



A Numbering System for MFS Transporter Proteins

Joanna Lee¹, Zara A. Sands² and Philip C. Biggin^{1*}

¹ Department of Biochemistry, University of Oxford, Oxford, UK, ² UCB Pharma S.A., Braine-l'Alleud, Belgium

OPEN ACCESS

Edited by:

Adrian Goldman,
University of Helsinki, Finland

Reviewed by:

Anastassios Papageorgiou,
University of Turku, Finland
Irina Tikhonova,
Queen's University Belfast, UK

*Correspondence:

Philip C. Biggin
philip.biggin@bioch.ox.ac.uk

Specialty section:

This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: 28 January 2016

Accepted: 17 May 2016

Published: 02 June 2016

Citation:

Lee J, Sands ZA and Biggin PC
(2016) A Numbering System for MFS
Transporter Proteins.
Front. Mol. Biosci. 3:21.
doi: 10.3389/fmolb.2016.00021

The Major Facilitator Superfamily (MFS) is one of the largest classes of secondary active transporters and is widely expressed in many domains of life. It is characterized by a common 12-transmembrane helix motif that allows the selective transport of a vast range of diverse substrates across the membrane. MFS transporters play a central role in many physiological processes and are increasingly recognized as potential drug targets. Despite intensive efforts, there are still only a handful of crystal structures and therefore homology modeling is likely to be a necessary process for providing models to interpret experiments for many years to come. However, the diversity of sequences and the multiple conformational states these proteins can exist in makes the process significantly more complicated, especially for sequences for which there is very little sequence identity to known templates. Inspired by the approach adopted many years ago for GPCRs, we have analyzed the large number of MFS sequences now available alongside the current structural information to propose a series of conserved contact points that can provide additional guidance for the homology modeling process. To enable cross-comparison across MFS models we also present a numbering scheme that can be used to provide a point of reference within each of the 12 transmembrane regions.

Keywords: homology modeling, LacY, alternating access, transport, transmembrane

INTRODUCTION

The Major Facilitator Superfamily (MFS) is the largest known superfamily of secondary transporters (Marger and Saier, 1993) expanding in recent years to contain 74 distinct families according to the Transporter Classification Database (TCDB; Saier et al., 2014). These transporters operate by uniport, symport, or antiport mechanisms that take advantage of the electrochemical gradient of the co-transported ion (in the case of symport or antiport) or the concentration of the ligand to instigate the transport cycle (Pao et al., 1998). MFS proteins transport a huge variety of ligands including monosaccharides, drugs, enzyme cofactors, peptides, oligosaccharides, nucleotides, iron chelates, and inorganic anions and cations (Burckhardt and Wolff, 2000; Saier and Paulsen, 2001; Guan and Kaback, 2006; Newstead et al., 2011).

MFS proteins have a conserved 12 transmembrane (TM) α -helix fold (Figure 1A; Reddy et al., 2012) that is comprised of two 6-TM helix bundles that are related by a pseudo two-fold axis of symmetry. The presence of both domains is thought to be functionally important for the transport mechanism, as the ligand binds to the central TM cavity at the interface between the two domains (Pazdernik et al., 1997; Figures 1B,C). The 12-TM topology appears to be the core fold, though the presence of additional helices is sometimes observed as for example seen in peptide transporters

Abbreviations: MFS, major facilitator superfamily.

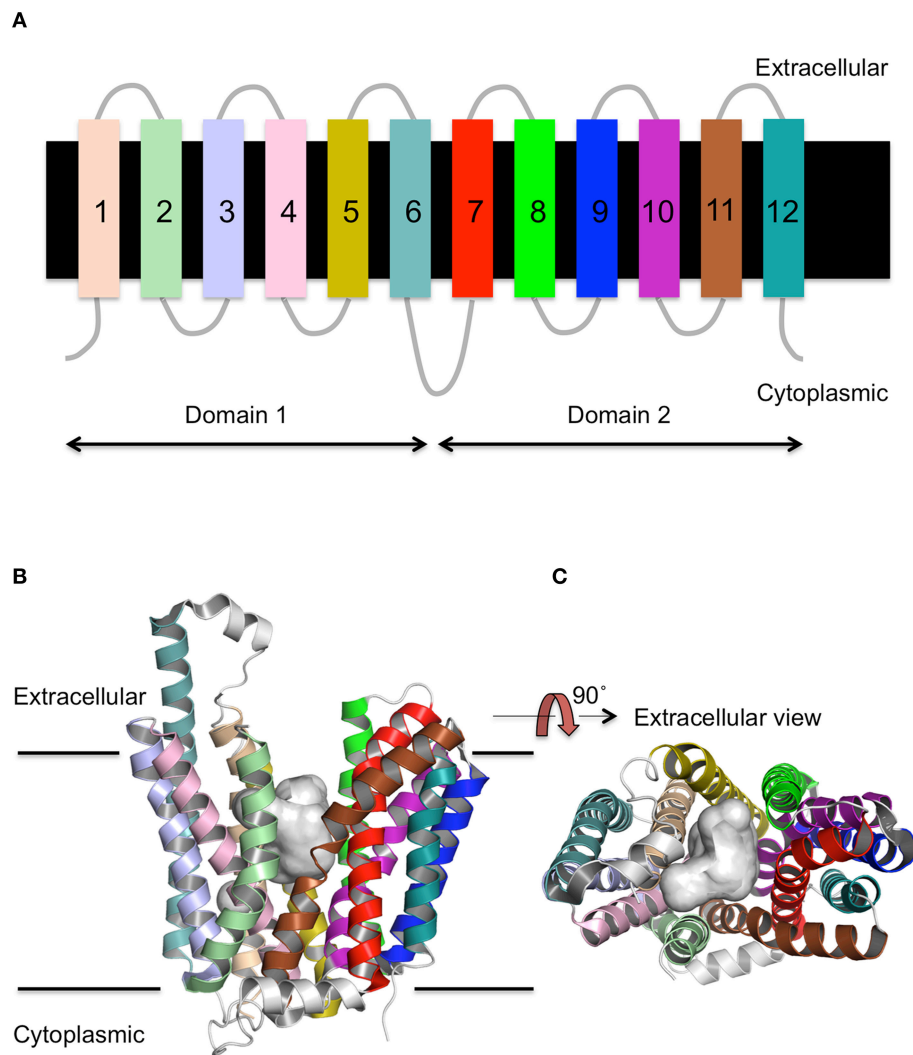


FIGURE 1 | Secondary and tertiary structure of MFS proteins. (A) The topology of the 12 TM helices. **(B)** The MFS fold as viewed from the side of the membrane, and **(C)** rotated through 90° as exemplified by FucP. The conserved 12 TM α -helix fold is arranged into two 6 TMH bundles (domains 1 and 2) and a cavity exists (gray surface) at the interface between them that is accessible to either the cytoplasm or extracellular region depending on the conformational state.

(Newstead et al., 2011). The arrangement of the helices within the distinct conformations of the X-ray crystal structures suggests that transmembrane helices (TMHs) 1, 4, 7, and 10 line the path of the substrate through the transporter, whilst TMHs 2, 5, 8, and 11 mediate the interface between the two domains (Yan, 2015).

There is some controversy as to the evolution of the six TM domains, which may have arisen through nucleotide duplication in the gene for either a 2-TM or 3-TM helix segment. The 3-TM helix repeat motif was proposed by Radestock et al. (Radestock and Forrest, 2011), who showed that an outward facing model of LacY generated via their approach had a backbone atom root mean squared deviation (RMSD) of 3.2 Å (over the 12 TM helices) when fitted to the crystal structure of FucP. Furthermore, alignments using this approach can improve the apparent sequence similarity when compared to alignment of the full sequences. This hypothesis has been extended in recent

years (Keller et al., 2014), culminating in a mix-and-match theory (Madej, 2015).

However, this triple helix motif hypothesis is disputed by (Västermark and Saier, 2014), who direct attention to the conserved MFS motifs in the cytoplasmic loops between TMHs 2–3 and 8–9. These loop motifs align best when using the entire protein sequence, rather than the triple helix motifs described by Radestock et al. (Radestock and Forrest, 2011). In addition to the motif on these cytoplasmic loops (GxlaDrxGR; Paulsen et al., 1996), there is a further motif present on TMH 4 that is comprised of at least one conserved glycine residue (Pascual et al., 2008).

