MFS Transporter Superfamily: Modelling and Dynamics

Joanna Lee

Structural Bioinformatics and Computational Biochemistry, The Department of Biochemistry and Lady Margaret Hall, Oxford

TRINITY 2015

A thesis submitted for the partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oxford.

Acknowledgements

Many thanks to both Dr Phil Biggin and Dr Zara Sands, #BRITISHCLUB and to everyone else who supported me. Especially Dr. DR Shorthouse.

Abbreviations

MFS – Major Facilitator Superfamily

MSA - Multiple Sequence Alignment

SV2A – Synaptic Vesicle 2A

TMH – Transmembrane Helix

TM - Transmembrane

MD – Molecular Dynamics

CV – Collective Variable

SV – Synaptic Vesicle

AED – Anti-Epileptic Drug

LEV - levetiracetam

Table of Contents

List of Figures

List of Tables

Abstract

The aims for this thesis were threefold. First, is it possible to develop a numbering protocol to aid homology modelling, despite the diversity of MFS sequences? Given these methods, apply these to improve the homology model of a putative MFS protein, SV2A. Finally, can Molecular Dynamics (MD) simulations be used to elucidate the transport mechanism in an MFS protein?

There are several solved X-ray crystal structures of secondary active transporter MFS proteins, all of which have a core 12 transmembrane (TM) helix fold. Despite this highly conserved fold, the sequence identity between the transporters is low (15 – 25 % identity between structures) and so this provides difficulties when exploring homology modelling of target proteins based on known structures of MFS transporters. To overcome this, exploration of both conservation of amino acid type in multiple sequence alignments (MSAs) and conservation of contacts between helices elucidate 'anchor points' in the structure. This analysis can be used to aid alignment between template and target in homology modelling, such as for SV2A, which is the binding site for an anti-epileptic drug, levetiracetam. The binding site was thought to be in the central cavity of the TM region of the MFS protein and so drug-protein interactions were explored using MD simulations.

Finally, MD simulations were conducted on FucP, the X-ray crystal structure of which is solved in an outward open conformation (open to the perplasmic cavity). The aim for this work was to investigate the movement from the outward open state to the occluded state and determine whether there is any effect on placing static hydrogen atoms on two key titrable residues in the central TM cavity, D46 and E135.

1 Introduction

The Major Facilitator Superfamily (MFS) is the largest superfamily of secondary active transporters. The two 6-transmembrane helix (TMH) domains transport small molecules and drugs across membranes using the alternating access mechanism. Here the TM cavity is alternatively open to either extracellular or cytoplasmic side of the lipid bilayer. Therefore small molecules are able to bind and unbind to the transporter on one side of the bilayer at any given time. For symport and antiport MFS transporters, the ligand is transported alongside a co-transported charged species. The canonical MFS proteins are ligand/H^+ symporters, such as LacY, which co-transport a proton with the ligand. For these proteins, the presence of charged side chains in the TM cavity are implicated in the proton translocation pathway.

1.1 The Major Facilitator Superfamily

As the largest known superfamily of secondary active transporters, the Major Facilitator Superfamily (MFS) has expanded in recent years since its definition by Marger et al. (Marger & Saier 1993), now containing 74 families according to the Transporter Classification Database (TCDB) (Saier, Reddy, Tamang, & Västermark, 2014). These transporters operate by uniport, symport or antiport mechanisms, which 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 transport cycle (Pao et al. 1998). The ligand is incredibly diverse across the

superfamily, ranging from simple monosaccharides, drugs, enzyme cofactors, peptides to vitamins, oligosaccharides, nucleotides, iron chelates and organic and inorganic anions and cations (Saier & Paulsen 2001; Newstead 2011; Guan & Kaback 2006; Shayeghi et al. 2005; Burckhardt & Wolff 2000).

1.1.1 The structure of MFS proteins

With many X-ray crystal structures solved for MFS transporters (44 in April 2015), it is established that MFS proteins have a conserved 12 transmembrane (TM) α-helix fold shown in Figure 1.1 (A) (Reddy et al. 2012) that is formed of two 6-TM helix bundles which are related by a pseudo two-fold symmetry (Figure 1.1 (B)). The presence of both domains is thought to be functionally important for the transport mechanism, as the ligand binds to a central TM cavity on the interface of the two domains (Pazdernik et al. 1997). Deviation from the 12 TM helices tends to be the presence of additional helices, for example peptide transporters contain two additional helices (Newstead et al. 2011).

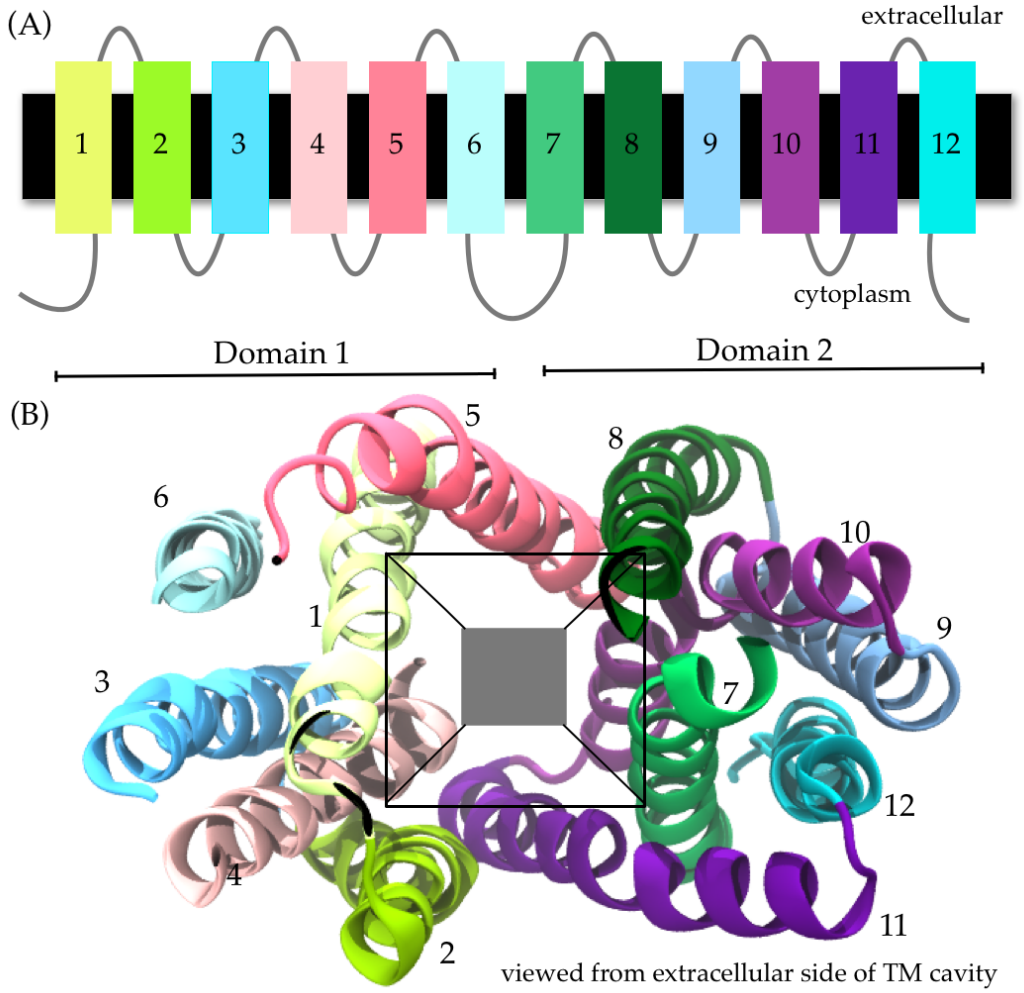


Figure 1.1 The Major Facilitator Superfamily secondary and tertiary structure. (A) The topology of the 12 TM helices. (B) The MFS fold as viewed from the extracellular side of the membrane. The example protein is FucP. The conserved 12 TM α-helix fold is arranged into two 6 TMH bundles and between the two domains a cavity exists (grey box), which is accessible to either the cytoplasm or extracellular region. Predicting the packing of the helices and specific helix-helix contacts is challenging with low sequence identity between template structure and target protein.

There is some controversy as to the evolution of the domains containing 6 TM helices, which were formed by nucleotide duplication in the gene for either a 2-TM helix or a 3-TM helix segment. The 3-TM helix repeat motif is proposed by Radestock et al. (Radestock & Forrest 2011), who compared LacY and FucP to show that a mix and match comparison of the four helix triples in each protein gave lower backbone RMSD values. Furthermore they have improved sequence homology between the two proteins in comparison with alignment of the full sequences.

However this is disputed by Västermark and Saier (Västermark & Saier 2014) who direct attention to the conserved MFS motifs in the loops between TM helices 2 – 3 and 8 – 9 which are contained within the cytoplasm. They demonstrate the motifs align best when using the whole protein sequences, rather than the helix triples described by Radestock et al..

In addition to the motif on these cytoplasmic loops (GxlaDrxGR) TMHs 2 – 3 and 8 – 9 (Paulsen et al. 1996), there is a further motif present on TMH 4 (Figure 1.2 (A) and (B) respectively). This is at least one, and often a series of conserved glycine residues (Pascual et al. 2008). These provide structural markers within the 12 TMH fold of MFS proteins that can be used in homology modelling.

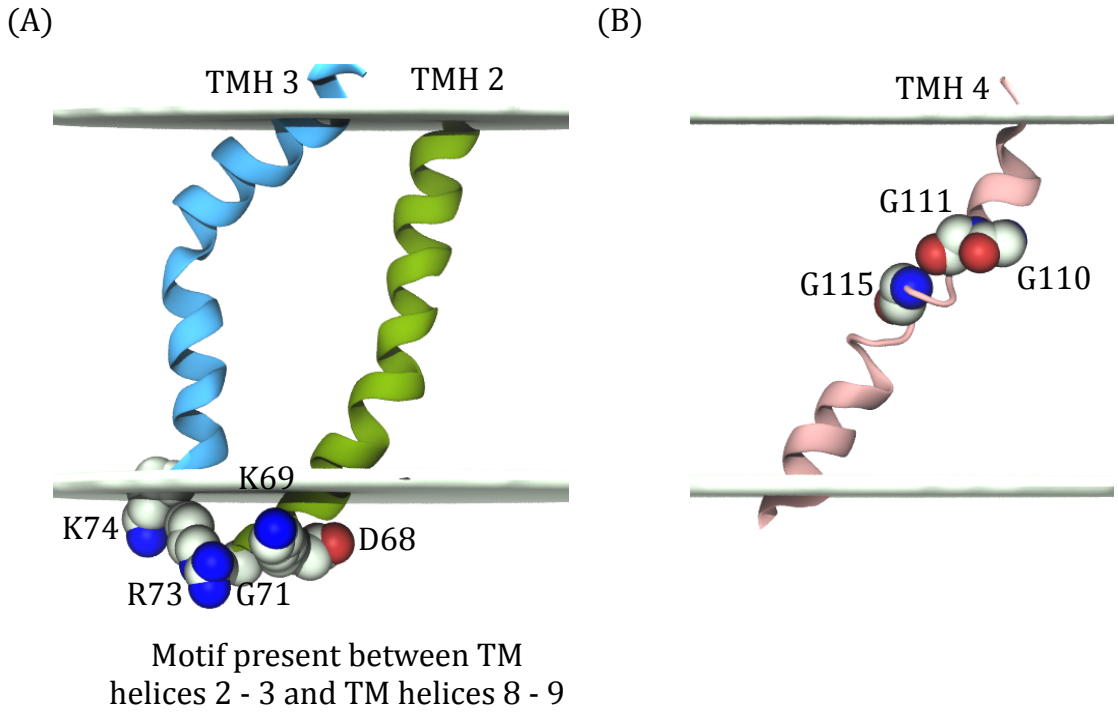


Figure 1.2 The position of conserved motifs in MFS transporters, on the FucP crystal structure (3O7Q). The (A) positive charged motif A on Loops 2 – 3 and 8 – 9 and (B) glycine residues in TM helix 4 are present in all of the superfamily proteins.

1.1.2 The transport mechanism of MFS proteins

The mechanism of these secondary active transporters is thought to operate via the alternating access mechanism as shown in Figure 1.3 (Eraly 2008; Smirnova et al. 2011). In this transport cycle, the MFS protein undergoes a series of conformational changes to allow the passage of a ligand from the extracellular side into the cytoplasm. The ligand first binds to the outward open, apo protein. The ligand and co-transported substrate binding then instigates conformational changes through occluded states to an inward open state that allows the ligand to move from the central TM cavity between the two domains of MFS transporters into the cytoplasm.

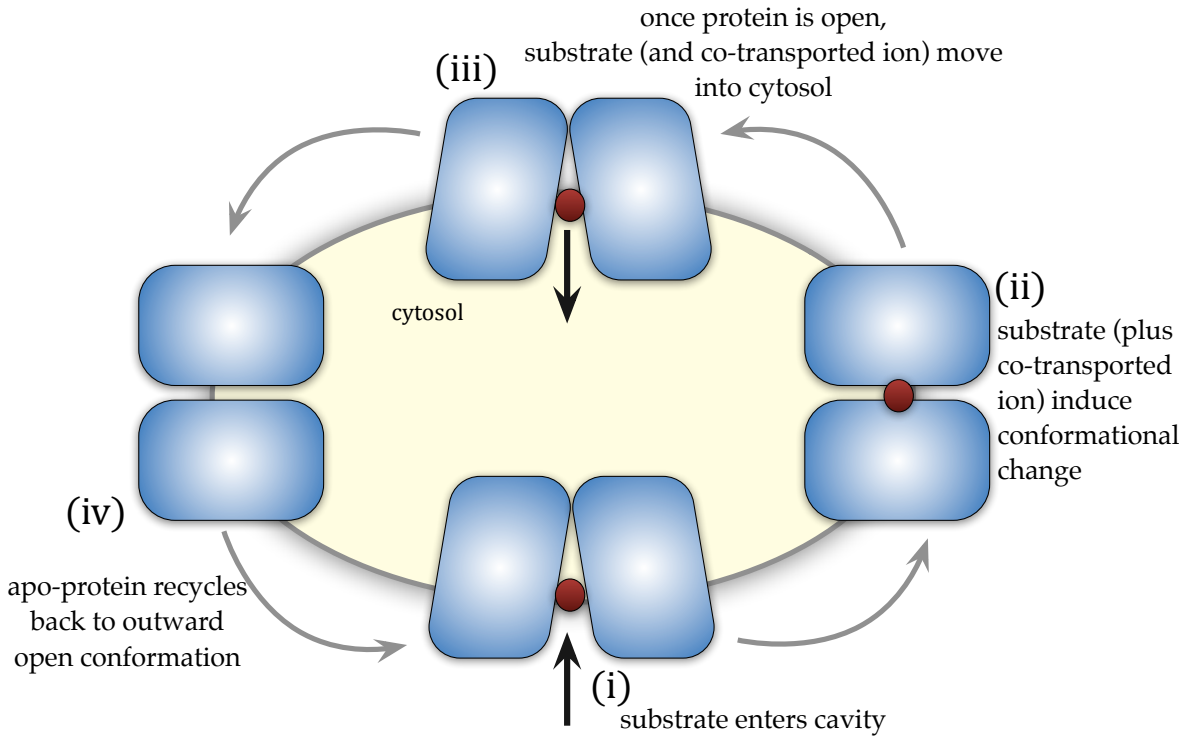


Figure 1.3 The conformations predicted in MFS transport cycling according to the Alternating Access Mechanism (AAM). (i) The ligand binds to the outward open protein. Once both substrate and co-transported ion are bound to the central TM cavity, the protein moves to an occluded state (ii) and then to an inward open (iii) state. Substrate passes into the cytosol and the apo protein can cycle back to outward open (iv). The terminology used in the subsequent results chapters refers to the outward open state as being open to either the synaptic vesicle (SV) for SV2A or the periplasm for FucP.

The paradigm for MFS proteins is LacY, which operates by a symport mechanism, that transports both sugar (predominantly lactose or galactose) and H^+ in the same direction across the membrane either into or out of the cytoplasm and both are necessary for transport to occur (Guan & Kaback, 2006; Madej, 2014). The protein uses the electrochemical gradient of H^+ across the lipid membrane to drive sugar transport against the sugar concentration gradient (uphill transport). Conversely the absence of the H^+ gradient will cause downhill sugar influx or efflux and drive uphill H^+ transport, generating the electrochemical gradient of H^+. Experiments to show this were conducted in a right-side out (RSO) membrane vesicle which uses the hydrolysis of ATP by F_1/F_0 ATPase to generate the H^+ electrochemical gradient and the opinion is that it is sugar binding, rather than the H^+ translocation which instigates transport (Guan & Kaback 2009; Smirnova et al. 2008; Smirnova et al. 2012).

In addition, the stoichiometry of ligand/H^+ transport was investigated in the proton-dependent oligopeptide (POT) transporter, $PepT_{st}$, which transports di- and tri-peptides with distinct binding sites (Lyons et al. 2014). The coupling mechanism of H^+ to both of these types of ligands was investigated by Parker et al. (Parker et al. 2014) and this showed that tri-peptides had 1:3 stoichiometry with H^+ whilst di-peptides had either 4 or 5 H^+ to each ligand. The internal pH of a liposome was monitored with a pH sensitive fluorophore. Since $PepT_{st}$ is electrogenic, like LacY, the acidification of the inside of the liposome can be measured directly to determine stoichiometry with the caveat that protonation of ligand is not taken into account (Kottra et al. 2002).

1.1.2.1 The position of key residues in MFS transporters

In MFS proteins, the most studied region is that of the ligand binding site which is situated between the two domains in the central TM cavity. LacY has been extensively investigated and the key residues involved in either lactose binding or H^+ translocation are shown in Figure 1.4. The ligand binding site is mediated by E269, E325, H322, R302 and Y236 (black stick representation) which are predominantly charged, despite a neutral ligand. The proton translocation pathway is mediated by E269, E126, R144 and K358 (orange stick representation). This shows that the LacY central cavity is fairly hydrophilic in character, at least at the point of ligand binding and that transport is mediated by TM helices 5, 7, 8, 10 and 11. The position of residues involved in LacY transport are similar to other MFS proteins with each having a cluster of residues in the central TM cavity that describe the ligand binding site (Madej et al. 2013; Madej & Kaback 2013).

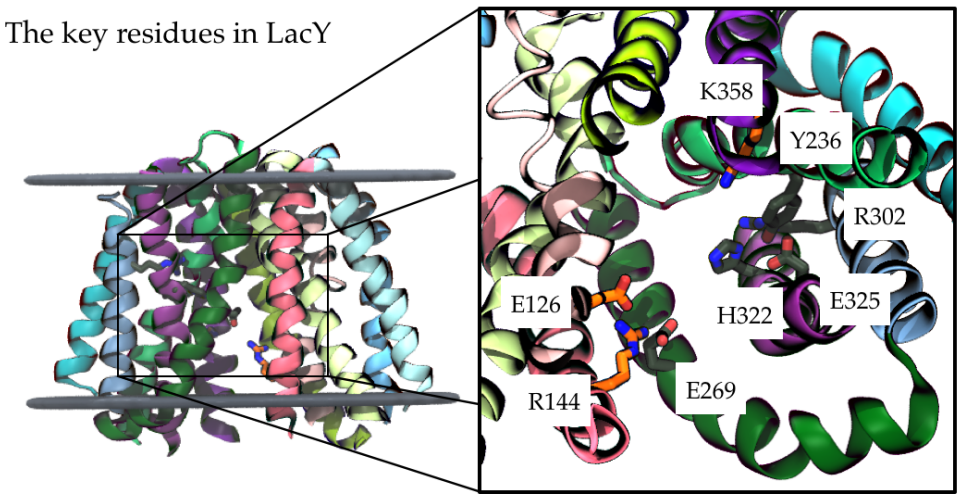


Figure 1.4 The position of key residues in the MFS transporter, LacY. The inset is a close up of the ligand binding site of LacY: residues identified as key for ligand binding and/or transport in LacY. The grey stick representations are residues involved in ligand binding only Y236, R302, E325 and H322. Orange stick representations are those residues involved in proton translocation: E126, K358 and R144.

The role of the loop regions is less well explored, however a recent publication by Masureel et al. (Masureel et al. 2014) determined that key, titrable residues were necessary in to loop regions of LmrP in order to change conformation. DEER experiments defined 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 and so it was discovered that at low pH LmrP would reside in the inward open state, whereas high pH closed the extracellular side. Like LacY, LmrP is thought to operate as a drug/H^+ symport transporter and these results suggest that the pH conditions of the environment at either side of the transporter can affect the transport by biasing towards conformations.

1.1.2.2 Water and the transport mechanism

Since the transport mechanism (for symport or antiport MFS proteins) utilises the electrochemical gradient of protons, it is expected that the transporters remain impermeable to water movement across the bilayer. It is probable that the proton translocation through the transporters occurs via water wires that connect titrable residues within the TM cavity (as described for Cl^- channels (Zeidel et al. 1992)), since these residues are too distant for direct transfer. These should be transient otherwise the transporters would become leaky, however some leakage has been shown to occur in the uniporters, GLUT1 and GLUT2 (Fischbarg et al. 1990; Zeuthen et al. 2007).

Molecular dynamic (MD) simulations of several secondary active transporters, including the MFS GlpT, have shown that the transporters become transiently open to water passage (Li et al. 2013). These events only allow single water molecules to pass across the membrane and relate to the orientation of side chains that create a

hydrophobic constriction within the TM cavity. It is difficult to interpret whether this is intrinsic to the transport pathway. However, targeted MD simulations (replicating the movement of the protein from the outward open to inward open state) of the XylE X-ray crystal structure saw several waters co-transported with the sugar moiety (Wisedchaisri et al. 2014) implying that water plays a role in transport.

1.2 Linking evolutionary conservation to structure

1.2.1 Conservation of the MFS fold

Whilst there is still a small proportion of X-ray crystal structures solved for the MFS proteins, homology modelling can be used to predict the structure of novel proteins (Kasho et al. 2006). This can be a powerful tool for drug discovery, the quality of the model is important in order to have confidence in screening for potential drugs in the correct binding site. Therefore the structure needs to be accurate and this is correlated with the sequence identity to the template structure used to build the homology model. For the case of MFS proteins, sequence identities are very low between families and they sit within the 'twilight zone' (Rost 1999). They have a highly conserved 12 TM helix fold and the relatively long length of their sequences means that this can be conserved despite low sequence identities (Figure 1.1 and see Table 1.1 for the sequence identity between MFS structures). This creates a problem when constructing models of MFS proteins since the lower the sequence identity between template and target protein, the less confidence there is in the pairwise alignment. Consequently an alignment is not able to properly convey the structural information required to ensure that the correct helix packing contacts occur in the target protein relative to their actual positions in the template structure without additional constraints on the alignment.

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	

Table 1.1 The sequence identity (%) between 10 MFS proteins for which there is an X-ray crystal structure. The percentage identity was computed from CLUSTAL Omega pairwise alignments and they range from 18 to 25 %.

1.2.2 Existing prediction tools

1.2.2.1 Secondary structure prediction

The prediction of secondary structure has been a long standing quest, which has resulted in several successful programmes that are specific to TM helices (therefore is relevant to MFS proteins), such as HMMTop (Tusnády & Simon 2001), PSIPRED (Jones 1999a), JPRED (Cole et al. 2008), SOSUI (Hirokawa et al. 1998) and TMHMM (Krogh et al. 2001). These use a range of algorithms that either predict helix position using scoring associated with sequence alignments (PSIPRED) or Hidden Markov Models (TMHMM, HMMTop) that exploit the helix-loop-helix motif common to many TM proteins. These are randomly optimised (since there are many local minima solutions to the problem) and the best state chosen from these. All programmes exploit the characteristic that TM proteins are generally hydrophobic in character on faces interacting with lipid tails and which led to early exploitation of hydrophobicity scales to predict helix positions (von Heijne 1992).

1.2.2.2 Tertiary structure prediction

These tools have had reasonable success and focus has shifted to 3D or tertiary structure predictions (Heijne 1994). In particular, there has been considerable advancement using helix-helix contact prediction within the membrane, such as FILM3 and EvFold (Nugent & Jones 2012; Marks et al. 2011; Hopf et al. 2012). These programmes exploit the conservation of residue-residue contacts within multiple sequence alignments (MSAs) by using *de novo* algorithms that rely only on the quality of the MSA and with impressive results in EvFold (Koehler Leman et al. 2015). The evolutionary coupling between contact points conservation of residue sites within MSAs is exploited in this thesis to develop a numbering scheme in MFS transporters (like that described below in GPCRs in Chapter 1.2.3). It is likely that these, over specific residue conservation, will provide a better tool for homology modelling in target MFS proteins.

1.2.2.3 Homology modelling

Due the improvement of 3D structure predictors, as described in Chapter 1.2.2.2, homology modelling is a key method to exploration of drug-target interactions in systems with no solved X-ray crystal structure (indeed there are structure predictions competitions for GCPRs (Michino et al. 2009)). However the efficacy of this technique relies on the availability of suitable templates so that misalignments do not disrupt sections of, or the whole of, the final structure of the target model (Marti-Remon et al. 2000). In order to have high accuracy, at least 50 % sequence identity is required between template and target, giving about 1 Å deviation in backbone, whereas below 30 % identity gives rapidly increasing errors in alignment that results in an incorrect target model (Baker & Sali 2001; Tramontano 1998). This means that automated template-target alignment is not sufficient to obtain a good

model of MFS proteins (whose identities are below 30 %) and that at the least a secondary structure alignment can improve the model (Forrest et al. 2006).

1.2.3 Using conservation to number residues in MFS helices

The research done by Ballesteros and Weinstein into the structure of GPCR proteins is testament to the power of using multiple sequence alignments (MSAs) to predict structure (Ballesteros & Weinstein 1992a; Almaula et al. 1996; Ballesteros et al. 1998; Zhou et al. 1994; Pardo et al. 1992; Konvicka et al. 1998; Javitch et al. 1998; Ri et al. 1999; Ballesteros & Weinstein 1992b). The culmination of their research meant that they could inform the structure of class A GPCRs by providing an anchor point common to each helix across the whole class of proteins (Ballesteros & Weinstein 1995). This has direct application to MFS proteins since the patterns used for these conserved sites on the helices of GPCRs can potentially be used in the same way for MFS model development, improving the template-protein pairwise alignment.

A multiple sequence alignment of several GPCRs was sufficient to pick out patterns of conservation at each site with regards to the chemical composition of the amino acid side chain (Ballesteros & Weinstein 1995). The chemical type is dependent on its environment and whether it is at a helix-helix interaction, associated with lipid tails or associated with either the lipid head groups or solvent. As such, hydrophobic residues are more likely to be at points between helices or lipid tails and polar or charged residues at lipid head groups and in solvent accessible positions. Furthermore TM proteins generally follow the positive-inside rule, with cytosolic loops containing more positively charged residues than extracellular loops (von Heijne 1992).

Analysis of integral membrane protein structures has also determined that the packing values of α-helices were smaller than their counterparts in globular proteins (Eilers et al. 2000). The occluded surface area (i.e. regions within the TM domains of the protein that are not solvent accessible) was computed between atoms, giving a distribution of distances between the helices, up to a cut-off that defines the atom as being solvent accessible. The average value between TM helices was found to be lower than that in globular proteins. Glycine and small hydrophobic residues mediate the tighter packing of helices, as do serine and threonine. In fact the hydroxyl residues tend to be positioned between helices in TM proteins, unlike in globular proteins where they are found on the surface. Further research has implied that the presence of the serine and threonine in α-helix TM proteins could have a functional role, by changing the local structure and dynamics (Del Val et al. 2012). A high propensity to form intrahelical hydrogen bonds was found for the hydroxyl residues. These are suggested to be involved in inducing local kinks or solvation or flexibility in the helices during MD simulations.

The conservation of residues according to chemical properties can be used in multiple sequence alignments (MSA) analysis, to guide the position of contacts between helices. Small or hydrophobic residues predominantly mediate these contacts, but hydroxyl 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.

Possibly the most useful outcome from the work by Ballesteros and Weinstein was the numbering system (Table 1.2) of Class A GPCRs. First, the residue is assigned a

number corresponding to its helix and the second number corresponds to its position relative to the most conserved residue on that helix. The most conserved residue is labeled as 50 and subsequent residues are counted up or down from this residue. For example, the most conserved residue on TM helix 6 of rhodopsin is P267 and so its corresponding number is P6.50. This provides a structural marker in the protein to relate further residues back to, since the conserved residue remains constant throughout the class A GPCRs.

TMH	100 % conserved in neurotransmitter GPCR	AA identifier in β_2ADR
1	Asn	N 1.50 (51)
2	Asp	D 2.50 (79)
3	Arg	R 3.50 (131)
4	Trp	W 4.50 (158)
5	Pro	P 5.50 (211)
6	Pro	P 6.50 (288)
7	Pro	P 7.50 (323)

Table 1.2 The numbering scheme used in Class A GPCRs. The table is taken from Ballesteros and Weinstein (Ballesteros & Weinstein 1995). The most conserved residue on each helix is numbered x.50 and then each residue of the helix is numbered up or down from 50. (AA = amino acid, βγADR is the GPCR protein, β_2-adrenergic receptor).

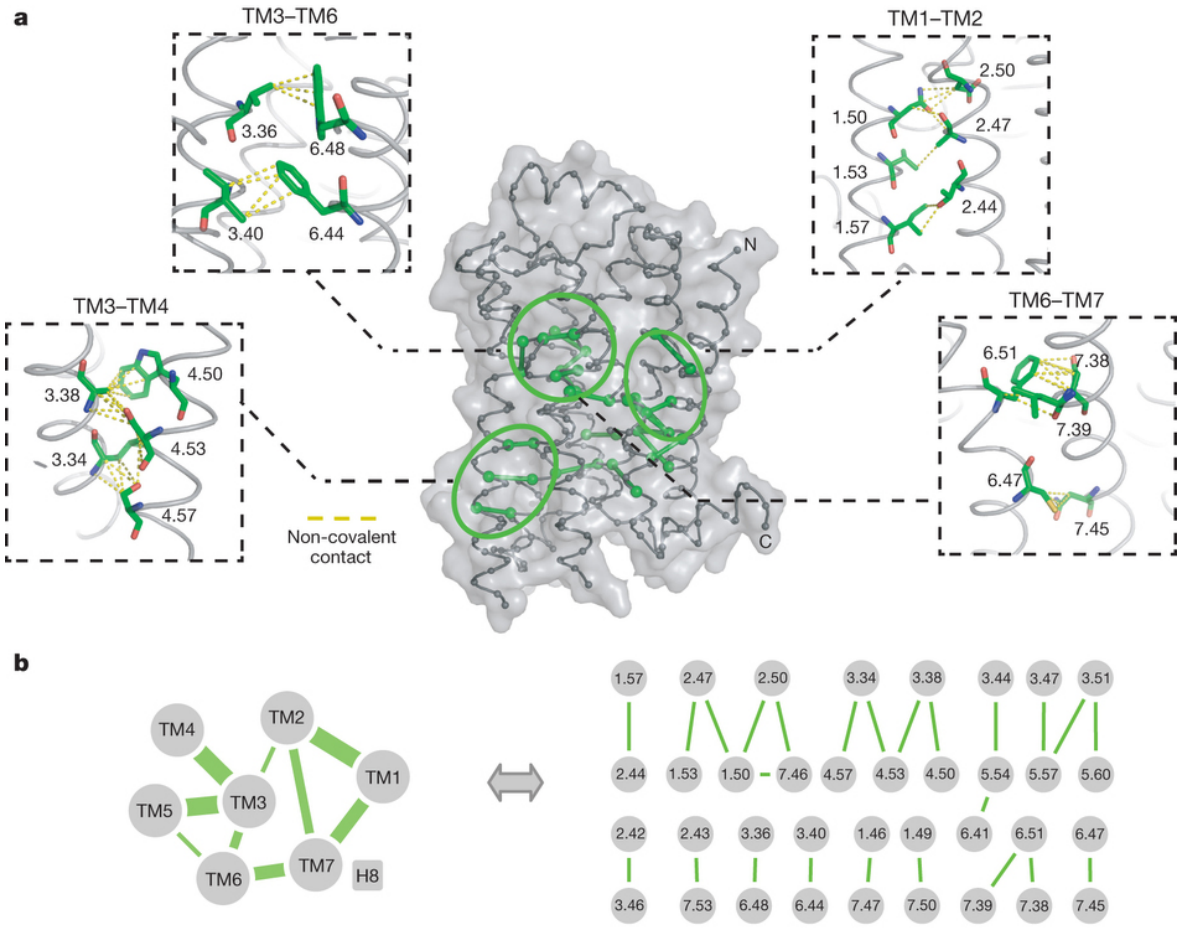


Figure 1.5 The network of non-covalent intra-helical contacts in GPCRs , taken from Venkatakrishnan et al. (Venkatakrishnan et al. 2013). (a) 24 residues in topologically equivalent positions in all GPCRs mediate the contacts, with (b) showing the relative abundance (thickness of the line) of these contacts between helices in a network map.

Recently research showed that 24 contacts between GPCR α-helices were identified that are mediated by a network of non-covalent interactions and are conserved in all GPCRs (Figure 1.5) (Venkatakrishnan et al. 2013). The network analysis in Figure 1.5 indicates that these contacts are predominantly between TMHs 1 – 2, 1 – 7, 3 – 4, 3 – 5 and 6 – 7. Defining similar contacts within MFS proteins will help provide insights into which intrahelical interactions are important and which change during the transport cycle. Therefore, as well as using the multiple sequence alignment (MSA) analysis to better inform template-target alignment for homology modeling, a more general numbering system was investigated. This will be similar to the

numbering of GPCRs by Ballesteros and Weinstein, picking out equivalent structural positions in MFS TM helices to aid the alignment process.

A recent update to the GPCR numbering was introduced by Isberg et al. (Isberg et al. 2015) that takes into account the bulging or kinking of helices and which would be a useful additional analysis in MFS proteins that have multiple kinks that potentially relate to conformation, such as seen in XylE and LacY (Vastermark et al. 2015). This could also inform the changes in the helix packing in MFS proteins according to the conformation and give a metric to help define the state required in the target model. In addition, the multiple states in MFS proteins (Figure 1.3) vary in conformation much more than the relatively small movements of GPCRs. This provides an added challenge in the analysis of MFS contacts for numbering.

1.3 Computational methods applied to MFS proteins

Despite an increasing number of X-ray crystal structures of MFS proteins, these structures are limited to predominantly bacterial proteins, although the first mammalian structure was solved for GLUT1 (Deng et al. 2014). Since so few human structures are available for MFS proteins, homology models are required for drug discovery. In this thesis, a homology model of SV2A was developed to study interactions of it with a class of drugs, racetams, which include anti-epileptic drugs.

1.3.1 Homology modelling of SV2A

1.3.1.1 SV2A structure and function

SV2A is a member of the Major Facilitator Superfamily (MFS) of transporters and is found in the membrane of both synaptic (SV) and dense-core vesicles (Tanner et al.

1996). It is one of three isoforms found throughout the brain that are implicated in Ca^{2+}-dependent vesicular neurotransmitter release (Chang & Sudhof 2009; Tanner et al. 1996). There is a proposed binding site for the protein synaptotagmin in the amino terminal domain (Figure 1.6 (A)), so SV2A is thought to mediate synaptotagmin presence at the SV membrane and consequently can indirectly mediate formation of the SNARE complex (Xu & Bajjalieh 2001). There is no proven change in ATP uptake by SV2A, yet a putative ATP binding site is present (Yao & Bajjalieh 2009) and recently SV2A has been shown to transport galactose into hexose transport-deficient EBY.VW4000 yeast cells (Madeo et al. 2014).

Despite a putative sugar transport mechanism in SV2A and so a proposed sugar binding site in the central TM cavity (Figure 1.6 (B) and (C)), this does not help elucidate its anti-epileptic mechanism. SV2A knock-out mice die within three weeks of birth, with seizures beginning within seven days, which relates to the start of SV2A expression (Crevecoeur et al. 2013). Kindling experiments (repeated stimulation to induce seizure) indicated up-regulation of SV2A in rats (Matveeva et al. 2007), yet patients with temporal lobe epilepsy show decreased SV2A expression (Feng et al. 2009). A similar effect is seen in other chronically epileptic animals (van Vliet et al. 2009). There is some evidence that SV2 proteins are involved in calcium homeostasis, though the exact role is unclear since SV2B knock-out mice show changed cytoplasmic calcium levels at the presynaptic terminal (Wan et al. 2010).

There is currently no structure for SV2A, which is predicted to have 12 transmembrane (TM) α-helices (Bajjalieh et al. 1994), arranged in two 6-TM α-helix bundles with pseudo two-fold symmetry and three extra-membranous domains (Figure 1.6 (A)). The amino domain is 170 amino acids in size and the loop between TMHs 6 and 7 is approximately 60 amino acids in size, both of which are on the

cytosolic side of the membrane (Figure 1.6). A final loop at TMHs 7 – 8 is within the synaptic vesicle and is the binding site of BoNT toxins (Benoit et al. 2014). Since the racetam binding site is thought to be analogous to MFS ligand binding sites within the TM domains (Shi et al. 2011), it is the TM domains which are the primary focus for this work (Figure 1.6 (B) and (C)).

