. 2008), where we first used TOPCONS (Tsirigos et al. 2015) result which predicts extracellular, intracellular or transmembrane regions for each MFS sequence. This resulted in all the sequences having 25 sections corresponding to extracellular, transmembrane and intracellular regions. The section of sequences corresponding to the transmembrane regions was aligned separately in MAFFT (Katoh et al. 2002; Katoh and Standley 2013) with gap opening penalty of 4 and using PHAT substitution matrix, while all other

Table 2 Families which are distant members of the MFS

S. no.	TCDB ID	Family	Substrate
1.	2.A.2	The Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	Mostly glycosides with monovalent cations such as H ⁺ or Na ⁺
2.	2.A.12	The ATP:ADP Antiporter (AAA) Family	ATP, ADP, UMP, GMP etc.
3.	2.A.17	The Proton-dependent Oligopeptide Transporter (POT/PTR) Family	Oligopeptides
4.	2.A.48	The Reduced Folate Carrier (RFC) Family	Folates, reduced folates, their derivatives and methotrexate
5.	2.A.57	The Equilibrative Nucleoside Transporter (ENT) Family	Nucleosides, xenobiotics
6.	2.A.60	The Organo Anion Transporter (OAT) Family	Bromosulfobromophthalein, prostaglandins, conjugated and unconjugated bile acids (taurocholate and cholate, respectively), steroid conjugates such as estrone-sulfate and dehydroepiandrosterone-sulfate
7.	2.A.71	The Folate-Biopterin Transporter (FBT) Family	S-adenosylmethionine, biopterin, folate etc.
8.	2.A.100	The Ferroportin (Fpn) Family	Fe ²⁺
9.	9.B.111	The 6 TMS Lysyl tRNA Synthetase (LysS) Family	Proteins have Lysyl tRNA synthetase domain terminal to the MFS domain

The families which exclusively include transporter proteins with 12 TMHs are highlighted in bold

Fig. 1 A typical MFS fold comprised of 12 transmembrane helices (TMHs), which are organized into N- and C-terminal domains. The domains are pseudo-twofold symmetrical to each other. Each domain is made up of two inverted repeats comprising of three consecutive

TMHs. The reference structure used in the figure is of the glycerol-3-phosphate:phosphate antiporter (GlpT) of *E. coli*, PDB ID: 1P24 (Huang et al. 2003)

Table 3 Details of the proteins with known 3D structures used for structural alignment

S. no.	PDB ID	Protein name and structure detail
1.	1pv6	Crystal structure of lactose permease
2.	1pv7	Crystal structure of lactose permease with TDG
3.	2v8n	Wild-type Structure of Lactose Permease
4.	2cfp	Sugar Free Lactose Permease at acidic pH
5.	2cfq	Sugar Free Lactose Permease at neutral pH
6.	4oaa	Crystal structure of <i>E. coli</i> lactose permease G46W,G262W bound to sugar
7.	2xut	Crystal structure of a proton-dependent oligopeptide (POT) family transporter
8.	4lep	Structural insights into substrate recognition in proton-dependent oligopeptide transporters
9.	4ikv	Crystal structure of peptide transporter POT
10.	1pw4	Crystal structure of the Glycerol-3-Phosphate Transporter from <i>E. coli</i>
11.	3o7p	Crystal structure of the <i>E. coli</i> Fucose:proton symporter, FucP (N162A)
12.	3o7q	Crystal structure of a Major Facilitator Superfamily (MFS) transporter, FucP
13.	3wdo	Structure of <i>E. coli</i> YajR transporter
14.	4iu8	Crystal structure of a membrane transporter
15.	4iu9	Crystal structure of a membrane transporter
16.	4 m64	3D crystal structure of Na ⁺ /melibiose symporter of Salmonella typhimurium
17.	4gby	The structure of the MFS proton:xylose symporter Xyle bound to D-xylose
18.	4gbz	The structure of the MFS proton:xylose symporter Xyle bound to D-glucose
19.	4gc0	The structure of the MFS proton:xylose symporter Xyle bound to 6-bromo-6-deoxy-D-glucose
20.	4j05	Crystal structure of a eukaryotic phosphate transporter
21.	4ja3	Partially occluded inward open conformation of the xylose transporter Xyle from <i>E. coli</i>
22.	4ja4	Inward open conformation of the xylose transporter Xyle from <i>E. coli</i>
23.	4pyp	Crystal structure of the human glucose transporter GLUT1
24.	4qiq	Crystal structure of D-xylose-proton symporter