MFS transporters are thought to operate via an alternating access mechanism (Kasho et al., 2006; Erally, 2008; Kaback et al., 2011; Smirnova et al., 2011, 2014). In this mechanism, the MFS protein undergoes a series of conformational changes to allow

the passage of a ligand from the extracellular solution into the cytoplasm or vice versa. In MFS proteins, the most studied region is that of the ligand-binding site that is situated between the two domains in the central TM cavity. The paradigm has been the lactose permease (LacY) that operates by a symport mechanism and transports both sugar (predominantly lactose or galactose) and H^+ in the same direction across the membrane (Guan and Kaback, 2006; Madej et al., 2014). The mechanism of sugar transport in LacY has been extensively investigated and the key residues involved in either lactose binding or H^+ translocation are well-characterized. The ligand-binding site is formed by E269, E325, H322, R302, and Y236, which are predominantly charged residues at neutral pH. Thus, the LacY central cavity is fairly hydrophilic in character, at least at the point of ligand binding and transport is mediated by TMs 5, 7, 8, 10, and 11. The positions of residues involved in substrate transport are similar

in other MFS proteins (Madej and Kaback, 2013; Madej et al., 2013).

The role of the loop regions in MFS proteins is less well-explored. However, Masureel et al. (2014) recently determined that key, titratable residues, were necessary in the loop regions of LmrP in order for conformational change to proceed. Double electron electron resonance (DEER) experiments have defined the distance between helix pairs in LmrP and the effect of pH on the relative distances was measured. The relative abundance of each distance correlated to the conformation of the protein. It was shown that at pH 5, LmrP would reside in the outward-closed state, but at pH 8 an outward-open state was preferred. Like LacY, LmrP is thought to operate as a drug/ H^+ symporter and these results suggest that the pH conditions of the environment can affect the movement of substrate by shifting the conformational equilibrium of the protein.

TABLE 1 | The sequence identity (%) between 10 MFS proteins for which there is an X-ray crystal structure.

Protein	LacY (1PV6)	FucP (3O7P)	XylE (4GBY)	GlpT (1PW4)	PepT (2XUT)	EmrD (3GFP)	YajR (3WDO)	PiPT (4J05)	NarK (4JR9)	GLUT1 (4PYP)
LacY (1PV6)										
FucP (3O7P)	19									
XylE (4GBY)	22	22								
GlpT (1PW4)	21	19	23							
PepT (2XUT)	23	24	19	24						
EmrD (3GFP)	21	18	20	21	25					
YajR (3WDO)	22	21	25	20	23	22				
PiPT (4J05)	22	21	24	19	22	23	22			
NarK (4JR9)	21	22	22	23	23	24	23	21		
GLUT1 (4PYP)	21	23	29	21	20	22	20	22	24	

The percentage identity was computed from CLUSTAL Omega pairwise alignments and they range from 18 to 25.

TABLE 2 | The MFS proteins used for determining key residues that are conserved across the superfamily in this work.

Protein	MFS family	UniRef50/UniProt ID	Substrate	Transport direction	Species
FucP	MFS_1	P11551	Fucose/ H^+ symport	Into cell	<i>E. coli</i>
GlpT	MFS_1	P08194	Glycerol-3P/Pi antiport	Into cell	<i>E. coli</i>
EmrD	DHA12	P31442	Multidrug/ H^+ antiport	Out of cell	<i>E. coli</i>
YajR	DHA12	P77726	Multidrug/ H^+ antiport	Out of cell	<i>E. coli</i>
LacY	lacy_symport	P02920	Galactose/ H^+ symport	Into cell	<i>E. coli</i>
PepT	POT	Q8EKT7	Peptide/ H^+ symport	Into cell	<i>Streptococcus thermophilus</i>
YgbH	POT	P75742	Peptide/ H^+ symport	Into cell	<i>E. coli</i>
GlcPse		Q8CQA7	Glucose/ H^+ symport	Into cell	<i>Staphylococcus epidermis</i>
gkPOT	POT	Q5KYD1	Peptide/ H^+ symport	Into cell	<i>Geobacillus kaustophilus</i>
GLUT1	SP	P11166	Glucose uniport	Into cell	<i>Homo sapiens</i>
PiPT	SP	A8N031	Phosphate/ H^+ symport	Into cell	<i>Piriformospora indica</i>
NRT1.1	NRT1/PTR	Q05085	Nitrite/ H^+ symport	Into cell	<i>Arabidopsis thaliana</i>
NarU	NNP	P37758	Nitrate/nitrite symport	Into cell	<i>E. coli</i>
XylE	SP	P0AGF4	Xylose/ H^+ symport	Into cell	<i>E. coli</i>
MelB	Glycoside-pentoside-hexuronide: cation symporter	P30878	Melibiose/ Na^+ symport	Into cell	<i>Salmonella typhimurium</i>

The cluster containing each protein sequence in UniRef50 was downloaded and aligned using MUSCLE (Edgar, 2004b).

There are 104 identified human MFS proteins in the Genomic Transport database (Ren et al., 2004) and most are classified as solute carriers (SLC). Many of these are of interest from a pharmaceutical point of view (Lin et al., 2015). For example, the SLC18 protein is a potential marker for Parkinson's disease (Giboureau et al., 2010) and glutamate transporters are being investigated as drug targets for neuropsychiatric and neurodegenerative diseases (Hinoi et al., 2005). Probenecid is an SLC22 (organic anion transporter) inhibitor whose co-administration is used in cases of nephrotoxicity (Devineni et al., 2015). Furthermore, SV2A is the drug target for levetiracetam, a successful anti-epileptic drug (Klitgaard et al., 1998; Löscher et al., 1998; Lynch et al., 2004).

Although, there are an increasing number of X-ray crystal structures for MFS proteins (Yan, 2016), there are still many examples where there is no structural information. In that case one has to use homology modeling (Kasho et al., 2006) to provide a working model. This can be a powerful tool for drug discovery, but the quality of the model is important in order to have confidence in screening for potential compounds in the binding site. The biggest factor in determining the quality of a model is the sequence alignment. For MFS proteins, sequence identities are very low between families and are said to sit within the so-called “twilight zone” (Rost, 1999). As a consequence achieving a reliable alignment is not trivial and even hidden Markov models (HMMs) struggle in this regard. Nevertheless, good starting models can provide valuable information for the design of experimental strategies, such as site-directed mutagenesis, in order to understand the molecular basis of solute (and co-solute) transport.

One strategy to improve confidence in the alignment is to focus on “anchor” points—residues that are highly conserved or have additional evidence that they are likely to be located at important regions within the protein. Identification of such points can in turn also help to provide a means to compare structures or models from diverse sequences with the same fold. Such an approach was implemented by Ballesteros and Weinstein (and recently updated; Isberg et al., 2015) with respect to GPCRs (see Table S1) and has been shown to aid homology modeling of GPCRs in drug design programs (Zhou et al., 1994; Almaula et al., 1996; Javitch et al., 1998; Pascual et al., 2008).