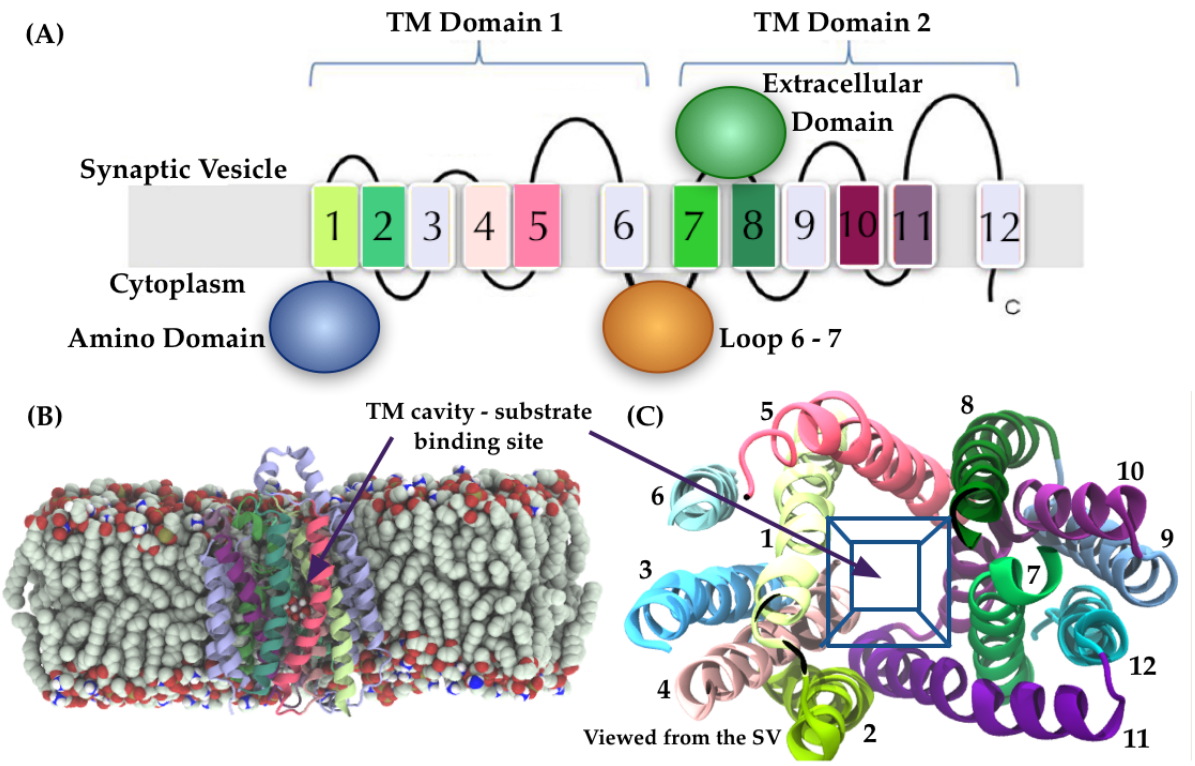


Figure 1.6 The topology of SV2A. (A) SV2A has two transmembrane (TM) domains, each containing 6 α-helices. There is a large Amino Domain (blue) of approximately 170 amino acids, a long loop in the cytoplasm (orange) of 60 amino acids and a further extracellular domain between TMHs 7 and 8 (green) of 120 amino acids. (B) A model of the TM domains of SV2A inserted into a POPC bilayer. A sugar is bound to the central cavity and putative substrate binding according to the position of ligands in other MFS proteins. (C) The arrangement of the 12 TM helices in SV2A and MFS transporters. There is a pseudo two-fold symmetry between the two domains. The central cavity runs down the interface of the domains and is the putative ligand binding site during transport cycling.

1.3.1.2 SV2A and racetams

Epilepsy is a debilitating neurological disorder, characterised by a spectrum of seizures that take both chronic and acute characteristics (Fisher et al. 2005). It is still

not adequately treated, with approximately 30 % of patients unable to control their seizures (Löscher et al. 2013). The cause of seizures is vast, ranging from Alzheimer's and Parkinson's diseases to hard blows to the head or brain infections and many cultures still stigmatise the illness, making it difficult to document (Beghi & Hesdorffer 2014). At a biochemical level, epilepsy is described as the hyperexcitation of brain signaling, causing the excessive release of neurotransmitters between synapses, leading to symptoms that range from loss of concentration or focus (absence seizure) to full body muscle spasms (clonic seizures) (Semah et al. 1998).

One drug administered to epileptic patients for its anti-seizure properties is levetiracetam (LEV (Figure 1.7 (A)); (S)-α-ethyl-2-oxo-pyrrolidine acetamide; KEPPRA™ (Löscher et al. 1998; Klitgaard et al. 1998)), which is a second-generation anti-epileptic drug (AED) (Gambardella et al. 2008). This comes from a family of compounds, racetams, which were found to have a neuroprotective effect that is related to binding affinity (Kaminski et al. 2008) and further research has seen the development of a second drug which is currently passing through stage III trials, brivaracetam (BRIV, Figure 1.7 (B) (Pack 2014)).

Studies in rats in the 1990s led to the identification of the target for the action of LEV being the synaptic vesicle protein, SV2A (Lynch et al. 2004) and it remains the only AED on the market which targets SV2A although the potential of racetam based compounds is shown through the potency of BRIV as an anti-epileptic drug (Matagne et al. 2008; Sargentini-Maier et al. 2008). The racetam family of compounds is of particular interest because the mode of action of LEV is not well characterised, but it has a different activity profile to other AEDs which act along (amongst others) the GABAergic pathway (Noyer et al. 1995; Klitgaard 2001;

Margineanu & Klitgaard 2009; Sills et al. 1997). LEV is also well-tolerated, with potential as an add-on drug alongside other AEDs (Harden 2001; Otoul et al. 2007).

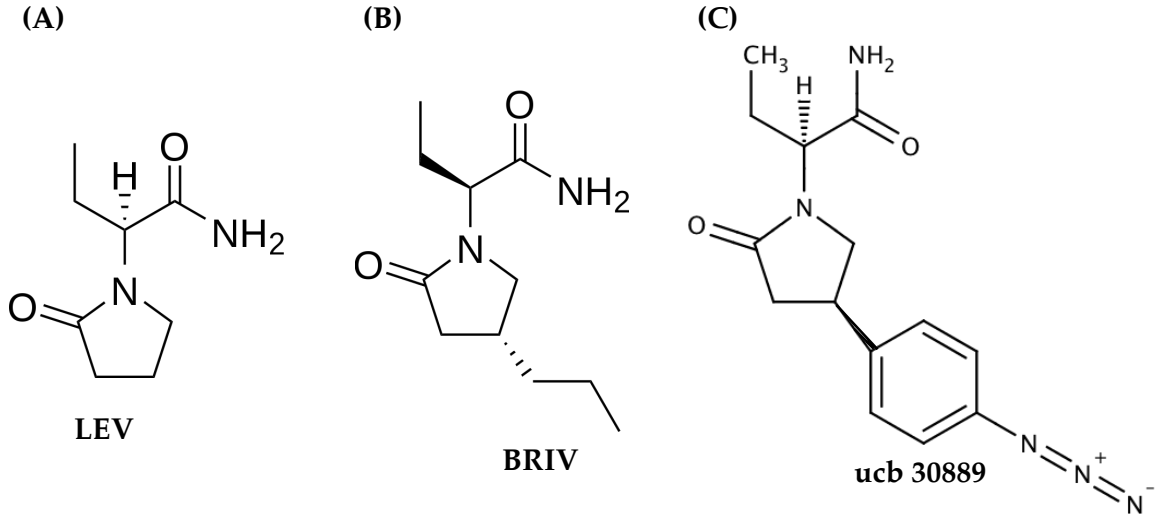


Figure 1.7 Three drugs from the racetam class of drugs. (A) Levetiracetam (LEV) has been on the market since 1999 as KEPPRA™ and (B) Brivaracetam (BRIV) is currently in stage III trials (Pack 2014). (C) The LEV and BRIV homologue, ucb 30889 is the radioligand used to determine binding constants of racetams to SV2A (Shi et al. 2011).

In homology modelling, for proteins with low sequence identity, the first step is to predict secondary structure using existing prediction tools. This information can guides the template-target alignment, which is then used as input for modelling software, such as Modeller (Eswar, Webb, Marti-Renom, Madhusudhan, Eramian, M.-Y. Shen, et al. 2007). Therefore the quality of this alignment is critically important and must convey structural information from template to target to ensure the correct packing of the predicted secondary structure. To aid this, the position of key structural motifs, common to the superfamily of proteins (Figure 1.2) can be used as anchor points in template – target alignments (Elofsson & von Heijne 2007).

1.3.2 FucP and MD simulation

1.3.2.1 FucP structure and function

FucP has previously only been studied in *E. coli*, however, research by Becerra et al. shows that it is present in the probiotic *Lactobacillus rhamnosus* genome (Becerra et al. 2015). The study recognises FucP as a component of a gene cluster, and the regulation of its operon is switched on by the presence of L-fucose, and by the absence of glucose. This permits the metabolism of L-fucose, feeding into the glycolysis pathway (Figure 1.8). The presence of L-fucose upregulates the transporter FucP, as well as FucI, FucK, FucU, FucA and FucR which are respectively an L-fucose isomerase, an L-fuculose kinase, an L-fucose mutarotase, an L-fuculose-1-phosphate aldolase, and a transcriptional regulator.

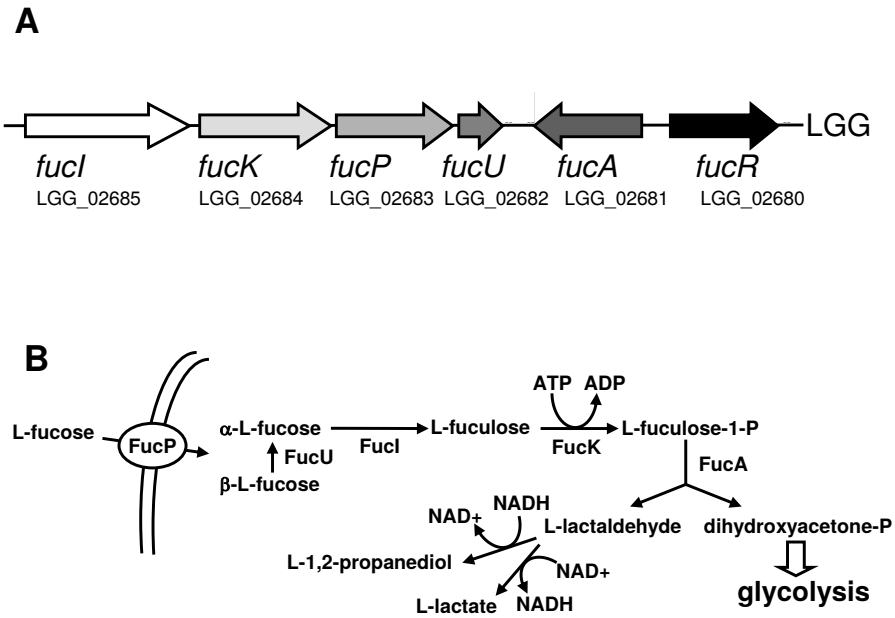


Figure 1.8 The L-fucose pathway as described in a figure taken from Becerra et al. (Becerra et al. 2015). The fuc operon is regulated by the transcription protein fucR and encodes fucPIKUA (A). The L-fucose pathway is an alternative carbon source in some bacteria and is switched off when glucose is present. (B) Describes the enzymatic action for the transcribed fuc proteins that feed products from the fucose pathway into the glycolysis pathway.

L-fucose (Figure 1.9 (A)) is commonly found as a component of glycans on the surface of eukaryotic cells; the Glycocalyx (Becker & Lowe 2003) and is present in highly glycosylated mucin proteins (Coyne et al. 2005). It is unsurprising that microbiota in the gastrointestinal tract have evolved to utilize it as a carbon source. Research into the pathogen *Camplyobacter jejuni* determined that its competitive advantage was the ability to metabolise L-fucose in a neonatal piglet infection model (Stahl et al. 2011). Whilst FucP is known to transport other sugars (D-arabinose and L-galactose), L-fucose is thought to be the primarily transported substrate. The X-ray structure of FucP was solved with a single molecule of the detergent *n*-nonyl-β-D-glucopyranoside (β-NG), in the cavity. The authors predict the site of the glycan ring of the detergent to be the binding site for L-fucose, near E135 and N162 (Figure 1.9 (B)).

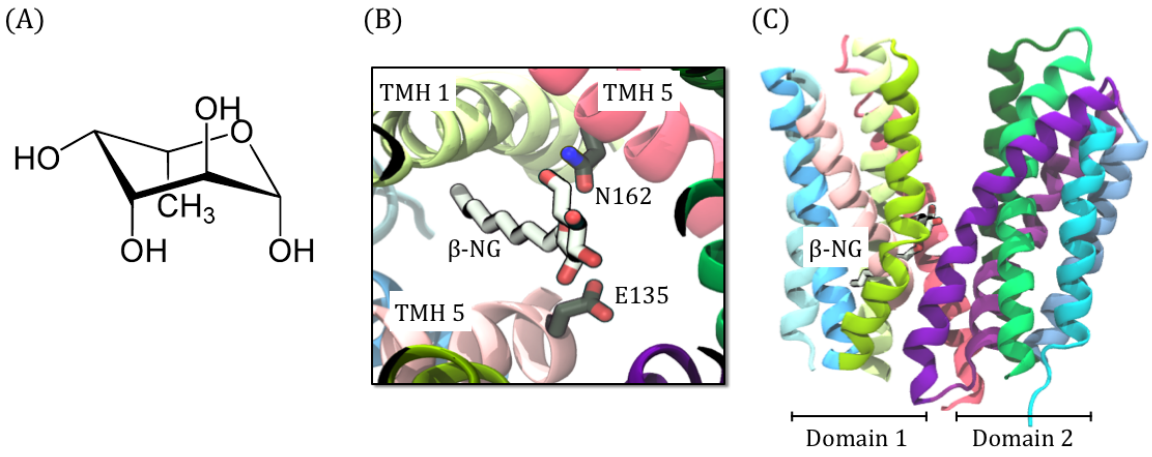


Figure 1.9 The X-ray crystal structure of FucP and ligands. (A) L-fucose is the key transported molecule by FucP, though it can also transport D-arabinose and L-galactose (these are not inducers). (B) The binding site of the co-crystallised β-NG detergent molecule which is thought to be the substrate binding site. (C) The position of the ligand binding site in the central cavity of FucP. The resolution of the crystal structure is 3.1 Å, so (B) gives the proposed binding site.

This predicted binding site is situated towards the closed end of the transporter's structure, nearer to the cytoplasmic side of the lipid bilayer (Figure 1.9 (C)). At this site, the transporter is closed, so water is not able to permeate through the cavity. This is important for the transport mechanism – distinguishing FucP from a channel protein. The transport mechanism in FucP, first proposed by the authors of the X-ray crystal structure, Dang et al., suggests that the solute interactions with this deep binding site, as well as proton translocation to E135, are key to instigating the conformational changes from outward to inward open states and thus the control of solute passage into the cell (Dang et al. 2010).

Subsequent to the paper by Dang et al. (Dang et al. 2010), a paper was published in 2013 that compares functional residues of LacY to residues in equivalent sites in FucP, as shown in Figure 1.10 (A) (Madej et al. 2013). In the figure, the equivalent residue in the LacY structure is labeled above the plot, which shows the measure of transport given a mutation in FucP (x-axis). The colour of the residues represents no effect (green), an increase (blue) and a decrease or abolition (red) of transport. In FucP residues, the equivalent positions have similar impact of sugar transport in the respective systems. The position of these residues in FucP is shown on the crystal structure in Figure 1.10 (B). The majority of these make up the fucose binding site around E135: W38, Y74, L134, N139, N162, R312 and Y365.

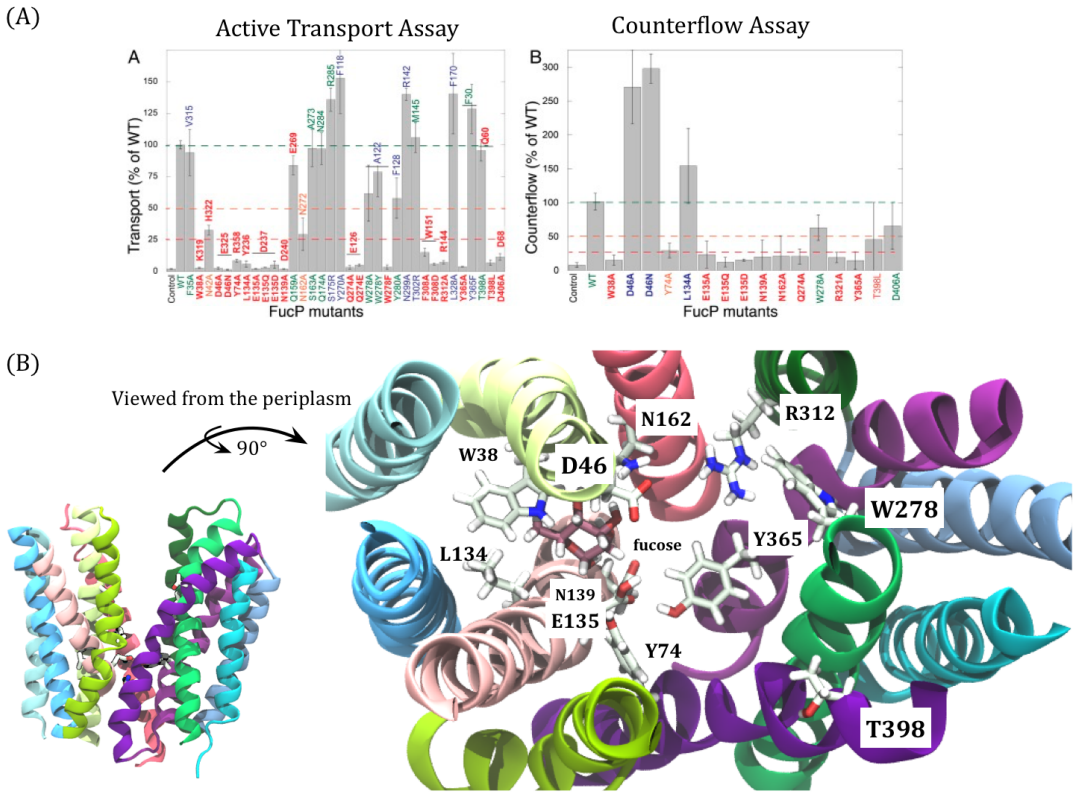
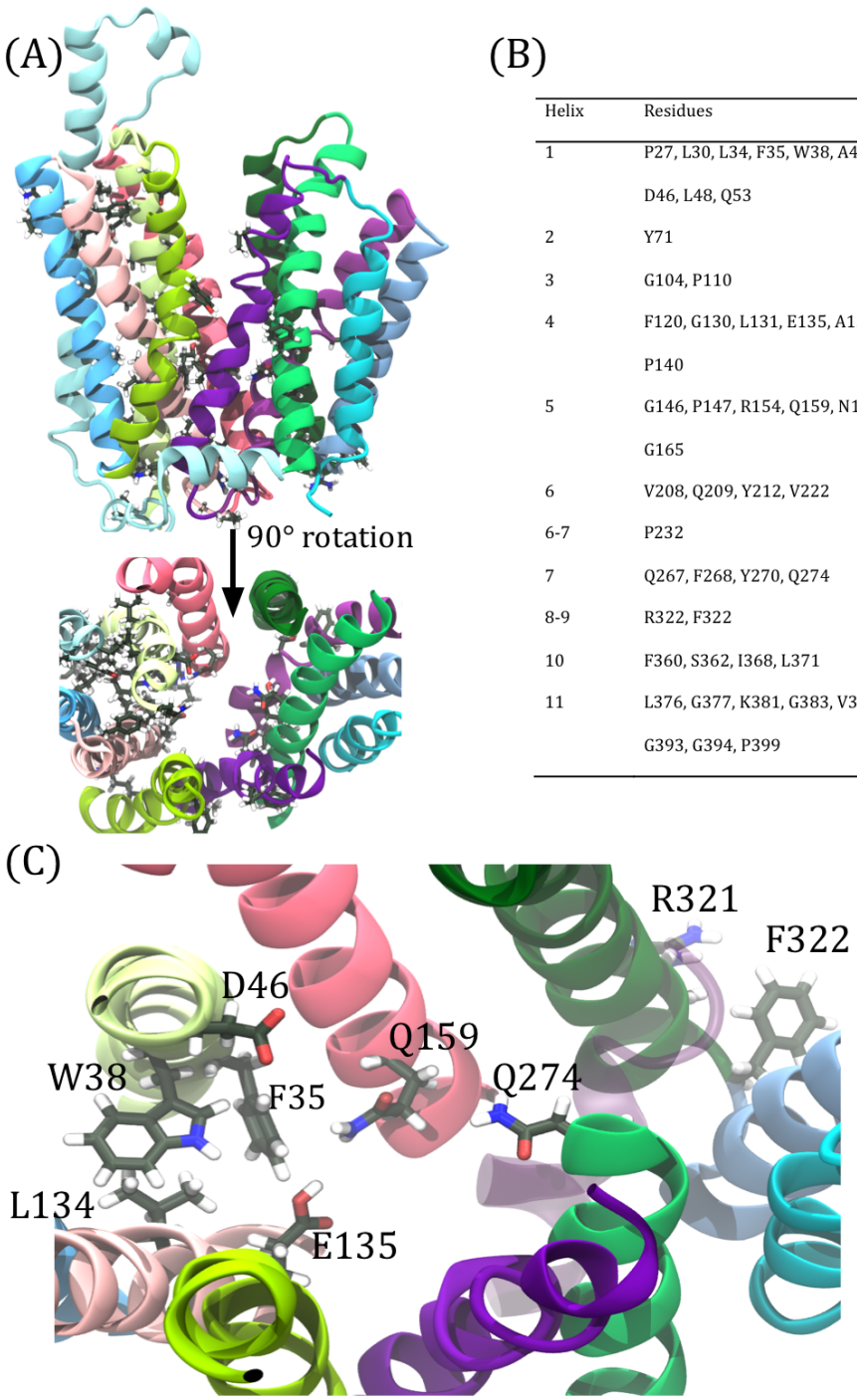


Figure 1.10 Key residues in FucP. (A) A series of single point mutations were tested in FucP that were chosen according to their relative position in FucP (x-axis) compared with functionally important residues in LacY (labels above the bars). The plots are taken from Madej et al (Madej et al. 2013) and the left one describes the active transport assay results for each mutation and the right is the results of counterflow assays. (B) The position in the FucP crystal structure of those residues from (A) that have the greatest effect on substrate transport, as viewed from the periplasm.

The X-ray crystal structure of FucP was solved in the outward open state compared with other MFS proteins, open to the periplasm (Dang et al. 2010). This unique MFS conformation differs from subsequent solved structures, which have more occluded, outward conformations (YarJ: 3WDO and GLUT1: 4PYP). This has led to the hypothesis that the conformation may not be physiological, and may be influenced by crystal packing artifacts (Andrec et al. 2007; Hinsen 2008). Molecular Dynamics (MD) simulation is well positioned to address this hypothesis, by exploring the conformations the protein accesses during the course of the simulation.

Further to exploring the outward conformation of FucP, the role of both D46 and E135 protonation was investigated as well as those residues identified as functionally important on mutation in transport assays (Madej et al. 2013). Figure

1.11 (B) lists these residues, which are highly conserved in FucP, according to the TM helix on which they reside. Many of the functionally important residues are also highly conserved (D46, E135, Q159, Q274, W38, L134, R321), however there is a notable exception from this list, R312. The arginine is within the cavity and reduces transport to below 10 % activity on mutation to alanine (Figure 1.10 (A)).



Helix	Residues
1	P27, L30, L34, F35, W38, A41, L44, D46, L48, Q53
2	Y71
3	G104, P110
4	F120, G130, L131, E135, A138, P140
5	G146, P147, R154, Q159, N162, G165
6	V208, Q209, Y212, V222
6-7	P232
7	Q267, F268, Y270, Q274
8-9	R322, F322
10	F360, S362, I368, L371
11	L376, G377, K381, G383, V388, G393, G394, P399

Figure 1.11 The conservation of key residues in FucP. (A) and (B) The most conserved residues for FucP, as given in the MSA analysis of Chapter 3, mapped on to the crystal structure of FucP. These residues are predominantly in domain 1 or line the central TM cavity, with positively charged residues in the cytoplasmic loops. (C) The single point mutations of FucP (Madej et al. 2013) indicate that these residues are both highly conserved and functionally important: F35, W38, D46, L134, E135, Q159, Q274, R321 and F322.

A closer investigation of where these conserved residues are located indicates that those with known functional importance are predominantly located in the centre of the cavity (Figure 1.11 (C)). This central area contains residues W38, L134, E135, Q159, N162 and Q274, and is predicted to be the ligand binding site within the cavity. D46 is distinct from this site, as described by Dang et al. (Dang et al. 2010) and so fits well with its role in proton translocation, rather than ligand binding (the residue is 96 % conserved in a multiple sequence alignment search of FucP). A notable exception in the list of conserved residues is W278, the residue predicted to interact with D46 when the transporter is in the occluded and inward open states. It is present in 74 % of the MSA, indicating that an aromatic residue in that position is important but the presence of alanine (24 %) implies less evolutionary pressure for W278 over the charged residues.

One rather interesting omission from this list is the presence of the conserved MFS motif in loop 2 – 3. A positively charged residue is conserved in this position (either K – 90 % or R – 7 %), but the motif is not conserved. In the analogous loop of domain 2 (loop 8 – 9), a conserved arginine is 99 % present at R321 (Figure 1.11). This implies that FucP is possibly non-typical of MFS proteins, however this is speculation as the scope of MD is not sufficient to fully investigate the role of the motif during the MFS transport cycle. There is, however, a conserved glycine on TMH 4 that fits with the position of the glycine motif, and glycine residues are also conserved in TMH 10 (as shown in Figure 1.2 (B)).

1.3.2.2 Using MD simulation to probe conformation

Recent advances in both computational power and algorithm efficacy (for example, the use of GPUs in GROMACS) mean that Molecular Dynamics (MD) simulations are now capable of application to TM protein systems with explicit lipid bilayer.

However it is still not feasible to measure the conformational changes associated with the MFS transport cycle, which is on the ms timescale in comparison to MD simulation which at best can reach the ns/μs timescale. Therefore biased simulations are required to sample the MFS conformation space, which are capable of obtaining the free energy landscape for the transport cycle. An example of this is the free energy landscape obtained by Moradi et al. in ABC transporters (Moradi & Tajkhorshid 2013a; Li et al. 2015; Moradi & Tajkhorshid 2013b).

In these papers, driven and adaptive-bias metadynamic simulations determined free energies. The latter, adaptive-bias MD is history-dependent. Given a short simulation, this takes a snapshot with a conformation that is closest to the desired final conformation and restarts a further short simulation from this position. The process is repeated until the desired end conformation is achieved. This relied on knowing the final conformation of the protein, which is decided from the initial driven MD simulations between the alternatively open states to periplasm and cytoplasm. This is determined using steered MD that moves the initial conformation into a known or predicted final conformation, using a set of collective variables along which the bias is applied.

Specific to MFS transporters, the conformations of LacY, an inward open MFS protein, were investigated by D'Rozario et al. and Stelzl et al. (Stelzl et al. 2014; Pendse et al. 2010). Sampling between inward open and an outward open model of LacY based on the model determined by Radestock and Forrest (Radestock & Forrest 2011). The subsequent conformational space sampled in MD simulations clearly delineated the inward, occluded and outward open conformations of LacY. The occluded LacY conformations were restrictive to both ligand and also to water, a key characteristic that prevents MFS transporters becoming channel-like in character. Further, D'Rozario determined that the motions in TM helices 4, 5, 10 and

11 dictated conformation dynamics in GlpT. A further application of the repeat-swap motifs of Radestock et al. was applied to $PepT_{so}$ (Fowler et al. 2015) and the subsequent simulations defined the gating helices at each side of the TM cavity of the MFS transporter. At both the cytoplasmic and periplasmic side of the cavity, TM helices 1, 2, 4, 5, 7, 8, 10 and 11 are involved in the gating mechanisms and minimal distance between TM helices 1, 2 and 7, 8 described closure of the periplasm and minimal distance between TM helices 4, 5 and 10, 11 described closure of the cytoplasm.

Biased and unbiased MD simulations were applied to FucP in the later chapter of this thesis. The incentive for this work was to explore the dynamics of FucP in an outward and occluded state that compared the effect of protonation in the central TM cavity in both states. A study by Liu et al. (Liu et al. 2015), published in August 2015 explored the transition of FucP from outward through occluded to inward facing in biased simulation. Here, the researchers look exclusively at E135 only and not at the effect of D46 nor the change in the water content in the cavity. In the relevant chapter, this work is compared with that conducted for this thesis and the scope of both assessed.

1.4 Aims

The aim for this thesis was to perform analysis of sequence conservation in MFS transporters and use this to inform structure prediction in homology modelling. To this end, the first aim for the work was to use the conservation of both residue sites within a multiple sequence alignment and contact points within helices in the tertiary MFS fold. This will produce an MFS numbering system. This can then be used as a comparison tool between MFS proteins and inform template-target

alignment for homology modelling. These techniques were applied to SV2A, producing a model that explored the drug-protein interactions for the anti-epileptic drug, LEV, and further discern the binding site in SV2A. A binding hypothesis for LEV was developed and experimentally tested.

The second aim for this thesis was to explore the dynamics of FucP, an MFS protein with a solved X-ray crystal, using molecular dynamics (MD) simulations. The transport cycle predicted for FucP implies that its ligand binds to the outward open state (the X-ray crystal structure). This as well as protonation of titrable residues in the central transmembrane (TM) cavity induces movement through an occluded to an inward open state and allowing ligand to pass into the cytosol. The conformations of FucP during the outward to occluded states of the transport cycle were explored using both biased and unbiased simulations.

2 Methods

This chapter will provide the detailed methods used within this thesis, giving explanations of techniques for the subsequent results chapters. A brief overview of methods will also be given at the start of these chapters and the corresponding sections of this chapter referenced for clarity. This chapter is split into two sections, the first will address sequence analysis and the methods used to extract conservation of residues, leading up to model building, and the second part will detail the theory behind the all-atom molecular dynamics (MD) simulation techniques used in this research.

2.1 Sequence analysis and model construction

In Chapter 3 of this thesis, the analysis of Major Facilitator Superfamily (MFS) multiple sequence alignments (MSAs) was performed to help inform structure and model construction of SV2A. The most important factor in homology modelling is sequence alignment such that structural information can be conveyed from template to target protein. In order to do this with confidence, there needs to be sufficient sequence identity between the two sequences (Rost 1999). To this end, both conserved sites within the MSAs and conserved contacts within the X-ray crystal structures were explored. The aim was to produce a numbering scheme for MFS proteins that will aid current and future homology modelling endeavours.

2.1.1 Identifying the most relevant sequences

For the first step, one has to choose the MFS sequences to be included in the MSAs. The high sequence diversity of MFS proteins (X-ray crystal structures typically have 15 – 25 % identity, as shown in Chapter 1, Table 1.1) meant that this step was not trivial and so the initial idea was to use curated alignments, such as those for the families of MFS proteins as defined by Pfam (Finn et al. 2014). However this was not successful in this study since the identities between family members were low, which meant there was a lack of confidence in translating the conservation to structurally relevant information, as described by Tramontano who assessed efficacy of MSAs (Tramontano 1998). The greater the sequence identity, the less likely the alignment will stray as there is less risk of incorrect gap insertion and so high sequence identity is preferable for aligning MFS proteins, with more than 50 % identity describing high accuracy in alignments (Baker & Sali 2001).

Given these difficulties, a data set was sought that provided almost unambiguous alignments but had enough variation to be useful for predicting core, conserved sites key to the structure of MFS proteins. The UniRef50 dataset provides this; it clusters the sequences in the UniProt Knowledgebase and Archive according to overlap with one another in sequence space (i.e. can multiple 3 amino acid words match between both sequences and cover the length of both with 50 % identity), rather than using a direct alignment algorithm. The algorithm CD-HIT (Li & Godzik 2006), used to assign the clusters, does so by taking the sequences in length order starting with the longest sequence which represents the first cluster. Subsequent sequences are checked against this sequence and if they have sufficient short word matches (3 amino acids) to count as identical (these are indexed and matches counted rather than determining a direct alignment), they are placed in the 100 % cluster (UniRef100) and further clusters are formed from these to make up 90

% identical (UniRef90) or 50 % identical (UniRef50) clusters. The UniRef50 was deemed to have sufficient similarity to convey structural information in subsequent multiple sequence alignments and so overcoming the issue that less than 30 % identity was insufficient to correctly align sequences (Doolittle 1986; Rost 1999).

For each X-ray crystal structure available for MFS proteins, a search was performed against the UniRef50 clustering of protein sequences (Suzek et al. 2007). The UniRef50 database is available at www.uniprot.org and is automatically updated fortnightly. In addition, redundant sequences were removed, streamlining the process and removing unnecessary noise from the MSAs by removing duplicated entries of sequences. Thus each sequence in the subsequent alignments of MFS proteins does not have bias towards multiple occurrences of the same sequence.

2.1.2 Multiple sequence alignment using MUSCLE

The computational cost for building an exact MSA is too high. For n sequences of length L, the calculation of the phylogenetic tree requires $2^n - 1$ combinations to be examined, which means that the search space grows exponentially with the number of sequences and so becomes computationally prohibitive (Sankoff & J., Cedergren 1983). To overcome this, heuristic multiple alignments were performed using scoring functions to determine alignments. These score the position of a particular amino acid in sequence 1 against that in sequence 2 and also contain a gap penalty such that above a certain threshold of scoring between two amino acid types a gap is instead used (Phillips 2006). Therefore the resulting MSAs are not exact and are prone to differences according to the alignment program used and the specific gap penalties.

The alignment program used in this research was MUSCLE (Edgar 2004) and MSAs were constructed using a local version and the default options. MUSCLE was chosen for its speed and it is also statistically indistinguishable from other alignment algorithms T-Coffee and MAFFT (Edgar 2004). The 240 PAM matrix (Müller et al. 2002) was employed by the MUSCLE algorithm, which describes the probability of amino acids occurring in equivalent sites in two sequences. In addition, an occupancy factor was implemented so that more highly occupied sites in the MSA are encouraged to align in the algorithm rather than awarding a gap. Refinement iterations were subsequently used and an MSA score compared the iterations, and further iterations were performed if the score continued to improve. At each step, as well as for the initial alignment, the programme computed the phylogenetic tree to improve the alignment by rearranging subtree branches, before assigning a score. A typical MSA in this analysis contained from 70 to 5000 sequences (Figure 9.1, Figure 9.2, Figure 9.3 and Figure 9.4).

2.1.3 Choosing the homology model

The homology models of SV2A were constructed from template-target alignments that were the input to Modeller 9.10 (Eswar, Webb, Marti-Renom, Madhusudhan, Eramian, M. Shen, et al. 2007). The template structure provided restraints for the target model and the subsequent 3D model was derived by optimisation of the molecular probability density function, such that atoms did not overlap. This optimisation employs a function that operates in Cartesian space such that target coordinates are improved using simulated annealing Molecular Dynamics (MD) simulations and energy minimisation. The quality of the 100 models produced in each production run of Modeller 9.10 (Eswar, Webb, Marti-Renom, Madhusudhan, Eramian, M. Shen, et al. 2007) were assessed by their DOPE score and the top solution was defined as that with the best DOPE score (Shen & Sali 2006). The score

is based on minimising the probability density functions assigned to the position of the atoms in the protein model.

This model was then checked using QMEAN (Benkert P. Biasini & Schwede 2011) whose scoring indicated an accepted range for membrane proteins, providing further validation of the chosen model. The QMEAN score is considered to be a range that can determine the viability of either models or X-ray crystal structures, according to a 'degree of nativeness' criteria that is quantified by a z-score. The scores are computed using a distance-dependent interaction potential of mean force based on the C-α atoms in the model, analysing the backbone geometry and solvation potential (Benkert P. Biasini & Schwede 2011; Benkert et al. 2009). Further optimisation of the models was done using MD simulations to assess the influence of the membrane on behavior of the model and the conformational stability.