sections were aligned with a gap opening penalty and extension penalty of 8 and 4, respectively, and using BLOSUM62 substitution matrix. These section-wise alignments were later joined and resulting alignment was further used in the study. MAFFT software was used to carry out profile–profile alignment. The multiple structural alignments were carried out using the program STAMP (Russell and Barton 1992) that is provided as part of VMD (Humphrey et al. 1996). To visualize the conservation patterns that emerge from the multiple sequence alignment (MSA) of nucleotide and amino acid sequences, we employed sequence logos which are an established method to display consensus sequences (Schneider and Stephens 1990). Herein, the characters representing the sequence are stacked on top of each other for each position in the aligned sequences. The height of each letter is made proportional to its frequency, and the letters are sorted such that the most common one occupies the top position. The height of the entire stack is then adjusted to signify the information content of the sequences at that position and measured in bits (Schneider and Stephens 1990). Information content at position k (in bits) is given by

$$I(k) = \log_2(20) + \sum_{i=1}^{20} p_i \log_2(p_i),$$

where p_i is the amino acid frequency at that position.

WebLogo 3 (<http://Weblogo.threeplusone.com/create.cgi>) (Crooks et al. 2004) was used to derive the sequence logos out of the MSA.

Calculation of relative entropy (RE) and cumulative relative entropy (CRE) scores of MFS family

Relative Entropy (RE) scores were calculated by comparing the amino acid probability distribution for each column of the multiple sequence alignment with that of the background distribution.

RE or the Kullback–Leibler divergence (KL divergence) was originally introduced by Solomon Kullback and Richard Leibler in 1951 as the direct divergence between two distributions (Kullback and Leibler 1951). It is used to measure the difference of an amino acid distribution P from some background distribution P_{null} . The RE score of a column i is defined as

$$\text{RE}_i = \sum_{x=1}^{20} p_i(x) \log \frac{p_i(x)}{p_{\text{null}}(x)},$$

where P_{null} is the background probability of amino acid x which is generally calculated as the probability of finding an amino acid x in all available protein sequences, i.e., protein sequences in Swiss-Prot database. The high scoring residues were further mapped onto the structure of *Escherichia coli* lactose permease (LacY), PDB ID: 1pv6:A (Abramson et al. 2003) by carrying out profile–profile alignment between the sequence based and structural alignment.

CRE calculation for DHA1 and DHA3 families was performed using previously described methodology (Kapoor et al. 2010).

Hannenhalli and Russel created the CRE method for identification of Specificity Determining Residues (SDRs) provided that there is an alignment and subfamilies are known (Kapoor et al. 2010; Nagata et al. 2017). For an alignment position i , the CRE is calculated as

$$\text{CRE}_i = \sum_{x=1}^{20} p_i(x, y_1) \log \frac{p_i(x, y_1)}{p_i(x, y_2)} + \sum_{x=1}^{20} p_i(x, y_2) \log \frac{p_i(x, y_2)}{p_i(x, y_1)},$$

where $p_i(x, y_1)$ and $p_i(x, y_2)$ denote the probabilities of amino acid x in the subfamily y_1 and subfamily y_2 at position i of the alignment, respectively.

All these algorithms have been implemented using PERL scripts written in-house. These scripts can be made available from the corresponding author upon a reasonable request.

Phylogenetic analysis of MFS transporters

FastTree software (version 2) with default amino acid substitution model was used to generate phylogenetic tree (Price et al. 2009). FastTree is known for its ability to handle large alignments and run speeds which are faster than other software (<http://meta.microbesonline.org/fasttree/#Performance>).

Results and discussion

Retrieval and downstream filtering of sequences

A total of 855 MFS sequences were collected from TCDB using the identifier 2.A.1 which were further run through TOPCONS web server (<http://topcons.cbr.su.se/>) to find topology of the MFS proteins. TOPCONS identified a total of 356 sequences which possess 12 transmembrane helices (TMHs). To limit the complexity, we used only 12 TMH MFS sequences for the present study. These 356 sequences segregate into 20 families depending upon their function (highlighted in bold in Table 1). Since, our scoring methods rely on the number of sequences used and there were 6 main functional groups which included sufficient number of sequences for further analyses, we considered only these

groups for our study and the final number of sequences used for analyses stood at 322. For details about the 6 families, kindly refer to Table 4.