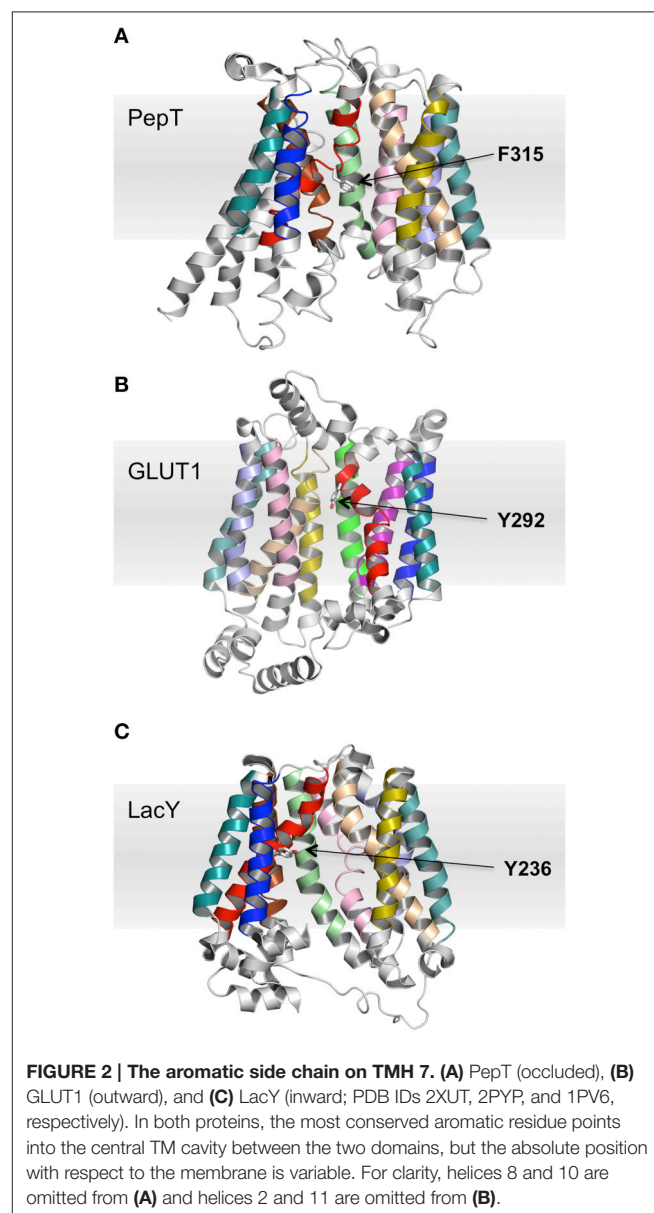
The conservation of residues according to distinct physicochemical properties can be used in multiple sequence alignment (MSA) analysis to guide the position of contacts between helices. Small or hydrophobic residues predominantly mediate these contacts, but hydroxyl groups or aromatic residues can also be present at helix–helix contact points. Buried hydrophobic residues within the TM region that are not solvent accessible tend to be the most highly conserved regions of TM proteins (Eyre et al., 2004). These properties can be used to predict the contacts between TM helices in MFS proteins. Indeed, this approach was utilized for GPCRs to show a network of contacts that are predominantly between TMHs 1–2, 1–7, 3–4, 3–5, and 6–7 (Venkatakrishnan et al., 2013). Defining similar contacts within MFS proteins will help provide insights into which intrahelical interactions are important and which change during the transport cycle.

The vast number of sequences now available, thanks to modern sequencing techniques, allows us to generate MSA for proteins with high sequence similarity to MFS transporters for which the crystal structure is known. These alignments were then compared with close contacts within all the current MFS structures to identify conserved interaction points and thus develop a set of rules to aid homology modeling of MFS transporters.

METHODS

Multiple Sequence Alignment

Since MFS proteins have low sequence identity with each other (Table 1), a MSA of the whole superfamily was not feasible for determining the conservation of residues that relate to



structure (Tramontano, 1998). It is highly unlikely that the alignment would be optimal. It was difficult to define at what point sequences become sufficiently similar to allow us to infer structural information (Rost, 1999). We initially explored the use of Pfam (Finn et al., 2014), but this turned out to be too diverse in terms of clan members, as indeed has been discussed when compared to the Transporter Classification Database (Chiang et al., 2015).

Given that we now have many more sequences available, we thus decided to use the UniRef50 clusters (Suzek et al., 2007) for each protein solved by X-ray crystallography (Table 2). These are sequences that have at least 50% identity and we took the view

that as long as the identity was above 50%, the alignment quality would be sufficient (Baker and Sali, 2001) to provide useful alignments. MSAs were then constructed using the MUSCLE algorithm (Edgar, 2004a) incorporated into a python script.

Analysis of MSA

Conservation at each site was determined using an in-house R script (available upon request) according to the chemical groupings of the amino acids: hydrophobic (M, A, V, I, L, C, Y, F, W), polar (S, T, N, Q), positive charge (R, H, K), negative charge (D, E), aromatic (W, F, Y), glycine (G), or proline (P). Data

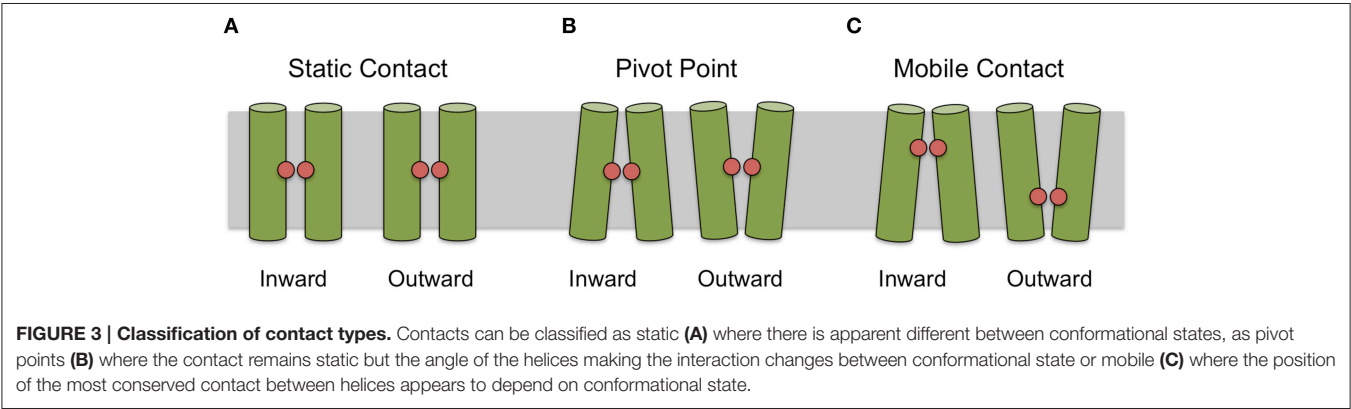


TABLE 3 | Analysis of the position of the most conserved contacts between helices (summarized by a series of heat maps in Figure 5 and Figures S1–S3).

Residue involved in contact	Helix–Helix	Description
STNQ	1–2	Keeps periplasmic end of helices in contact regardless of state
STNQ	2–11	Keeps periplasmic end of helices in contact regardless of state
STNQ	5–8	In the center of the TM region, and state dependent
STNQ	7–11	In the center of the TM region, and state dependent
Small hydrophobic	1–5	Contacts are at a pivot point in the helices, with each conformation having a different tilt to bilayer normal
Small hydrophobic	1–6	The contact is in the middle of the helix and TMH 1 has different tilts for each state
Small hydrophobic	2–4	The contact is in the middle of the helices and small changes in tilt
Small hydrophobic	2–11	The contact is along length of helices in occluded and inward structure but only at cytoplasmic end in outward structure
Small hydrophobic	3–4	Mediates packing
Small hydrophobic	3–6	Mediates packing
Small hydrophobic	5–8	The contacts are in the center of TM region
Small hydrophobic	7–11	In cytoplasmic half of the helices, outward state TMH 7 bends away from TMH 11 at periplasmic end to occluded and inward
Small hydrophobic	8–10	The contacts are along the helices—mediates packing
Small hydrophobic	9–10	The contacts are along the helices—mediates packing
Small hydrophobic	9–12	The contacts are along the helices—mediates packing, all three states have straight helices
Large hydrophobic	2–4	The position of the contact is in the center of the TM region, and the helices rock around that central point
Large hydrophobic	3–6	At the periplasmic end of the TM region. Inward structure has a further contact at cytoplasmic end of the TM region
Large hydrophobic	8–10	The contacts are in the center of the TM region, with a small difference in TMH 10 tilt between the conformations
Glycine	1–5	The contact is in the center of the helices, and is a pivot point for the helices which have different tilts in each state
Glycine	2–4	The contact is in the center of the helices, not much difference in tilt implying a packing motif
Glycine	9–10	The contacts are along the length of the helices, indicating packing mediators

The analysis depicts either static contact points (green) that do not change between different conformational states; contacts that move according to conformational state (blue), and contacts that constitute distinct pivot points (purple) around which the rest of the protein moves.

were visually represented using the heatmap.2 package in R. This analysis was applied to the MFS proteins listed in **Table 2**.

Contact Prediction and Analysis

The contacts between helices were defined by distance (side chain atoms within 7 Å of the adjacent helix backbone atoms). The VMD package (Humphrey et al., 1996) was used to find all contacts within crystal structures based on this definition. We analyzed 44 MFS crystal structures for contacts between TMH pairs (see Figures S1–S3). The resolution of the crystal structures ranges from 1.9 Å (4IKV:POT peptide transporter) to 4.2 Å (4JA4: xylose transporter). Although, higher resolution structures may define side-chain positions more accurately we do not anticipate this to dramatically influence the results of the contact analysis. Heat maps were constructed which showed the presence or absence of the particular residue type contact for all potential helix–helix interactions. We then assessed the location within the bilayer of conserved positions according to the OPM (Orientations of Proteins in Membranes) database (Lomize et al., 2006).