2.2 Molecular dynamics theory

The aim of molecular dynamics (MD) is to accurately describe the atomic motions within a molecular system. In order to solve this, the application of Schrödinger's equation is required (and so applying the particle-wave duality of all atoms to ensure the most accurate representation of the system). Here both the positions and velocities of the atoms in the system are solved *ab initio,* without any external input. However this is computationally unfeasible for anything greater than a few atoms, therefore approximations are required that balance accuracy with what is computationally feasible. These come in the form of empirical parameters that predict the positions of atoms according to recorded physical properties established using NMR and IR spectra and Quantum Mechanics (QM) calculations. Such constants are described in force-fields, which are a set of constants that describe the

potential energy of a molecular system at a given set of atomic coordinates and these can then be simulated using classical mechanics. The classical interpretation becomes a feasible approximation in protein systems, whose relative mass means that the system is predominantly particle-like rather than wave-like and so the wave properties of the system can be ignored.

With the increase in computational power over the last decades, MD simulation is a reasonably time efficient method and it provides key insights to dynamics that are experimentally tricky to investigate. The system is approximated using a force-field (in this case Optimised Potentials for Liquid Simulations – OPLS (Jorgensen & Tirado-rives 1988)) and the MD algorithm then applies Newton's second law of motion. Each particle is assumed to have a mass, charge and initial velocity, which is dependent on the interactions it maintains with those particles near it in space (Equation 2.1).

$$F = \frac{dv}{dt}m$$

Equation 2.1 Newton's second law of motion. This is applied to force-field constants in MD simulations (see Figure 2.2 for MD workflow).

2.3 Force-fields

Each force-field contains a list of constants that are used in the MD protocol whose validation comes from experiments. In this thesis, the OPLS force-field is used (Jorgensen et al. 1996; Kaminski et al. 2001a; Felts et al. 2002), where the partial charges on each atom, bond angles and lengths were taken from NMR and IR spectra and Quantum Mechanics (QM) calculations. The OPLS force-field was chosen as it is specifically designed for proteins within lipid membranes and the

Berger lipid parameters (a united – atom model of lipids applied alongside OPLS all atom force-field for the protein and solvent) were able to most accurately replicated experimental parameters at the point of this research (Berger et al. 1997; Tieleman et al. 2006).

One of the weaknesses of the most commonly used protein force-fields, such as OPLS is that they are unable to account for polarisability of bonds. Instead OPLS, like most force-fields, attributes atoms with a charge that is set for the duration of the simulation. So it is not possible to see phenomenon such as bond breaking, which is a statistically likely event, and instead the atoms remain statically bound in their original states.

2.3.1 Bonded interactions

In MD simulations, the bonded interactions, namely those interactions applied between two atoms that are covalently bonded to each other, are based on an input list that defines the 2-, 3- and 4-body (atom) interactions. The first of these describes the bond stretching (Figure 2.1 (A)) between two atoms and is approximated using a harmonic potential, most commonly equated to that of a simulated spring (Equation 2.2), which is dependent on the distance between the atoms (b_{ij}) and the equilibrium bond distance (b_0) and the strength of the spring (k).

$$\sum V_{bonds} = \sum 0.5 \; k_{ij}(b_{ij} - b_0)^2$$

Equation 2.2 Bonded interactions in system. b_{ij} is the bond length between atoms i and j at a given time point and the equilibrium bond length is b_0. (see Figure 2.1 (A))

The bond angle (Figure 2.1 (B)) is defined between three atoms, where the central atom is attached to a further two atoms by a harmonic potential. The harmonic potential, which describes the energy involved in the deviation of the angle, is given in Equation 2.3 where the deviation of each bond angle is computed from the reference angle θ_0.

$$\sum V_{angles} = \sum 0.5\ k_{ijk}(\theta_{ijk} - \theta_0)^2$$

Equation 2.3 The angles between bonded atoms. θ_{ijk} is the fluctuation of a bond angle between atoms i, j and k from the equilibrium bond angle θ_0. (see Figure 2.1 (B))

That final bonded interactions that are applied in force-fields are for proper and improper dihedrals (both 4-body interactions as show in Figure 2.1 (C) and (D)). The dihedral angles maintain the correct orientation across long chains, most commonly applied across the peptide bond, whereas the improper dihedrals maintain planar groups as planar. These are described by Equation 2.4 and Equation 2.5 respectively.

$$\sum V_{dihedrals} = \sum k_\psi[1 + \cos(n_\psi - \gamma)]$$

Equation 2.4 The proper dihedral angles describing the angle in Figure 2.1 (C).

$$\sum V_{improper} = \sum k_\psi(\psi - \psi_0)$$

Equation 2.5 The improper dihedral angles describing the angle in Figure 2.1 (D).

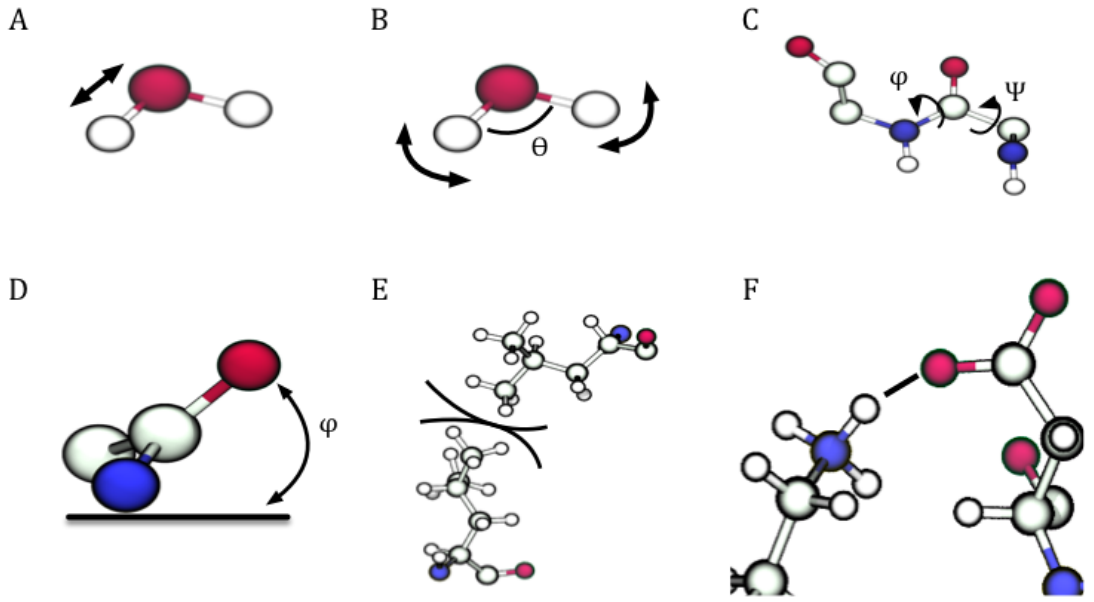


Figure 2.1 Bonded and non-bonded interactions. The bond stretch between two atoms is described by a harmonic potential (A). The bond angle between three atoms is described by a harmonic potential (B). The proper (C) and improper (D) torsion angles. The van der Waals forces (E) are the non-bonded interactions between atoms and (F) Coulombic interactions are electrostatic forces due to charge or permanent dipole between the 2 atoms.

2.3.2 Non-bonded interactions

There are two forms of non-bonded interactions, the first form is Coulombic interactions, which describe the interactions between charged species (Figure 2.1 (F)) and is defined by a Coulombic potential in Equation 2.6. The charges are pre-assigned within the force-field used and OPLS does not cater for polarisability in bonds.

$$\sum V_{Coulomb} = \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Equation 2.6 The non-bonded Coulombic interactions, where ε_0 is the dielectric constant. q_i and q_j are the charges and r_{ij} is the radius between atoms i and j. (seeFigure 2.1 (F))

The Lennard-Jones interactions are used to approximate the van der Waals interactions which are attractive and repulsive forces between all atoms. These are

described in MD using Equation 2.7, where α_{ij} is the collision diameter between two particles i and j.

$$\sum V_{vdw} = \sum 4\varepsilon_{ij}\left[\left(\frac{\alpha_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\alpha_{ij}}{r_{ij}}\right)^{6}\right]$$

Equation 2.7 The form of the van der Waals interactions. (see Figure 2.1 (E))

The use of long-range cut offs (which ignores any interactions between atoms that are beyond a cut-off distance) in the form of a radius applied to both the Leonard-Jones and Coulombic interactions reduces the computational demand in the energy calculation. This is system specific but the use of periodic boundary conditions mimic a bulk fluid environment and minimise edge effects (Leach, 1996). In simulations, this means that a correction is necessary so that the protein cannot interact with its mirror image, since simulation set up assumes infinite repetition of the simulation box in all directions.

In order to fully account for the electrostatic interactions in the system, particle-mesh Ewald (PME) is implemented. This speeds up the calculations of long range interactions (Essmann et al. 1995) where the interaction is split into two components, a short range potential, and a long range potential. The short range potential converges in real space, whereas the long range potential converges in Fourier space. Since the two components convergence is fast, it is possible to approximate them with only a small loss in accuracy, whilst gaining in efficiency.

2.4 Simulation Protocol

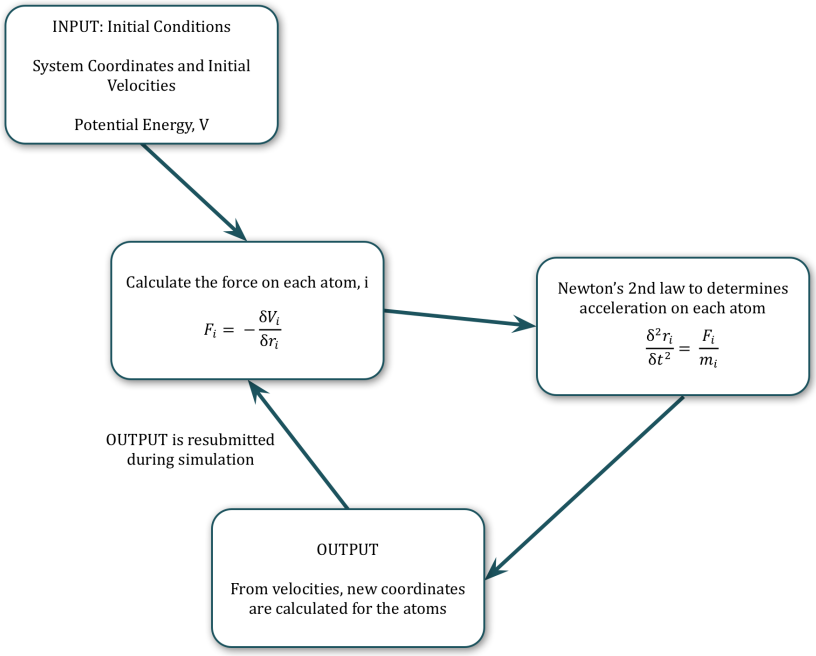


Figure 2.2 Workflow for simulation in GROMACS. Newton's second law is applied to determine new velocities after a small time step. Potential and kinetic energies are recalculated as well as new coordinates determined from the atom velocities, giving new outputs and this process is repeated until the simulation end.

2.4.1 Gromacs 4.5.4 and 4.6.5

Two different versions of GROMACS were used between Chapter 4 (GROMACS 4.5.4) and Chapters 5 and 6 (GROMACS 4.6.5) (Pronk et al. 2013; Abraham et al. 2015). In the earlier version (4.5.4) a group scheme was used which computed neighbour lists or pairs that were within the specified cut-off distance and the interactions between these were included in the system. However the drawback to this is that certain atoms may move across the cut-off boundary without being included in the current list because the lists are only updated every n steps. In the later version, GROMACS 4.6.5, this is addressed by implementing the Verlet cut-off scheme, which provides a pair list greater in size than the number of pairs within

the cut-off distance to cater for proportionally more interactions, though it is not totally inclusive.

2.4.2 Temperature and Pressure Coupling

In order to sample the energy surfaces during the simulation, an *NVT* protocol is adopted, which means the temperature of the system is held constant (canonical ensemble). This is implemented using the velocity-rescaling scheme (Bussi et al. 2007), which mimics the presence of an external heat bath at temperature T_0 that slowly corrects the system's temperature back to T_0 using Equation 2.8. The advantage of the velocity scheme over Berendsen thermostat is that has a further stochastic term, which corrects the distribution of kinetic energy.

$$\frac{\boldsymbol{dT}}{\boldsymbol{dt}} = \frac{(\boldsymbol{T_0} - \boldsymbol{T})}{\boldsymbol{\tau}}$$

Equation 2.8 The correction to an external heat bath at temperature T_o for the system, using the velocity-rescaling scheme.

In order to sample constant pressure simulations (NPT), a Parrinello-Rahman (Parrinello & Rahman 1981; Nosé & Klein 1983) approach is used which rescales the coordinates and box vectors at every step as given by Equation 2.9.

$$\frac{\boldsymbol{dP}}{\boldsymbol{dp}} = \frac{(\boldsymbol{P_0} - \boldsymbol{P})}{\boldsymbol{\tau_p}}$$

Equation 2.9 Scaling of the pressure to remain constant, according to the Parrinello-Rahman approach in NPT simulations.

2.4.3 Water Model

In this research, the Transferable Intermolecular Potential 3P (TIP3P) water model (Jorgensen et al. 1983).

2.4.4 System Set Up in GROMACS

Since the proteins explored in this thesis are all embedded in bilayers, the chosen shape of the simulation box was cuboid with the approximate dimensions 11 x 11 x 12 nm. The height of the box was particularly important to prevent loop regions of the protein from interacting with each other in the adjacent box as a result of the periodic boundary conditions implemented in GROMACS.

The first step in system set up is to ensure the correct protonation states are assigned to atoms in the system, ensuring the correct charges and bonds are described. Unless specifically specified, any acidic or basic residues with charges are left in their default charge states and the N- and C-termini residues are charged. The ligand parameters were first determined using topolbuild, and those atoms not identified by the programme were manually added to the itp file for the ligand, as described using the OPLS-AA atom parameters deemed most appropriate.

The protein and ligand complex was embedded within either a POPC or POPE (taken from the lipidbook: lipidbook.bioch.ox.ac.uk) bilayer using the g_membed feature of GROMACS. This resized the protein around its centre-of-mass by the factor 0.5 in the xy-plane (membrane plane) and overlapping lipids were removed. Intramolecular interactions in the protein were turned off and an MD step was performed, followed by incremental resizing to its initial state, and further lipids were removed at each step.

The system was then solvated in GROMACS (genbox) and ions (0.15 M NaCl) were added by randomly replacing water molecules (genion). The Berger lipid parameters were used to describe the lipid bilayers (Berger et al. 1997).

2.4.5 Energy minimisation

In order to check for incorrect geometries or steric clashes, each system was first relaxed using an energy minimisation step. A steepest descent algorithm was applied to the simulation, such that after the application of force F the forces were computed and checked against the previous step. They are accepted and the new atom geometries are taken if the energies in the new step were less than the previous, otherwise they were rejected. Each simulation was terminated when the absolute force values for each step are below the specified value: 1000 $kJmol^{-1}nm^{-1}$ or at the maximum number of steps was reached.

2.4.6 Equilibration

Two stages were used to equilibrate the system, the first of which applies the *NVT* ensemble region (constant number of particles, volume of system and temperature). Under these conditions the temperature should plateau. The second stage invoked a *NPT* ensemble regime (constant number of particles, pressure and temperature). Here a Parrinello-Rahman barostat (Parrinello & Rahman 1981; Nosé & Klein 1983) was added. Using this protocol, the box dimensions are updated according to the offset of the current pressure to the initial pressure. The Parrinello-Rahman protocol is best used once the system is close to equilibrium, since pressures far from this will cause large fluctuations in box size and can potentially

crash the system. Both stages were run for 1 ns simulation (method taken from Bevan laboratory protocol: www.bevanlab.biochem.vt.edu).

After successful completion of these stages, the system is ready for MD simulation. As stated previously, two versions of GROMACS were applied in this thesis, 4.5.4 and subsequently 4.6.5 after it's release in 2014. The advantage of 4.6.5 is the increased computational efficiency, since it is optimised to run on GPUs as non-bonded interactions can be calculated using the GPUs and PME interactions on the CPU.

2.4.7 Steered molecular dynamics

Molecular dynamics simulations can also be biased so as to force the system between conformational states of choice. In Chapter 6 a steered MD simulation using the PLUMED plugin for GROMACS (Bonomi et al. 2009) was implemented. A collective variable (CV) is defined and the protocol will then pull the system until the CV obtained a desired state. For the steered MD in this thesis, the CV was defined as two C-α atoms that were pulled to a specific distance (5 Å). At each time step the distance of the CV was recorded as well as the work placed on the system to bias it towards the target distance. The bias on the system is in the form of a harmonic potential that has a constant speed (which was user defined) and takes the form of Equation 2.10. There is a second term relating to a constant force, which can also be applied to the reference atoms to push the system towards the reference position s_0.

$$V_{smd}(s) = \frac{1}{2}k(s - s_0)^2 + m(s - s_0)$$

Equation 2.10 The harmonic potential applied to bias the system in the PLUMED plugin for GROMACS where k is the spring constant and m is the force applied in the steered MD. s_0 is the reference position which position s moves toward.

2.4.8 Analysis of simulations

Analysis of all MD simulations were performed by either the GROMACS toolkit, or in-house tcl (for VMD (Humphrey et al. 1996)), R, python or perl scripts.

3 Sequence analysis for MFS proteins

The aim for this chapter of research was to aid construction of MFS protein models. This was done by careful analysis of the conservation patterns of amino acids in multiple sequence alignments (MSAs), as well as contact preservation in MFS structures. The use of these techniques is well documented for GPCRs, 7 helix membrane-spanning proteins, which resulted in the prevalent numbering system for residues on helices of Class A GPCRs by Ballesteros and Weinstein in the early 1990s. More recently Venkatakrishnan et al. determined a set of structurally conserved sites within the GPCR fold that have topologically equivalent amino acids. This type of analysis is applied here to MFS transporters, producing a numbering scheme based on conservation in sequence and contacts between helices.

3.1 Introduction

The increasing number of protein structures solved by X-ray crystallography and NMR have applications in drug-target elucidation, for instance the structures can be used to perform virtual screening to support drug discovery programmes (Kitchen et al. 2004). The MFS transporters are no exception, with 44 distinct structures solved (as of April 2015) of which 29 have substrates bound to the central cavity between the two transmembrane (TM) domains. This is applicable to MFS transporters, which are implicated as components of the multidrug resistance in many species (Bambeke et al. 2000; Shaheen et al. 2015; Kumar et al. 2013) and in addition, an MFS protein, SV2A, contains the binding site for the anti-epileptic drug, LEV (Lynch et al. 2004).

A limiting factor to the use of homology modelling in MFS proteins is the low sequence identity between family members (Table 1.1). This means that MFS proteins fall into the 'twilight zone' where the highly conserved 12 TM helix fold is not conserved in sequence (Rost 1999). The result is that obtaining the perfect template-target alignment is impossible using current methods and other metrics are required to inform this alignment. A successful method that has been used to inform the alignment of disparate protein family members is the numbering scheme for helical residues in class A GPCRs, where the most conserved residue of each helix is labeled as 50 and then neighbouring residues are numbered accordingly (for example, the most conserved residue on TM helix 6 of rhodopsin is P267 and so its corresponding number is P6.50). This aids alignment of template and target in the helices of GPCRs (Figure 1.5).

3.1.1 Aims

The inspiration for the work in this chapter comes from the numbering system implemented for Class A GPCRs by Ballesteros and Weinstein (Ballesteros & Weinstein 1995), which is widely used. For MFS proteins, the development of a numbering system according to residue conservation could help to inform homology modelling of target MFS proteins. Investigation of both sequence conservation and conservation of helix-helix contacts throughout the whole MFS fold would provide structural 'anchor points' to help align templates to MFS query targets.

3.2 Methods

3.2.1 Multiple sequence alignment

Since MFS proteins have low sequence identity (Table 1.1), a multiple sequence alignment (MSA) of the whole superfamily was not feasible for determining conservation of residues that relate to structure since there is little confidence that the optimal alignment has been obtained (Tramontano 1998). It is difficult to define at what point the sequences become sufficiently similar and thus be able to infer structural information (Rost 1999). The general consensus is that so long as identity is above 50 %, the alignment quality will be high (Baker & Sali 2001) and so a method is required for MFS proteins that can exploit this.

Consequently MSAs were constructed using the MUSCLE algorithm (Edgar 2004) using sequences downloaded from UniRef50 clusters (Suzek et al. 2007). This was chosen because it clusters proteins with greater than 50 % sequence identity and removes redundant sequences (see Chapter 2.1.1). Therefore it provides a cluster of sequences that are at least 50 % identical to the investigated MFS. For each MFS protein with an X-ray crystal structure (Table 3.1), the UniRef50 cluster containing it was downloaded. The sequences were aligned using an automated python script (Appendix 9.1).

3.2.2 Analysis of MSA

Using an in house R script, conservation at each site was determined 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 was visually represented using the heatmap.2 package in R. This analysis was applied to the MFS proteins listed in Table 3.1.

Protein	MFS family	UniRef50 ID	Number proteins in cluster	Substrate	Direction of Transport	Species
FucP	MFS_1	UniRef50_P11551	752	fucose/H^+ symport	into cell	*E. coli*
GlpT	MFS_1	UniRef50_P08194	2154	glycerol-3P/Pi antiport	into cell	*E. coli*
EmrD	DHA12	UniRef50_P31442	794	multidrug/H^+ antiport	out of cell	*E. coli*
YajR	DHA12	UniRef50_P77726	608	multidrug/H^+ antiport	out of cell	*E. coli*
LacY	lacy_ symport	UniRef50_ P02920	1164	galactose/H^+ symport	into cell	*E. coli*
PepT	POT	UniRef50_Q8EKT7	10	peptide/H^+ symport	into cell	*Streptococcus thermophilus*
YgbH	POT	UniRef50_ P75742	503	peptide/H^+ symport	into cell	*E. coli*
GlcPse	MFS_1	UniRef50_ O07563	45	glucose/H^+ symport	into cell	*Staphylococcus epidermis*
gkPOT	POT	UniRef50_ Q5KYD1	35	peptide/H^+ symport	into cell	*Geobacillus kaustophilus*
GLUT1	SP	UniRef50_ P11166	698	glucose uniport	into cell	*Homo sapiens*
PiPT	SP	UniRef50_A8N031	2	phosphate/H^+ symport	into cell	*Piriformospora indica*
NRT1.1	NRT1/ PTR	UniRef50_ Q05085	320	nitrite/H^+ symport	into cell	*Arabidopsis thaliana*
NarU	NNP	UniRef50_P37758	2540	nitrate/nitrite symport	into cell	*E. coli*
XylE	SP	UniRef50_ P0AGF4	595	xylose/H^+ symport	into cell	*E. coli*
MelB	Glycoside-pentoside-hexuronide: cation symporter	UniRef50_ P30878	377	melibose/Na^+ symport	into cell	*Salmonella typhimurium*

Table 3.1 The MFS proteins used for determining key residues that are conserved across the superfamily. The cluster containing each protein sequence in UniRef50 was downloaded and aligned using MUSCLE.

3.2.3 Contact prediction and analysis

The contacts between helices are defined by distance (side chain atoms within 7 Å of the adjacent helix backbone atoms or side chain atoms that do not include hydrogen atoms, which are not present in X-ray crystal structures). For each MFS protein with an X-ray crystal structure (44 at the time of the analysis), contacts mediated by amino acid chemical type were noted (the contact between TMHs 1 and 5 is shown in Figure 3.1). Heat maps were constructed which showed the presence or absence of the particular residue type contact for all potential helix-helix interactions (see Appendix 9.3).

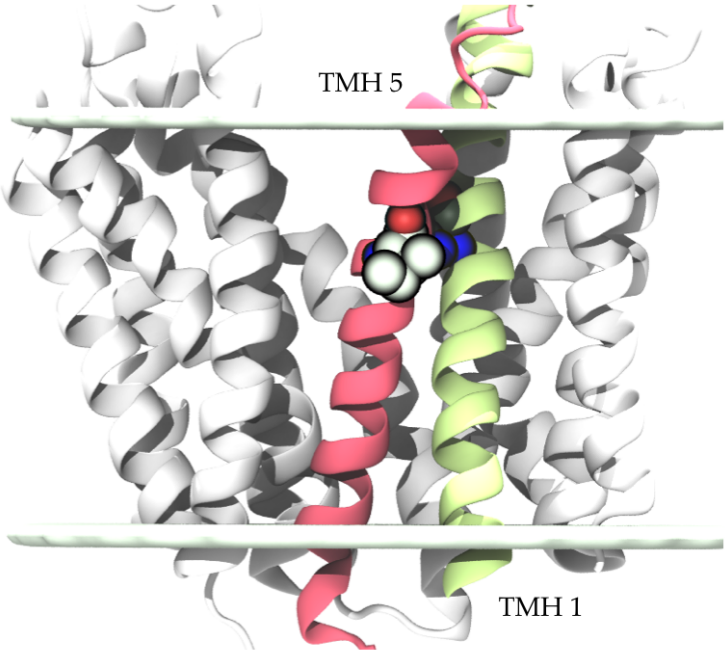


Figure 3.1 The contact point between TM helices 1 and 5 in FucP. The amino acid chemical types were determined from the multiple sequence alignments of the proteins in Table 3.1.

3.3 Results and Discussion

To improve the confidence in models built for novel MFS proteins, it is important to have a good alignment between the template protein (in this case, an MFS protein with a high resolution structure) and the target. The diversity in sequence between the families of MFS proteins makes this step non trivial, since the more distant two sequences are, the less able programmes are at getting the best, structurally relevant, alignment (Iantorno et al. 2014). Therefore, constructing a map of key residues that are conserved in all MFS proteins can be used to improve the alignment as this gives anchor points at key structural points in the tertiary fold of the protein to ensure the correct template-target alignment.

3.3.1 Multiple Sequence Alignment Analysis

Intuitively, the most obvious way to determine highly conserved regions of MFS proteins is to construct a superfamily-wide alignment of proteins, or use an existing MSA to then analyse. However, the compelling argument against using this approach for MFS proteins is the degree of sequence diversity (Table 1.1), since it is well documented that sequence identities below 30 % cannot be aligned with confidence (Baker & Sali 2001; Tramontano 1998).

As sequences become increasingly large, the time it would take to obtain the alignment grows exponentially, so it is unfeasible to determine the correct MSA (Wang & Jiang 1994). Instead, heuristic approaches are used to obtain the MSA, for example using progressive alignments which rely on a scoring function to give the best result (Edgar & Batzoglou 2006). The problem occurs with more distant sequences, as it is not trivial to align structurally similar regions since the sequences diverge at these points. For these, gaps become more prevalent and these distort the local alignments masking key regions. This is compounded by the issue that there is

no universal method for benchmarking MSA algorithms and so it is difficult to determine which is best in a particular situation (Iantorno et al. 2014).

With modern sequencing techniques, there has been an explosion in the number of protein sequences available and so there are multiple entries for proteins in different species. This means that analysis of MFS proteins can be narrowed, as there are sufficient numbers of highly homologous sequences for each protein. Here, the UniRef50 clusters contain proteins that have at least 50 % identity to the query protein. Each cluster is likely to only contain sequences from the same family within the superfamily and the high similarity between sequence means that gaps are less likely to obscure the results.

3.3.2 Conserved positions in MSAs

The highly conserved MFS fold led to the hypothesis that there would be further structural positions that can aid homology modelling, in addition to the identified motifs for MFS transporters (Figure 1.2). To explore this further, MSAs were analysed for each MFS protein that has an X-ray crystal structure. The conservation of amino acids and also amino acids grouped by chemical properties (see Chapter 3.2.2) and the occurrence rate for each, at each site in the MSA, was recorded as a percentage. The conservation for each protein was compared in order to determine whether conserved sites were present in all the explored MSAs. As the structures were known, the positions of helices could be placed onto the MSAs and so direct structural comparisons were made between each site in the MSA and the protein structure.

Those MFS proteins with X-ray crystal structures were used for the analysis in order to easily relate the conservation back to structure. The next step in the analysis,

after obtaining the conserved sites common to each of the MSAs for the investigated MFS proteins, was to determine whether the conserved residues were found at the same position on the helix relative to its position in the bilayer. The bilayer position was taken from the OPM (Orientations of Proteins in Membranes) database (Lomize et al. 2006) and the position of every conserved site common to each investigated MFS protein is recorded in Table 3.2. Images of each helix with both residue and bilayer position (according to the OPM database) are shown in the figure, which broadly suggests conservation of the position of the conserved site residues detected on each helix.

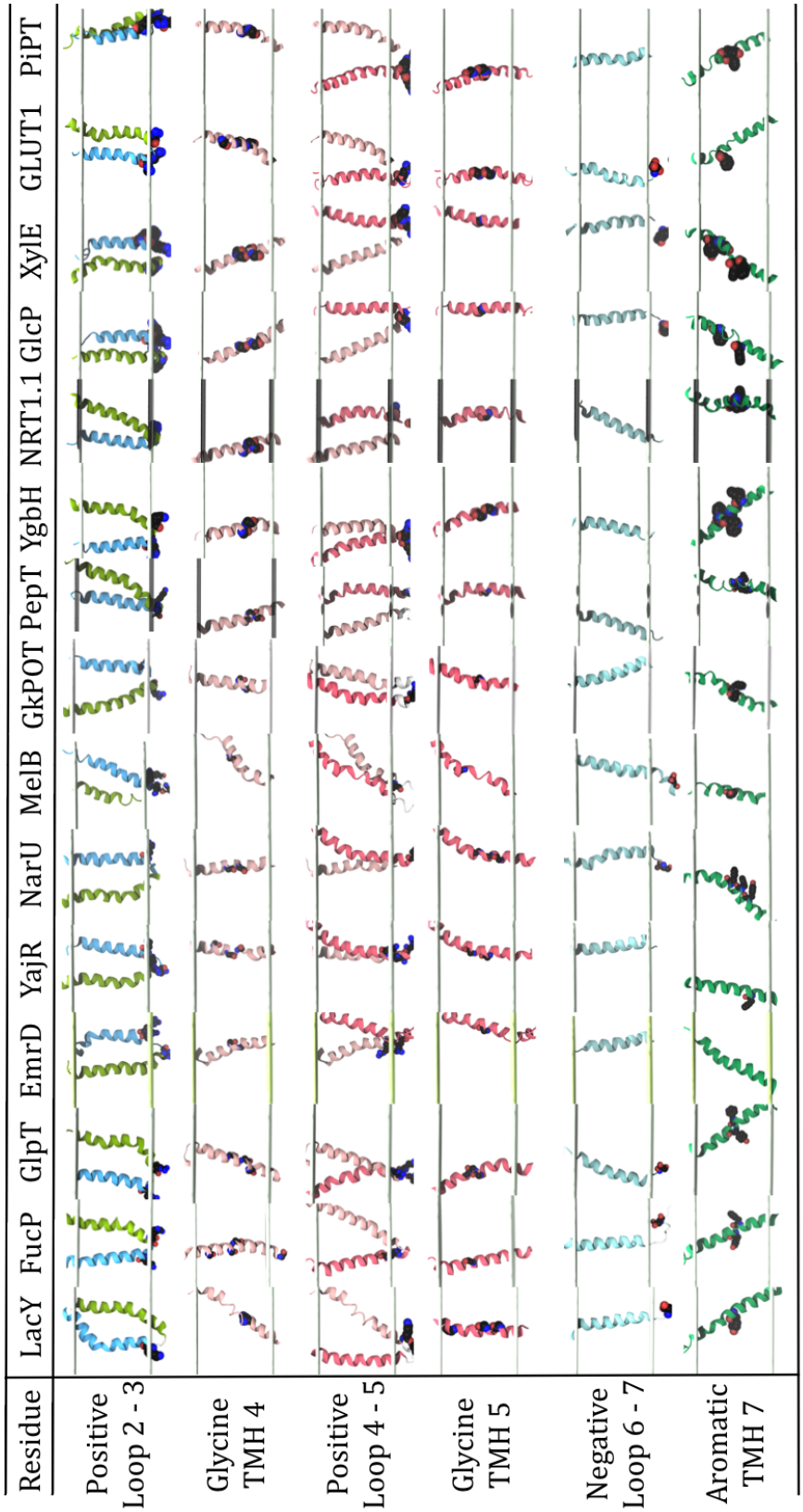


Table 3.2 The position of conserved residues taken from the analysis of multiple sequence alignments (MSAs) of homologous sequences to the X-ray crystal structure proteins. The conserved residue is represented by a black space fill representation, each helix is coloured according to Figure 1.1 and the bilayer is shown as white lines. Each conserved residue is conserved across all the MFS proteins, with some omissions, for example the TMH 7 aromatic in EmrD and many of the negative residues in Loop 6 – 7 and the C-terminal loop are not near to the end of the preceding helix. The residues also have spatially similar positions on the respective helices.

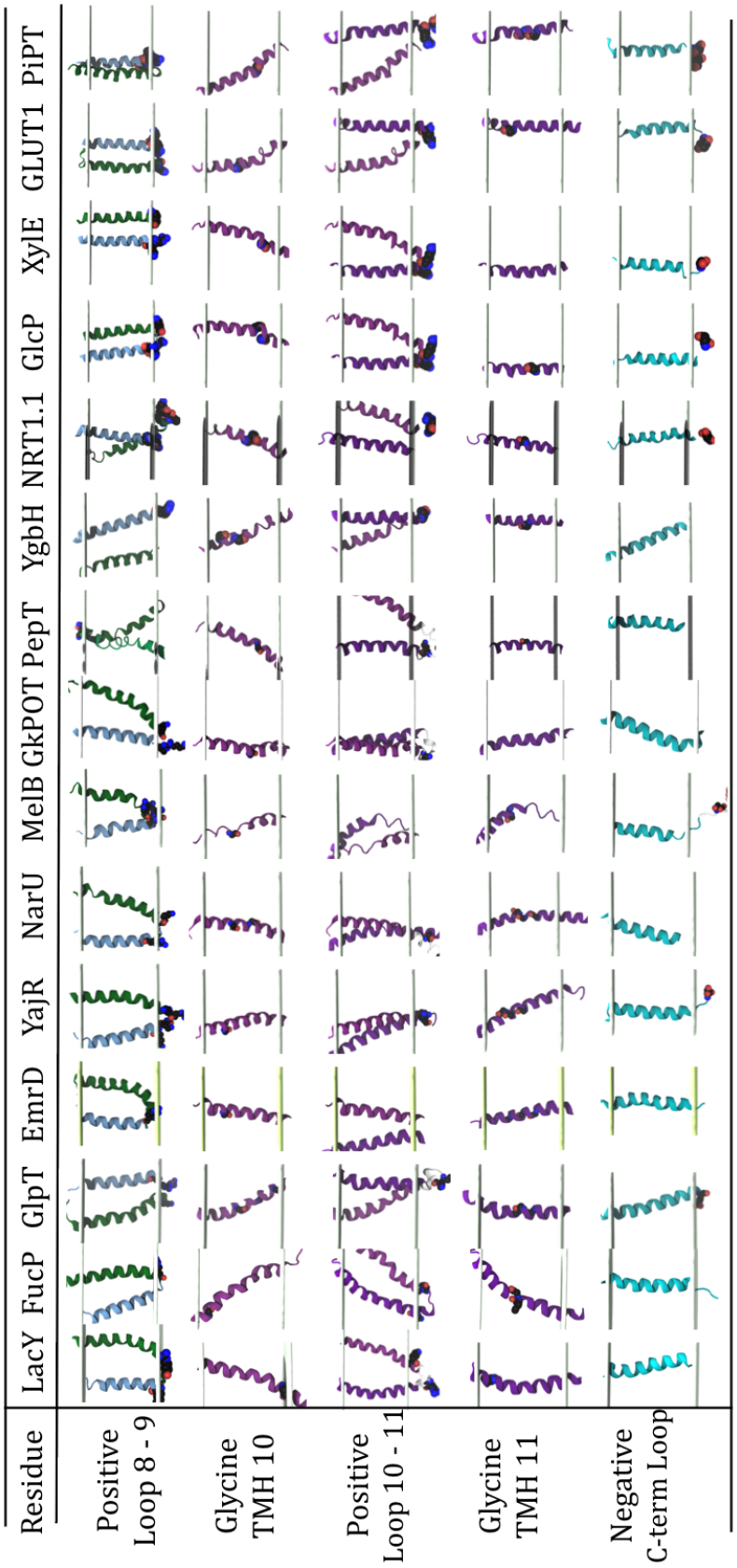


Table 3.2 Continued. The position of conserved residues taken from the analysis of multiple sequence alignments (MSAs) of homologous sequences to the X-ray crystal structure proteins. The conserved residue is represented by a black space fill representation, each helix is coloured according to Figure 1.1 and the bilayer is shown as white lines. Each conserved residue is conserved across all the MFS proteins, with some omissions, for example the TMH 7 aromatic in EmrD and many of the negative residues in Loop 6 – 7 and the C-terminal loop are not near to the end of the preceding helix. The residues also have spatially similar positions on the respective helices.