Multiple sequence alignment (MSA)

All 322 sequences were cut into three sections: cytoplasmic, transmembrane, non-cytoplasmic section and these sections were aligned separately using a strategy developed in house and elaborated in “Methods” section. To check for the accuracy of alignment, we also performed profile–profile alignment. For this, we took alignment file obtained using the structural alignment program in VMD (24 known structures of MFS proteins containing 12 TMHs were used for the analyses; details in Table 3) as one profile and MFS multiple sequence alignment file as another profile. Sequence logo for the entire protein stretch is not shown since the most conserved pattern was identified in transmembrane portion and it implies that the cytoplasmic and extracellular regions assumed sequence variation during the course of evolution. WebLogo shown in Fig. 2 represents the conserved residues in TM region of the MFS proteins. Some noteworthy residues with very high degree of conservation are arginine (alignment position 16) and glycine (alignment position 20) present in TMH 4 and glycine (alignment position 15) in TMH 5. For the MFS proteins, such high conservation in transmembrane region indicates importance of the residues in maintaining the structure and function of the proteins. The presence of charged residues is probably required to support the essential movement of ions. Notably, this arginine in case of Mdr1p (R215) of *C. albicans* has been found to be essential for drug/H⁺ antiport mechanism (Redhu et al. 2016). Further, the homologous arginine (R112) in *E. coli* multidrug resistance transporter MdfA is critical for conferring antibiotic resistance (Heng et al. 2015). Glycine due to its physico-chemical nature is plausibly conserved at certain regions in the protein structure to introduce flexibility, which is essential for conformational changes during substrate/ion movement. Interestingly, in case of Lactose permease (LacY) of *E. coli*, substitution of these glycine residues by cysteine leads to altered transport profile (Jung et al. 1995).

Prediction of fold-specific residues: results of RE calculation

Fold-specific residues, are residues which are responsible for maintaining the overall fold of the protein. Such residues usually tend to remain conserved across the MFS protein alignment. To predict the fold-specific residues, relative entropy calculations (RE) were performed. RE calculations identify those residues whose probability distributions are significantly different from their background probability distribution. RE, therefore, predicts residues that are not

Table 4 Details of the MFS families harboring 12 TMHs which are used for sequence and phylogenetic analyses

S. no.	Family	TCDB ID	Domain	Substrate	No. of members	Known structures
1.	The Sugar Porter (SP) Family	2.A.1.1	Bacteria, Archaea and Eukaryota	Glucose, xylose, fucose, arabinose, 2-deoxygalactose, 2-deoxyglucose, quinate, myoinositol	123	4PYP: D-Glucose facilitator from <i>Homo sapiens</i> (GLUT1)
2.	The Drug:H + Antiporter-1 (12 Spanner) (DHA1) Family	2.A.1.2	Bacteria, Archaea and Eukaryota	Pyridoxine, pyridoxal, pyridoxamine, amiloride, cycloheximide, chloramphenicol, tetracycline, daunomycin, ethidium bromide, benomyl, methotrexate, fluconazole, quinolone etc	92	2GFP: The Multi-drug Transporter EmrD from <i>Escherichia coli</i>
3.	The Drug:H + Antiporter-3 (12 Spanner) (DHA3) Family	2.A.1.21	Bacteria and Archaea	Macrolide (erythromycin; oleandomycin; azithromycin; telithromycin), tetracycline etc	32	
4.	The Monocarboxylate Transporter (MCT) Family	2.A.1.13	Eukaryota	Lactate, pyruvate, mevalonate, branched chain oxo acids, β -hydroxybutyrate, γ -hydroxybutyrate, butyrate, acetoacetate, acetate and formate, anti-tumor agents, 3-bromopyruvate, dichloroacetate etc	21	
5.	The Anion:Cation Symporter (ACS) Family	2.A.1.14	Bacteria and Eukaryota	Hexuronate, dipeptide (e.g., Gly-Leu), allantate, ureidosuccinate, allantoin etc	30	
6.	The Organic Cation Transporter (OCT) Family	2.A.1.19	Eukaryota	Monoamine neurotransmitters, l-carnitine, α -ketoglutarate, cAMP, cGMP, prostaglandins, urate etc	24	