RESULTS AND DISCUSSION

MSA Analysis

Many of the previous observations regarding conserved sites (Pao et al., 1998) are reinforced by the MSA analysis here and we will not go into detail, but just emphasize some of the key points. Two of the conserved sites highlighted by the MSA analysis are the glycine residues on TMH 4 and its symmetrically equivalent helix in domain 2, TMH 10. In every crystal structure these glycine residues are conserved. There is predominantly only one glycine in TMH 10, but there are between one and four conserved glycine residues in TMH 4. A similar pattern continues for all conserved sites common to the helices in domains 1 and 2. This implies that even if the domains arose from an evolutionary repeat (Reddy et al., 2012) it was very distant in time and so the domains have evolved to have considerable dissimilarity between them. Presumably this reflects the need for certain MFS proteins to become specialized and enable them to bind different substrates. Intragenic duplication of two TM domains and subsequent evolution has been argued as the most likely explanation for the variation observed within the MFS proteins (Reddy et al., 2012). Similarly, the conserved aromatic residue at the center of TMH 7 does not exist as a conserved residue in TMH 1. In LacY, Y236 (on TMH 7) is known to be functionally important and so it could be that there is functional evolutionary pressure that maintains the chemical nature of the residue at this position. Mutation of this residue has previously been shown to impair transport (Guan and Kaback, 2006).

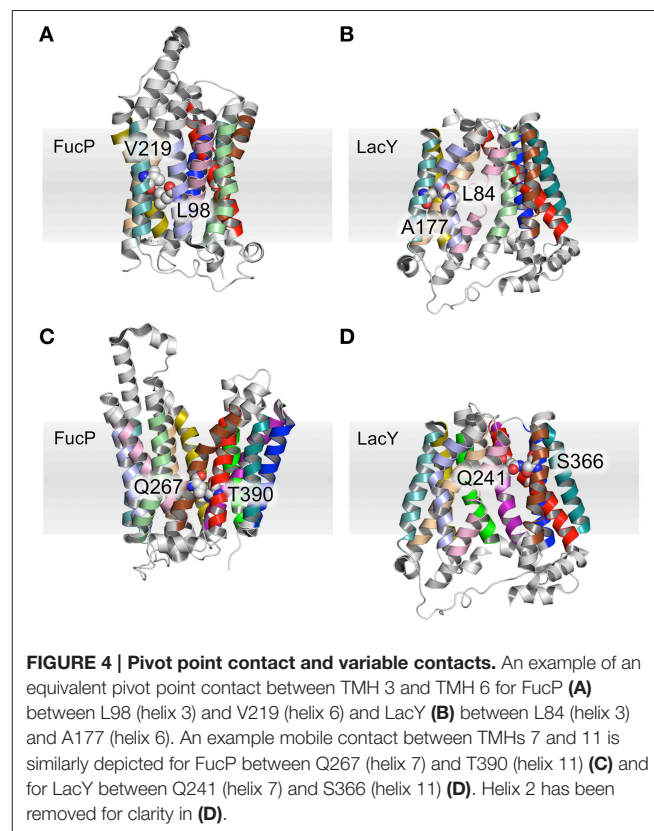
There are several positions that reflect the expected symmetry of the domains. A glycine at the center of TMH 5 is also present in TMH 11 and the positive residue in loop 2–3 is present in loop 8–9. The positive residue in loop 2–3 is part of a conserved motif known to exist in the MFS proteins (Pao et al., 1998) and the positive residue of that motif is mirrored in domain 2 (loop 8–9). Positively charged residues also

exist in loops 4–5 and 10–11, indicating that MFS transporters conform to the positive-inside rule (von Heijne, 1992). A negative charge at the end of TMH 6 is also mirrored in TMH 12. However, its position was not always conserved at the level of the lipid head groups, but instead in loop 6–7 or near the N-terminus.

There are some apparent anomalies. In particular, EmrD does not have the aromatic residue in TMH 7 or the positive charge in loop 10–11. With regards to the conserved aromatic residue on TMH 7, in some proteins there is more than one conserved aromatic residue. In order to identify which is the functionally or structurally relevant residue, the position of these aromatic residues was related to the percentage conservation. Those that are most conserved are also those pointing into the central cavity, as seen in PepT (2XUT) and GLUT1 (4PYP). These residues have similar orientations to the functionally important residue in LacY (Y236; **Figure 2**). A similar issue arises with TMHs 4, 5, 10, and 11, which may contain more than one conserved glycine and so these positions cannot always be used in isolation to aid homology modeling.

To What Extent Are Contacts Conserved?

For some helices it was not possible to identify a uniquely conserved position from sequence analysis alone. Therefore, we manually examined the 44 crystal structures for helix–helix contacts (see Methods Section). The resulting contacts can be broadly classed into three groups: (1) those that mediate the packing of helices and where the relative orientations of the



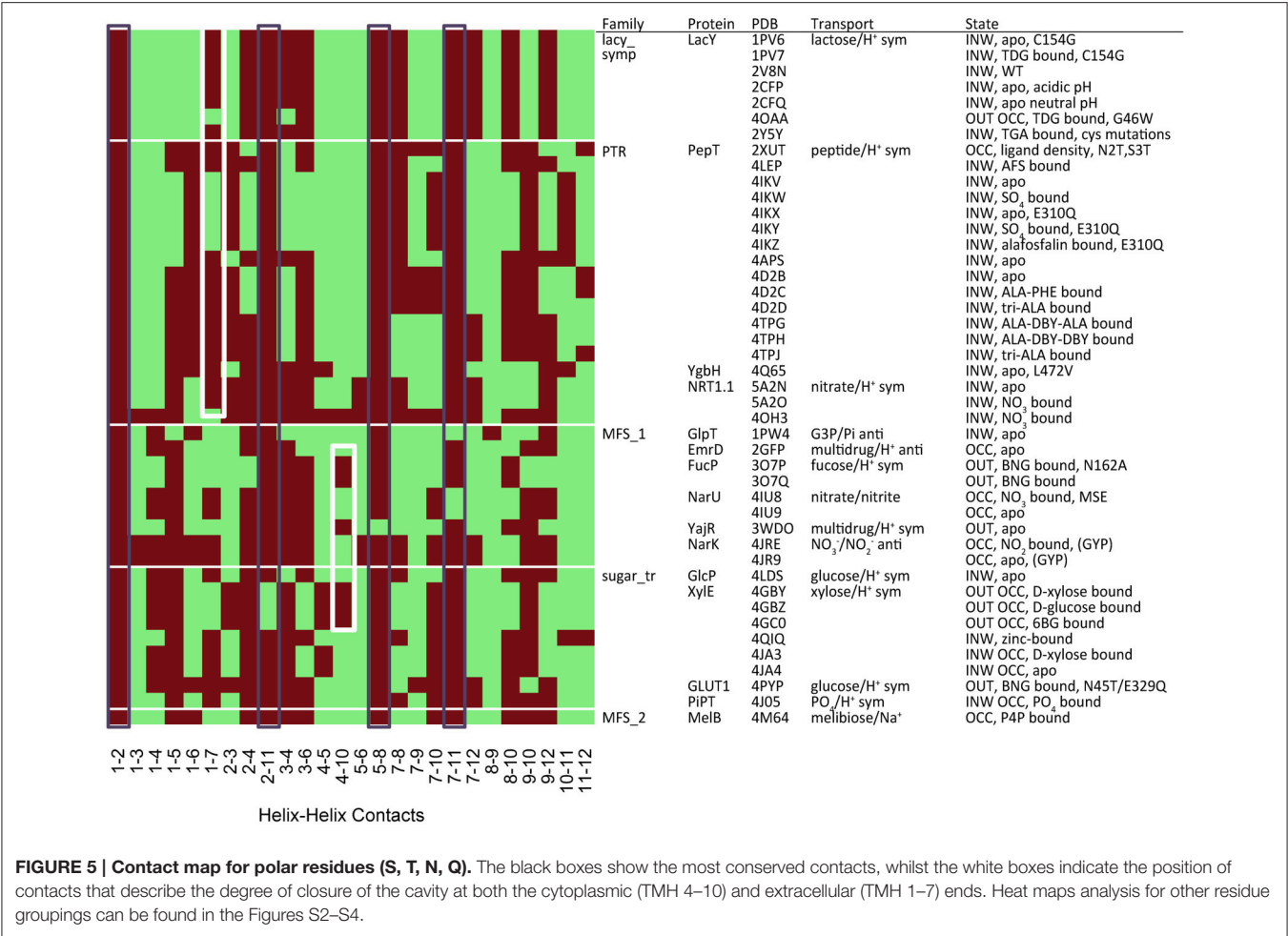


FIGURE 5 | Contact map for polar residues (S, T, N, Q). The black boxes show the most conserved contacts, whilst the white boxes indicate the position of contacts that describe the degree of closure of the cavity at both the cytoplasmic (TMH 4–10) and extracellular (TMH 1–7) ends. Heat maps analysis for other residue groupings can be found in the Figures S2–S4.