Two of the conserved sites highlighted by the MSA analysis (Table 3.1) are the glycine residues on TMH 4 and its symmetrically equivalent helix in domain 2, TMH 10. In every crystal structure these glycine residues were conserved in the helix, though the number of glycine residues varied. For example, there was predominantly only one glycine in TMH 10, but there were between one and four conserved glycine residues in TMH 4. A similar pattern continued for all conserved sites common to the helices in domain 1 and 2. This implies that even if the domains arose from an evolutionary repeat (which is believed to be the case, see Chapter 1.1.1), this was very distant in time and so the domains have evolved to have considerable dissimilarity between them. Similarly, the conserved aromatic residues at the centre of TMH 7 did not exist as a conserved residue in TMH 1. In LacY, Y236 (on TM helix 7) is known to be functionally important and so it could be that there is a functional pressure which maintains the presence of this conserved site as the removal of the residue impairs transport (Guan & Kaback 2006).

Despite these discrepancies, there were a series of residues that did exist in equivalent positions between the two domains. A glycine at the centre of TMH 5 was also present on TMH 11 and the positive residue in loop 2 – 3 was present in loop 8 – 9. The positive residue at loop 2 – 3 was part of a conserved motif known to exist in the MFS proteins (see Figure 1.2 (A)) and the positive residue of that motif was found in domain 2 (loop 8 – 9). Positively charged residues also existed in loops 4 – 5 and 10 – 11, indicating that MFS transporters conform to the positive-inside rule (von Heijne 1992). The final conserved residue was a negative charge at the end of TMH 6 and 12. However its position was not always conserved at the lipid head group regions of the TM helices but instead in the loop 6 – 7 or the N-terminal loop. Since these residues were not found near to the helices, it is not thought to be integral to the TM fold.

It is worth noting here that there are a few anomalies in the conservation of these key residues, in particular EmrD does not have the aromatic residue on TMH 7 and the positive charge in loop 10 – 11. With regards to the conserved aromatic residue on TMH 7, in some proteins there was more than one conserved aromatic residue. In order to identify which was the functionally or structurally relevant residue, the position of these aromatic residues was related to the percentage conservation of these points. Those which were most conserved were also those pointing into the central cavity, as seen in PepT (2XUT) and GLUT1 (4PYP) (Figure 3.2) and so these described the key conserved residue on TMH 7. These residues had similar orientations to the functionally important residue in LacY (Y236). A similar issue arises with TMHs 4, 5, 10 and 11, which may contain more than one conserved glycine and so these cannot always be used in isolation to aid homology modelling. Therefore, it becomes informative when used in conjunction with the contact analysis below.

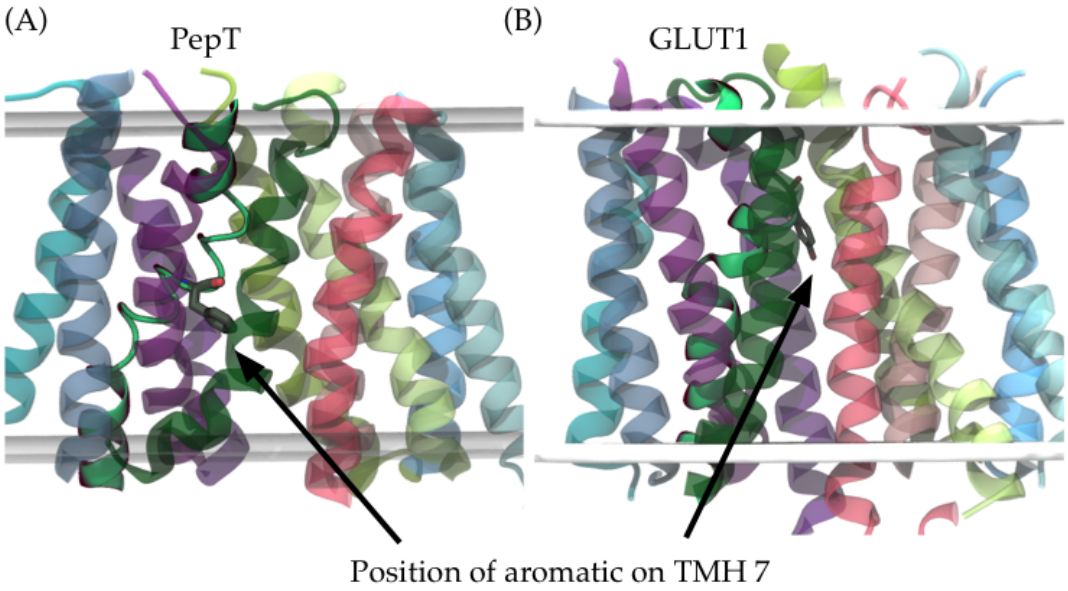


Figure 3.2 The position of the aromatic side chain on TM helix 7 for (A) PepT (occluded) and (B) GLUT1 (outward) (PDB IDs 2XUT and 2PYP respectively). In both proteins, the most conserved aromatic residue points into the central TM cavity between the two domains.

3.3.3 Contact analysis

The second stage of analysis looked at the conserved contacts in the X-ray crystal structures of MFS proteins. Thus, 44 structures were checked for helix-helix contacts between all 12 TM helices. The aim for this stage was to probe the structure to determine whether a series of contacts describe the distinct different conformations of MFS proteins. The resulting contacts can be broadly classed into two groups, those that mediate the packing of the helices and those contacts which act as pivot points, around which the protein moves throughout the alternating access mechanism. In Table 3.3 the contacts conserved across all 44 structures is given.

Residue involved in contact	Helix - Helix	Description
STNQ	1 – 2	No change in contact in different states, which keeps the periplasmic end of helices in contact
STNQ	2 – 11	Contact at periplasmic end and middle of the helices, which do not change with conformation
STNQ	5 – 8	The position of the contact is in the centre of the TM region, and the contact moves according to state
STNQ	7 – 11	The position of the contact is in the centre of the TM region, and the contact moves according to state
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

Table 3.3 Analysis of the position of the most conserved contacts between helices, as given by a series of heat maps (Figure 3.5 and Appendix 9.3). Each state is represented as: inward, LacY (1PV6); occluded, PiPT (4J05); outward, FucP (3O7Q) (see Figure 3.4).

Residue involved in contact	Helix - Helix	Description
Small Hydrophobic	2 – 4	The contact is in the middle of the helices and tilt remains constant – mediates packing
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	The contact is along the length of the helices – mediates packing
Small Hydrophobic	3 – 6	The contact is along the length of the helices – mediates packing
Small Hydrophobic	5 – 8	The contacts are in the centre of TM region
Small Hydrophobic	7 – 11	The contacts are in the 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 centre of the TM region, and the helices rock around that central point
Large Hydrophobic	3 – 6	The contacts are at the periplasmic end of the TM region and helices remain straight and inward structure has a further contact at cytoplasmic end of the TM region

Table 3.3 Continued. Analysis of the position of the most conserved contacts between helices, as given by a series of heat maps (Figure 3.5 and Appendix 9.3). Each state is represented as: inward, LacY (1PV6); occluded, PiPT (4J05); outward, FucP (3O7Q) (see Figure 3.4).

Residue involved in contact	Helix - Helix	Description
Large Hydrophobic	8 – 10	The contacts are in the centre of the TM region, with a small difference in TMH 10 tilt between the conformations
Glycine	1 – 5	The contact is in the centre 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 centre 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

Table 3.3 Continued. Analysis of the position of the most conserved contacts between helices, as given by a series of heat maps (Figure 3.5 and Appendix 9.3). Each state is represented as: inward, LacY (1PV6); occluded, PiPT (4J05); outward, FucP (3O7Q) (see Figure 3.4).

The first of the two groups of contacts are useful for model building as they are static entities that do not change with conformational state and so are structural anchor points (Figure 3.3 (A) and (B)). These are listed in Table 3.3 and include; glycine residues between TMHs 2 – 4 and 9 – 10, polar residues (S, T, N, Q) between TMHs 1 – 2 and 2 – 11, small hydrophobic residues (A, V, I, L) between TMHs 2 – 4, 3 – 4, 3 – 6, 8 – 10, 9 – 10 and 9 – 12, large hydrophobic residues (M, C, F, Y, W) between TMHs 2 – 4 and 8 – 10. There are four scaffold residues described in the literature (Doki et al. 2013; Yaffe et al. 2013), whose contacts remain unchanged across the conformations: TMHs 3, 6, 9 and 12 and the contact analysis agrees with this (TMHs 3 – 6, 9 – 12 which were both mediated by small hydrophobic residues and 3 – 6 large hydrophobic residues).

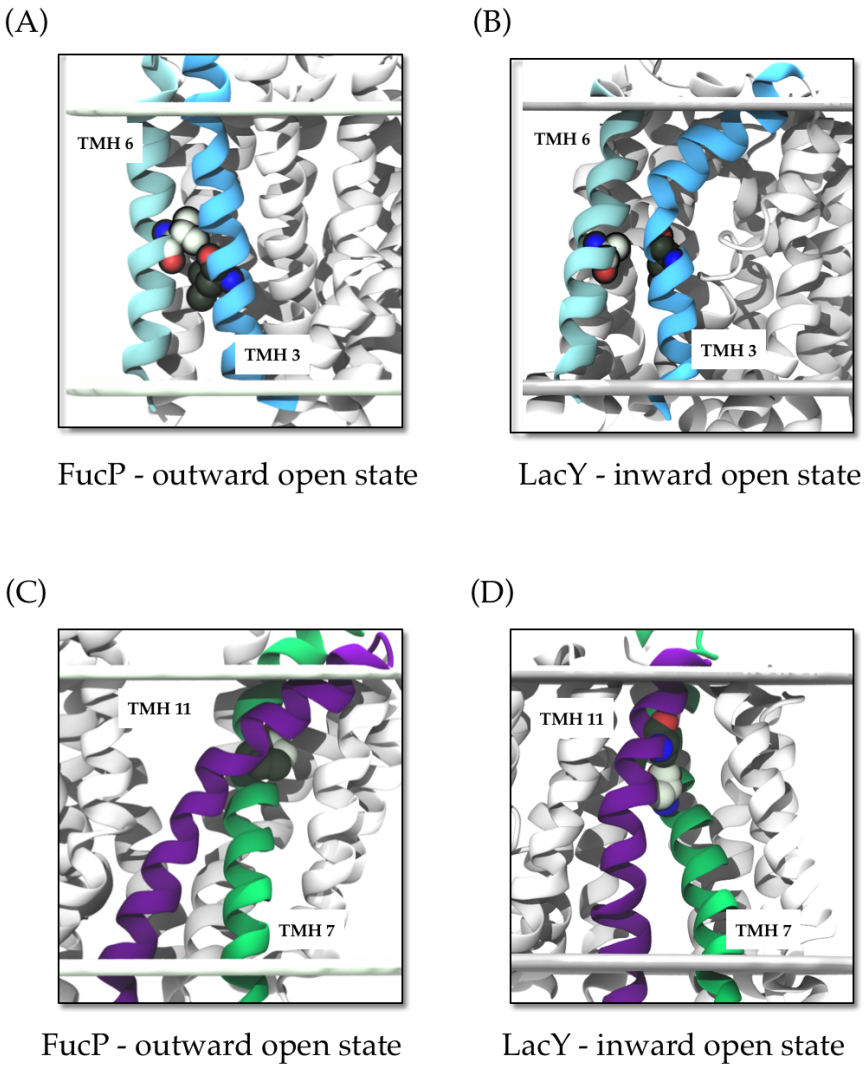


Figure 3.3 The static and moving contact types shown on FucP and LacY. (A) shows the static contact between TMHs 3 and 6 on the outward open FucP structure (3O7Q) and (B) shows the same contact in the inward open LacY structure (1PV6). The moving contact between TMHs 7 and 11 are similarly depicted on (C) FucP and (D) LacY. The angle between TMHs 7 and 11 changes, shifting the position of the contact, which does not occur between TMHs 3 and 6.

A further nine contacts mediate points in helix-helix interactions where the helices pivot between conformations, allowing them to move in the transport cycle (examples shown in Figure 3.3 (C) and (D)). The two domains are mediated by contacts between TMHs 2 – 11 and 5 – 8, so it is expected that these would have varied contacts according to the conformation. This is the case for 5 – 8, which has two contacts around which the helices pivot: polar residues and small hydrophobic residues. However, a series of contacts between TMHs 2 – 11 implied that the

contacts were reasonably static between the two helices, with only a small change in hydrophobic contacts. It was difficult to determine whether this was the case because the data was skewed towards occluded and inward open conformations (only two structures have on outward open state: FucP and GLUT1). Certainly the structures in Figure 3.4 suggest that the largest conformational change is outward to occluded, rather than 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 the remaining helices: 1, 4, 7 and 10.

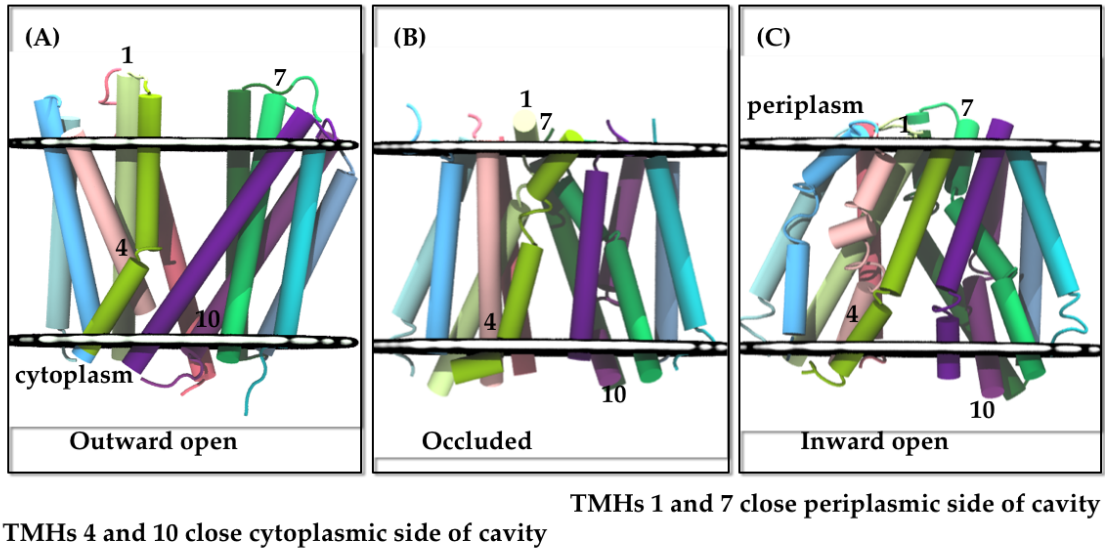


Figure 3.4 The X-ray crystal structures of (A) an outward (FucP), (B) an occluded (PiPT) and (C) and inward (LacY) facing conformation. These show that the key motions that describe the outward – inward movement are given by 8 cavity-lining helices, TMHs 1, 2, 4, 5, 7, 8, 10 and 11. In particular TMHs 1 and 7 close the periplasmic site of the protein and TMHs 4 and 10 close the cytoplasmic side of the TM cavity. What is also evident in these images is that a rigid body rotation of the two domains against each other is occurring, since the tilt relative to the bilayer of the four scaffold helices (3, 6, 9 and 12 which are all blue) does change from outward to occluded states.

The final conserved contacts are intra-domain and describe the movements that the helices undergo in the transport cycle. In domain 1, there was a rearrangement of the contacts with TMH 1: TMHs 1 – 5 and TMHs 1 – 6, which are mediated by small hydrophobic residues. In domain 2 the contacts involved in rearrangement of

helices are TMHs 7 – 11 (polar and small hydrophobic), where the largest motion was in TMH 7. Here TMHs 1 and 7 underwent the largest movements of contact positions of all the moving contacts. If the triple-helix model of MFS transporters is assumed correct for MFS proteins (Radestock & Forrest 2011), the similar helices to 1 and 7 would be TMH 4 and 10 and these latter helices do not have conserved contact movements in the same way as those of TMHs 1 and 7.

An important note to the analysis is 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. This does not appear to be the dominant motion conserved in the contacts, which show small motions in the contacts for TM helices 2 and 11 or TM helices 5 and 8. Instead TM helices 1 and 7 appear to move their contacts the most, and in fact move independently of their domains. 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, both further addressed in MD simulations in Chapter 6, which described the degree to which either side of the transporter is open. When TMHs 1 and 7 are interacting, the periplasmic side of the TM cavity is closed and TMHs 4 and 10 interacting imply that the cytoplasmic side is closed. Therefore, of all the residues it was these used to describe the degree of closure of either side of the TM cavity. It is possible to see such contacts in the polar residue contacts in Figure 3.5 (white boxes), where the TM helix 4 – 10 contact is present in the outward open X-ray crystal structures and TM helix 1 – 7 contact is present in inward open structures.

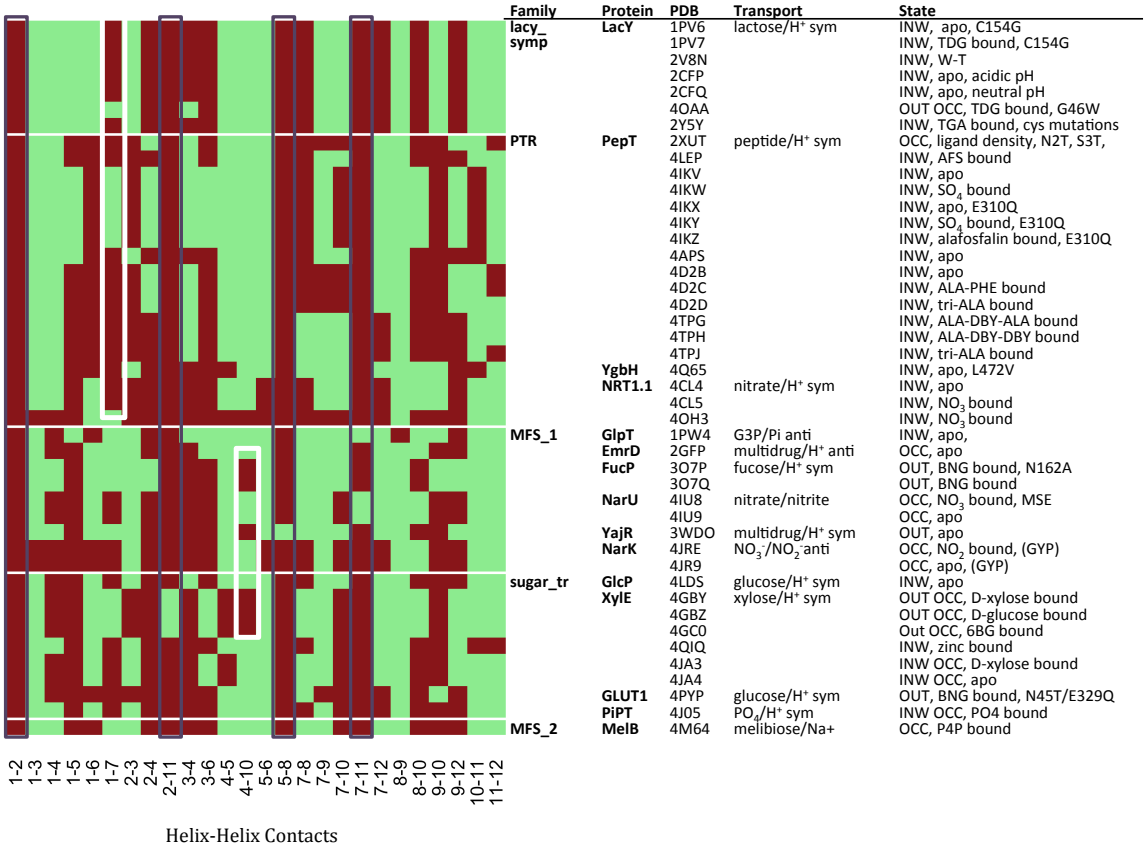


Figure 3.5 The contact map for polar residues (S, T, N, Q) in MFS X-ray crystal structures. The black boxes show the most conserved contacts, whilst the white boxes indicate the position of contacts that describe the degree of closure of the cavity at both the cytoplasmic (TMH 4 – 10) and extracellular (TMH 1 – 7) ends. This is an example of the heat maps used to determine the conserved contacts in MFS proteins (Appendix 9.3).

The contact analysis in the X-ray crystal structures added detail to the conservation patterns determined in the MSA analysis. For example, the glycine mediated contacts between TMHs 2 – 4, 1 – 5 and 9 – 10 helped to place the conserved glycine residues in the tertiary MFS structure. In an extension to this analysis, the contacts in Table 3.3 can also be used to aid packing of the helices in the MFS fold, by providing a framework in which to place the amino acid by chemical type; small and large hydrophobic, polar and glycine residues. This was used to aid the numbering of residue positions in MFS helices, in particular the static contacts.

3.3.4 Numbering MFS proteins

The final stage for the sequence analysis was to use all conservation information to determine a scheme for residue numbering on TM helices in MFS proteins. Five proteins were chosen to apply this protocol to: XylE, PepT, LacY, FucP and GLUT1. The resulting residues, which were numbered on each TM helix, are shown in Table 3.4 (only the canonical 12 TM helix fold was used for the numbering, so the extra two TM helices of PepT were ignored).

As a superfamily, the MFS sequences are highly diverse and so there was not enough confidence in superfamily alignments to give structural information. In Ballesteros and Weinstein's analysis (Ballesteros & Weinstein 1995), their numbering scheme was applicable to only one class, rather than all GPCR proteins. The aim for this research was to develop a method that extended this research, using information from conserved contacts, such that all MFS proteins could be compared. Therefore, a numbering scheme would contain those residues and contacts that exist across all MFS families investigated (these are listed in Table 3.1 and Table 3.2). A set of rules was implemented, which meant that cross-family conserved sites defined the x.0 point on helix x, and in subsequent helices the cross-family contact conservation was assigned as x.0.

The numbering of TM helices from 1 to 12 will be followed by a second number, which defines the distance between each residue and site 0. Here, site 0 is either the most conserved residue on that helix which is conserved across the superfamily (Chapter 3.3.2) or a contact conserved across the superfamily (Chapter 3.3.3). In cases where conserved residues across the superfamily do not exist (not present in Table 3.1), then the contacts are used as anchor points in the helices to number 0 (Table 3.2).

3.3.4.1 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. The second residue involved in these contacts (i.e. on the opposing helix) was used for numbering in TMHs 1, 2 and 9.
    - Thus positions on TMHs 1, 2 and 9 are defined this way.
3) For the remaining helices, the conserved superfamily-wide contacts were used. Residues were numbered by taking the most conserved residue within a contact.

Once x.0 is determined in a helix, the numbering will increase from 0 when moving towards 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 3.6). The numbering for XylE, PepT, LacY, FucP and GLUT1 helices are shown in Table 3.4.

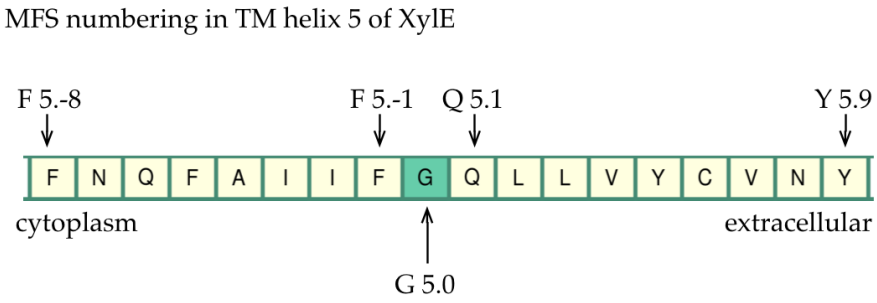


Figure 3.6 The numbering in TM helix 5 of XylE. The most conserved site is labeled G5.0 and then increasing negative values are given towards the cytoplasm and increasing positive values towards the extracellular cavity.

The majority of x.0 sites are conserved glycine residues, which are 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 was a polar-polar contact. The conserved aromatic residue on TMH 7 was used and numbered as 7.0. While this residue moved in its position on the helix relative to the bilayer position, it always faced into the TM cavity and so can define the rotation of TMH 7 in the bilayer. Pairs of contacts were used to describe the residues to number from in TMH pairs 2 and 4, 3 and 6 and these were glycine-glycine/large hydrophobic and small-small/large hydrophobic, respectively. For the remaining helices, residues were numbered by taking the most conserved residue within a contact, and these are described for each protein helix in Appendix 9.4. The contacts chosen for the numbering of MFS proteins were able to inform the rotation of the helices within the MFS fold by defining a point at which the helix interacts with another helix.

Protein	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
XylE	17-	57-	87-	126-	166-	201-	281-	316-	343-	371-	410-	443-
(4JA4)	33	74	104	145	183	218	301	333	361	390	427	463
	G25	G71	A92	G141	G174	A210	Y298	T329	G348	G388	Q415	A456
	G1.0	**G2.0**	**A3.0**	**G4.0**	**G5.0**	**A6.0**	**Y7.0**	**T8.0**	**G9.0**	**G10.0**	**G11.0**	**G12.0**
PepT	18-	53-	86-	110-	153-	178-	304-	335-	376-	404-	444-	483-
(2XUT)	39	75	105	129	172	195	324	352	393	424	464	500
	A22	G66	L93	G124	G160	F189	F315	A342	G379	G418	S449	A489
	A1.0	**G2.0**	**L3.0**	**G4.0**	**G5.0**	**F6.0**	**F7.0**	**A8.0**	**G9.0**	**G10.0**	**S11.0**	**A12.0**
LacY	10-	41-	75-	103-	143-	167-	222-	257-	289-	312-	352-	380-
(1PV6)	35	64	96	128	163	187	248	278	309	332	373	399
	G13	F55	L84	G111	G147	A177	Y236	T266	G296	G332	S366	L390
	G1.0	**F2.0**	**L3.0**	**G4.0**	**G5.0**	**A6.0**	**Y7.0**	**T8.0**	**G9.0**	**G10.0**	**S11.0**	**L12.0**
FucP	28-	64-	90-	119-	154-	210-	261-	301-	326-	348-	383-	412-
(3O7Q)	47	86	108	140	174	229	282	319	345	372	403	430
	N43	G73	L98	G132	G165	V219	Y270	V306	L328	G372	T390	A419
	N1.0	**G2.0**	**L3.0**	**G4.0**	**G5.0**	**V6.0**	**Y7.0**	**V8.0**	**L9.0**	**G10.0**	**T11.0**	**A12.0**
GLUT1	14-	64-	94-	119-	157-	187-	275-	306-	335-	366-	402-	431-
(4PYP)	36	91	112	140	176	206	295	326	354	387	426	449
	G27	G79	A103	G134	G167	A197	F291	I315	G340	G382	N415	L441
	G1.0	**G2.0**	**A3.0**	**G4.0**	**G5.0**	**A6.0**	**F7.0**	**I8.0**	**G9.0**	**G10.0**	**S11.0**	**L12.0**

Table 3.4 The numbering in MFS proteins. In the five proteins, XylE, PepT, LacY, FucP and GLUT1, the 0 point residue of each helix is defined as either the most conserved site across all the MSAs or the most conserved contact between helices and which is present in all crystal structures. This point in the helix is numbering 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 towards the cytoplasm are labeled as negative numbers and residues toward the periplasm are given positive numbers (for example TM helix 5 in XylE is FNQFAIIF**G**QLLVYCVNY giving: … , I5.-2, F5.-1, G5.0, Q5.1, L5.2, …).

The numbered sites for XylE, PepT, LacY, FucP and GLUT1 are shown in Figure 3.7. For each helix, the 0 position is in structurally similar positions, when compared to the rotation of the helices in the X-ray crystal structures. Importantly, the numbering scheme can be applied to target sequences now that a series of defined metrics can be looked for in a multiple sequence alignment. After obtaining an MSA for the target, the contacts relate to the most conserved amino acids of that chemical type on each helix. In their respective MSAs, contacts identified as key for the numbering scheme of XylE, PepT and LacY (given in the rules in Chapter 3.3.4.1) were subsequently utilized to determine FucP and GLUT1. Here, the contacts could be identified through sequence conservation of the amino acid type and the structure was used to check that the correct site was chosen for numbering. Therefore, a set of key contacts has been identified for the MFS numbering scheme and this can be identified in target MSAs, with the caveat that some numbered site occur one helix away from the numbered site in the other proteins (FucP 1.0, FucP 2.0, LacY 4.0, FucP and GLUT1 5.0, LacY and FucP 11.0 and XylE 12.0).

However there are two exceptions to this, which are the numbers for TMHs 1 and 7. The conserved aromatic residue on TMHs 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 contact and so its position is shifted by 1 helical turn according to the protein explored. The second site in a different position is number 1.0 in PepT and LacY, which differ from the sites on TMH 1 of XylE, FucP and GLUT1. The numbered residue in TMH1 changed according to the protein investigated as did the position. This can be accounted for by the flexibility of TMH 1 in the transport mechanism and so the contact point changes according to conformation. (The sites are shown in snapshots of the structures in Appendix 9.5).

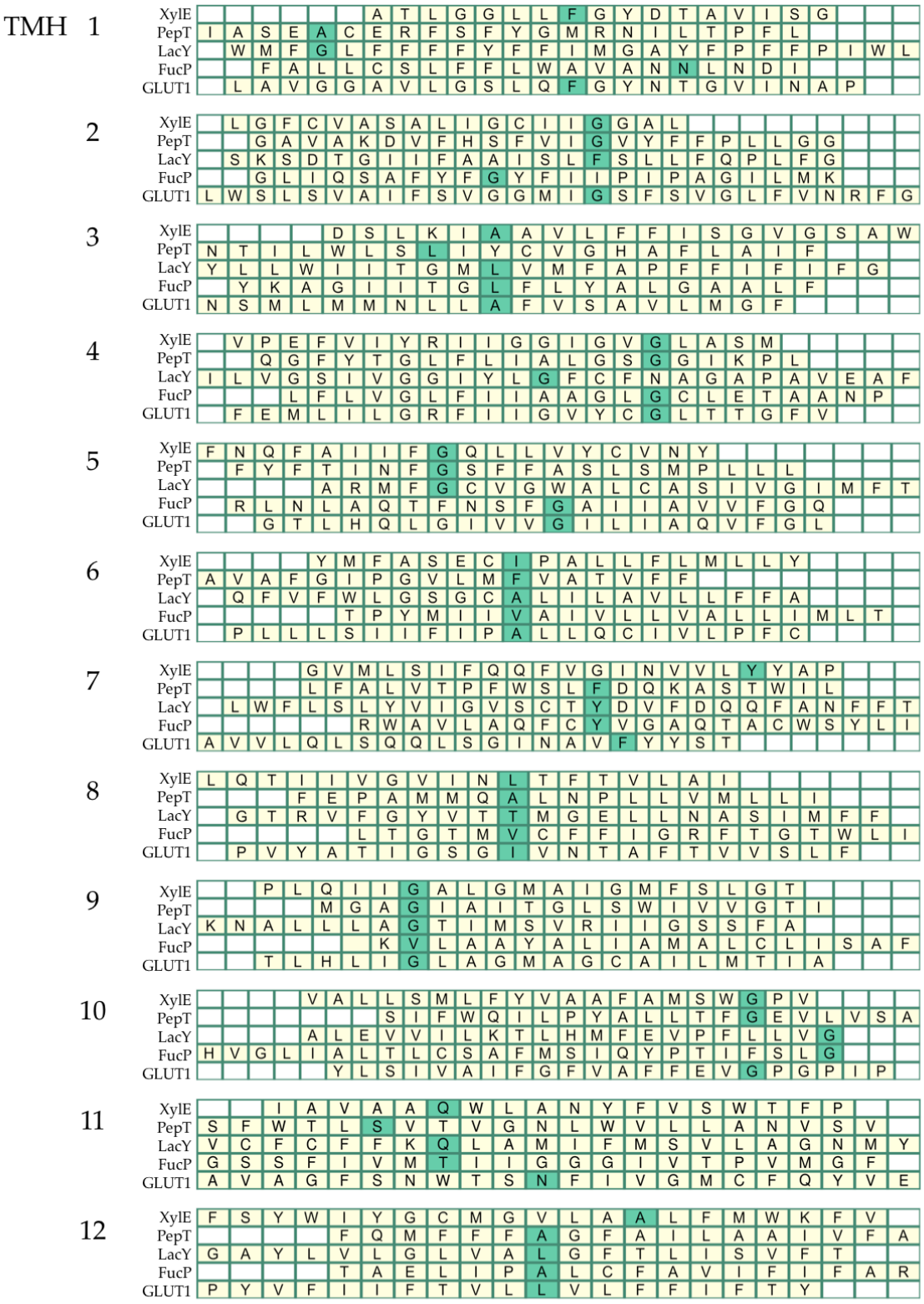


Figure 3.7 The position of the numbered residue on each helix. TM helices 1 – 12 respectively are shown for XylE, PepT, LacY, FucP and GLUT1 and each line of the heat maps corresponds to the helix in each crystal structure. The helix lengths are defined by those within the region of the bilayer given in the OPM database and helices are different lengths depending on the helix tilt and whether there are any kinks. The helices are aligned according to the rotation of the helix, so that numbered sites in the same structural position are aligned.

3.4 Conclusions

The aim for this work was to use conservation of both sequence and contacts within the 12 TMH fold of MFS proteins to inform homology modelling. A series of conserved glycine residues across all the multiple sequence alignments (MSAs) also had conserved position and were used to start a numbering scheme in MFS transporters, becoming G 4.0, G 5.0 and G 10.0 as well as an aromatic residue on TMH 7 (Y 7.0 in LacY). In addition, contacts that occur in all the MFS X-ray crystal structures (44 as of April 2015) provided a series of contacts mediated by amino acid side chains according to their chemical composition (for example small hydrophobic residues are alanine, valine, isoleucine and leucine). These provided a protocol for determining the number in the final helices: 1, 2, 3, 6, 8, 9, 11 and 12.

Further to this, the conservation of the contacts was compared with the conformation of the structures. The differences between the states highlighted that TMHs 1 and 7 contacts change the most between the structures, which ties with the hypothesis that these helices mediate gating of the periplasmic side of the TM cavity, as key cavity lining helices. Furthermore, the majority of contacts remained static across the different states. There are only small and helix specific (TMHs 1, 7 and 10 predominantly) rearrangements that take place on top of the rigid body rotation that each domain, thought to occur during the transport mechanism of an MFS transporter.

The success of the numbering scheme for MFS transporters comes from its utility for homology modelling. Now that a set of conserved metrics in each helix of MFS proteins has been identified, these can be identified in target proteins by exploring the MSA of that target. Therefore its real test is whether it can be used to inform the alignment of sequences for template-target proteins.

4 Modelling SV2A

The Major Facilitator Superfamily (MFS) protein, SV2A, is the target for the anti-epileptic drug, Levetiracetam (LEV). The drug is confirmed to bind with SV2A, but the mode of action of LEV remains unclear. In the absence of a structure of SV2A, a model of SV2A was constructed using methods in the previous chapter, in order to better understand the ligand-binding site. Due to its putative transport function and not knowing to which conformation in the transport cycle LEV binds, both an outward and inward facing model of SV2A (with respect to the cytosol) were explored, as was subsequent ligand docking of a LEV homologue. The results suggest a putative racetam binding site and inform further site-directed mutagenesis experiments.

4.1 Introduction

SV2A is a member of the MFS transporters and so is predicted to have the canonical 12 TM helix and undergo the alternating access mechanism in its putative galactose transport function (Madeo et al. 2014). As well as this, it is the binding site for the anti-epileptic drug (AED), LEV, which acts on a novel pathway to other AEDs (Lynch et al. 2004).

Tomography experiments indicate at least two major conformations of SV2A (Lynch et al. 2008) and LEV binding does not cause large-scale conformational changes in SV2A nor does it appear to lock the protein in a preferred state (Inward, Outward or Occluded). However saturation binding assays for ucb 30889 in the presence of ucb

1244283 showed a 5-fold increase in affinity, which suggests allosteric modulation of SV2A (Daniels et al. 2013), and so there is potentially more than one racetam binding site in SV2A. Work by Shi et al. (Shi et al. 2011) determined 13 residues important for binding the racetam, ucb 30889 (Figure 1.7 (C)), which were chosen according to their alignment to functionally relevant residues in LacY. The position of TM helices helped inform the alignment and these were predicted for SV2A using the consensus of three algorithms, SOSUI (Hirokawa et al. 1998), TMHMM (Krogh et al. 2001) and MINNOU (Cao et al. 2006). This was further refined using FUGUE (Shi et al. 2001) to identify structural environment features.

As an extension to this work, two alternate conformations of SV2A were considered; inward-open (based on GlpT: 1PW4) and outward-open (based on FucP: 3O7P) with respect to the cytosol. The aim for the research was to further characterize the racetam binding site, thought to be analogous to the ligand binding site in LacY, therefore in the central binding cavity of SV2A. On the basis of *in-silico* modelling and molecular dynamics (MD) simulations, three additional residues were found that line the binding pocket for ucb 30889, which were subsequently confirmed by site-directed mutagenesis in conjunction with binding affinity assays. As this work was conducted in conjunction with the previous chapter, the numbering scheme was not completed and so could not be implemented in the SV2A research.