identified by the tradition scoring methods. It was observed that residues identified as fold specific are scattered across the 12 TMHs consistent with a general requirement for membrane localization. Interestingly, TMH4 and TMH5 contain certain residues with very high RE score. Interestingly, in case of lactose permease of *E. coli*, both these helices have been implicated in substrate binding (Weinglass and Kaback 1999). Notably, in case of the antiporters of the MFS, TMH5 houses the well-known antiporter-motif or motif-C (Redhu et al. 2016). It is a glycine and proline rich motif (G X(8) G X(3) G P X(2) G G), which is specific to the 12 and 14 TMH member antiporters of the superfamily and not found in symporters and uniporters. Mutagenesis of the glycine residue of the GP dipeptide within motif-C has been shown to alter tetracycline resistance conferred by TetA(C), a tetracycline/H + antiporter (Varela et al. 1995). Much recently, mutational analyses of a MFS transporter of *C. albicans*, Mdr1, revealed that both these helices (TMH4 and TMH5) contains quite a significant number of residues which are important for the transport function (Redhu et al.

2018). To get an insight into the distribution of the high scoring residues on a 3D-structure, top 20 residues with the maximum scores were mapped on to the crystal structure of the lactose permease (LacY) of *Escherichia coli* (PDB ID: 1pv6) (Fig. 3). LacY, since long has been the model to study structural and mechanistic aspects of MFS transporters (Abramson et al. 2003; Kaback et al. 2011). Table 5 provides list of all the high scoring residues that have been highlighted in the structure. Interestingly, quite a number of these have already been reported to be important for the protein's structure and/or function. For instance, G115 and G386 when mutated to cysteine leads to reduced cell-surface expression of the protein. However, while G115C was characterized by a severely impaired transport profile, G386C mutation led to an altered, yet significant transport of lactose by the protein (Jung et al. 1995). E325 which has been established as a key player in proton translocation, also is among the top 20 high scoring residues deduced from our analysis (Guan and Kaback 2006; Andersson et al. 2012). Further, the top scoring residue N119 has also been recently

Fig. 2 Sequence logos generated using WebLogo3 (Crooks et al. 2004) to demonstrate conservation patterns which emerged from the multiple sequence alignment. The height of the entire stack of residues is the information measured in bits. While overall height of the stack represents sequence conservation at a particular position, the height of a symbol indicates frequency of the respective amino acid at that position. Color scheme used is hydrophobicity. All the hydropho-

bic residues (YVMCLFIW) are colored black, while neutral amino acid (SGHTAP) and hydrophilic residues (RKDENQ) are colored green and blue, respectively. The arrow above the sequence logos marks the residue numbers of the respective transmembrane helix (TMH) in *Candida albicans* Mdr1p (UniProtKB - Q5ABU7 (MDR1_CANAL)). Note: only the starting residue number is indicated for helices which have gaps in their alignment

Fig. 3 Tertiary structure of Lactose permease of *E. coli* (PDB ID: 1pv6:A) (Abramson et al. 2003) highlighting the high RE score residues. The right panel provides view of the structure from the extracellular side. UCSF Chimera was used for visualization

Table 5 Residues with maximum Relative Entropy (RE) scores. Reference structure to the is *E. coli* lactose permease, PDB ID: 1pv6:A

S. no.	RE score	Single letter code for amino acid	Position in the structure
1.	1.929622816	N	119
2.	1.85476317	G	150
3.	1.572738342	G	71
4.	1.500695803	G	115
5.	1.206093639	D	68
6.	1.162674489	L	72
7.	1.151250687	S	366
8.	1.028255826	F	55
9.	1.026309946	A	50
10.	0.9976300838	S	174
11.	0.9934374063	Y	382
12.	0.9727971209	C	117
13.	0.968186685	P	123
14.	0.9448318204	A	295
15.	0.9410380679	E	342
16.	0.9409698009	G	386
17.	0.9240998624	F	324
18.	0.9192998083	E	325
19.	0.9179856293	F	59
20.	0.8768442085	L	58

revealed to be involved in substrate interaction (Klauda and Brooks 2007; Zhuang and Klauda 2016). Some other high scoring residues with established structural or functional

importance include P123 (Guan and Kaback 2006) and D68 (Suárez-Germà et al. 2012). Hence, it is of much importance to evaluate the importance of residues that have high RE scores but not been studied yet.