TABLE 4 | Summary of proposed numbering scheme.

TM helix	How assigned
1	Contact with TMH 5 (either polar, glycine or small hydrophobic)
2	Contact with TMH 4 (most conserved glycine)
3	Contact with TMH 6 (most conserved small hydrophobic)
4	Contact with TMH 2 (most conserved glycine)
5	Contact with TMH 1 (either polar, glycine or small hydrophobic)
6	Contact with TMH 3 (most conserved small hydrophobic)
7	Contact with TMH 11 (most conserved polar)
8	Contact with TMH 5 (most conserved polar or small hydrophobic)
9	Contact with TMH 10 (most conserved glycine or small hydrophobic if glycine not present in sequence)
10	Contact with TMH 9 (most conserved glycine)
11	Contact with TMH 7 (most conserved polar residue)
12	Contact with TMH 9 (most conserved small hydrophobic)

In TMHs 4, 5, and 10 these correspond with the conserved glycine motifs present on those helices. In other TMHs, the residues are mostly in the same spatial location on the helix, but sometimes the position changes and thus they are easier to track via their conserved contacts rather than conservation alone.

helices is invariant; (2) those where the contact is conserved but the orientation of the helices changes according to different conformational states (we refer to these as pivot points); and (3) those whose position exhibits some apparent dependence on the conformational state of the protein. For example a particular contact may only be conserved in one particular conformational state (such as the inward open for example) and a different contact between those two helices may be present in an alternative conformational state. **Figure 3** summarizes these different classes of contacts schematically.

Table 3 summarizes the contacts conserved across all 44 structures. Helix 2 makes a contact with helices 1 and 11 at the cytoplasmic side of the membrane that is conserved regardless of conformational state and thus may be particularly useful in improving alignments in the future. Four scaffold residues (small hydrophobic) have previously been described in the literature (Doki et al., 2013; Yaffe et al., 2013). The contacts of these residues also remain unchanged across the conformations and they are therefore classified as pivot points. From our analysis, it is apparent that these mediate contacts between TMHs 3–6 and 9–12 (**Table 3, Figures 4A,B**).

TABLE 5 | The numbering in MFS proteins.

Protein	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
XylE (4JA4)	17–33 G25 G1.0	57–74 G71 G2.0	87–104 A92 A3.0	126–145 G141 G4.0	166–183 G174 G5.0	201–218 A210 A6.0	281–301 S285 S7.0	316–333 T329 T8.0	343–361 G348 G9.0	371–390 G388 G10.0	410–427 Q415 G11.0	443–463 A456 G12.0
PepT (2XUT)	18–39 A22 A1.0	53–75 G66 G2.0	86–105 L93 L3.0	110–129 G124 G4.0	153–172 G160 G5.0	178–195 F189 F6.0	304–324 T309 T7.0	335–352 A342 A8.0	376–393 G379 G9.0	404–424 G418 G10.0	444–464 S449 S11.0	483–500 A489 A12.0
LacY (1PV6)	10–35 G13 G1.0	41–64 F55 F2.0	75–96 L84 L3.0	103–128 G111 G4.0	143–163 G147 G5.0	167–187 A177 A6.0	222–248 Q248 Q7.0	257–278 T266 T8.0	289–309 G296 G9.0	312–332 G332 G10.0	352–373 S366 S11.0	380–399 L390 L12.0
FucP (3O7Q)	28–47 N43 N1.0	64–86 G73 G2.0	90–108 L98 L3.0	119–140 G132 G4.0	154–174 G165 G5.0	210–229 V219 V6.0	261–282 Q267 Q7.0	301–319 V306 V8.0	326–345 L328 L9.0	348–372 G372 G10.0	383–403 T390 T11.0	412–430 A419 A12.0
GLUT1 (4PYP)	14–36 G27 G1.0	64–91 G79 G2.0	94–112 A103 A3.0	119–140 G134 G4.0	157–176 G167 G5.0	187–206 A197 A6.0	275–295 Q283 Q7.0	306–326 I315 I8.0	335–354 G340 G9.0	366–387 G382 G10.0	402–426 N415 S11.0	431–449 L441 L12.0
EmrD (2GFP)	11–31 Q21 Q1.0	43–64 T55 T2.0	73–92 L77 L3.0	97–116 G109 G4.0	134–155 G140 G5.0	157–175 L169 L6.0	208–229 N210 N7.0	237–261 S246 S8.0	267–283 L279 L9.0	289–306 G295 G10.0	326–345 T334 T11.0	357–378 L374 L12.0
GlcP (4LDS)	7–31 G20 G1.0	41–64 G54 G2.0	74–92 I83 I3.0	95–115 G106 G4.0	134–154 G143 G5.0	159–178 V168 V6.0	244–266 N256 N7.0	275–295 N287 N8.0	305–325 G310 G9.0	335–355 G349 G10.0	376–395 S388 S11.0	400–418 A408 A12.0
GlpT (1PW4)	32–51 N47 N1.0	65–84 G77 G2.0	93–110 L100 L3.0	121–139 G131 G4.0	159–178 G168 G5.0	190–208 A197 A6.0	255–277 N262 N7.0	292–311 T306 T8.0	322–340 G325 G9.0	349–369 G363 G10.0	386–405 T388 T11.0	416–435 L431 L12.0
MelB (4M64)	11–29 G23 G1.0	46–63 W54 W2.0	79–98 L91 L3.0	107–129 G117 G4.0	147–169 G156 G5.0	178–196 L186 L6.0	233–252 N244 N7.0	272–291 N279 N8.0	295–312 G301 G9.0	332–347 G337 G10.0	368–385 T373 T11.0	415–432 L420 L12.0
NarU (4IU8)	37–56 L44 L1.0	73–91 G83 G2.0	101–119 L110 L3.0	130–147 G139 G4.0	167–187 G172 G5.0	211–229 V218 V6.0	254–277 S258 S7.0	290–307 S304 S8.0	316–334 N322 N9.0	347–367 G362 G10.0	405–423 S408 S11.0	431–454 V447 V12.0
NRT1.1 (5A2N)	38–57 G50 G1.0	70–91 G88 G2.0	100–119 A110 A3.0	147–166 G161 G4.0	193–213 G200 G5.0	218–237 L232 L6.0	342–364 Q358 Q7.0	381–399 S383 S8.0	421–439 G426 G9.0	462–481 G475 G10.0	500–520 S514 S11.0	542–561 L554 L12.0
PiPT (4J05)	39–56 A46 A1.0	74–92 G84 G2.0	106–123 I113 I3.0	133–151 G143 G4.0	173–195 G180 G5.0	209–227 A222 A6.0	313–333 N333 N7.0	358–377 N361 N8.0	385–404 G392 G9.0	415–436 G429 G10.0	452–469 N470 N11.0	481–499 L482 L12.0
YajR (3WDO)	14–33 L26 L1.0	51–69 G58 G2.0	79–98 L85 L3.0	102–119 G112 G4.0	136–155 G141 G5.0	165–183 I179 I6.0	215–234 N218 N7.0	250–268 V264 V8.0	279–298 G286 G9.0	303–325 G308 G10.0	340–361 S346 S11.0	368–386 A378 A12.0
YgbH (4Q65)	13–33 G27 G1.0	49–67 G65 G2.0	78–95 A83 A3.0	101–120 G115 G4.0	142–161 G150 G5.0	169–187 L182 L6.0	267–289 Q285 Q7.0	311–331 S315 S8.0	342–362 G351 G9.0	379–400 G388 G10.0	414–433 N427 N11.0	461–480 V470 V12.0

TM helices were taken as those predicted by the OPM database (Lomize et al., 2006). The x.0 residue of each helix is summarized in **Table 4**. The helix is numbered x.0, where x is the helix number and it is prefixed by the conserved residue, for example in TM helix 5, the most conserved residue is a glycine, therefore it is G5.0. Residues from this residue toward the cytoplasm are labeled as negative numbers and residues toward the periplasm are given positive numbers (for example TM helix 5 in XylE is FNQFAIFGQLLVYCVNY giving: ..., I5.-2, F5.-1, G5.0, Q5.1, L5.2, ...). The proposed numbering start points are indicated in bold for each helix from each protein.