4.2 Methods

4.2.1 Sequence analysis and model building

A final consensus transmembrane (TM) helix prediction was made using SOSUI (Hirokawa et al. 1998), HMMTOP (Tusnády & Simon 2001), JPRED (Cole et al. 2008)

and PSIPRED (Jones 1999b), with two templates used to construct the homology models in inward-open (GlpT – PDB: 1PW4 (Lemieux et al. 2005)) and outward-open (FucP – PDB: 3O7P (Dang et al. 2010)) conformations. The predicted TM helices of SV2A were aligned to those in GlpT and FucP respectively and models were generated using Modeller 9.10 (Eswar, Webb, Marti-Renom, Madhusudhan, Eramian, M. Shen, et al. 2007). This research was conducted in conjunction to the previous chapter. Therefore, refinement of these models used a multiple sequence alignment of SV2A to identify key conserved sites in MFS proteins to inform helix position. The numbering system was not tested on SV2A. The final structures were chosen according to the best DOPE score (Shen & Sali 2006) and further assessed using a QMEAN score (Benkert P. Biasini & Schwede 2011).

4.2.2 Molecular dynamics simulation

The final models were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer using the g_membed (Wolf et al. 2010) feature of GROMACS and an energy minimisation with a steepest descent algorithm until convergence with a force tolerance of 0.239 kcal mol^{-1} $Å^{-1}$ was performed. Sodium and chloride ions were then added to the systems to a concentration of 150 mM followed by two restrained MD runs whereby all heavy atoms were restrained by a harmonic potential of 2.39 kcal mol^{-1} $Å^{-2}$ for 1 ns. Finally, 80 ns of production runs were performed on three repeats that differed in their initial velocities only. As an additional check of the simulations we analysed the area per lipid for all simulations and observed that there was no significant difference compared to t = 0 or between runs. The areas per lipid (in $Å^2$) for the Inward-apo, Inward-ucb 30889, Outward-apo and Outward-ucb 30889 simulations were 65.58 ± 0.82, 68.63 ± 0.74, 68.21 ± 0.71 and 67.49 ± 0.71 respectively. Simulations were performed for inward and outward models both in the apo state and in complex with the radioligand (used in previous

functional assays) ucb 30889 as summarised in Table 4.1. MD simulations were carried out with GROMACS v4.5.4 (Hess et al. 2008) using the OPLS-AA (Jorgensen et al. 1996; Kaminski et al. 2001b) force-field and TIP3P water molecules (Jorgensen 1981). Production simulations were performed in an NPT ensemble maintained at 323 K and 1 bar pressure. The integration time step was set as 2 fs and a stochastic dynamics integrator (Van Gunsteren & Berendsen 1988) was used. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method (Essmann et al. 1995) with a 14 Å cut-off and 1 Å space grid. The Lennard-Jones potential used a cut-off of 9 Å, with a switch at 8 Å. The LINCS algorithm (Hess et al. 1997) was used in order to constrain bond lengths in both the lipid molecules and the protein. For the simulations with ucb 30889 bound, the protein coordinates were taken from the apo simulation at 80 ns timestep and Autodock Vina (Trott & Olson 2010) was used to dock the ligand into the cavity to provide the initial starting coordinates for the protein-ligand complex. The docking grid was defined as a rectangle that encompassed the TM cavity between the lipid head groups of the bilayer. The conformation of ligand chosen was that with the best score in Autodock Vina (Trott & Olson 2010) and energy minimisation of the complex was conducted as above, before MD simulation.

Name	Description
Inward-apo	Model based on GlpT template
Outward-apo	Model based on FucP template
Inward-ucb 30889	Inward apo model with ucb 30889 docked into the TM cavity
Outward-ucb 30889	Outward apo model with ucb 30889 docked into the TM cavity

Table 4.1 A summary of the simulations of SV2A. Both apo- and ucb 30889 docked to the central cavity of the TM region of SV2A. Simulations were 80 ns in length and had three repeats.

4.3 Results and Discussion

This work focused on a better understanding of the binding site for racetams in SV2A. Since there is no crystal structure of SV2A, a model was constructed exploiting its close homology with Major Facilitator Superfamily (MFS) transporters. Despite a low sequence identity between SV2A and solved MFS structures (the TM domains of SV2A have only 16 % identity with GlpT, PDB: 1PW4), and so are reasonably confident in it maintaining the 12 TM α-helix fold. This is totally conserved in all solved MFS structures. Here, the position of the 12 TM α-helices in SV2A are predicted using conserved residue positions so that the structural alignment of the SV2A sequence to that of the templates predicts with more confidence the tertiary packing of these α-helices. The LEV homologue, ucb 30889 (Figure 1.7 (C)) was then docked to the SV2A model to give insights into the racetam binding site.

4.3.1 The models of SV2A

Two models of SV2A were constructed, of outward and inward facing conformations with respect to the cytoplasm (Figure 1.3). The transport cycle of MFS proteins goes from apo outward open, where substrate (and any co-translated ions) binds to the TM cavity and this starts the movement to occluded and finally inward open to allow substrate to pass into the cytoplasm. This does not preclude cycling in the opposite direction, which is possible in LacY and is controlled by either the ligand or proton concentration gradient (Guan & Kaback 2006). Recent research on SV2A found it acts as a sugar transporter and restores growth in hexose transport-deficient EBY.VW4000 yeast cells (Madeo et al. 2014). Since the cells survive and presuming the transporter inserts correctly into the membrane, this suggests cycling from outward to inward of sugars by SV2A which agrees with the

cycle in Figure 1.3. So SV2A transport direction would be into the cytoplasm, suggesting that the outward facing state is the apo conformation to which sugar binds, similar to FucP (Dang et al. 2010), though it is unclear whether transport occurs while SV2A resides on the synaptic vesicle (SV) membrane or the synapse. If racetams were to influence substrate transport, it is likely that they would affect the relative stability (or longevity) of the states in Figure 1.3 or the ability of the substrate to bind to SV2A, possibly through binding into the substrate binding site itself, which is the suggestion given the positions of residues whose mutation affects racetam binding (Shi et al. 2011). Since it is not known which conformation this is more likely to happen in, both inward and outward models of SV2A were explored.

```
HMMTop       MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60
PSIPred      MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60
SOSUI        MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60
JPRED jnet   MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60
JPRED jhmm   MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60
JPRED jpssn  MEEGFRDRAAFIRGAKDIAKEVKKHAAKKVVKGLDRVQDEYSRRSYSRFEEEEDDDDFPA 60

HMMTop       PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120
PSIPred      PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120
SOSUI        PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120
JPRED jnet   PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120
JPRED jhmm   PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120
JPRED jpssn  PADGYYRGEGAQDEEEGGASSDATEGHDEDDEIYEGEYQGIPRAESGGKGERMADGAPLA 120

                                                          ______________
HMMTop       GVRGGLSDGEGPPGGRGEAQRRKDREELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180
PSIPred      GVRGGLSDGEGPPGGRGEAQRRKDREELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180
SOSUI        GVRGGLSDGEGPPGGRGEAQRRKDREELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180
JPRED jnet   GVRGGLSDGEGPPGGRGEAQRRKDRELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180
JPRED jhmm   GVRGGLSDGEGPPGGRGEAQRRKDREELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180
JPRED jpssn  GVRGGLSDGEGPPGGRGEAQRRKDREELAQQYETILRECGHGRFQWTLYFVLGLALMADG 180

             ____________________  ____________________________  ________
HMMTop       VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240
PSIPred      VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240
SOSUI        VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240
JPRED jnet   VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240
JPRED jhmm   VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240
JPRED jpssn  VEVFVVGFVLPSAEKDMCLSDSNKGMLGLIVYLGMMVGAFLWGGLADRLGRRQCLLISLS 240

             ____________________  ____________________________  _________
HMMTop       VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
PSIPred      VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
SOSUI        VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
JPRED jnet   VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
JPRED jhmm   VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
JPRED jpssn  VNSVFAFFSSFVQGYGTFLFCRLLSGVGIGGSIPIVFSYFSEFLAQEKRGEHLSWLCMFW 300
```

Table 4.2. The consensus prediction of secondary structure. The position of TM α-helices in SV2A (red) and β–sheets (blue) using HMMTop (Tusnády & Simon 2001), PSIPred (Jones 1999a), SOSUI (Hirokawa et al. 1998) and JPRED (Cole et al. 2008). The predicted TM helices for SV2A are indicated by black bars across the top the alignment.

```
                ____________________                ____________________
HMMTop          MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360
PSIPred         MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360
SOSUI           MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360
JPRED jnet      MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360
JPRED jhmm      MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360
JPRED jpssn     MIGGVYAAAMAWAIIPHYGWSFQMGSAYQFHSWRVFVLVCAFPSVFAIGALTTQPESPRF 360

HMMTop          FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420
PSIPred         FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420
SOSUI           FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420
JPRED jnet      FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420
JPRED jhmm      FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420
JPRED jpssn     FLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQRW 420

                                    ____________________________________
HMMTop          GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480
PSIPred         GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480
SOSUI           GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480
JPRED jnet      GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480
JPRED jhmm      GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480
JPRED jpssn     GVRALSLGGQVWGNFLSCFSPEYRRITLMMMGVWFTMSFSYYGLTVWFPDMIRHLQAVDY 480

HMMTop          AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540
PSIPred         AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540
SOSUI           AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540
JPRED jnet      AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540
JPRED jhmm      AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540
JPRED jpssn     AARTKVFPGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVT 540

                                                              _____
HMMTop          SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600
PSIPred         SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600
SOSUI           SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600
JPRED jnet      SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600
JPRED jhmm      SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600
JPRED jpssn     SSNTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMVYFV 600

                ____________________ ____________________ ____________
HMMTop          SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660
PSIPred         SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660
SOSUI           SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660
JPRED jnet      SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660
JPRED jhmm      SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660
JPRED jpssn     SFLGTLAVLPGNIVSALLMDKIGRLRMLAGSSVLSCVSCFFLSFGNSESAMIALLCLFGG 660

                ____________________ ____________________ _________
HMMTop          VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720
PSIPred         VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720
SOSUI           VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720
JPRED jnet      VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720
JPRED jhmm      VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720
JPRED jpssn     VSIASWNALDVLTVELYPSDKRTTAFGFLNALCKLAAVLGISIFTSFVGITKAAPILFAS 720

                ____________
HMMTop          AALALGSSLALKLPETRGQVLQ 742
PSIPred         AALALGSSLALKLPETRGQVLQ 742
SOSUI           AALALGSSLALKLPETRGQVLQ 742
JPRED jnet      AALALGSSLALKLPETRGQVLQ 742
JPRED jhmm      AALALGSSLALKLPETRGQVLQ 742
JPRED jpssn     AALALGSSLALKLPETRGQVLQ 742
```

Table 4.2 Continued. The consensus prediction of secondary structure. The position of TM α-helices in SV2A (red) and β–sheets (blue) using HMMTop (Tusnády & Simon 2001), PSIPred (Jones 1999a), SOSUI (Hirokawa et al. 1998) and JPRED (Cole et al. 2008). The predicted TM helices for SV2A are indicated by black bars across the top the alignment.

Prediction of TM helices is relatively well researched, with consensus predictions taken from multiple algorithms giving confidence in the helices in SV2A (Table 4.2). The positions of the α-helices in the SV2A sequence were further verified using a multiple sequence alignment of 758 sequences (Methods 4.2) that investigated the conservation of key chemical properties in the side chains, most notably hydrophobic conservation for helix prediction. The hydrophobic character of the lipid bilayer tails implies that evolutionary pressure would see conservation of hydrophobic side chains in TM regions of proteins. The heat map in Figure 4.1 was used in conjunction with secondary structure predictors to increase confidence in helix placement in the SV2A sequence. This information was used to structurally align the helices of SV2A to those in GlpT (inward facing conformation) and FucP (outward facing conformation), with the hydrophobic conservation patterns acting as a refinement step. The periodicity of hydrophobic conservation can be seen in the SV2A heat map (in particular for TM helices 1, 2 and 11), and this can help determine the helix face adjacent to the central TM cavity, which is solvent accessible and so is expected to have less well conserved hydrophobic side chains in the equivalent site in the alignment. Therefore hydrophobic conservation can pick out helix faces exposed to the solvent in the central cavity.

The quality of the 100 models produced in each production run of Modeller 9.10 were assessed with QMEAN (Benkert P. Biasini & Schwede 2011) and the scores indicated an accepted range for membrane proteins, providing confidence when picking the best model of each run using DOPE scoring (Shen & Sali 2006). The inward-facing model had a score of 0.301 (the range for all 100 models was between 0.301 and 0.401) which when compared to the GlpT template, which has a score of 0.525, was considered reasonable. The outward-facing model had a score of 0.381 (the range for all 100 models was between 0.317– 0.388) which when compared to the FucP template with a QMEAN score of 0.512 was also considered reasonable

(Benkert et al. 2009). The QMEAN score is a range that can determine the viability of either models or X-ray crystal structures, according to 'degree of nativeness' criteria that is quantified by a z-score. The scores are computed using a distance – dependent interaction potential of mean force based on the C-α atoms in the model, analysing the backbone geometry and solvation potential. If the model's QMEAN score is within or near to that range (since membrane proteins often sit slightly outside the range optimized for globular proteins), then it is considered a reasonable model.

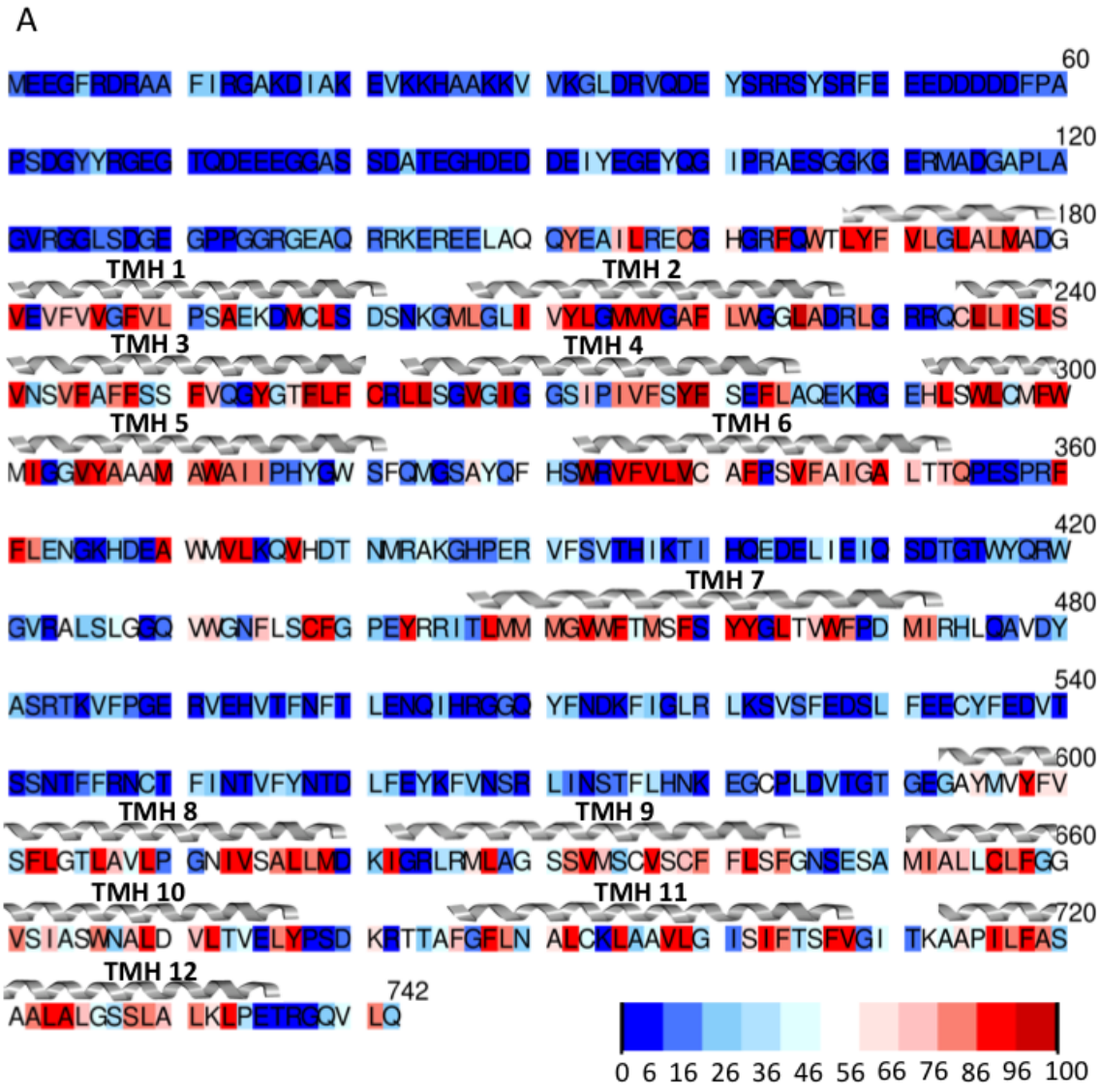


Figure 4.1 The conservation pattern of residues (M, A, V, I, L, C, Y, W and F) as ascertained by an alignment of 758 sequences from a BLAST search against rat SV2A. The degree of conservation is coloured from blue to red as a function of percentage. The position of the TM helices were taken from the consensus prediction of four algorithms (Table 4.2) are indicated.

4.3.2 The Inward and Outward models

The overlay of the Inward-apo (A) and Outward-apo (B) models with their respective templates is shown in Figure 4.2. The templates, GlpT and FucP were chosen as they have very open conformations to the cytoplasm and extracellular cavity respectively, representing states that facilitate substrate binding and

unbinding. As the proposed racetam binding site, the open states are those which would expose the binding cavity to the drug and so are most relevant to this study. Interestingly, the central cavity of the SV2A models was predominantly hydrophobic in character, with only three charged residues and one polar residue (Figure 4.3).

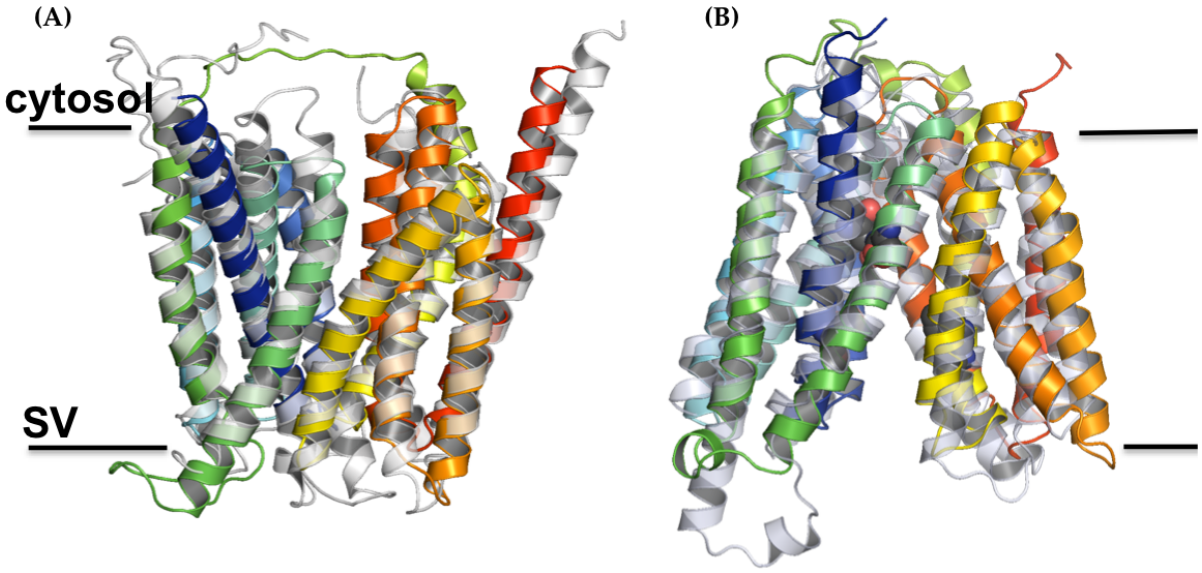


Figure 4.2 (A) Models of the Inward (based on GlpT – PDB ID: 1PW4 - template shown in grey) and (B) Outward (based on FucP - PDB ID: 3O7P - template shown in grey) SV2A protein.

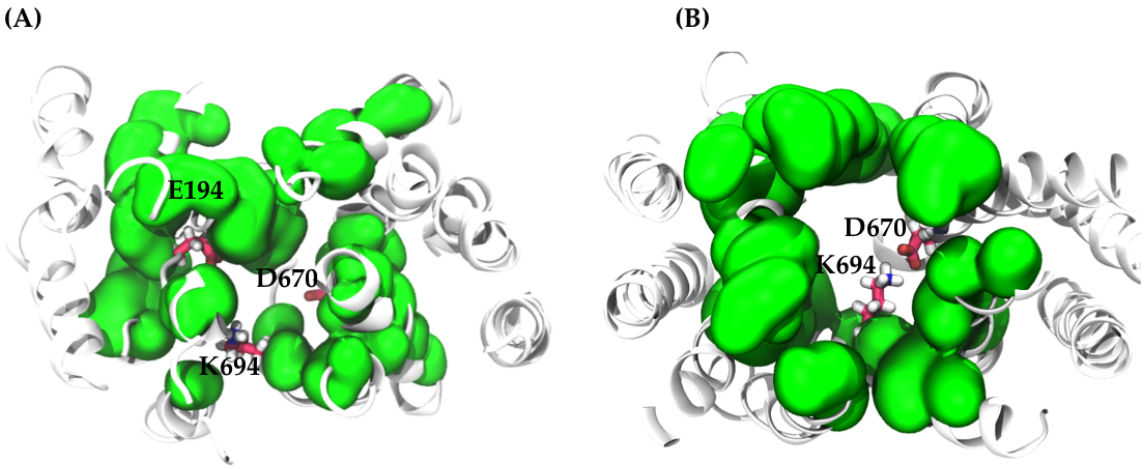


Figure 4.3 The composition of the TM cavity. (A) A space fill view of the cavity for the Inward-apo and (B) Outward-apo models, with hydrophobic residues coloured green. The cavity has three charged residues (E194, D670, K694) which are coloured red (stick representation) and the remaining residues are hydrophobic. E194 and K694 are highly conserved and D670 is 45 % conserved.

4.3.3 The Ligand Binding Site

Attempts to define the ligand binding site were done by Shi et al. (Shi et al. 2011), where 13 residues were identified as impacting IC_{50} scores of the binding of racetam, [^{3}H]ucb 30889 to SV2A. These residues were chosen according to their alignment with key functional residues within the cavity of the TM domain in LacY, the site of ligand binding within the LacY transport cycle (Smirnova et al. 2011). The views of both the inward and outward models of SV2A in Figure 4.3 (A) and (B) show a predominantly hydrophobic cavity, with three charged residues (E194, K694 and D670) and there is a further polar residue (N667). There are highly conserved hydrophobic residues within the cavity, as listed in Table 4.3 and these are within the putative binding site.

Residue	Conservation (%)	Hydrophobic Conservation (%)
SDM residues		
C297	28	36
W300	83	94
Y462	83	86
W666	60	83
K694	66 (R), 24 (K)	-
Charged residue in cavity		
E194	97	-
Residues in cavity		
W454	62	80
F688	25	88
Hydrophobic pocket		
V276	61	96
F280	31	80
L284	42	96
L296	41	76

Table 4.3 Conserved residues within the cavity of SV2A, as given from the MSA of 758 residues found in a BLAST search of SV2A.

This ties in well with SV2A as a putative sugar transporter as it is expected that there would be show charged residues in the ligand binding site, as seen in LacY (Abramson et al. 2003). Namely D670, K694 and N667 are charged or polar residues in the cavity. K694 and N667 are identified as important to ucb 30889 binding affinity, as well as a series of hydrophobic residues line the binding site: W300, C297, W666, M301 and Y462. This is similar to what is found in the FucP

transporter, which has three charged residues across its TM cavity, two of which are in the substrate binding cavity (E135 and R312). Further to this, there are a series of hydrophobic residues (Y74, Y365, L134, F35 and W38) and then three polar residues (N162, Q159 and N139), this comparison between FucP (A) and SV2A (B) is shown in Figure 4.4.

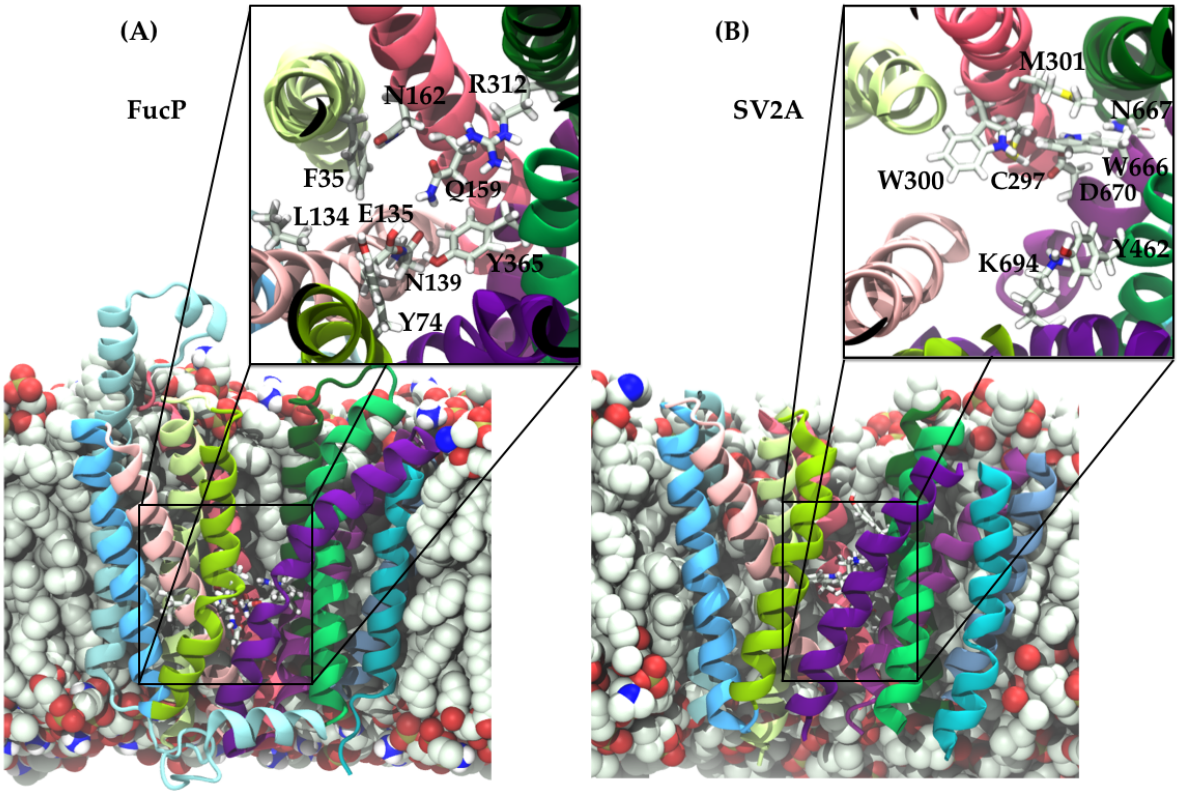


Figure 4.4 The ligand binding site in SV2A (A) and FucP (B). There is a similarity in composition, with two charged residues in each, as well as a collection of hydrophobic and aromatic residues. The binding cavity of FucP is slightly more polar than in SV2A, having three polar residues (N162, N139 and Q159) to one equivalent in SV2A (N667).

The models indicate that D670 is in close proximity to K694 and also to other residues already determined by Shi et al. (Shi et al. 2011) to impact racetam binding (W300 and Y462). However it is only 45 % conserved in the MSA, implying that its function is SV2A specific (since it is totally conserved in an alignment of SV2A

proteins only) rather than a structural role common to the wider family of proteins. D670 has not been postulated to be involved in ligand binding before, however its position correlates well with the position of charged residues in other MFS transporters (for example the E135 residue in FucP). A final residue of interest is E194. In the models, it is located towards the cytosolic side of the TM region and so is not expected to form part of the racetam binding site. However its conservation of 97 % implies an important but currently undefined role, but could potentially be part of the proton translocation pathway (Madeo et al. 2014).

The ligand [^{3}H]ucb 30889 (Figure 1.7 (C)) has been used in assays on site-directed mutagenesis (SDM) studies reported previously and so is used in the following racetam bound MD simulations of the Inward-ucb 30889 and Outward-ucb 30889. The ligand was docked to SV2A snapshots from the 80 ns time steps of the Inward-apo and Outward-apo simulations for a further 80 ns simulation time.

4.3.4 MD of apo and ligand bound SV2A

Molecular dynamics (MD) simulations were conducted of apo SV2A in a pure POPC bilayer for 80 ns. It is expected that SV2A will undergo conformational changes as its function as a putative transporter and so a certain amount of structural fluctuation is to be expected, however the RMSDs of the C-α atoms indicate that both systems remain below 5 Å (Figure 4.5 (A) and (B)). Helical character, as defined by DSSP (Kabsch & Sander 1983), was measured and showed that the apo-systems have greater than 60 % conservation of model helical character in all but three helices (TMHs 2, 4 and 6) for the inward and two helices (TMHs 2 and 4) for the outward model. This indicates that the helix packing of the tertiary structure is fairly robust (Figure 4.5 (C) and (D)).

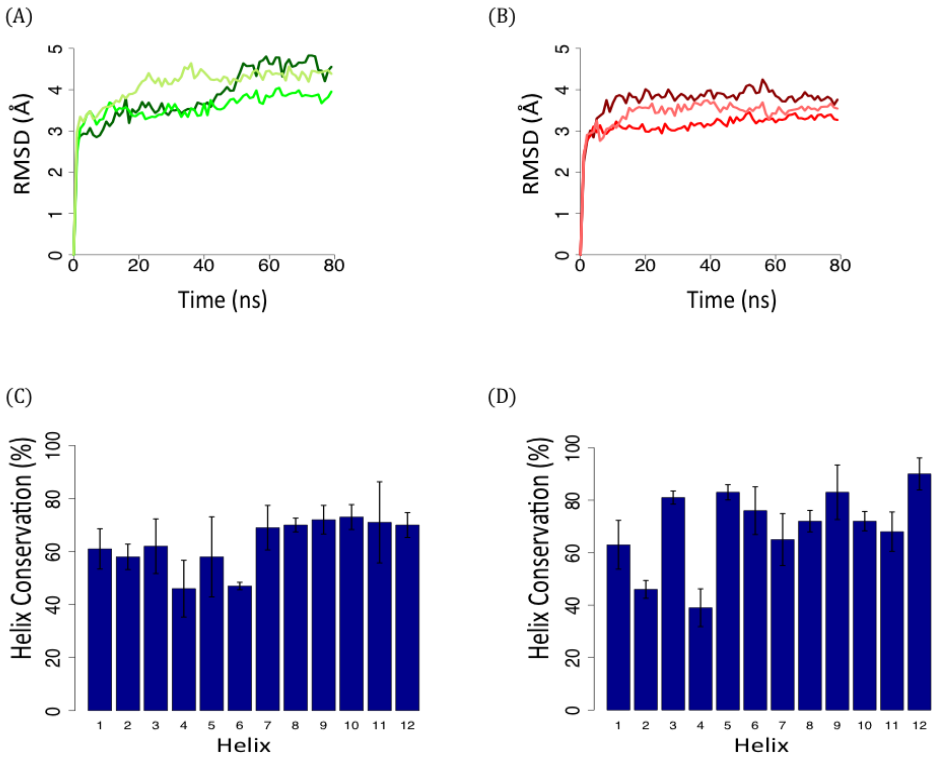


Figure 4.5 The RMSDs and helix conservation in inward and outward SV2A. (A) Change in RMSDs from 80 ns of the Inward-apo simulation (green) and (B) the Outward-apo simulation (red lines) from the initial snapshot. Helix conservation in the (C) Inward-apo and (D) Outward-apo simulations.

Simulation	Volume / Å^3
Inward-apo	3843 ± 158
Outward-apo	2929 ± 45
Inward-ucb 30889	3553 ± 103
Outward-ucb 30889	3263 ± 111

Table 4.4 The volume of the cavity in the four simulations.

The RMSD values of the TM C-α atoms show a plateauing (Figure 4.5 (A) & (B), Figure 4.6 (A) & (B)) but at slightly higher values in the ucb 30089 systems compared to the apo- SV2A (4.59 ± 0.08 Å for Inward-ucb 30889 and 4.14 ± 0.06 Å for Outward-ucb 30889). This suggests fluctuations in the conformation of SV2A on

ligand binding. The volume of the cavity was also analysed throughout the simulations and there was no significant change in volume between systems compared with those in each simulation (Table 4.4). The Inward-apo simulation had a volume of 3843 ± 158 Å^3 whereas the Inward-ucb 30889 simulation had a mean volume of 3263 ± 111 Å^3. The outward models had similarly low levels of fluctuation: 2929 ± 45 Å^3 and 3553 ± 103 Å^3 for the Outward-apo and Outward-ucb 30889 respectively. This indicates that the systems were stable and that ligand binding had little impact on the cavity volume.

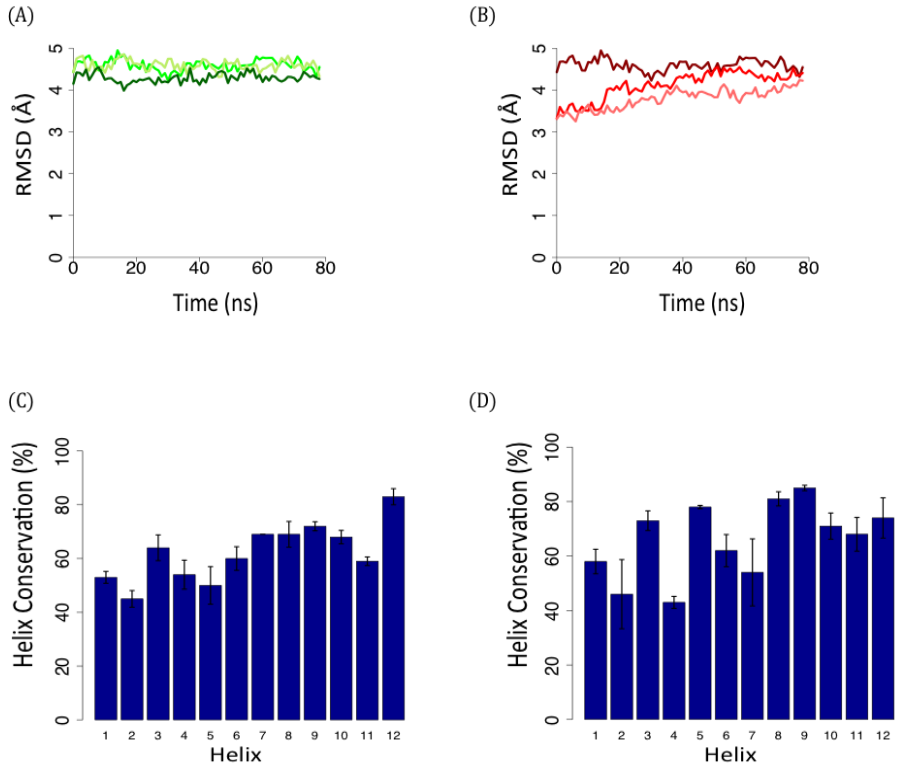


Figure 4.6 The RMSDs and helix conservation in inward and outward ucb 30889 bound SV2A. (A) Change in RMSDs from 80 ns of the Inward-ucb 30889 simulation (green) and (B) the Outward-ucb 30889 simulation (red lines) from the initial snapshot from the respective apo simulations. These do not fluctuate more than 1 Å from the final frame of the apo simulations. Helix conservation in the (C) Inward-ucb 30889 and (D) Outward-ucb 30889 simulations.