Phylogenetic analyses of 2.A.1 MFS family

MFS is an ancient protein superfamily, with members existing in almost all prokaryotes and eukaryotes. FastTree2, which shows good and fast performance with large alignments, was used for the tree construction. Interestingly, the protein family clusters on the basis of function, and not lineage. The resultant trees showed clustering of the sequences into six groups (Fig. 4). The clustering of branch is from points near the center of the tree. Group 1 and group 2 are the largest clade and Group 1 includes all the Sugar Porter (SP) Families which are colored red in the phylogenetic tree. The second group includes all the members of Drug: H⁺ Antiporter-1 (12 Spanner) (DHA1) Family and colored with purple in the phylogenetic tree. The third group is the Drug: H⁺ Antiporter-3 (12 Spanner) (DHA3) Family (colored light green) and it is to be noted that DHA1 and DHA3 families are coming together after the initial divergence from the center of the tree, suggesting that they are more closely related to each other than other MFS families. The clustering of DHA1 and DHA3 together also indicates their similar functional specificities. The fourth cluster is the Monocarboxylate Transporter (MCT) (lavender color) Family which branches from other MFS groups at a point that is somewhat distant from the center of the tree. The fifth

Fig. 4 Phylogenetic tree made using FastTree (Price et al. 2009) with representative sequences from the six functional groups: SP, DHA1, DHA3, OCT, ACS and MCT

cluster is the Anion:Cation Symporter (ACS) Family branch (colored cyan) also emerges from a point that is relatively distant from the center of the tree. The sixth cluster is the Organic Cation Transporter (OCT) Family branch (colored olive green) emerges far from the center of the tree. From the phylogenetic tree analysis, it is predicted that all these sequences are clustering on the basis of their functions and not evolutionary relationships. Table 4 summarizes the details about these families.

Cumulative relative entropy (CRE) calculations reveal a set of residues differentially conserved in DHA1 as compared to DHA3

Drug H⁺ antiporters utilize proton gradients to expel drugs. Within the MFS, there exist three families of Drug/H⁺ antiporters: DHA1, DHA2 and DHA3. As mentioned previously, only DHA1 and DHA3 comprise of 12 membrane spanning helices. These proteins are highly promiscuous as far as their substrates are concerned, and hence, contribute to multidrug resistance, i.e., resistance to two or more classes of xenobiotics. For instance, MdfA of DHA1 family has a wide substrate spectrum that includes benzalkonium, tetracycline, rifampin, daunomycin and puromycin (Edgar and Bibi 1997). In case of pathogenic fungi *C. albicans*, Mdr1 is a DHA1 transporter whose overexpression has been associated with multidrug resistance in clinical isolates (Sanglard et al. 1995; White 1997; Redhu et al. 2016). DHA3 includes drug translocators of bacteria and archaea. Mef(A) and TetA(P) are two key efflux pumps of this group which are involved in macrolide and tetracycline resistance, respectively (Bannam et al. 2004; Bley et al. 2011). Keeping in view the implications of DHA1 and DHA3 in multidrug resistance phenomenon, we utilized another information-theoretic measure to explore distinct sequence signatures in these families. In addition, since DHA3 is prokaryote-specific while DHA1 is not, it was of interest to explore if there exist certain differentially conserved residues in one family when compared with another. If true, such residues would plausibly be responsible for functional specificity. Hence, we carried out the CRE calculation using an already developed method by us (Kapoor et al. 2010). Top twenty residues with the maximum scores have been provided in Table 6. Figure 5 illustrates the RE and CRE results for TMH4 and TMH5 which had the maximum representation in the top 20 high scoring CRE residues. The noted residue among them is the arginine; R which had the maximum CRE score. As mentioned previously, this arginine in MdfA (R112) has been found to be essential for imparting antibiotic resistance. Actually, this residue and the closely followed glycine is part of the motif-B, which has been postulated by Heng and coworkers to be an important motif for coupling protonation status with substrate binding in MdfA (Heng et al.

Table 6 Residues with maximum Cumulative Relative entropy (CRE) scores in case of DHA1

S. no.	CRE score	Single letter code for amino acid residue
1.	2.483559062	L
2.	1.947490894	H
3.	1.927739021	D
4.	1.733056299	G
5.	1.436364617	G
6.	1.246116874	P
7.	1.162684149	Y
8.	1.10889475	A
9.	1.0898554	I
10.	1.041862322	T
11.	1.020104641	A
12.	1.003872814	A
13.	0.9328613477	Q
14.	0.9321445455	T
15.	0.8593531984	A
16.	0.8240044938	G
17.	0.8112236309	G
18.	0.8063260266	A
19.	0.7733401738	S
20.	0.7697734608	D

2015). Under such scenario, a relative poor representation or the absence of this residue from DHA3 family points to a different mechanism which might be operating in the latter.