A further 12 interactions were identified that provide either pivot or moving (according to the conformational states) helix-helix contacts (Table 3, and exemplified in Figures 4C,D). The two inter-domain contacts are mediated by interactions between TMHs 2–11 and 5–8. It is expected that these would have varied contacts according to the conformational state. This is the case for TMHs 5–8 where the helices move around two pivot point contacts formed by polar and small hydrophobic residues.

However, the nature of interactions between TMHs 2–11 is less clear-cut. Our analysis suggests that the small hydrophobic contact points are conformationally dependent, but that the polar contacts would be classified as invariant. We note here that the data is skewed toward occluded and inward open conformations (only two structures have on outward open state: FucP and GLUT1) and therefore caution should be applied in the interpretation here.

Certainly the structures so far suggest that the largest conformational change is from outward to occluded, rather than from occluded to inward. If this is a genuine trend, this implies that the motion of the TMHs 2, 5, 8, and 11 is less important to the transport cycle than helices: 1, 4, 7, and 10. Indeed, TMHs 7 and 10 in the PepT_{So} transporter have previously been highlighted as being dynamic (Newstead et al., 2011). In addition Doki et al. (2013) reported TMH 4 as being dynamic in the GkPOT transporter.

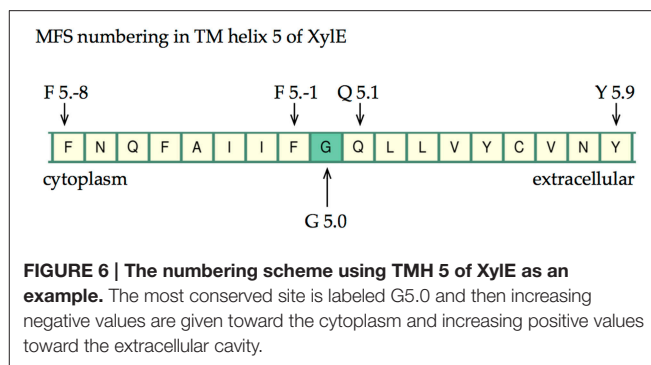
All the remaining conserved contacts (Table 3) are intra-domain (i.e., within TMHs 1–6 or within TMHs 7–12) and reflect the movements that the helices undergo in the transport cycle. In domain 1, there are two hydrophobic contact points made by TMH 1: TMHs 1–5 and TMHs 1–6. The former acts as a pivot point, but the latter shows variation according to the conformational state.

For the equivalent TMH in domain 2, (TMH 7), there is no 7–12 contact point (at least not as conserved as the 1–6 contact), but there are two TMH 7–11 contacts mediated by polar and small hydrophobic residues (Table 3). Both of these contact points are conformationally dependent though.

In terms of the conformationally dependent contacts, those formed by TMH 1 or TMH 7 exhibit the largest variations. A prediction that arises from the triple-helix repeat model (Radestock and Forrest, 2011) is that a similar variation would occur for TMHs 4 and 10, but this was not evident from our contact analysis.

There are nine contacts that we have assigned as static (Table 3). That is to say these contacts remain in the same position, regardless of the conformational state of the protein. They are predominantly between TMHs 3 and 4, TMHs 3 and 6, TMHs 8 and 10, TMHs 9 and 10, and TMHs 9 and 12. The invariant nature of these contacts, despite differing conformations, renders them as useful anchor points for supporting MFS homology modeling. Since the contacts remain in the same position in all MFS proteins, they serve as a guide to predict the orientation of helices in MFS targets through sequence alignment.

An important caveat to note concerns the way in which the helices move in the transport cycle. The theory is that



the domains could move through a rigid body rotation (Shi, 2013) and therefore there ought not to be too many movements in the relative position of helices within domains. Conversely, the inter-domain contacts should move as the domains rock against each other. However, our analysis shows that small movements occur in the contacts between TMHs 2 and 11 or between TMHs 5 and 8. This suggests that there are some independent rearrangements of the helices on top of the predicted rigid body movement of domains.

There are two metrics that describe the degree to which either side of the transporter is open. When TMHs 1 and 7 interact, the periplasmic side of the TM cavity is closed and when TMHs 4 and 10 interact, the cytoplasmic side is closed. Therefore, we used these residues to describe the degree of closure of either side of the TM cavity. It is possible to see such contacts in the heat-map of polar residue contacts (Figure 5, white boxes), where the TMH 4–10 contact is present in the outward open X-ray crystal structures and the TMH 1–7 contact is present in the inward open structures. Heat maps for other contact types are shown in Figures S1–S3.

A Suggested Numbering System to Help MFS Modeling

Considering the identification of conserved residues alongside the identification of conserved contact points, we postulated whether it would be possible to use this information to provide common reference points within the MFS topology that could be useful for guiding homology modeling of these proteins. Furthermore, it should facilitate structural comparison across the entire family. Inspired by the Ballesteros and Weinstein approach to GPCRs (Table S1), we thus devised a set of rules that allow cross-family conserved sites to be compared. The most conserved positions are defined as x.0, where x is the TMH (Marger and Saier, 1993; Pazdernik et al., 1997; Pao et al., 1998; Burckhardt and Wolff, 2000; Saier and Paulsen, 2001; Guan and Kaback, 2006; Newstead et al., 2011; Radestock and Forrest, 2011; Reddy et al., 2012; Keller et al., 2014; Saier et al., 2014; Yan, 2015). In cases where conserved residues across the superfamily do not exist (i.e., not present in Table 3, Tables S2–S15), then the conserved contact points are used to define position 0.