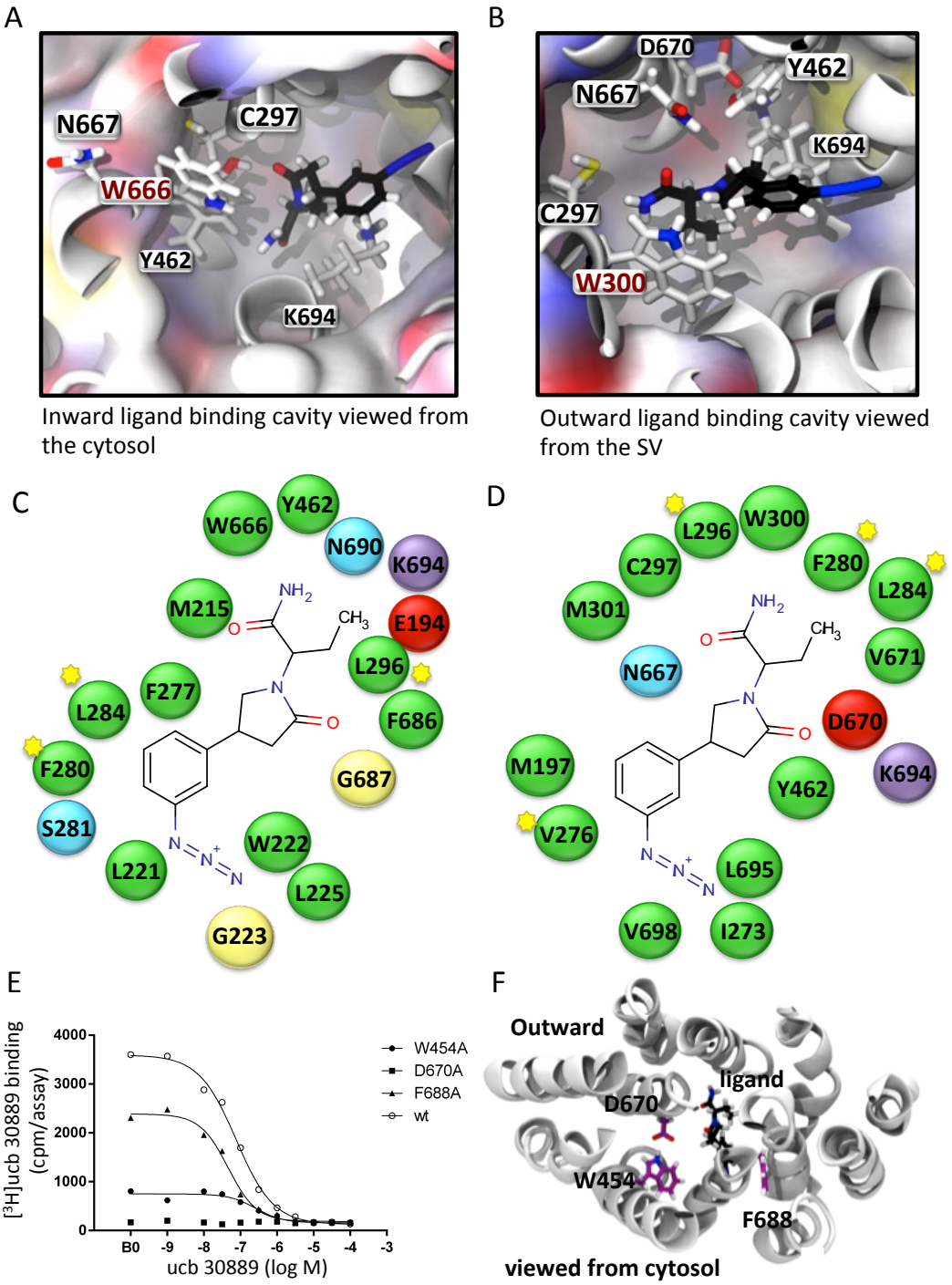


Figure 4.7 The ligand binding site in inward and outward SV2A. The ligand binding sites in the Inward-apo model of SV2A (A) and the Outward-apo model (B) from simulation (60 ns). The ligand (black stick) was docked to a snapshot the apo-model after 80 ns simulation. Key residues identified by mutagenesis are highlighted as stick representations. Schematic interaction maps of the docked ligand, generated via MOE with an interaction cut-off of 6 Å are shown for the inward (C) and outward (D) models. Residues starred are conserved hydrophobic residues common to both the inward and outward ligand binding pockets. (E) Affinity of ucb 30889 for recombinant rat SV2A (wt and mutants). A concentration range of ucb 30889 was incubated with 5 nM of [^{3}H]ucb 30889 during 120 min at 4°C. B0 is the binding of [^{3}H]ucb 30889 in the absence of any competing compound. Data are representative of three independent experiments. pIC_{50} values were calculated from untransformed raw data by non-linear regression using a model describing a sigmoidal dose-response curve with variable slope and are reported in Table 3. The position of the mutants with respect to the ligand in the outward model is shown in (F).

Three residues were picked out for investigation by site-directed mutagenesis, W454, F688 and D670 (Figure 4.7 (E) and (F)). The first, W454 was not directly interacting with the ligand binding site and though it could potentially impact on ligand binding the extent of such will be minimal. Conversely F688 was found at the cytosolic side of the TM cavity and so should have no impact on the racetam binding. Finally D670 was selected since it is one of two charged residues within the TM cavity (K694 is the other) and it is within the ligand binding site. In fact, in MD simulation, the carboxy oxygen atoms of D670 and the amino hydrogen atoms of K694 was less than 3.5 Å for 35% of the simulation time (with a mean of 4.5 ± 2.6 Å). It is possible that the mutation of D670 to alanine would have a large impact on racetam binding.

It is worth noting that these are two independently constructed models. Each was based on GlpT or FucP and so provide an inward and outward open structure. Therefore the analysis is designed to assess whether one of the other is a more likely conformation for SV2A, since the direction of transport is not known. Nor is it clear whether SV2A transports into the synapse cytoplasm or between the cytoplasm and inside the synaptic vesicle.

4.3.5 Experimental conclusions

A collaboration with the experimentalist at UCB Pharma, Dr Veronique Daniels, meant that the three residues previously mentioned could be tested using site directed mutagenesis studies. The wild type and three mutant SV2As (W454A, D670A, F688A) were prepared using HEK cells and competition binding experiments were conducted to get the IC_{50} values for each against the radioligand [^{3}H]ucb 30889. These experiments agreed with the hypothesis put forward in the previous section (Chapter 4.3.4), with D670A completely abolishing ucb 30889

binding. F688A had no impact and W454A may have some effect on racetam binding, but not a large amount (Figure 4.7 (E)). One caveat must be given with regards to the D670A data, since mutations can manifest their influence on binding through indirect changes as well as direct changes. Indeed, this has been explicitly demonstrated for SV2 proteins where mutant proteins can for example become trapped in the endoplasmic reticulum, presumably reflecting a misfolded state (Nowack et al. 2010). Another caveat that we should raise at this point is that the mutations are performed in HEK cells and thus the influence of any vesicle proteins on drug binding will also be absent.

However, given that the D670A mutant abolishes binding and the importance of interactions between ligand and TM helix 5 (W300A, C297alkylated are each reported to abolish binding by Shi et al. (Shi et al. 2011)), the position in the outward open model is much more conducive to interaction with the ligand and thus we take this also to favour the Outward-apo model over the Inward-apo model (Figure 4.7 (A) and (B)).

Rat recombinant SV2A	pIC_{50} (mean ± S.D.)
Wildtype	7.1 ± 0.05
W454A	6.7 ± 0.14
D670A	/
F688A	7.1 ± 0.19

Table 4.5 Affinity of ucb 30889 for rat recombinant SV2A (wildtype and mutants) labeled with [^{3}H]ucb 30889 (n = 3). Data taken from Lee et al. (Lee et al. 2015).

4.4 Conclusions

This work has used homology modelling based on templates to explore two conformations of SV2A: inward open and outward open states with respect to the cytoplasm. The use of sequence analysis techniques has improved confidence in the model and also provided insight into residues within the ligand binding site that have functional relevance. In particular D670 has been found to abolish the binding of the racetam, ucb 30889 in site directed mutagenesis studies, and the conservation of E194 and K694 imply key structural roles (K694 has been previously identified as impacting racetam binding (Shi et al. 2011)). Further to this, 80 ns of MD simulation indicates substantial flexibility in both the apo and ucb 30889 bound systems, longer simulation time would be of interest to further explore the impact of this flexibility on the ligand binding site. The agreement of the Outward-ucb 30889 poses with experimental data suggests that the outward state is the preferred ligand binding conformation. One omission from this model is the insertion of the BoNT binding domain between TMHs 7 and 8, who structure was solved after the completion of this work (Benoit et al. 2014) but which is likely to impact TM domain stability. Therefore a suggestion for further work is exploring through MD simulation the updated model with this extracellular domain attached.

5 MD simulations of FucP

FucP was the only MFS transporter whose crystal structure had been solved in an outward open state. Since there were limitations to the model of SV2A, which was specifically for the investigation of the drug binding site, the logical next step was to investigate dynamics of an X-ray crystal structure. This can be done using molecular dynamics (MD) simulations. It was found that the protein retains its outward open state, and that lipid molecules snorkel between two transmembrane domains, which is thought to be non-physiological and would hinder the transport mechanism where the domains rock against each other in a rigid-body motion. During the transport cycle of FucP, it is hypothesized that D46 and E135 are protonated and this is key to the co-transport of L-fucose and H^+. Therefore the effect of protonation of D46 and E135 is investigated in MD simulation.

5.1 Introduction

At commencement of this research, FucP was the only MFS transporter whose X-ray crystal structure was solved in an outward open conformation (Dang et al. 2010), though structures existed of more occluded states which are partially outward open (Kumar et al. 2014; Jiang et al. 2013). As predicted for other MFS proteins, FucP is thought to transport its ligand via an alternating access mechanism (Jardetzky 1966) where the rigid body movement of the two domains of the protein against each other allow the cavity between the domains to become alternatingly open to either

the cytoplasm or the periplasm (see Chapter 1, Figure 1.3). More specifically to FucP, it is thought to operate as a L-fucose/H^+ symport transporter and is hypothesized to have the transport mechanism described in Figure 5.1. Here, the protonation of either D46, E135 or both (see Figure 5.2) is required, as is substrate binding, in order to cycle from outward to occluded and finally to inward open.

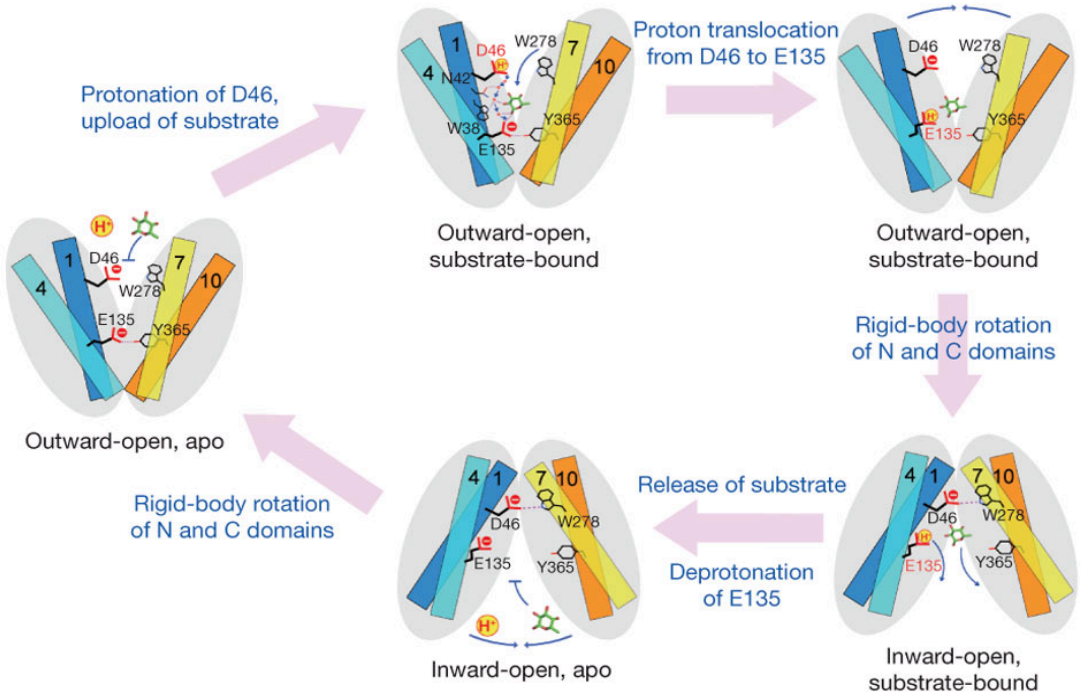


Figure 5.1 The proposed alternating access mechanism in FucP as proposed by Dang et al. (Dang et al. 2010). Movement from the outward to inward open conformation is facilitated by the presence of fucose and protonation of at least one of D46 or E135, allowing transport of fucose into the cell. The substrate is co-translocated with a proton in the symport mechanism and it is the proton translocation from D46 to E135 that destabilizes the outward open conformation, permitting closure of the periplasmic side, and subsequent opening of the cytoplasmic side of the transporter for substrate release into the cell.

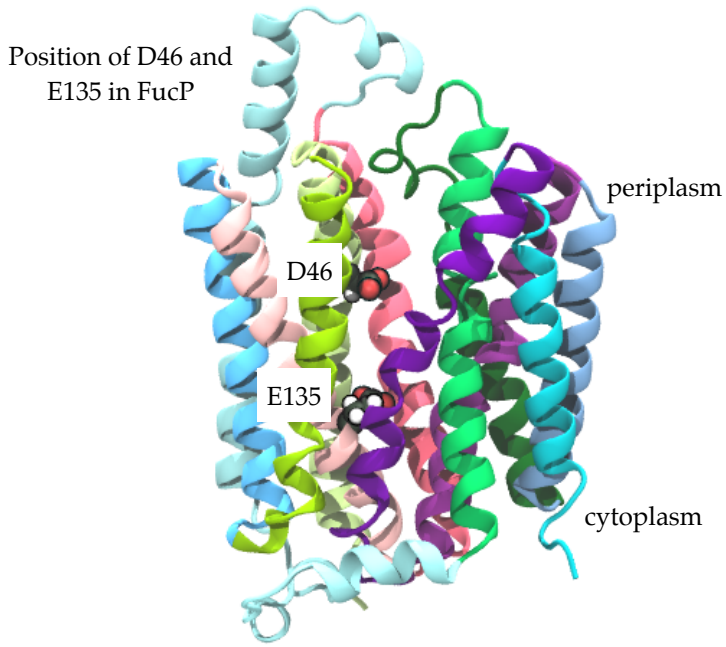


Figure 5.2 The position of D46 and E135 in the central TM cavity of FucP. D46 is closer to the periplasmic side and E135 closer to the cytosolic side of the cavity.

The logical next step after exploring the structure of MFS transporters in previous chapters is to now look at the mechanism for the function of FucP through MD simulations. The X-ray crystal structure of FucP is investigated to determine the impact of key functional and structural residues on the conformational dynamics of FucP with specific focus on D46 and E135 (Figure 5.2). Further to this, E135 is thought to be one of the residues involved in fucose binding and so docking studies explore this (Figure 1.11).

The stability of the outward open structure of FucP appears to be mediated by crystal contacts (Figure 5.3). Here, the loop between TMHs 5 and 6 interacts with loop 11 – 12 in the adjacent protein, inferring the possibility that this interaction stabilises the conformation of the X-ray crystal structure. During MD simulations of LacY, this was seen to move into a more occluded state (Stelzl et al. 2014) and this is predicted to happen in FucP. If this were to be the case, then the protein would

undergo the first transition of the transport cycle from outward open to occluded with ligand and co-transported proton bound or at least move to a more occluded state.

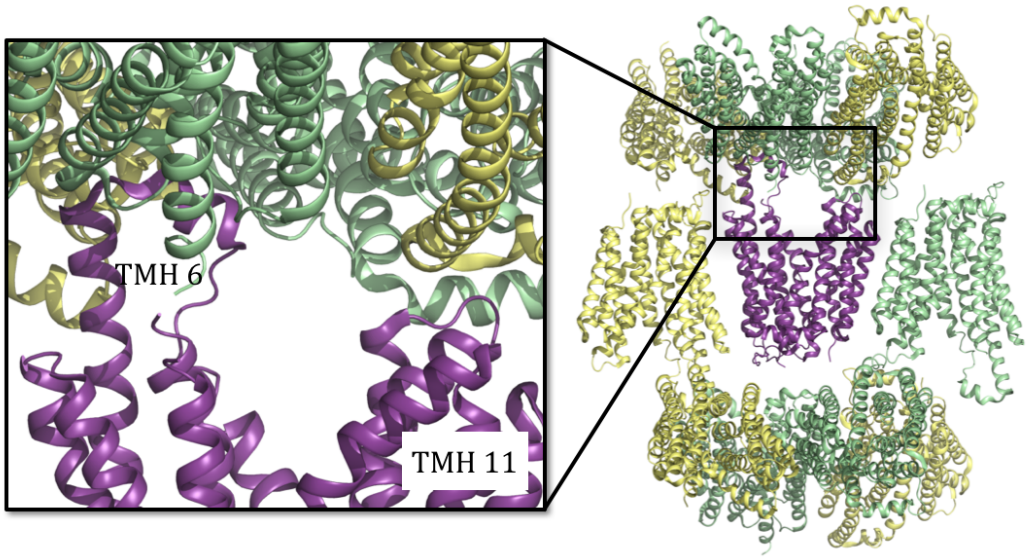


Figure 5.3 The crystal contacts in the FucP X-ray crystal. The elongated helices of TMHs 5 and 6 appear to be interacting with loop 11 and 12 in the protein of the subsequent subunit, which implies that the conformation is being artificially stabilised in the crystal. This is something that can be tested using MD.

5.1.1 Aims

The aims for the remaining two chapters are to investigate the dynamics of FucP in an outward open structure. This chapter explores Molecular Dynamic (MD) simulations of FucP in lipid bilayers, with particular focus on the protonation state of the two key residues implicated in the proton translocation pathway, D46 and E135 (see Figure 5.1). The structure and hypothetical mechanism of transport proposed by Dang et al. (Dang et al. 2010) raises several questions. What is the stability of this open state in a bilayer? Is it possible to test the influence of

protonation states on the dynamics of the system? To what extent does the presence of ligand affect the dynamics of FucP?

5.2 Methods

5.2.1 Molecular dynamics simulation

The FucP X-ray crystal structure was embedded in either a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) or a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) bilayer using the g_membed (Wolf et al. 2010) feature of GROMACS, and an energy minimisation with a steepest descent algorithm until convergence with a force tolerance of 0.239 kcal mol^{-1} $Å^{-1}$ was performed. Subsequent simulations were set up as described in Table 5.1. Sodium and chloride ions were then added to the systems to a concentration of 150 mM followed by two restrained MD runs whereby all heavy atoms were restrained by a harmonic potential of 2.39 kcal mol^{-1} $Å^{-2}$ for 1 ns. Finally, production runs were performed on three repeats that differed in their initial velocities only for time lengths given in Table 5.1. MD simulations were carried out with GROMACS v4.6.5 (Abraham et al. 2015) using the OPLS-AA (Jorgensen et al. 1996; Kaminski et al. 2001b) force-field and TIP3P water molecules (Jorgensen 1981). Production simulations were performed in an NPT ensemble maintained at 323 K and 1 bar pressure. The integration time step was set as 2 fs and a stochastic dynamics integrator (Van Gunsteren & Berendsen 1988) was used. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method (Essmann et al. 1995) with a 14 Å cut-off and 1 Å space grid. The Lennard-Jones potential used a cut-off of 9 Å, with a switch at 8 Å. The LINCS algorithm (Hess et al. 1997) was used in order to constrain bond lengths in both the lipid molecules and the protein. For ligand bound

simulations, Autodock Vina (Trott & Olson 2010) was used to dock the ligand into the cavity to provide the initial starting coordinates for the protein-ligand complex. The docking grid was defined as a 10 x 10 x 8 Å rectangle centred at the residue E135. The conformation of ligand chosen was that which best agreed with the position of the detergent in the X-ray crystal structure (Figure 1.9 (B)) and energy minimisation of the complex was conducted as above, before MD simulation.

First, FucP was inserted into a POPC bilayer and then compared to insertion in a POPE bilayer. These were used to check the dynamics of the X-ray crystal structure once embedded within a bilayer, and so were conducted on the apo protein, without any explicit changes to protonation states of side chains. Subsequent to this, two residues key to the H^+ translocation pathway (Figure 5.2), D46 and E135 were protonated in two apo systems. Finally L-fucose was docked into the central TM cavity of the unprotonated and E135 protonated systems. All systems are detailed in Table 5.1.

Name	Description
Apo-pc	Apo-FucP with no protonation on D46 or E135 and in POPC bilayer (3 x 100 ns)
Apo-pe	Apo-FucP with no protonation on D46 or E135 and in POPE bilayer (1 x 200 ns)
Apo-pe2	Apo-FucP with no protonation on D46 or E135 and in POPE bilayer (3 x 100 ns) - after rotation of the side chains on TM helices 2, 5, 8 and 11
Apo-pe-eh	Apo-FucP with E135 protonated and in POPE bilayer (3 x 200 ns) - coordinates of the Apo-pe2 were re-inserted into the bilayer and E135 was protonated
Apo-pe-dh	Apo-FucP with D46 protonated and in POPE bilayer (3 x 200 ns) - coordinates of the Apo-pe2 were re-inserted into the bilayer and D46 was protonated
Fuc-pe-up	FucP with L-fucose docked to central cavity of final snapshot from Apo-pe2 - no protonation of D46 or E135 and in POPE bilayer (3 x 20 ns)
Fuc-pe-eh	FucP with L-fucose docked to central cavity of final snapshot from Apo-pe2 - protonation of E135 and in POPE bilayer (3 x 100 ns)

Table 5.1 Summary of simulations of FucP with protonation states and ligand presence listed.

5.3 Results and Discussion

The X-ray crystal structure of FucP, solved in an outward open conformation, has a separation between the two domains. The presence of crystal packing contacts, when looking at the symmetry repeats of FucP (Figure 5.3), implies that the open state could be an artifact, with the large loop 5 – 6 interacting with the loop 11 – 12 of the neighbouring crystal structure. The effect of this is investigated in MD

simulations of the protein structure in two pure POPC or POPE bilayers and with two different protonation states.

5.3.1 Initial MD simulations

In the first instance, FucP was placed in a pure POPC bilayer and the system simulated for 3 x 100 ns in a solvent box with 0.15 M NaCl (Apo-pc). Simulations of other MFS proteins in open conformations (LacY, which is inward open) have revealed movements of the protein to more occluded states within this time frame (Stelzl et al. 2014; Andersson et al. 2012; Holyoake & Sansom 2007), leading to the hypothesis that FucP will undergo similar transitions to an occluded state from the outward open conformation.

On closure of the periplasmic side of the TM cavity to an occluded state, it is expected that the TM helices 2 and 5 would move closer to TM helices 11 and 8 respectively in domain two. However, this was not the case for FucP in the POPC bilayer (Figure 5.4 (A)). Instead a lipid molecule inserts between TM helices 2 and 11 and this then migrates across the domain interface, implying that the periplasmic side of the TM cavity remains open (certainly lipid accessible). This remains between TMHs 2 and 11 from the first 10 ns of each simulation and two further lipid molecules subsequently move between the two helices. In the X-ray crystal structure of FucP, the space between the periplasmic side of the TM helices 2 and 11 is large enough that a lipid head group could easily fit between the helices and the simulation set up for the Apo-pc system indicates that this space remains there.

To compare with a different lipid type, a 200 ns simulation was conducted with FucP now inserted into a POPE bilayer (Apo-pe in Figure 5.4 (B)). The phosphoethanolamine (POPE) head group differs from the phosphocholine (POPC)

since it is has a negative charge at physiological pH (approximately 7.5 (Padan et al. 1981)), whereas the zwitterion of phosphocholine is neutral overall. There was no movement of lipid molecules across the interface between the two domains, though a head group of one lipid did move between TM helices 2 and 11 and so the TM helices remained apart at the periplasmic end. By comparison, the POPE was also the logical pure bilayer of choice when investigating FucP, since it is the most common lipid in *E. coli* inner membranes (Shokri & Larsson 2004) and so all subsequent MD simulations.

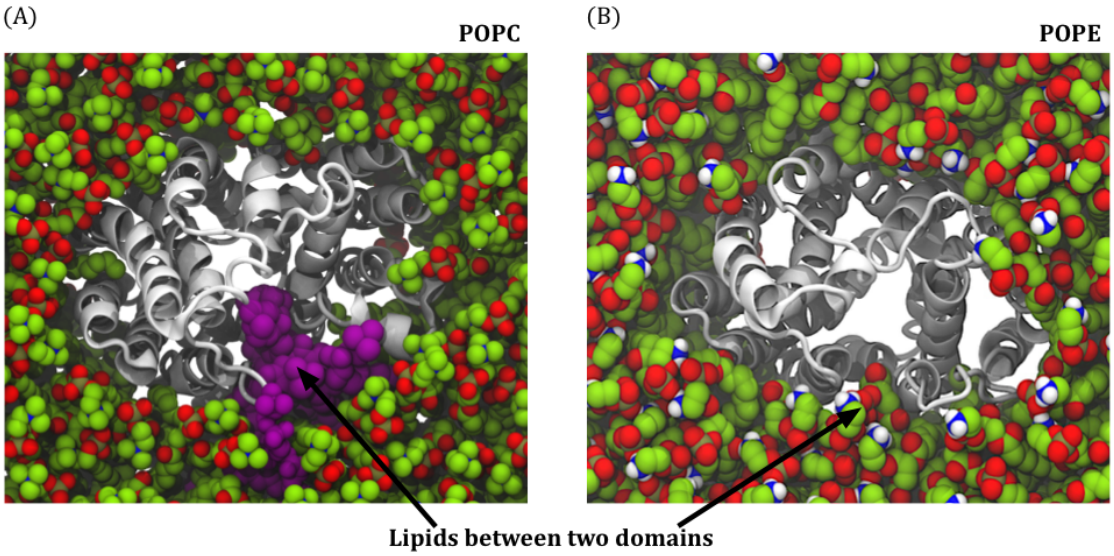


Figure 5.4 The insertion of lipid molecules between the two domains of FucP in simulations of unprotonated FucP with (A) POPC (Apo-pc) and (B) POPE (Apo-pe), as viewed from the periplasmic side of the lipid bilayer. As indicated by the three purple molecules, lipid insertion is more prominent by the POPC molecule, such that the lipid molecules move between the two domains. To a lesser degree, this is still present in POPE. One hypothesis for this is that the open structure is the cause, as the distance between the two domains is sufficient for lipids to move between (Figure 5.5).

The points of contact between the two domains are TM helices 2 and 11 and then TM helices 5 and 8 (Figure 5.5 (A) and (B)), such that in the occluded state these are interacting along the whole length of each helix. On closer examination of FucP, the packing of side chains between TM helices 5 and 8 is close and does not allow lipid

entry as seen in the simulations. Equally, TM helices 2 and 11 have close packing of side chains at the cytoplasmic side of the bilayer, however at the periplasmic side the bilayer TM helix 11 is bent back towards its domain, creating a gap between the 2 domains that allows lipid snorkeling. To overcome this, different side chain rotamers were selected in PyMol on TM helices 2 and 11 to minimize the gap between the two helices and then re-inserted the modified FucP into a POPE bilayer for all subsequent simulations (Apo-pe2, Apo-pe-eh and Apo-pe-dh systems as well as Fuc-pe-up and Fuc-pe-eh), the position of the helices is shown in Figure 5.5 (C) and (D). No snorkeling of lipid molecules was evident in these simulations, which implies that crystal contacts stabilised TMH 11 in the conformation and this caused the large distance between TMHs 2 and 11 (this is not present in other crystal structures).

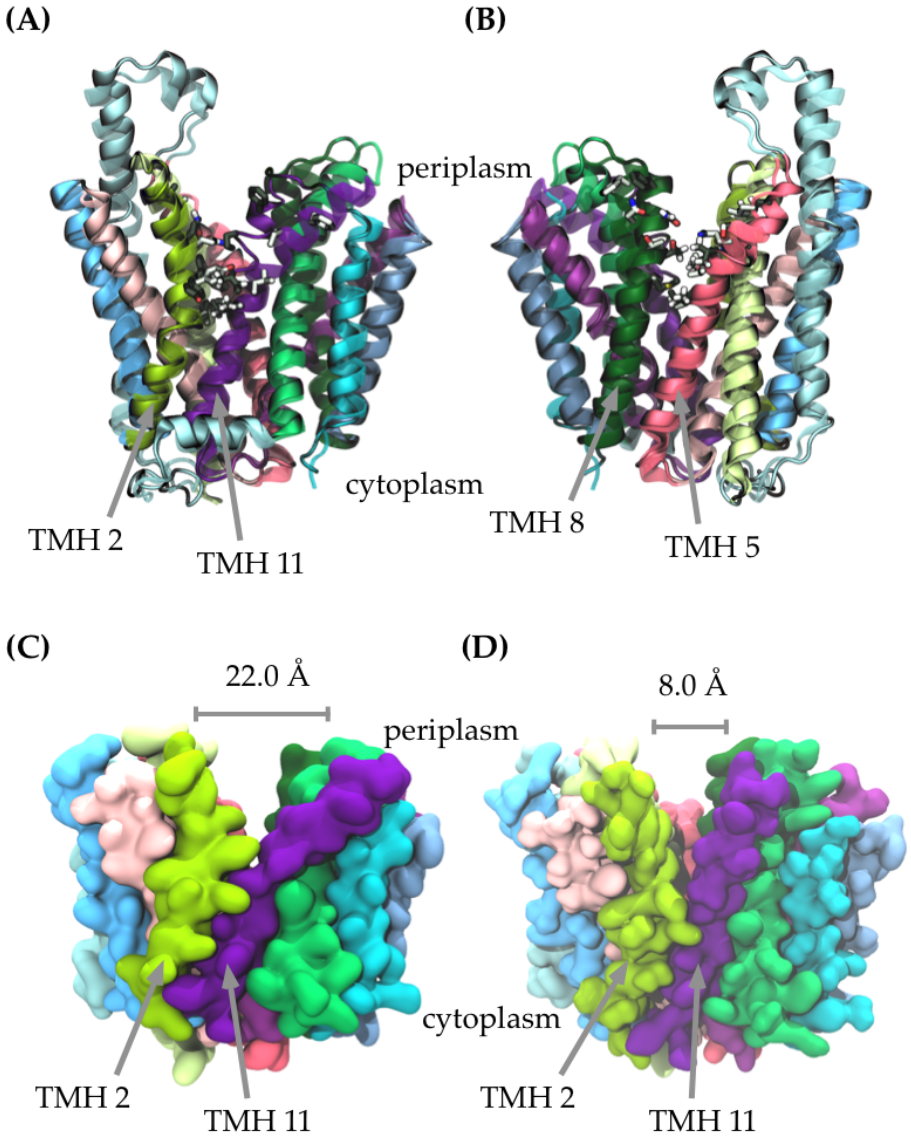


Figure 5.5 The interface between the two domains. The interface is mediated by contacts between TMHs 2 – 11 (A) and TMHs 5 – 8 (B). The solid helices are a snapshot from an Apo-pe2 simulation with interfacial residues shown as black sticks and in comparison the X-ray crystal structure is transparent, with the same residues represented by white sticks. The residues between the domains are closer in space in the Apo-pe2 simulation, over the X-ray crystal structure. A volume representation of the X-ray crystal structure (C) and a snapshot from an Apo-pe2 simulation (D), which show that the domains move together in the simulation. The grey lines represent the distance between two C-α atoms, one on TMH 2 and the second on TMH 11 at the periplasmic end of the helices.

5.3.2 The outward facing conformation

The first objective was to determine to what degree the outward open state of the X-ray crystal structure of FucP was due to crystal contacts. Already, the snorkeling of lipids between TM helices 2 and 11 was shown to occur and which has not been documented for other MFS proteins in MD simulation. In fact, moving the side chains to reduce the space between the helices, this prevented the lipids moving between TM helices 2 and 11. Subsequent to this and shown in Figure 5.5 (D), in the Apo-pe2 MD simulation, TM helices 2 and 11 move closer together. To measure whether this can be explained as part of a concerted closure of the outward open state rather than the movement of the two helices, principal component analysis (PCA) was used.

This technique describes the motions of a simulation as a set of linearly uncorrelated eigenvectors, which can be used to determine the contribution to the overall simulation for each of these eigenvectors. The analysis was applied to the Apo-pe2 simulation repeats. Here, the first eigenvector contributes to 60 % of the simulation motion. This vector describes the movement of the periplasmic loops of each domain towards each other (Figure 5.6 (A)). The dominant motion is the movement of loop 5 – 6 towards domain 2 (reflected in the RMSF plot in Figure 5.8 (B)) and while this does show a rearrangement of the domain interactions, it does not result in an occluded conformation (Figure 5.6 (B)).

This analysis shows is that apo FucP remained in an outward open state within the timescale of the simulations and while in a pure POPE bilayer. The Apo-pe2 simulation showed some occlusion of the protein, but not closure of the periplasmic side of the protein, as defined by the TM cavity becoming solvent inaccessible from the periplasm (Figure 5.6 (B)). Therefore, the next step was to investigate the two

titrable key residues (D46 and E135 shown in Figure 5.2) in the central TM cavity and the effect these have on protein dynamics.

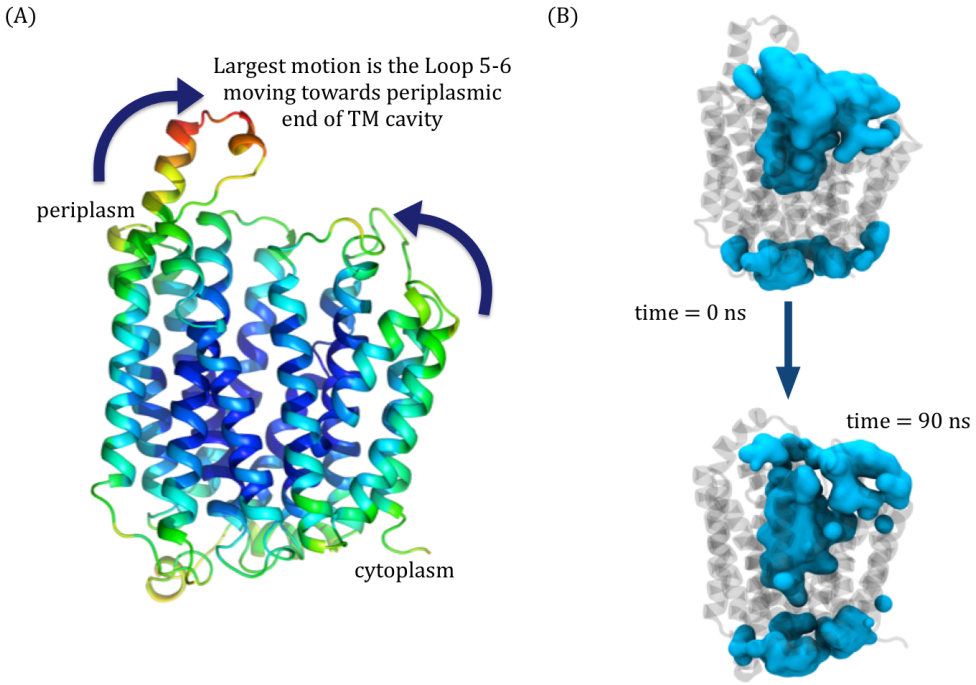


Figure 5.6 B-factor representation of the first eigenvector using principal component analysis (PCA), which shows that the periplasmic loop regions are most mobile in Apo-pe2. The first eigenvector shows that the most dominant motion in the simulation is a movement of loop regions together at the periplasmic end (A). However the protein remains in an outward open conformation, as shown by the solvent within the TM cavity in (B).

5.3.3 Protonation of D46 and E135

The are two titrable residues D46 and E135 which are thought to be important for the cycling of FucP since the transport of fucose exploits the electrochemical gradient via the co-transportation of protons. Therefore it is predicted that at least one of these residues must be protonated in order to allow for the conformational change from outward to inward open and permitting substrate passage into the cell, as see in the schematic in Figure 5.7.

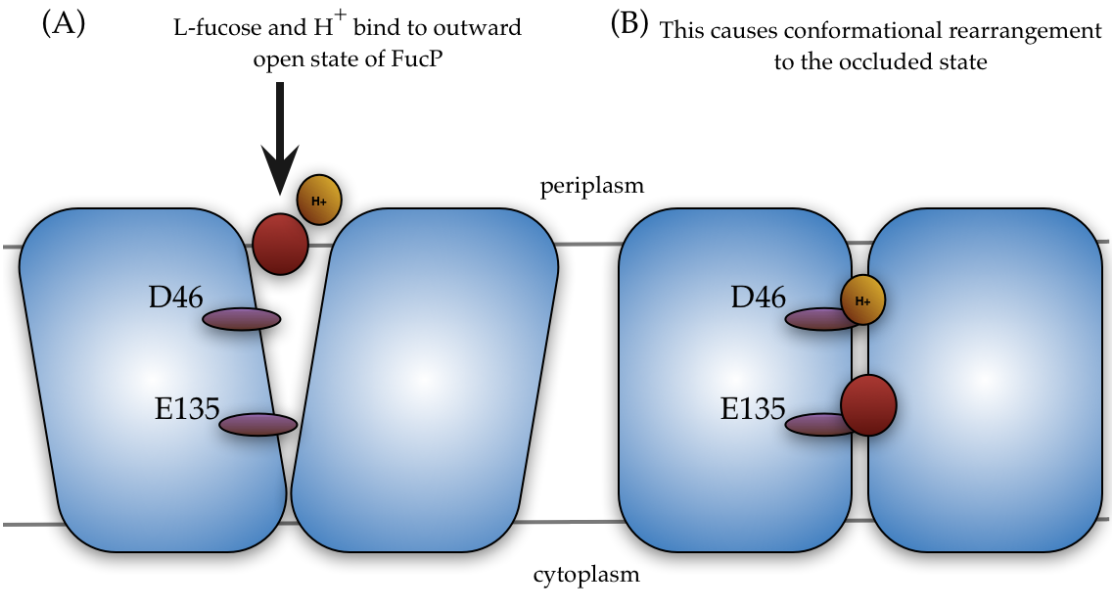


Figure 5.7 A schematic of the first two stages in the transport cycle of FucP. Both the ligand and protons are required to instigate the movement from outward open to occluded states. It is unclear which of the two residues would be protonated during the transport cycle, or in fact both (more details are given in Chapter 6).