Comparative analyses of MFS families under 2.A.1 with families which are distant members of the MFS

Apart from the MFS families listed under TCDB id 2.A.1, there are nine families which are distant members of the MFS and comprise proteins with transport function. Table 2 contains details about these families. Five families out of these nine possess 12 TMHs. These include families: #2.A.2: Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family, #2.A.12: ATP:ADP Antiporter (AAA) Family, #2.A.17: Proton-dependent Oligopeptide Transporter (POT/PTR) Family, #2.A.48: Reduced Folate Carrier (RFC) Family, #2.A.60: Organo Anion Transporter (OAT) Family. A total of 89 sequences could be retrieved from these families. We also performed phylogenetic analyses of TCDB #2.A.1 with these families (Fig. 6). The first clade as described before is 2.A.1 and is the largest clade denoted with red color. It shows closeness to the second clade TCDB #2.A.2 (GPH) which is colored purple and third clade with TCDB #2.A.60 which is colored olive green. A small portion of clade 2.A.1 shows closeness to fourth clade with TCDB #2.A.17 (colored with magenta color) and to fifth clade with

Fig. 5 WebLogo and graphical representation of RE and CRE calculations performed with DHA1 and DHA3 families of MFS highlighting the presence of differentially conserved residues within TMH4 (top panel) and TMH5 (lower panel). The height of the entire stack of residues is the information measured in bits. While overall height of the stack represents sequence conservation at a particular position, the height of a symbol indicates frequency of the respective amino acid at that position. Color scheme used is hydrophobicity. All the

hydrophobic residues (YVMCLFIW) are colored black, while neutral amino acid (SGHTAP) and hydrophilic residues (RKDENQ) are colored green and blue, respectively. The double-headed arrow above the sequence logos marks the residue numbers of the respective transmembrane helix (TMH) in *Candida albicans* Mdr1 (UniProtKB - Q5ABU7 (MDR1_CANAL)) and Cmr of *Corynebacterium glutamicum* (UniProtKB—Q79VC7 (Q79VC7_CORGL)) for DHA1 and DHA3 families, respectively

Fig. 6 Phylogenetic tree made using FastTree (Price et al. 2009) to analyze relatedness of MFS family (2.A.1) members with other families present in the superfamily, viz. #2.A.2: Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family, #2.A.12: ATP:ADP Antiporter (AAA) Family, #2.A.17: Proton-dependent Oligopeptide Transporter (POT/PTR) Family, #2.A.48: Reduced Folate Carrier (RFC) Family, #2.A.60: Organo Anion Transporter (OAT) Family

TCDB #2.A.12 which is colored with light green. The sixth clade with TCDB #2.A.48 and colored cyan shows closeness to TCDB #2.A.12.

Conclusion

MFS, being the quite diverse protein superfamily includes proteins with definite physiological and clinical implications. With the increased inflow of sequenced genomes, huge numbers of MFS proteins are being identified. To initiate functional characterization of these proteins, prior knowledge of their putative function is a prerequisite. Multiple sequence alignment serves as a valuable tool not only to predict function but also to derive phylogenetic relationships via tree construction. In the present study, we performed an extensive analysis of the MFS family members using established computational methods to provide an essential framework for their detailed characterization. The main highlights of the current study include function-based clustering of the sequences of MFS superfamily, prediction of putative fold-specific residues and the revelation of certain residues in DHA1 which are differentially conserved when compared with DHA3. Keeping in view that some of these residues have already been highlighted as functionally important in the family, they serve as essential factors to be kept in mind before initiating any targeted approach against these proteins. Thus, the current results lay a suitable platform for future investigations on proteins of this superfamily.

Acknowledgements The authors would like to thank HPCF facility of SC & IS, JNU for providing computational facility. PV would like to acknowledge UPE-II (project ID 105) for financial support.

Compliance with ethical standards

Conflict of interest The authors would like to declare that no conflicts of interest exist.