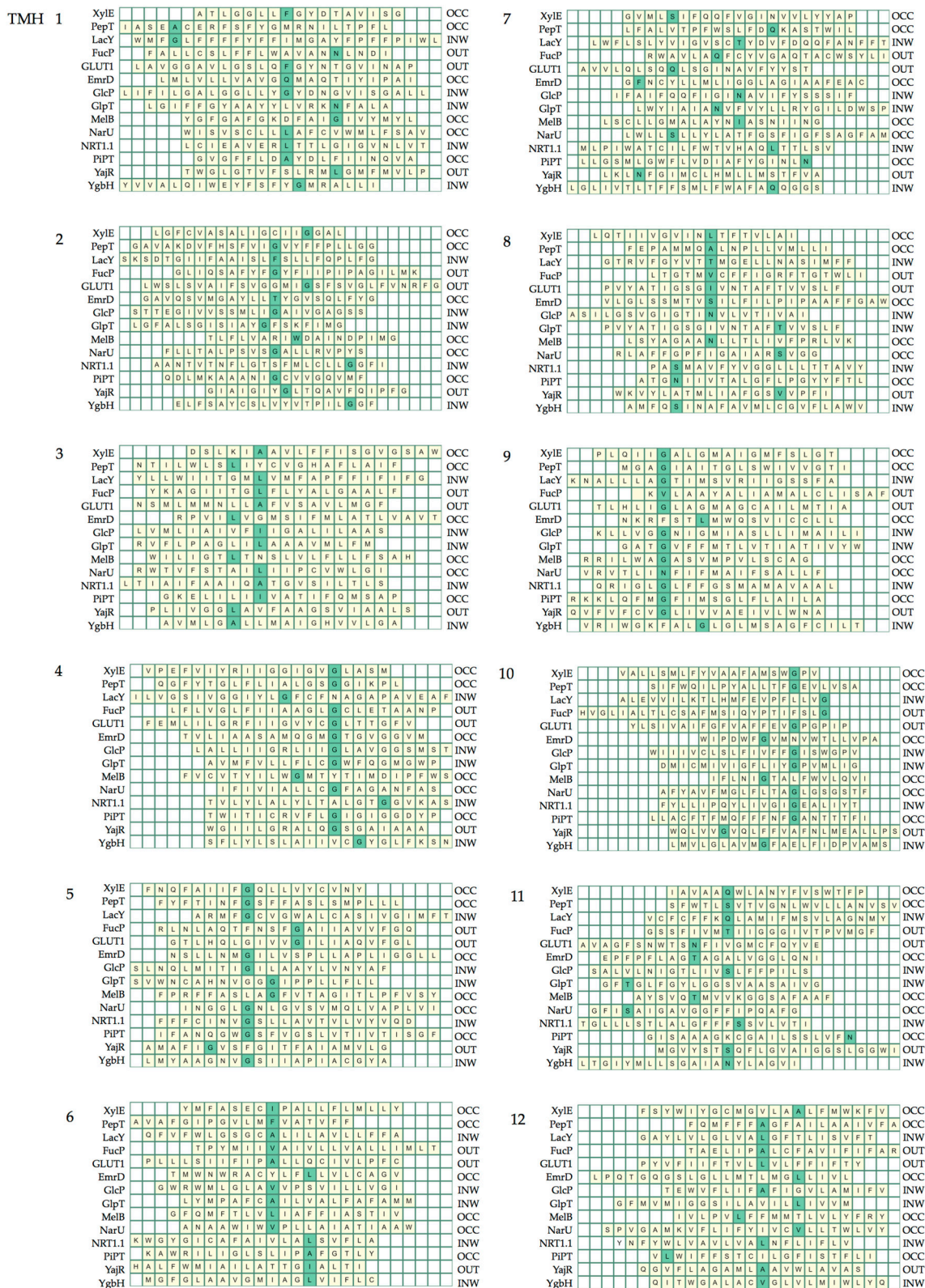


FIGURE 7 | The position of the numbered residue on each helix. Each line of the heat maps corresponds to the helix in each crystal structure. The helix lengths are defined by those within the region of the bilayer given in the OPM database and helices are different lengths depending on the helix tilt and whether there are kinks. The helices are aligned according to the rotation of the helix such that numbered sites in the same structural position are aligned.

Rules for Numbering

- (1) Residues that are conserved across all MFS families are assigned as x.0. This defines x.0 for TMHs 4, 5, and 10 via conserved glycine residues.
- (2) The conserved glycine residues in TMHs 4, 5, and 10 are also involved in conserved contacts with TMHs 2, 1, and 9, respectively (see **Tables 3, 4** and Tables S2–S15). The second residue involved in these contacts (i.e., on the opposing helix) was used for numbering in TMHs 1, 2, and 9. Positions on TMHs 1, 2, and 9 are defined this way.
- (3) For the remaining helices, the conserved superfamily-wide contacts were used (**Table 3** and Tables S2–S15). The x.0 position was defined as the most conserved residue within a contact.

Table 5 summarizes the assignment. Note that the conserved contact approach actually recapitulates the identification of the conserved residues in helices 4, 5, and 10. Once x.0 is determined in a helix, the numbering will increase from 0 when moving toward the extracellular side of the TM region and will decrease from 0 when moving to the cytoplasmic side of the TM region (an example is shown in TM helix 5 of XylE in **Figure 6**). The majority of x.0 sites are conserved glycine residues as identified in TMHs 4, 5, and 10 or the subsequent residue involved in the conserved contact for these glycine residues in TMHs 2, 1, and 9, respectively. The conserved glycine in TMH 11 was not suitable to use for the numbering because the position moved according to the crystal structure investigated. Therefore, 11.0 corresponded to the contact between TMHs 7 and 11, which is a polar-polar contact. The numbering for all helices from all the structures considered is shown in **Table 4** and illustrated in **Figure 7**. For each helix, the x.0 position is in structurally similar positions when compared to the orientation of helices in the X-ray crystal structures.

A couple of interesting observations arise upon further inspection of known MFS crystal structures. The first is that the conserved aromatic residue on TMH 7 of XylE (Y7.0) is closer to the periplasmic side of the helix compared with LacY, PepT, FucP, and GLUT1. This is most likely because it is not involved in a critical contact. The second is that position 1.0 in PepT and LacY, differs in location when compared to XylE, FucP, and GLUT1. This can be accounted for by the flexibility of TMH 1 in the transport mechanism that enables the contact points to change according to conformation.

From our analysis, it appears that there is less symmetry than might be expected in the contacts between the two 6-TMH domains. Whilst, the TMH 2 and 4 contact defines the numbering in domain one, the numbering of TMH 8 comes from an interaction with TMH 5. Similarly, it is a contact with TMH 10 that defines which residue is the anchor point in TMH 9, whereas in domain one, a contact between TMHs 3 and 6 defines the numbering in both those helices. Whilst this does

not disagree with the inverted repeat topology of Radestock and Forrest (2011), since that investigates the conservation between whole helices rather than single residues on them, it does imply a subtlety in the conservation of contacts. It is impossible to say whether this is directly caused by evolutionary pressure, but it does pose the question of whether there is a slightly different role for each domain in the alternating access mechanism.

CONCLUSIONS

Alignment of MFS family proteins is made difficult by their low sequence similarity, which in turn makes homology modeling difficult. The problem is compounded by the fact that many structures now exist in different states. In this work we have tried to ascertain which positions may be structurally common via contact analysis of the structures. By combining contact analysis with conservation analysis we have suggested a way to identify “anchor” points on each of the TMHs that should aid the modeling process. We have found that the majority of contacts remain static across the different conformational states. Our analysis would suggest that these contact positions would be particularly intolerant to mutation, but a systematic study would be required to fully address that. There are only small and helix specific (TMHs 1, 7, and 10 predominantly) rearrangements that take place on top of the rigid body rotation of each domain. To facilitate cross-family comparison we have also devised a numbering scheme, similar in essence to that proposed for GPCRs. We anticipate that by exploiting this analysis, homology modeling of MFS proteins should be improved.

AVAILABILITY OF SOFTWARE

The R-script used to perform the analysis is available on request from the authors.

AUTHOR CONTRIBUTIONS

PB and ZS designed the research. JL performed the research and developed the computational methods. JL, ZS, and PB wrote the manuscript. All authors approved the final version and accept accountability for its accuracy.

FUNDING

JL is a BBSRC-funded (BB/F01709X/1) student in receipt of additional financial support from UCB BioPharma SPRL.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00021>