Using unbiased MD simulation, the effect of different protonation states of D46 and E135 on the dynamics on FucP was investigated. The final snapshot from an Apo-pe2 simulation was taken and either the D46 or E135 residue was protonated for Apo-pe-dh and Apo-pe-eh systems respectively. The limitation of atomistic MD simulation is that it can only simulate a fixed protonated state, so is not capable of investigating proton translocation between the two residues.

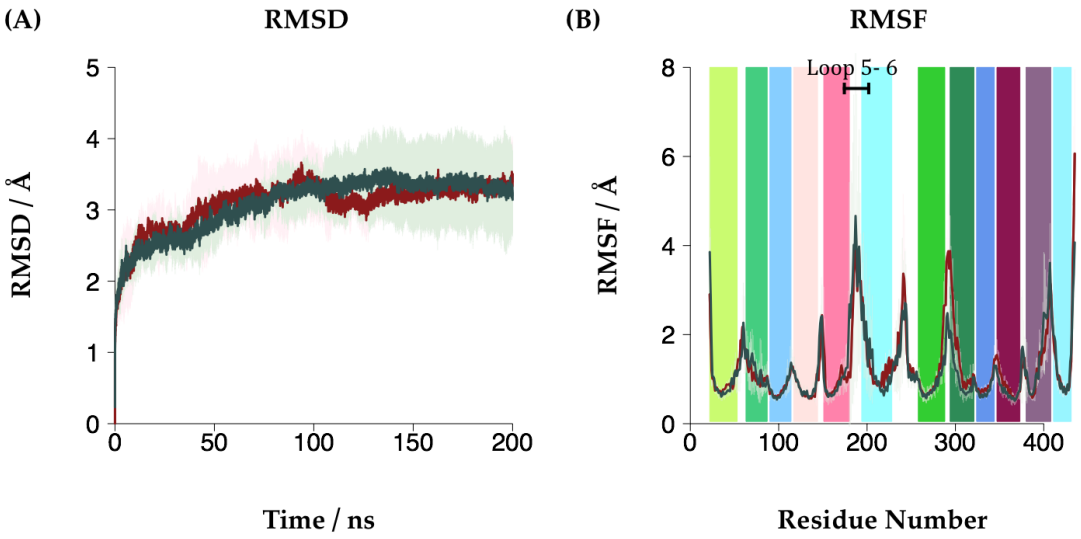


Figure 5.8 The RMSD (A) and RMSF (B) for the Apo-pe-dh (■) and Apo-pe-eh (■) MD simulations. The colour of the bars in (B) represent the position of the 12 TM helices, coloured according to Figure 1.1 (A). The RMSDs are not significantly different for either the D46 or the E135 protonated state. The RMSF profiles are similar, each showing the largest fluctuations in Loop 5 – 6, which are also seen in the principle component analysis (Figure 5.6). All measurements are three repeats for each simulation.

The overall fluctuations are not significantly different between the two protonation states, as indicated in the RMSDs and RMSFs in Figure 5.8. The RMSDs equilibrate to approximately 3 Å for the averages of three repeats for the two systems (Figure 5.8 (A)), equally the RMS fluctuations (Figure 5.8 (B)) are equivalent across the averages of the three systems. An exception to this is one repeat for Apo-pe-eh, where the slightly higher RMSD values relate to the larger movements of the loop 5 – 6 and this is reflected in the RMSF values for this repeat. At loop 5 - 6, the RMSF value for this repeat of Apo-pe-eh reached 8.2 Å, significantly higher than any other simulation, though in all simulations this loop has the largest fluctuations.

Second to those of loops 5 – 6 and the fluctuations of loops 7 – 8 was high, which is in accordance with the PCA of the Apo-pe2 system. In the PCA, the most dominant motion was the perplasmic loops of each domain moving towards each and in the Apo-pe-eh and Apo-pe-dh systems, the periplasmic loops 5 – 6 and 7 - 8. In order to

establish whether these fluctuations relate to the domains are closing together to form an occluded structure, three metrics were measured: the distance between the centre-of-mass of loop 5 – 6 and the periplasmic loops of domain 2; the distance between the periplasmic side of TMHs 2 and 11 and TMHs 5 and 8 (Figure 5.9). These metrics define the concerted motion required to close the outward open transporter into the occluded state (Figure 5.7), and so reducing their distances implies moving to an occluded state.

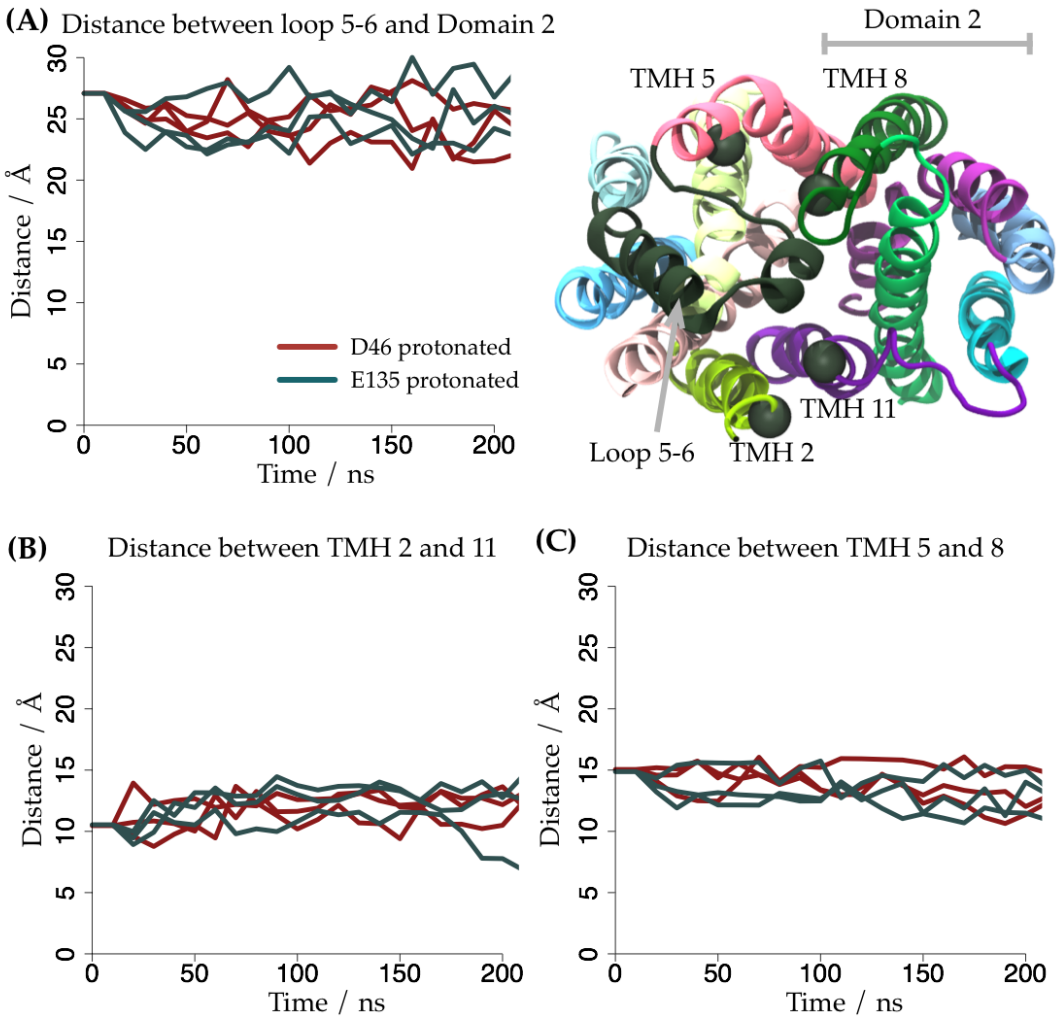


Figure 5.9 The three distances here show the degree of openness of the peripasmic side of the TM cavity in FucP. (A) The distance between loop 5 – 6 and the periplasmic loops of domain 2 is reduced, but not significantly different between the two systems: Apo-pe-dh (■) and Apo-pe-eh (■) MD simulations. The distances between TMHs 2 – 11 (B) do not change over the simulation. TMHs 5 – 8 (C) distance at the periplasmic side of the TM cavity show a slight trend to closure except one repeat each for the Apo-pe-up and Apo-pe-eh simulations. All measurements are the average over three repeats.

The distance between loop 5 – 6 and the periplasmic loops of domain 2 was reduced from 26 Å to approximately 19 Å by 100 ns for the Apo-pe-eh system and while the Apo-pe-dh system moved to a similar distance, it had more fluctuation (Figure 5.9 (A)). This could be caused by the protonation of D46, which is solvent accessible and near the periplasmic side of the TM cavity. Thus, removing the charge impacted the hydrophilicity of the cavity and hence increased the likelihood of the protein fluctuations into a more occluded state due to the cavity being less favourable for water. There was no change in TMHs 2 and 11 distances, which does not support this hypothesis (Figure 5.9 (B)), however this could be due to the kink in TMH 11 which means it is interacting more with domain 2 helices. The position of TMH 11 could be a small well in the conformational profile of FucP. Interestingly, there is a small decrease in distance in both systems between TMHs 5 and 8 which is not significantly different to the distance between TMHs 2 and 11 by the end of the simulations (Figure 5.9 (C)).

The final metric measured was the volume of the cavity, chosen to describe the degree to which the system becomes occluded. This metric is relevant to the hypothesis that the transporter's mechanism requires a constant 'dry' patch in the cavity, which distinguishes it from a channel. During the transport cycle, both protonation and ligand is required and the presence of the 'dry' patch is essential to stop the transporter from leaking protons. Hence this dry patch, which is initially near the cytoplasmic side of the TM cavity, will move towards the periplasmic side of the cavity. Therefore, as the domains close around the substrate, the number of water molecules in the cavity will change depending on the z position (axis running through the TM cavity). Interestingly, no significant difference between the volumes of the cavity between the two charged states (D46H and E135H (Figure 5.10)) is observed and all three show a small reduction over the simulation.

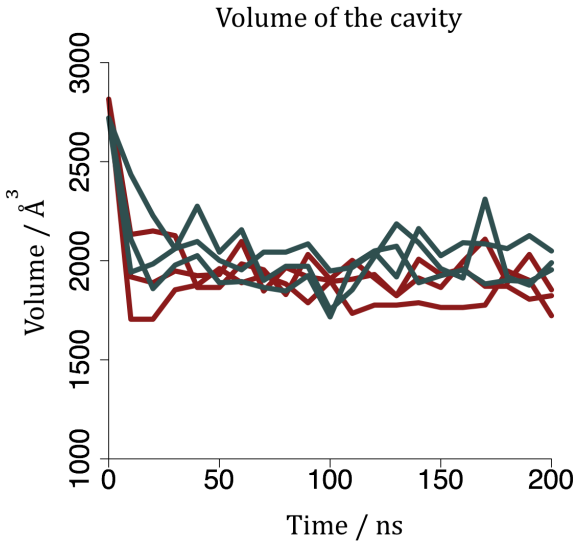


Figure 5.10 The volume of the TM cavity in FucP MD simulations: Apo-pe-dh (■) and Apo-pe-eh (■). Each simulation shows a rapid reduction in the volume from ~ 2700 Å^3 to ~ 2000 Å^3 in 10 ns and volumes then stablise around this point for the remainder of the simulations. All measurements are the average over three repeats.

The occluded state is not achieved in the apo simulations, after 200 ns of MD simulation, though some occlusion is observed. From these simulations, the effect of protonation is not immediately evident. This could be due to the timescale of the simulations (that are too short to capture the large conformational changes associated with transportation) or that the ligand is required in addition to the proton for efficient cycling of the transporter.

5.3.4 The effect of protonation on predicted pK$_a$ in FucP

The approximate physiological pH for *E. coli* cells is 7.5 (Padan et al. 1981) and so was compared with the predicted pK$_a$ values for D46 and E135 (using Propka (Søndergaard et al. 2011)). In Figure 5.11, the pK$_a$ values were predicted for the

snapshot at every 10 ns and the value gives the point at which exactly half of the acid species are dissociated and so a pH above the predicted pK_a implies that the species is mainly in the form A^- (assuming that the central TM cavity has a similar physiological pH to the cytoplasm). In the X-ray crystal structure, the predicted pK_as for D46 and E135 are 6.61 and 7.34 respectively. Since physiological pH is approximately 7.5, this implies that there is a 50 % probability of E135 being protonated and D46 is much more likely to be unprotonated. However both pK_a values are well above the model value of 3.8 and 4.5 respectively for aspartic acid and glutamic acid and so are weaker acids than expected.

In comparison, there is a third charged residue, R312, in the TM cavity and which is not suggested to be part of the proton translocation pathway. The predicted pK_a is 9.65 and while this is much lower than the model value (12.5), it indicates that R312 is likely to always be protonated and explains why the residue is not thought to be tirable and part of the translocation pathway. The pK_a value for R312 remains above 9.5 throughout the Apo-pe-dh and Apo-pe-eh simulations.

The effect on the predicted pK_a is explored in apo FucP systems during the simulation time (Figure 5.11) and the predicted pK_a for both D46 (A) and E135 (B). The pK_a for D46 is slightly higher in the Apo-pe-dh simulations, where D46 is already protonated and similarly the pK_a for E135 is slightly higher in the Apo-pe-eh simulations. This implies that in both cases, the system has accommodated the proton at each position, making it more likely that the D46 and E135 remain protonated in their respective protonated systems. However all pK_a values remain

around the physiological pH and so there is approximately 50 % probabilities for protonation.

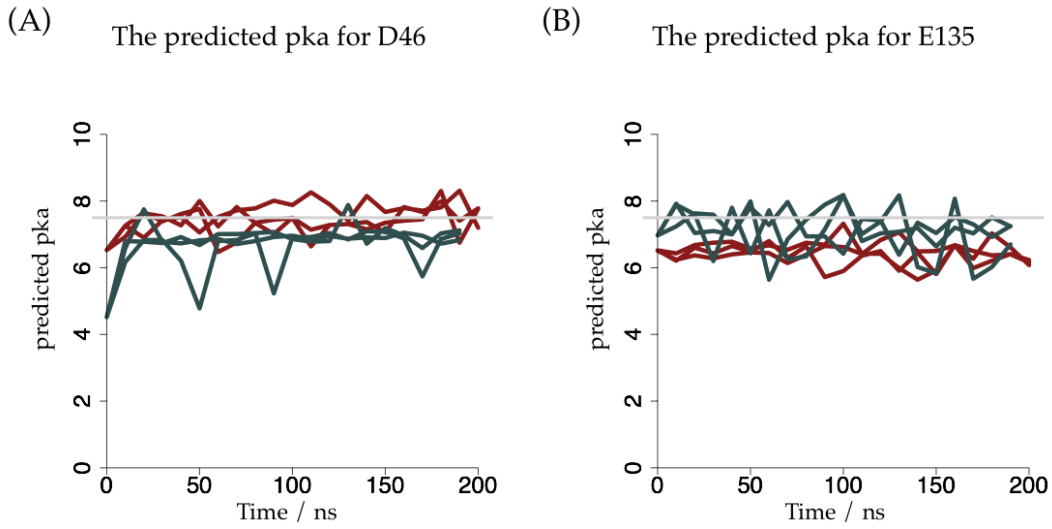


Figure 5.11 The predicted pK_a values for (A) D46 and (B) E135 during the simulations: Apo-pe-dh (■) and Apo-pe-eh (■). The physiologically relevant pK_a is given by the grey line (Padan et al. 1981).

5.3.5 The ligand binding site in FucP

The final simulations explored in this chapter were Fuc-pe-up and Fuc-pe-eh. The protonation state of E135 is potentially important because it is part of the ligand binding site and so could affect ligand binding, as well as proton translocation. L-fucose was docked to the central TM cavity of FucP (in the same position as the sugar ring in β–NG in Figure 1.9 (B)). Analysis of the metrics in Figure 5.9 for the fucose bound simulations determined no significant change in conformation and FucP remained outward open. Therefore it appears that the timescale of the MD simulations is not sufficient to capture the outward to occluded conformation.

The residue contacts with fucose were determined in both the Fuc-pe-eh and Fuc-pe-up simulations. In Figure 5.12 those residues interacting with fucose for at least 80 % of all simulation time defined the binding site: L131, E135, Q159, N162, I391 and Y365. Of these residues, E135, Q159, N162 and Y365 are already known to impact functional assays when mutated (Madej et al. 2013), however two further residues are identified; L131 and I391. Further, the residues involved in fucose interactions do not change between the Fuc-pe-eh and Fuc-pe-up simulations.

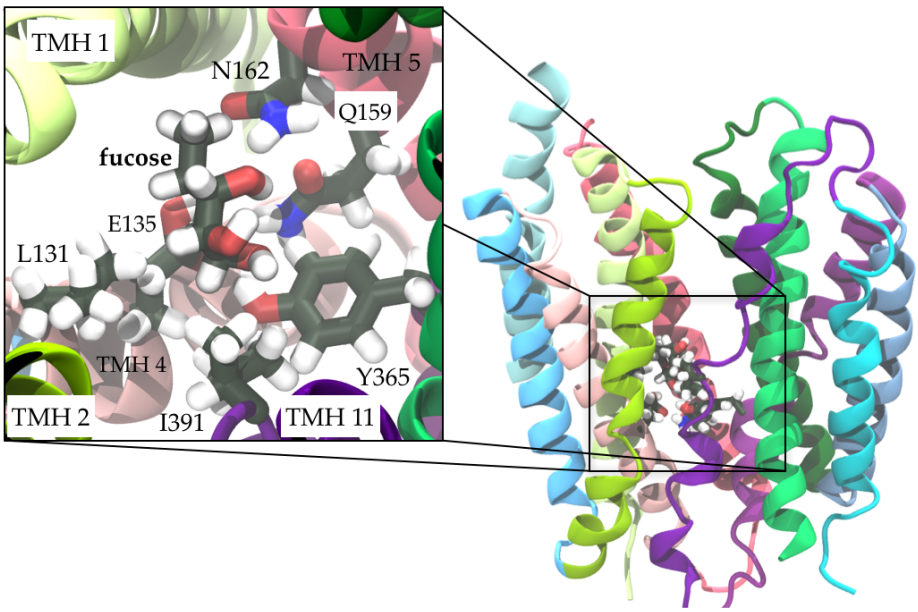


Figure 5.12 The ligand binding site of FucP during MD simulation. The duration of interactions between fucose and residues was measured in the three repeats of Fuc-pe-up and Fuc-pe-eh and those interactions which exist for more than 80 % of simulation time in both protonation states (E135 protonated and charged) are: L131, E135, Q159, N162, I391 and Y365. Of these residues, E135, Q159, N162 and Y365 are functionally important, disrupting transport on mutation (Madej et al. 2013).

5.4 Conclusions

The X-ray crystal structure of FucP was solved in 2010 by Dang et al. (Dang et al. 2010) in an outward conformation, open to the periplasm. The structure is atypical

of subsequent MFS proteins, whose structures are more occluded (Quistgaard et al. 2013; Newstead et al. 2011; Pedersen et al. 2013), and the crystal contacts indicate that interactions between loops 5 – 6 and 11 – 12 and the subsequent unit cells could be stabilising the conformation. MD simulations of the apo and substrate bound FucP indicate small motions towards a more occluded state. However, the periplasmic side of the protein is still solvent accessible and so remains open. The next stage in this work is to explore sampling of the path to an occluded conformation using biased MD simulation.

6 Exploring conformations in FucP

Given that substantial conformational changes were not seen in unbiased MD, the aim for this final results chapter was to explore the outward to occluded states of FucP. This was done using steered MD simulations predict the first conformational changes of the transport cycle, biasing the protein structure so that it moves from an outward open to an occluded conformation. Different protonation states of D46 and E135, the only two titrable residues in the central cavity of FucP, indicated that the occluded conformation was most stable when D46 was protonated.

6.1 Introduction

This chapter aims to build on the MD simulations described in Chapter 5, which investigated the dynamics of FucP in unbiased simulations. To this end, biased steered MD (SMD) techniques were used to force the system along the predicted conformational space into an occluded state. The system was further tested for stability in that occluded state using unbiased MD. The transport cycle of FucP is investigated in the presence of L-fucose, which is one of the key sugars transported as well as the protonation state of the two titrable residues in the central cavity, E135 and D46 (Figure 5.2). The study which originally described the FucP transport pathway predicted one proton will be translocated with one L-fucose molecule (Dang et al. 2010). However recent studies imply that the stoichiometry for ligand/H^+ transport in POT transporters is not necessarily 1:1 (Parker et al. 2014) and so D46H, E135H and D46H/E135H states were explored.

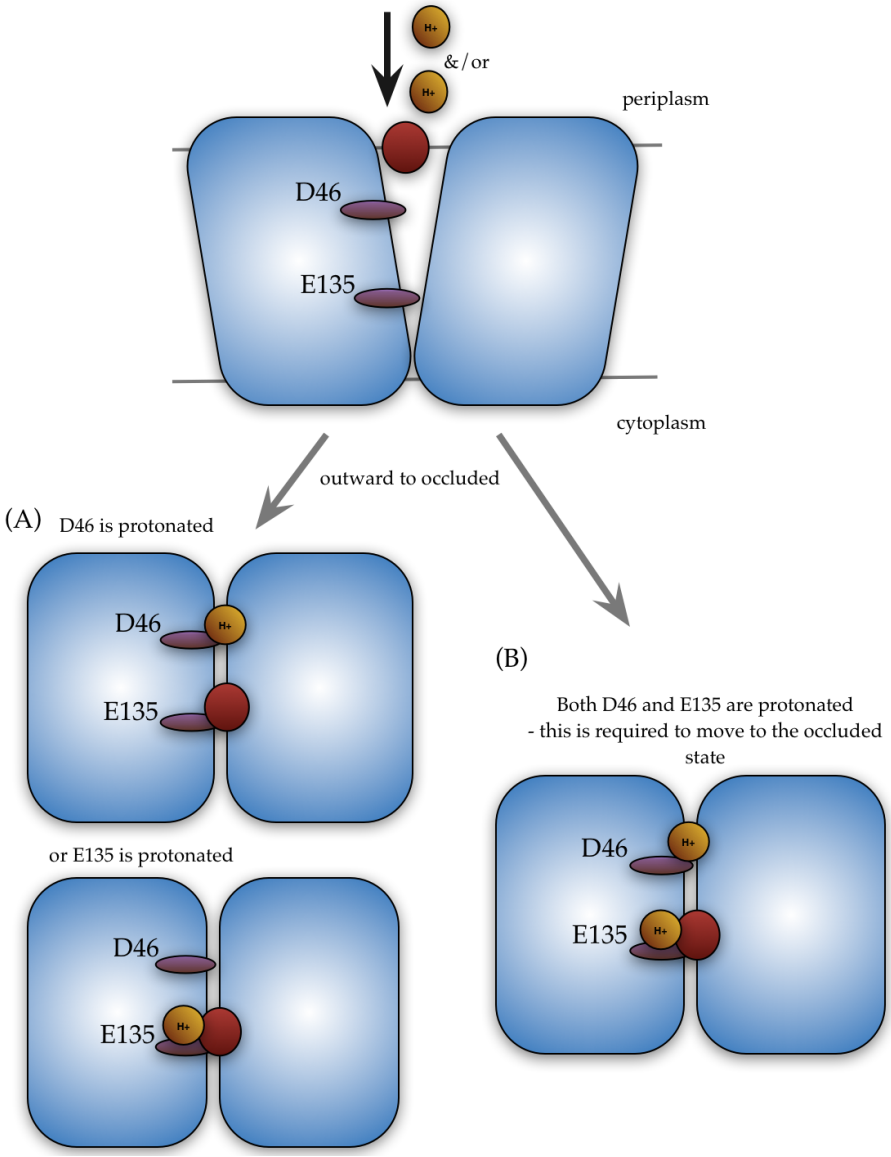


Figure 6.1 The different protonation states in the FucP alternating access mechanism. Movement from the outward to inward open conformation is facilitated by the presence of fucose and protonation of at least one of D46 or E135 (A) but it is possible that both residues could be protonated (B), allowing transport of fucose into the cell. The outward to occluded transition is shown here as that is explored using steered MD in this chapter.

6.1.1 Biased MD to probe conformations in proteins

Like MFS transporters, ABC transporters also operate by the alternating access mechanism and the mechanism has been explored through computational methods by Moradi and Tajkhorshid (Moradi & Tajkhorshid 2013b; Moradi & Tajkhorshid 2013a). They investigate a more sophisticated set of non-equilibrium MD methods

than will be used for FucP, but demonstrate the utility of biased simulations to investigate conformation space in transporters. The aim for these simulations is to drive the conformations along defined reaction coordinates, namely collective variables, with the caveat that deciding these pathways are non-trivial. For the alternating access mechanism, this can be reduced to a set of distances between atoms or RMSD values of the protein structure from a target conformation (Schlitter et al. 1993; Izrailev et al. 1997). In the case of ABC transporters, Moradi and Tajkhorshid used the distance between the centre of mass of the NBDs as the metric (Moradi & Tajkhorshid 2013b), and a similar process can be used in MFS transporters and in particular FucP. Here, two atoms at the periplasmic side of the TM cavity on TMHs 1 and 7 were chosen and a force was applied to pull these together and enter an occluded state, as has been done in GlpT (Moradi & Tajkhorshid per commun.).

6.2 Methods

6.2.1 Steered molecular dynamics

The steered MD (SMD) simulations were used with the PLUMED 1.3 plugin (Bonomi et al. 2009) in GROMACS v4.6.1 (Abraham et al. 2015). The plugin required a definition of a collective variable, which pulling two C-α atoms on TM helices 1 and 7 to a distance of 5 Å. The system was biased towards obtaining that distance by application of a force (kJ/mol) and velocity (nm/ns) on the collective variable. Five different conditions were first tested (Table 6.1) and simulation 4 was used for further analysis since this run used the lowest force and velocity which still pulled the collective variable to 5 Å.

Simulation	Force κ (10^3 kJ/mol)	Velocity (nm/ns)
1	100	0.001
2	10	0.001
3	100	0.0005
4	10	0.0005
5	10000	0.0001

Table 6.1 The forces and velocities applied to pull the collective variable to 5 Å in the steered MD simulations. This pulled the two domains of FucP into an occluded conformation.

6.2.2 Molecular dynamics simulation

Snapshots taken from the Fuc-pe systems (Chapter 5) were used for the UP and E135H systems. D46 was then protonated in these snapshots to obtain D46H and D46H/E135H systems (see Table 6.2). Water molecules and sodium and chloride ions were then added to the systems to a concentration of 150 mM followed by two restrained MD runs whereby all heavy atoms were restrained by a harmonic potential of 2.39 kcal mol^{-1} $Å^{-2}$ for 1 ns. Finally, 200 ns of production runs were performed on three repeats that differed in their initial velocities only. As an additional check of the simulations we analysed the area per lipid for all simulations and observed that there was no significant difference compared to t = 0 or between runs. MD simulations were performed with GROMACS v4.6.5 (Abraham et al. 2015) using the OPLS-AA (Jorgensen et al. 1996; Kaminski et al. 2001b) force-field and TIP3P water molecules (Jorgensen 1981). Production simulations were performed in an NPT ensemble maintained at 323 K and 1 bar pressure. The integration time step was set as 2 fs and a stochastic dynamics integrator (Van Gunsteren & Berendsen 1988) was used. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method (Essmann et al. 1995) with a 14 Å cut-off and 1 Å space grid. The Lennard-Jones potential used a cut-off of 9 Å, with a

switch at 8 Å. The LINCS algorithm (Hess et al. 1997) was used in order to constrain bond lengths in both the lipid molecules and the protein. For ligand bound simulations, the coordinates for the ligand were taken from snapshots described above.

Name	Description
UP	L-fucose bound FucP with no protonation on D46 or E135 (3 x 200 ns)
D46H	L-fucose bound FucP with D46 protonated (3 x 200 ns)
E135H	L-fucose bound FucP with E135 protonated (3 x 200 ns)
D46H/E135H	L-fucose bound FucP with D46 and E135 protonated (3 x 200 ns)

Table 6.2 Summary of simulations of FucP with protonation states. L-fucose is present in all and protein is embedded in a POPE bilayer.

6.3 Results and Discussion

The first aim for this chapter was to determine the atoms upon which to apply the collective variable, such that biased SMD could move the outward-open state of FucP into an occluded conformation. Success was defined as the two domains moving together in a rigid body motion to close the periplasmic side of the protein. Only one C-α atom defined at the periplasmic end of TMHs 1 and 7 was required to do this. The importance of the protonation states for D46 and E135 are then explored in both the SMD and the unbiased MD simulations of FucP with L-fucose in the central cavity between the two domains.

6.3.1 Steered MD simulation of FucP

Defining the atoms on which to apply the collective variable (CV) is not trivial as your decision pushes the simulation further and further from an unbiased system and so care must be taken to still obtain data that is physiologically relevant. To this end, Li et al. (Li et al. 2015), who has explored the conformation landscapes of GlpT, showed that two key TM helices define the closure of the periplasmic side of the cavity, namely TMHs 1 and 7 (Moradi & Tajkhorshid, per commun.). An advantage to using fewer atoms is that the simulations are more efficient and so ONE C-α was chosen on TMHs 1 and 7 and these were pulled to 5 Å (Figure 6.2).

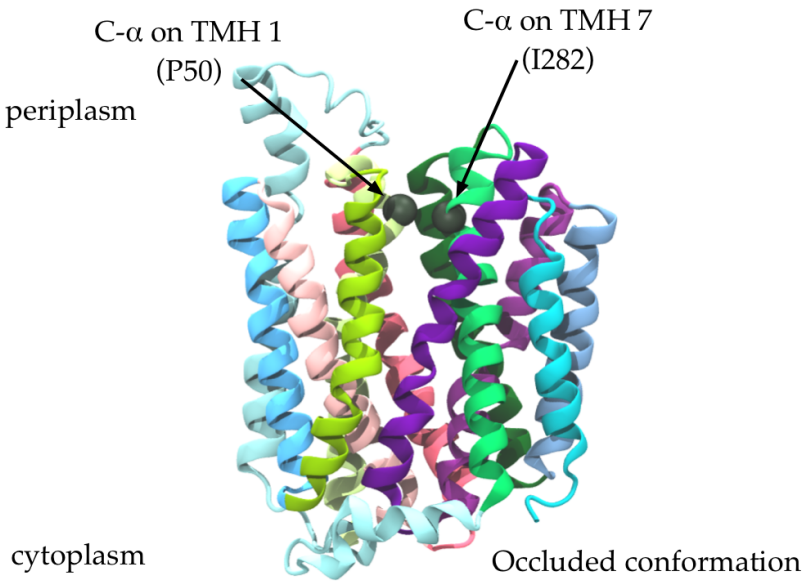


Figure 6.2 The C-α atoms on P50 and I282 were pulled to within 5 Å in the steered MD simulations of FucP. A range of forces and velocities were used to pull the two domains of FucP into an occluded state for the four protonation states: UP, D46H, E135H and D46H/E135H (Table 6.1).

In order to test this, the set of simulation variables used in the SMD (listed in Table 6.1) were repeated three times for each of the four protonation states. In total, five

different simulation conditions were used to pull the CV in the steered MD and it was found that simulation 5 ($\kappa = 10 \times 10^{7}$ kJ/mol, $v = 10 \times 10^{-5}$ nm/ns) did not pull the CV to 5 Å. Since this described the limit at which FucP is pulled into an occluded state, simulation 4 ($\kappa = 10 \times 10^{3}$ kJ/mol, $v = 5 \times 10^{-4}$ nm/ns) was the chosen conditions for further analysis.

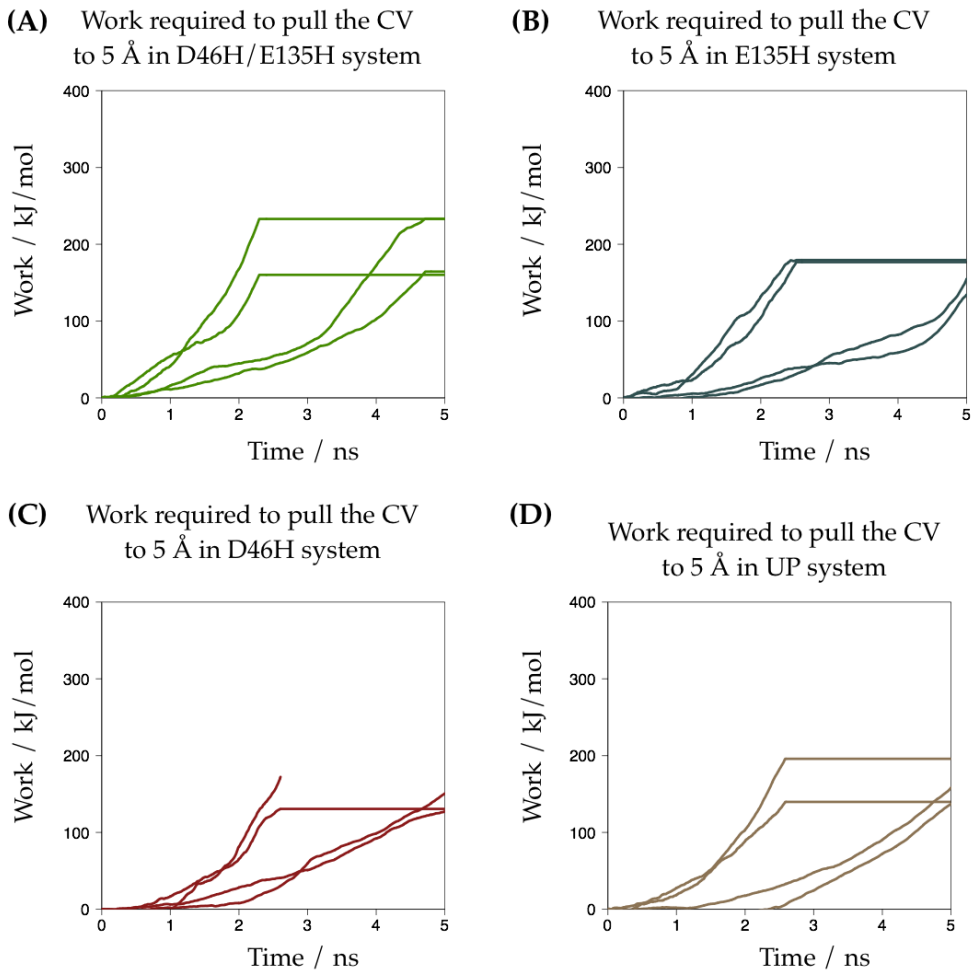


Figure 6.3 The SMD of the four protonation states with $\kappa = 10 \times 10^{3}$ kJ/mol, $v = 5 \times 10^{-4}$ m/s. The protonations states do not affect the range of work done on the system. (A) D46H/E135H (■), (B) E135H (■), (C) D46H (■) and (D) unprotonated (UP) (■). In (C) the fourth experiment exploded and so finished recording work at 2.4 ns.

The initial conditions are described in Figure 6.3 (A to D) which indicates that there was a wide spread of values for the work required to pull the CV to 5 Å. Further to this, there does not appear to be a clear trend regarding the different protonation states and the work required on the system. The work is around 200 kJ/mol regardless of protonation state and simulations with smaller force take longer to reach this work and subsequently take longer to attain a CV distance of 5 Å.

As well as pulling the C-α atoms to within 5 Å, the aim for the simulation was to move FucP into an occluded conformation and in order to do this a rigid body movement of the two domains is thought to occur. Remarkably, this was found to be the case (Figure 6.4 (A)) with the initial and final snapshots showing that. The comparison of occluded FucP (final snapshot of simulation 4) with an occluded X-ray crystal structure (XylE: 4JA4) indicates that the FucP system has moved into an occluded state (Figure 6.4 (B)).

Further to the movement into the occluded state, the interactions that maintained the structure of the domains and permitted a rigid body rotation were investigated. The domain structure appears to be maintained through the SMD simulations by a series of intra-domain interactions (Figure 6.4 (C)): in domain 1 N42 – Y101 (2.0 Å) and N45 – Y212 (2.2 Å) and in domain 2 Y280 – E415 (1.7 Å) and G347 – Y284 (2.0 Å). These interactions appear to be sufficient to permit a rigid body movement of each domain.