References

- Abramson J, Smirnova I, Kasho V et al (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301:610–615. <https://doi.org/10.1126/science.1088196>
- Andersson M, Bondar AN, Freitas JA et al (2012) Proton-coupled dynamics in lactose permease. *Structure* 20:1893–1904. <https://doi.org/10.1016/j.str.2012.08.021>
- Bannam TL, Johannesen PA, Salvado CL et al (2004) The *Clostridium perfringens* TetA(P) efflux protein contains a functional variant of the Motif A region found in major facilitator superfamily transport proteins. *Microbiology* 150:127–134. <https://doi.org/10.1099/mic.0.26614-0>
- Bley C, van der Linden M, Reinert RR (2011) mef(A) is the predominant macrolide resistance determinant in *Streptococcus pneumoniae* and *Streptococcus pyogenes* in Germany. *Int J Antimicrob Agents* 37:425–431. <https://doi.org/10.1016/j.ijantimicag.2011.01.019>
- Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>
- Deng D, Xu C, Sun P et al (2014) Crystal structure of the human glucose transporter GLUT1. *Nature* 510:121–125. <https://doi.org/10.1038/nature13306>
- Diallinas G (2016) Dissection of transporter function: from genetics to structure. *Trends Genet* 32:576–590. <https://doi.org/10.1016/j.tig.2016.06.003>
- Edgar R, Bibi E (1997) MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J Bacteriol* 179:2274 LP–L2280
- Finn RD, Coghill P, Eberhardt RY et al (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Forrest LR, Krämer R, Ziegler C (2011) The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta Bioenerg* 1807:167–188. <https://doi.org/10.1016/j.bbabi.2010.10.014>
- Gaur M, Puri N, Manoharlal R et al (2008) MFS transportome of the human pathogenic yeast *Candida albicans*. *BMC Genom* 9:579. <https://doi.org/10.1186/1471-2164-9-579>
- Guan L, Kaback HR (2006) Lessons from lactose permease. *Annu Rev Biophys Biomol Struct* 35:67–91. <https://doi.org/10.1146/annurev.biophys.35.040405.102005>
- Heng J, Zhao Y, Liu M et al (2015) Substrate-bound structure of the *E. coli* multidrug resistance transporter MdfA. *Cell Res* 25:1060
- Huang Y, Lemieux MJ, Song J et al (2003) Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* 301:616–620. <https://doi.org/10.1126/science.1087619>
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:27–28,33–38
- Jardetzky O (1966) Simple allosteric model for membrane pumps. *Nature* 211:969–970
- Jiang X, Smirnova I, Kasho V et al (2016) Crystal structure of a LacY-nanobody complex in a periplasmic-open conformation. *Proc Natl Acad Sci USA* 113:12420–12425. <https://doi.org/10.1073/pnas.1615414113>
- Jung K, Jung H, Colacurcio P, Kaback HR (1995) Role of glycine residues in the structure and function of lactose permease, an *Escherichia coli* membrane transport protein. *Biochemistry* 34:1030–1039. <https://doi.org/10.1021/bi00003a038>
- Kaback HR, Smirnova I, Kasho V et al (2011) The alternating access transport mechanism in LacY. *J Membr Biol* 239:85–93. <https://doi.org/10.1007/s00232-010-9327-5>
- Kapoor K, Rehan M, Lynn AM, Prasad R (2010) Employing information theoretic measures and mutagenesis to identify residues critical for drug-proton antiport function in mdr1p of *Candida albicans*. *PLoS One*. <https://doi.org/10.1371/journal.pone.0011041>
- Kasho VN, Smirnova IN, Kaback HR (2006) Sequence alignment and homology threading reveals prokaryotic and eukaryotic proteins similar to lactose permease. *J Mol Biol* 358:1060–1070. <https://doi.org/10.1016/j.jmb.2006.02.049>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066
- Klauda JB, Brooks BR (2007) Sugar binding in lactose permease: anomeric state of a disaccharide influences binding structure. *J Mol Biol* 367:1523–1534. <https://doi.org/10.1016/j.jmb.2007.02.001>
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 28:89–110