REFERENCES

- Almula, N., Ebersole, B. J., Ballesteros, J. A., Weinstein, H., and Sealfon, S. C. (1996). Contribution of a helix 5 locus to selectivity of hallucinogenic and nonhallucinogenic ligands for the human 5-hydroxytryptamine_{2A} and 5-hydroxytryptamine_{2C} receptors: direct and indirect effects on ligand affinity mediated by the same locus. *Mol. Pharmacol.* 50, 34–42.
- Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. *Science* 294, 93–96. doi: 10.1126/science.1065659
- Burckhardt, G., and Wolff, N. A. (2000). Structure of renal organic anion and cation transporters. *Am. J. Physiol. Renal Physiol.* 278, F853–F866.
- Chiang, Z., Vastermark, A., Punta, M., Coghill, P. C., Mistry, J., Finn, R. D., et al. (2015). The complexity, challenges and benefits of comparing two transporter classification systems in TCDB and Pfam. *Brief. Bioinform.* 16, 865–872. doi: 10.1093/bib/bbu053
- Devenini, D., Vaccaro, N., Murphy, J., Curtin, C., Mamidi, R. N. V. S., Weiner, S., et al. (2015). Effects of rifampin, cyclosporine A, and probenecid on the pharmacokinetic profile of canagliflozin, a sodium glucose co-transporter 2 inhibitor, in healthy participants. *Int. J. Clin. Pharmacol. Ther.* 53, 115–128. doi: 10.5414/CP202158
- Doki, S., Kato, H. E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., et al. (2013). Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11343–11348. doi: 10.1073/pnas.1301079110
- Edgar, R. C. (2004a). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C. (2004b). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Eraly, S. (2008). Implications of the alternating access model for organic anion transporter kinetics. *J. Membr. Biol.* 226, 35–42. doi: 10.1007/s00232-008-9137-1
- Eyre, T. A., Partridge, L., and Thornton, J. M. (2004). Computational analysis of α -helical membrane protein structure: implications for the prediction of 3D structural models. *Prot. Eng. Des. Sel.* 17, 613–624. doi: 10.1093/protein/gzh072
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucl. Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Giboureau, N., Som, I. M., Boucher-Arnold, A., Guilloteau, D., and Kassiou, M. (2010). PET radioligands for the Vesicular Acetylcholine Transporter (VACHT). *Curr. Top. Med. Chem.* 10, 1569–1583. doi: 10.2174/156802610793176846
- Guan, L., and Kaback, H. R. (2006). Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* 35, 67–91. doi: 10.1146/annurev.biophys.35.040405.102005
- Hinoi, E., Takarada, T., Tsuchihashi, Y., and Yoneda, Y. (2005). Glutamate transporters as drug targets. *Curr. Drug Targets CNS Neurol. Disord.* 4, 211–220. doi: 10.2174/1568007053544093
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD - Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F. H., et al. (2015). Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharm. Sci.* 36, 22–31. doi: 10.1016/j.tips.2014.11.001
- Javitch, J. A., Ballesteros, J. A., Weinstein, H., and Chen, J. (1998). A cluster of aromatic residues in the sixth membrane-spanning segment of the dopamine D₂ receptor is accessible in the binding-site crevice. *Biochemistry* 37, 998–1006. doi: 10.1021/bi972241y
- Kaback, H. R., Smirnova, I., Kasho, V., Nie, Y., and Zhou, Y. (2011). The alternating access transport mechanism in LacY. *J. Membr. Biol.* 239, 85–93. doi: 10.1007/s00232-010-9327-5
- Kasho, V. N., Smirnova, I. N., and Kaback, H. R. (2006). Sequence alignment and homology threading reveals prokaryotic and eukaryotic proteins similar to lactose permease. *J. Mol. Biol.* 358, 1060–1070. doi: 10.1016/j.jmb.2006.02.049
- Keller, R., Ziegler, C., and Schneider, D. (2014). When two turn into one: evolution of membrane transporters from half modules. *Biol. Chem.* 395, 1379–1388. doi: 10.1515/hsz-2014-0224
- Klitgaard, H., Matagne, A., Gobert, J., and Wülfert, E. (1998). Evidence for a unique profile of levetiracetam in rodent models of seizures and epilepsy. *Eur. J. Pharmacol.* 353, 191–206. doi: 10.1016/S0014-2999(98)00410-5
- Lin, L., Yee, S. W., Kim, R. B., and Giacomini, K. M. (2015). SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug. Discov.* 14, 543–560. doi: 10.1038/nrd4626
- Lomize, M. A., Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2006). OPM: Orientations of prototins in membranes database. *Bioinformatics* 22, 623–625. doi: 10.1093/bioinformatics/btk023
- Löscher, W., Hönack, D., and Rundfeldt, C. (1998). Antiepileptogenic effects of the novel anticonvulsant levetiracetam (ucb 1059) in the kindling model of temporal lobe epilepsy. *J. Pharmacol. Exp. Ther.* 284, 474–479.
- Lynch, B. A., Lambeng, N., Nocka, K., Kensel-Hammes, P., Bajjalieh, S. M., Matagne, A., et al. (2004). The synaptic vesicle protein SV2A is the binding site for the antiepileptic drug levetiracetam. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9861–9866. doi: 10.1073/pnas.0308208101
- Madej, M. G. (2015). “Comparative sequence–function analysis of the major facilitator superfamily: the mix-and-match method, Chap. 24,” in *Methods Enzymol*, ed K. S. Arun (Academic Press) 557, 521–549. doi: 10.1016/bs.mie.2014.12.015
- Madej, M. G., Dang, S., Yan, N., and Kaback, H. R. (2013). Evolutionary mix-and-match with MFS transporters. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5870–5874. doi: 10.1073/pnas.1303538110
- Madej, M. G., and Kaback, H. R. (2013). Evolutionary mix-and-match with MFS transporters II. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4831–E4838. doi: 10.1073/pnas.1303538110
- Madej, M. G., Sun, L., Yan, N., and Kaback, H. R. (2014). Functional architecture of MFS d-glucose transporters. *Proc. Natl. Acad. Sci. U.S.A.* 111, E719–E727. doi: 10.1073/pnas.1400336111
- Marger, M. D., and Saier, M. H. (1993). A major superfamily of transmembrane facilitators that catalyze uniport, symport and antiport. *Trends Biochem. Sci.* 18, 13–20. doi: 10.1016/0968-0004(93)90081-W
- Masureel, M., Martens, C., Stein, R. A., Mishra, S., Ruyschaert, J.-M., McHaourab, H. S., et al. (2014). Protonation drives the conformational switch in the multidrug transporter LmrP. *Nat. Chem. Biol.* 10, 149–155. doi: 10.1038/nchembio.1408
- Newstead, S., Drew, D., Cameron, A. D., Postis, V. L. G., Xia, X., Fowler, P. W., et al. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* 30, 417–426. doi: 10.1038/emboj.2010.309
- Pao, S. S., Paulsen, I. T., and Saier, M. H. (1998). Major Facilitator Superfamily. *Microbiol. Mol. Biol. Rev.* 62, 1–34.
- Pascual, J. M., Wang, D., Yang, R., Shi, L., Yang, H., and De Vivo, D. C. (2008). Structural signatures and membrane helix 4 in GLUT1: Inferences from human blood-brain glucose transport mutants. *J. Biol. Chem.* 283, 16732–16742. doi: 10.1074/jbc.M801403200
- Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microb. Rev.* 60, 575–608.
- Pazdernik, N. J., Cain, S. M., and Brooker, R. J. (1997). An analysis of suppressor mutations suggests that the two halves of the lactose permease function in a symmetrical manner. *J. Biol. Chem.* 272, 26110–26116. doi: 10.1074/jbc.272.42.26110
- Radestock, S., and Forrest, L. R. (2011). The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* 407, 698–715. doi: 10.1016/j.jmb.2011.02.008
- Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I., and Saier, M. H. (2012). The major facilitator superfamily (MFS) revisited. *FEBS J.* 279, 2022–2035. doi: 10.1111/j.1742-4658.2012.08588.x
- Ren, Q., Kang, K. H., and Paulsen, I. T. (2004). TransportDB: a relational database of cellular membrane transport systems. *Nucl. Acids Res.* 32, D284–D288. doi: 10.1093/nar/gkh016
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Prot. Eng.* 12, 85–94.
- Saier, M. H. Jr., and Paulsen, I. T. (2001). Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* 12, 205–213. doi: 10.1006/scdb.2000.0246
- Saier, M. H. Jr., Reddy, V. S., Tamang, D. G., and Vastermark, Å. (2014). The transporter classification database. *Nucl. Acids Res.* 42, D251–D258. doi: 10.1093/nar/gkt1097

- Shi, Y. (2013). Common folds and transport mechanisms of secondary active transporters. *Annu. Rev. Biophys.* 42, 51–72. doi: 10.1146/annurev-biophys-083012-130429
- Smirnova, I., Kasho, V., and Kaback, H. R. (2011). Lactose permease and the alternating access mechanism. *Biochemistry* 50, 9684–9693. doi: 10.1021/bi2014294
- Smirnova, I., Kasho, V., and Kaback, H. R. (2014). Real-time conformational changes in LacY. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8440–8445. doi: 10.1073/pnas.1408374111
- Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi: 10.1093/bioinformatics/btm098
- Tramontano, A. (1998). Homology modeling with low sequence identity. *Methods* 14, 293–300. doi: 10.1006/meth.1998.0585
- Västermark, Å., and Saier, M. H. (2014). Major Facilitator Superfamily (MFS) evolved without 3-transmembrane segment unit rearrangements. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1162–E1163. doi: 10.1073/pnas.1400016111
- Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., and Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature* 494, 185–194. doi: 10.1038/nature11896
- von Heijne, G. V. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive inside rule. *J. Mol. Biol.* 225, 487–494. doi: 10.1016/0022-2836(92)90934-C
- Yaffe, D., Radestock, S., Shuster, Y., Forrest, L. R., and Schuldiner, S. (2013). Identification of molecular hinge points mediating alternating access in the vesicular monoamine transporter VMAT2. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1332–E1341. doi: 10.1073/pnas.1220497110
- Yan, N. (2015). Structural biology of the major facilitator superfamily transporters. *Annu. Rev. Biophys.* 44, 257–283. doi: 10.1146/annurev-biophys-060414-033901
- Yan, N. (2016). Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* 38, 151–159. doi: 10.1016/j.tibs.2013.01.003
- Zhou, W., Flanagan, C., Ballesteros, J. A., Konvicka, K., Davidson, J. S., Weinstein, H., et al. (1994). A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol. Pharmacol.* 45, 165–170.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ZS is an employee of UCB BioPharma SPRL.

Copyright © 2016 Lee, Sands and Biggin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.