- Kumar H, Finer-Moore JS, Jiang X et al (2018) Crystal structure of a ligand-bound LacY-nanobody complex. *Proc Natl Acad Sci USA*. <https://doi.org/10.1073/pnas.1801774115>
- Lee J, Sands ZA, Biggin PC (2016) A numbering system for MFS transporter proteins. *Front Mol Biosci* 3:1–13. <https://doi.org/10.3389/fmolb.2016.00021>
- Nagata S, Imai J, Makino G et al (2017) Evolutionary analysis of HIV-1 pol proteins reveals representative residues for viral subtype differentiation. *Front Microbiol* 8:2151. <https://doi.org/10.3389/fmicb.2017.02151>
- Pedersen BP, Kumar H, Waight AB et al (2013) Crystal structure of a eukaryotic phosphate transporter. *Nature* 496:533–536. <https://doi.org/10.1038/nature12042>
- Pirovano W, Feenstra KA, Heringa J (2008) PRALINE™: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics* 24:492–497. <https://doi.org/10.1093/bioinformatics/btm636>
- Price MN, Dehal PS, Arkin AP (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>
- Redhu AK, Shah AH, Prasad R (2016a) MFS transporters of *Candida* species and their role in clinical drug resistance. *FEMS Yeast Res* 16:1–12. <https://doi.org/10.1093/femsyr/fow043>
- Redhu AK, Banerjee A, Shah AH et al (2018) Molecular basis of substrate polyspecificity of the *Candida albicans* Mdr1p multidrug/H(+) antiporter. *J Mol Biol* 430:682–694. <https://doi.org/10.1016/j.jmb.2018.01.005>
- Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* 14:309–323. <https://doi.org/10.1002/prot.340140216>
- Saier MH, Reddy VS, Tamang DG, Västermark Å (2014) The transporter classification database. *Nucleic Acids Res* 42:D251–D258. <https://doi.org/10.1093/nar/gkt1097>
- Sanglard D, Kuchler K, Ischer F et al (1995) Mechanisms of resistance to azole antifungal agents in *Candida albicans* isolates from AIDS patients involve specific multidrug transporters. *Antimicrob Agents Chemother* 39:2378–2386. <https://doi.org/10.1128/AAC.39.11.2378>
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100. <https://doi.org/10.1093/nar/18.20.6097>
- Shi Y (2013) Common folds and transport mechanisms of secondary active transporters. *Annu Rev Biophys* 42:51–72. <https://doi.org/10.1146/annurev-biophys-083012-130429>
- Suárez-Germà C, Loura LMS, Domènech Ò et al (2012) Phosphatidylethanolamine–lactose permease interaction: a comparative study based on FRET. *J Phys Chem B* 116:14023–14028. <https://doi.org/10.1021/jp309726v>
- Sun L, Zeng X, Yan C et al (2012) Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* 490:361–366. <https://doi.org/10.1038/nature11524>
- Sun J, Bankston JR, Payandeh J et al (2014) Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* 507:73
- Tsirigos KD, Peters C, Shu N et al (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 43:W401–W407. <https://doi.org/10.1093/nar/gkv485>
- Varela MF, Sansom CE, Griffith JK (1995) Mutational analysis and molecular modelling of an amino acid sequence motif conserved in antiporters but not symporters in a transporter superfamily. *Mol Membr Biol* 12:313–319
- Weinglass AB, Kaback HR (1999) Conformational flexibility at the substrate binding site in the lactose permease of *Escherichia coli*. 96:11178–11182
- White TC (1997) Increased mRNA levels of ERG16, CDR, and MDR1 correlate, with increases in azole resistance in *Candida albicans* isolates from a patient infected with human immunodeficiency virus. *Antimicrob Agents Chemother* 41:1482–1487
- Yan N (2013) Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem Sci* 38:151–159. <https://doi.org/10.1016/j.tibs.2013.01.003>
- Yan N (2015) Structural biology of the major facilitator superfamily transporters. *Annu Rev Biophys* 44:257–283. <https://doi.org/10.1146/annurev-biophys-060414-033901>
- Yin Y, He X, Szewczyk P et al (2006) Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* 312:741–744. <https://doi.org/10.1126/science.1125629>
- Zhang XC, Zhao Y, Heng J, Jiang D (2015) Energy coupling mechanisms of MFS transporters. *Protein Sci* 24:1560–1579. <https://doi.org/10.1002/pro.2759>
- Zhuang X, Klauda JB (2016) Modeling structural transitions from the periplasmic-open state of lactose permease and interpretations of spin label experiments. *Biochim Biophys Acta Biomembr* 1858:1541–1552. <https://doi.org/10.1016/j.bbamem.2016.04.008>