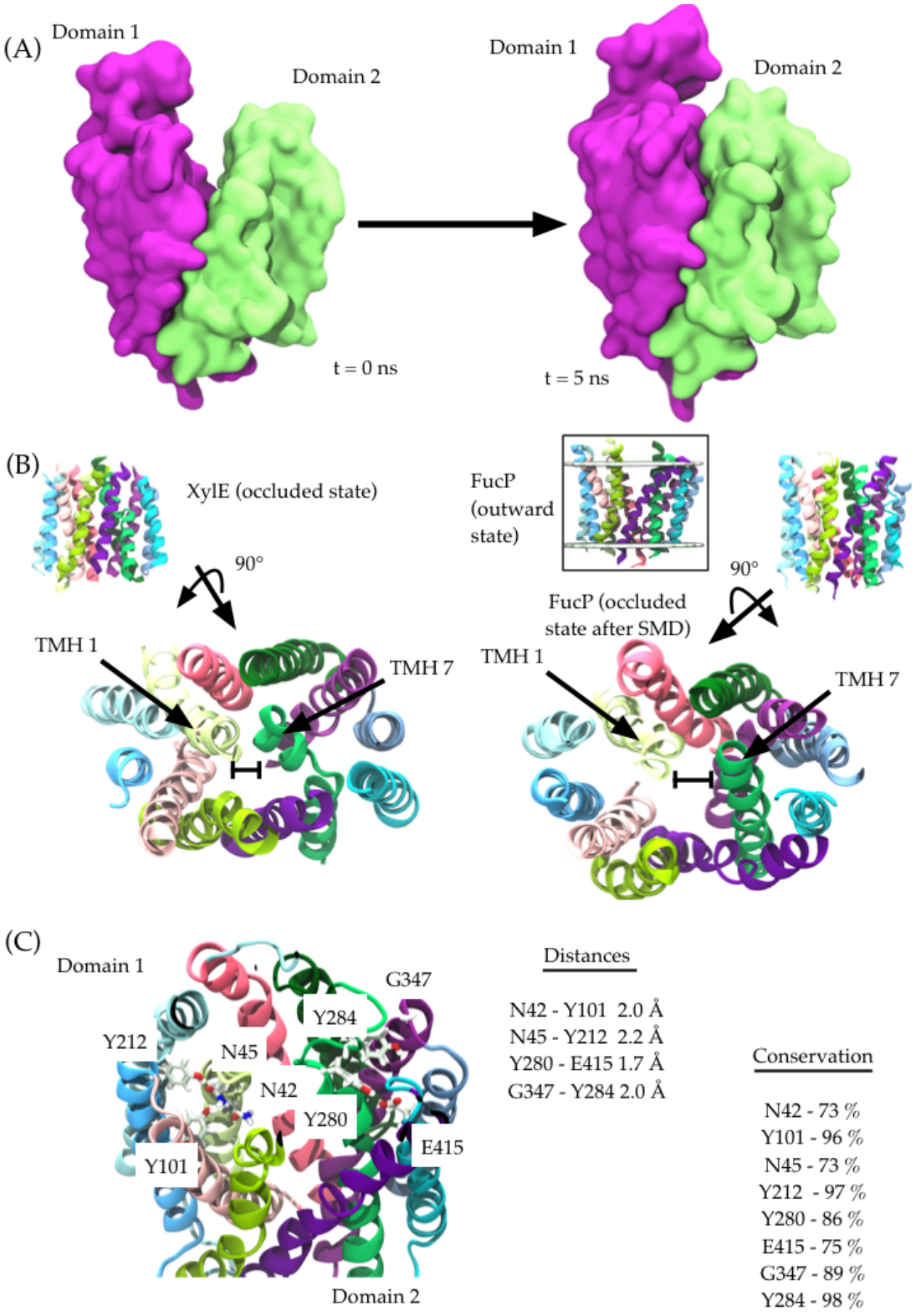


Figure 6.4 The motion from outward to occluded states during steered MD simulation. (A) The surface representations of the whole protein shows that the two domains move in a rigid body fashion to the occluded state (t = 5 ns) from the outward state (t = 0 ns) for D46H/E135H system (this occurs in all repeats and protonation states). (B) Comparison of the occluded FucP (t = 5 ns) with the occluded X-ray crystal structure, XylE (4JA4) and the periplasmic ends are in similar conformations (C) Interactions in each domain which are maintaining their fold through the SMD. There are four hydrogen bonds in the simulation that maintain the close packed character of the two domains: N42-Y101, N45-Y212 in domain 1 and Y280-R415, G347-Y284 in domain 2.

The four protonation states were explored as it was hypothesized that the presence of a proton, as well as fucose, is key to the destabilisation of the outward open state to allow movement through the occluded to the inward open state. Interestingly there appears to be little difference between either the final conformations or the work profiles (Figure 6.5) of the four states. One notable trend, though it is small, is that the D46H/E135H system (green lines in Figure 6.5) requires more work for the CV to reach 5 Å and the UP system requires less. The caveat here is that there is a lot of overlap between the four protonation states and so the trend is not clear.

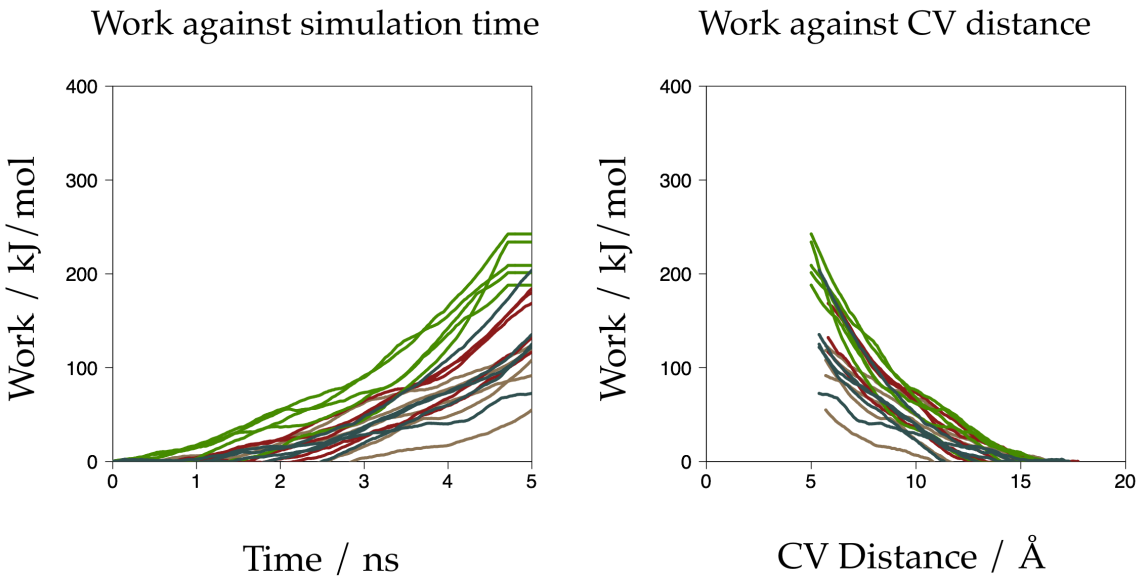


Figure 6.5 Steered MD simulations of the four protonated systems of FucP. The work applied on the system for each of the simulations (3 repeats of 8 SMD simulation variables - Table 6.1): (A) D46H/E135H (■), (B) E135H (■), (C) D46H (■) and (D) unprotonated (UP) (■) states.

It is possible that the short timescale (5 ns) of the SMD simulations means that there is little chance to capture any change in dynamics on the protonation of either D46 or E135. Therefore unbiased simulations of the occluded states were conducted to further explore this and determine whether protonation affects the state.

6.3.2 Unbiased MD simulation of occluded FucP

To explore the four different protonation states (unprotonated (UP), D46H, E135H and D46H/E135H), unbiased MD simulations of length 200 ns, with three repeats, were set up using as starting coordinates the final snapshot of the steered MD simulations above (Chapter 6.3.1). The system was re-equilibrated before the production run (Chapter 6.2.2). The C-α RMSD values between these four snapshots ranged from 2.4 – 2.8 Å, implying that the structures were in similar conformations after the steered MD.

The dynamics of the occluded FucP systems were assessed using RMSD and RMSF (Figure 6.6). Relative to the re-equilibrated system, the C-α RMSD values were computed at each time point in the MD simulations for each of the three repeats of the four protonation states (Figure 6.6 (A)). Excepting one simulation of D46H/E135H, the RMSD values remained below 2.4 Å and this anomalous simulation had a jump in RMSD value at 150 ns to approximately 3 Å. Compared with the other two D46H/E135H simulations, movement of TMHs 7, 8 and 11 into the TM cavity, explains the increased RMSD values from 150 – 200 ns. This does not result in large motions in the simulation, only that there appear to be fewer water molecules in the periplasmic side of the TM cavity.

The RMSD values between the final snapshots of the four occluded FucP simulations was 3.1 – 3.9 Å. This implies that the relative motions between the different protonation states are greater than the motions within each MD repeat (which is given in Figure 6.7 (A)), though this requires further investigation.

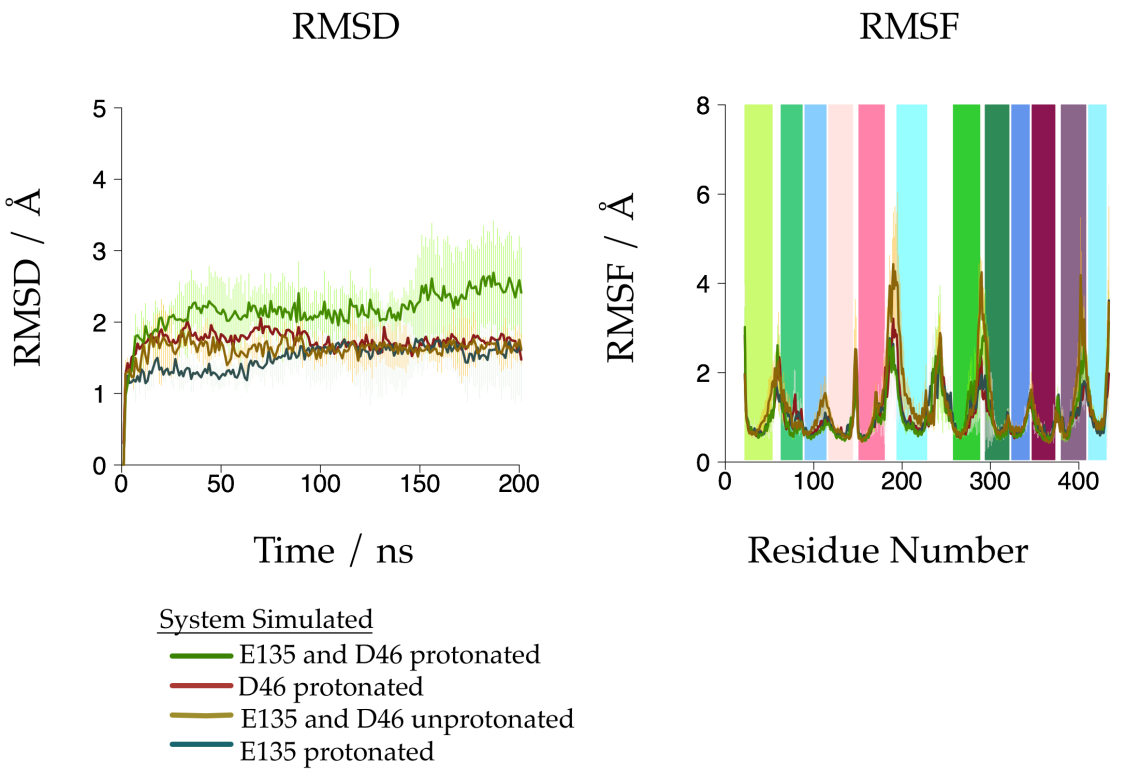


Figure 6.6 (A) The RMSDs and (B) RMSFs for FucP in an occluded conformation and for the four protonation states. Each simulation is represented by: D46H/E135H (■), D46H (■) and E135H (■) and UP (■). The RMSD for each system indicate the largest motions are in the D46H/E135H system. The RMS fluctuations are similar in all but the 5 – 6 and 7 – 8 loops, which are greater in the UP system.

The relative fluctuations of each residue were explored using RMSF (Figure 6.6 (B)). In all protonated states, the largest fluctuations are seen in the three loops: loop 5 – 6, loop 6 – 7 and loop 7 – 8. The fluctuations in loop 5 – 6 and 7 – 8 can be explained since it is one of the dominant motions involved in the closure of the periplasmic side of the TM cavity. Therefore it is most likely the systems relaxing in the occluded state after having two TM helices pulled together, especially since there was no large difference in comparing the four protonation states. Further to this, loop 5 – 6 is reasonably large (30 amino acids) so is expected to have the greatest fluctuations of the loops and the predominant motion is the loop closing over the TM cavity, interacting with domain 2 loops (predominantly loop 7 – 8).

The distance was computed between the centres of mass of the C-α atoms in loop 5 - 6 and the periplasmic loops in domain 2 (Figure 6.7 (A)). This shows that the distance between loop 5 – 6 and domain 2 is larger in the UP system compared to the other three systems. Counter to expectation, the smallest distance between the domains was in E135H. Here, it was expected that the decrease in hydrophilicity in D45H and D46H/E135H systems would intuitively reduce the number of waters in the periplasmic side of the TM cavity and so would result in the most occluded state. However it does not appear to impact the position of the loops and suggests that there is no further rigid body rotation of the domains.

A further two distance metrics were used to define the degree of openness of the periplasmic side of the TM cavity, which measure the distance between pairs of helices lining the interface between domains 1 and 2 in FucP. The distances between TMHs 2 and 11 (Figure 6.7 (B)) and between TMHs 5 and 8 (Figure 6.7 (C)) remain stable across all protonation states, which imply that the occluded state remained stable.

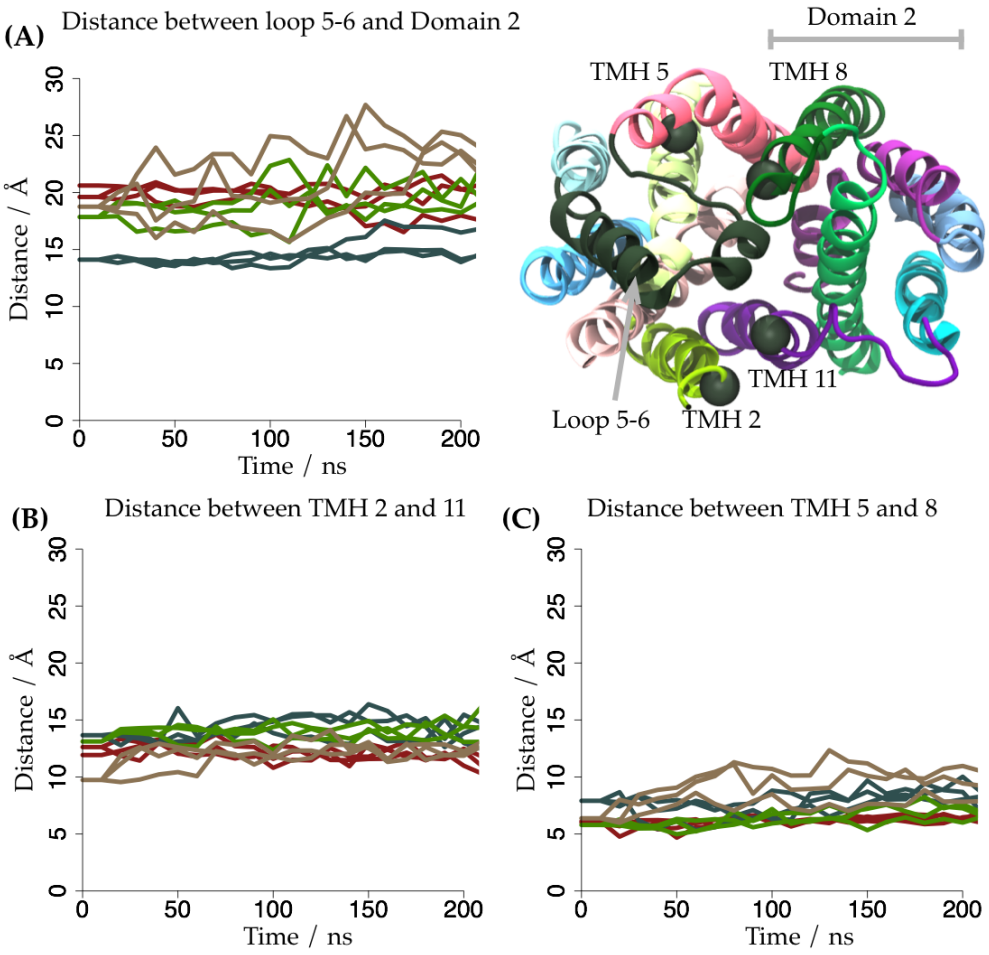


Figure 6.7 The distances between the two domains of FucP. The distance between the two loops from separate domains: (A) Loop 5 – 6 and domain 2 loops. The adjacent helices that define the interface between the two domains of FucP: (B) TMH 2 and TMH 11; (C) TMH 5 and TMH 8. Each simulation is represented by: D46H/E135H (■), D46H (■) and E135H (■) and UP (■).

The centres of mass between E135 and R312 and D46 and W278 are shown in Figure 6.8 (A) and (B). At the original proposal for the FucP transport mechanism, the reduced distance between D46 and W278 is thought to describe an occluded state, from the outward open state (Dang et al. 2010). Since E135 is closer to the cytoplasmic side of the TM cavity, the distance between it and R312 was used as a check to ensure the cytoplasmic side of the transporter remained occluded. The distance was constant in all states excepting the E135, which saw fluctuations of the E135 – R312 distance ranging 10 – 14 Å (UP, D46H an D46H/E135H distances were

below 10 Å – see Figure 6.8 (B)). However this does not coincide with the large scale conformational changes required to open the cytoplasmic side of the transporter.

Conversely, the distance between D46 and W278 differed between the four protonation states (Figure 6.8 (A)). The UP and D46H/E135H systems both contained the largest fluctuations in the distance (ranging from 5 – 12 Å), whilst E135H had the smallest fluctuations (remaining below 5 Å). The majority of this motion can be attributed to the two residues sampling different rotamer conformations and this implies that the interaction of D46 and W278 is not steady in these simulations. It is not clear from this analysis whether D46 and W278 create a constriction point in the TM cavity. The hypothesis is that a dry patch at the D46 - W278 z – axis position of the TM cavity will prevent the proton leakage across the membrane and therefore enter a conformation that is favourable for opening the cytoplasmic side of the TM cavity and allowing ligand passage into the cytosol. Therefore investigation into the number of waters within the cavity can help describe this.

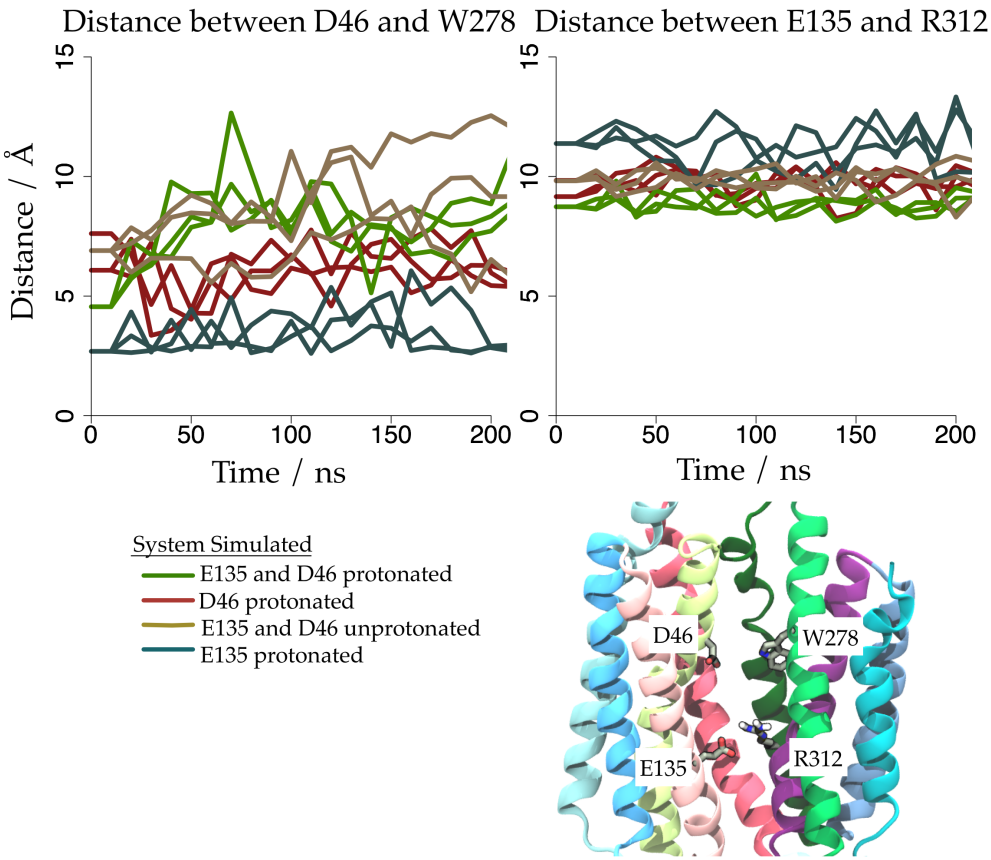


Figure 6.8 The distances between D46 and W278 (A) and E135 and R312 (B). The centres-of-mass of the side chains were determined as a metric for how much fluctuation there is within the central cavity of FucP. The distance between E135 and R312 remains the same regardless of protonation, which implies that little rearrangement is occurring at the ligand binding site. However the distance between D46 and W278 is slightly greater when D46 and E135 are protonated, which ties in well with the solvent accessibility of the TM cavities in each system.

6.3.3 The predicted pK_a of each FucP system

Similar to the predicted pK_a values for FucP in Chapter 5.3.4, the pK_a values for the four protonation states of occluded FucP remained below physiological pH of approximately 7.5 (Padan et al. 1981). This suggests that, at most, there is a 50 % probability of either residue being protonated (Figure 6.9 (A) and (B)). For D46, the E135H and UP systems predict an approximate pK_a value of 6 and so these systems suggest that D46 will most probably be charged. Similarly the D46H and UP

systems predict an approximate pK_a value of 6, suggesting E135 would be charged (Figure 6.9 (B)). Again, the MD simulations appear to be adapting to the protonation state given to the residues, with D46H and D46H/E135H having an increased likelihood of D46 protonation and with E135H and D46H/E135H an increased likelihood of E135 protonation.

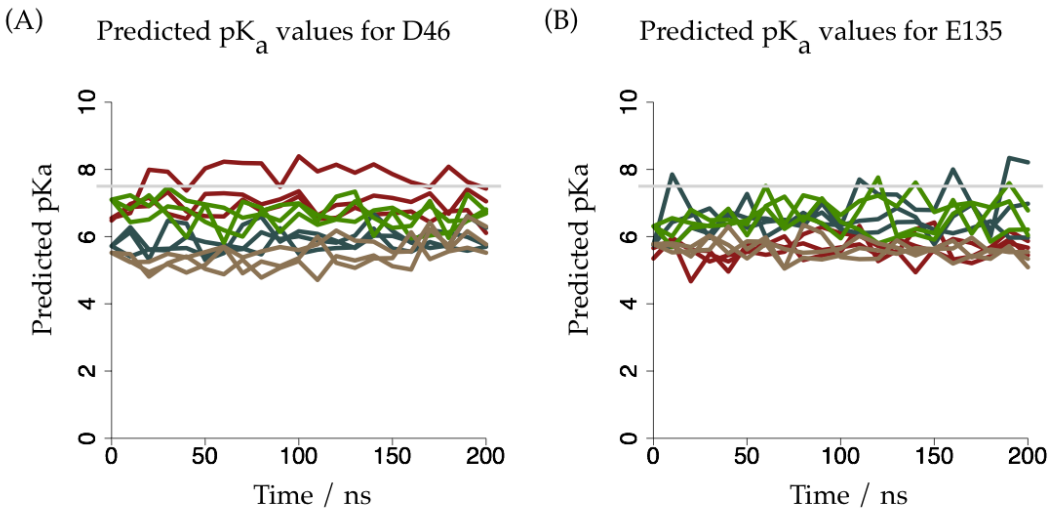


Figure 6.9 The predicted pK_a values for (A) D46 and (B) E135 during the simulations: Apo-pe-up (■), Apo-pe-dh (■), Apo-pe-eh (■) and Apo-pe-edh (■). The physiological pK_a is given by the grey line at 7.5 (Padan et al. 1981). In all simulations and for both D46 and E135, the predicted pK_a remains below the physiological pH (excepting the pK_a for D46 in one repeat of D46H).

6.3.4 How dry is the TM cavity in occluded FucP?

The number of water molecules in the TM cavity gives an idea of how 'open' each system is at both the periplasmic and cytoplasmic side and what is more, it also describes whether the TM cavity retains its hydrophobic constriction throughout the simulations. In Figure 6.10 the number of waters at every 1 Å slice in the central TM cavity was counted for the four systems by counting the number of water oxygen

atoms in that slice of the cavity. With E135 charged (D46H and UP systems), there were transient water wires across the hydrophobic constriction, whereas the protonation of E135 significantly reduced their occurrence. This water permeability has been demonstrated in GlpT, amongst transporters in other superfamilies, by Li et al. (Li et al. 2013). It was found that the permeability only occurred with ligand bound to the TM cavity and its transient nature implied that it is caused by the fluctuations in side chains. Further to this, it is hypothesized that the ligand may even be co-transported with water, as suggested by MD simulations of XylE (Wisedchaisri et al. 2014). In FucP, the constriction in the cavity near E135 is composed of Y365 and the position of its side chain controls the movement of waters through the cavity. Any protonation in the system also led to fewer waters present at the D46 position of the TM cavity, with a transient dry patch seen there in the D46H system.

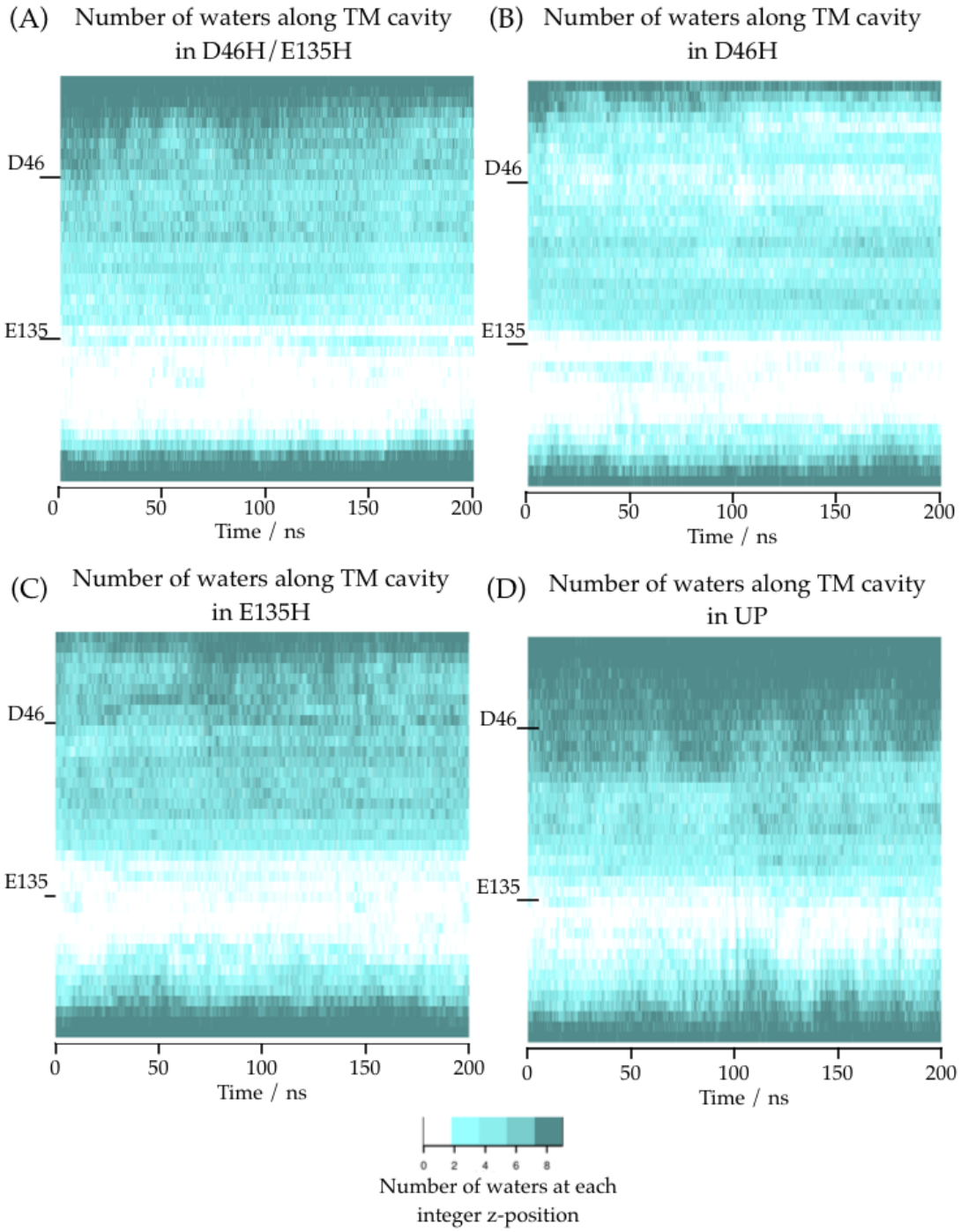


Figure 6.10 The heat maps show the numbers of waters along the TM cavity for the occluded, fucose bound FucP systems. Each heat map represents the (A) D46H/E135H, (B) D46H, (C) E135H and (D) UP states. In D46H and UP, there were transient water wires across the 'dry' patch of the TM cavity.

These observations imply that protonation of at least one of D46 or E135 is required to maintain the dry cavity. The timescale of the simulations was not long enough to fully capture this, but the trend of the UP system implied that this is the most solvent accessible TM cavity (Figure 6.10 (D)). The protonation of E135 appeared to promote the 'dry' patch at the E135/Y365 position of the cavity and D46 protonation reduces the number of water molecules at the D46/W278 position. With regards to the transport cycle in FucP, this implies that the occluded conformation requires the protonation of either E135 or D46. This is key to the closure of the periplasmic side of the TM cavity and the subsequent drying of the cavity. To maintain its role as a transporter, it is hypothesized that the hydrophobic constriction should be maintained. In order to do this from outward to inward open, the constriction is required to move from the E135/Y365 nearer to the cytoplasmic side of the cavity to the D46/W278 position closer to the periplasmic side. The occluded simulations indicate that this hypothesis could be the case, but further exploration of the inward open conformation is needed to verify it.

6.3.5 Comparison of work to Liu et al., 2015

A recent paper by Liu et al. (Liu et al. 2015) investigated the effect of protonation of E135 on FucP during accelerated MD simulations. In both E135 and E135H systems, they used temperature (310 – 500 K) to enhance sampling of the conformational landscape. Each run lasted for 30 ns and two distances at the periplasmic and cytoplasmic side of the TM cavity defined the collective variables to decide the snapshot to take forward into the subsequent run. Therefore the conformations of FucP sample from outward to occluded then to inward open. A series of snapshots in distinct conformations (measured by the two distances) then allowed further simulations to obtain free energy calculations along the reaction pathway. Their suggestion is that there is a stable occluded state, and to move to the inward open

conformation, 7 kcal/mol and 6 kcal/mol is required for the E135 and E135H systems, respectively. So protonation of E135 is required to move into the inward open conformation and then stabilised the state.

This neglects the effect of D46 protonation on the conformation of FucP, which aids the drying of the TM cavity in the occluded simulations of FucP in this thesis. The hypothesis from the simulations is that the protonation of D46 is required to stabilise the occluded conformation and the removal of the charge is sufficient to dry the cavity at D46/W278. It is this constriction that enables the opening of the cytoplasmic side of the TM cavity and allows fucose to move into the cell.

6.4 Conclusions

The suggestion from the unbiased MD on the four protonation states implies that protonation of D46 appeared to stabilize the occluded conformation with respect to the degree of openness at the periplasmic side of the TM cavity. Taking the number of waters within the cavity and also measuring distances between helices at the periplasmic opening obtained measurement of the occlusion of FucP. Analysis of the number of water molecules in the TM cavity led to the hypothesis that E135H stabilised the hydrophobic constriction at the E135/Y365 position. Further to this, the protonation of D46 led to fewer waters at the D46/W278 position in the TM cavity. Hence, the hypothesis is that the position of protonation on either D46 or E135 stabilises or induces the hydrophobic constriction at this point, maintaining transporter function of FucP.

The suggestion from this is the D46 needs to be protonated in order to move into the occluded conformation. However the role of E135 is less clear and its protonation in

MD simulations of the occluded FucP is not sufficient to determine whether E135 protonation alone is required to open the cytoplasmic side of the TM cavity, or whether D46 needs to be protonated too. Since E135 is also within the ligand binding site, the suggestion is that the sugar moiety also plays a part. It is also possible that the proton moves between the D46 and E135 via a water wire, but this cannot be captured using the MD simulation techniques in this work and it is not immediately evident whether one or two protons are required for the transport cycle of FucP.

7 Conclusions and Future Directions

7.1 Major Conclusions

The analysis of conservation of amino acids according to chemical properties has determined a series of sites in the multiple sequence analysis of all MFS proteins with X-ray crystal structures. This research took inspiration from to the work conducted for GPCR proteins, first by Ballesteros and Weinstein who numbered residues in class A GPCRs (Ballesteros & Weinstein 1995) and then the analysis by Venkatakrishnan et al. (Venkatakrishnan et al. 2013) who determined 24 conserved contacts across GPCR structures. This provides metrics to aid homology modelling for MFS proteins, but also determined key structural points within the TM fold that relate to either static of moving contacts.

The helix-helix contacts in MFS proteins were assessed and it was found that a substantial number of them remained static and present in all conformations of structures. Notably, the contacts for TM helices 1 and 7 showed the greatest positional movement according to which state the protein was in, indicating that the motion of these helices went beyond the rigid body rotation of the two 6 TM-helix domains against each other, as predicted to occur in the alternating access mechanism. This observation ties in well with a hypothesis that the outward to occluded transition is modulated by the position of TMHs 1 and 7, and the closure of the extracellular side of the TM cavity by the same helices coming together (Moradi & Tajkhorshid per commun.). The analysis agrees with that structural research conducted by Fowler et al. in a POT transporter, which suggested that

TMHs 1, 2, 7 and 8 defined the closure of the periplasmic side of the TM cavity (Fowler et al. 2015).

Two homology models of SV2A in the inward and outward open states were constructed using GlpT and FucP as templates. These models were informed by the early workings of Chapter 3, though a complete numbering system was not ready in time for the models. MD simulations of both the apo and ucb 30889 bound systems determined that the drug binding site was similar to that proposed by Shi et al. (Shi et al. 2011) and an additional residue was found to impact drug binding, D670. This provided insights into the proposed drug binding site, however extension of this work relies upon an X-ray crystal structure for SV2A or further experimental work to better characterise its function. Recently, it was shown to transport galactose, but the role of this function on the synaptic vesicle membrane remains unclear (Madeo et al. 2014).

The effect of protonation of both D46 and E135 was explored in biased and unbiased MD simulations of FucP. The occluded states of FucP, derived through steered MD of the outward open state by pulling together TMHs 1 and 7, showed that the protonation of D46 and E135 caused a drying of the cavity. The proposed mechanism for transport in FucP is that a 'dry patch' is maintained at all times in the TM cavity and that this travels from between E135 and Y365 to D46 and W278. With D46 protonated, the cavity contained fewer water molecules at the D46/W278 position, though the timescales are not sufficient to see a complete hydrophobic constriction, something that is expected in longer simulations.

Previous MD simulations saw a transient wetting of the TM cavity, allowing water to pass through the transporter (Li et al. 2013). This has been demonstrated experimentally in GLUT1 and GLUT2 (Zeuthen et al. 2007; Zeidel et al. 1992) and

was present in the occluded FucP systems where E135 was charged. It is hypothesized that the ligand is required for this water conductance to occur and that it is part of the random fluctuations of the transporter as it moves into a more open state to allow release of ligand. It is possible that waters are co-transported with the ligand, as shown in MD simulations of XylE (Wisedchaisri et al. 2014).

The aim for the FucP chapters was to explore the conformations FucP underwent in the transport cycle, focusing on unbiased MD simulations of the outward open and occluded states. While writing up this thesis, Liu et al. published research into free energy calculations of the FucP transitions from outward to inward open using accelerated MD simulations (Liu et al. 2015). Their focus was on protonation of E135, which they found to affect the transition from occluded to inward open. This has not been explored in this thesis, but is the next logical step to the research to explore occluded to inward open FucP states.

7.2 Future directions

One major direction for the numbering of MFS proteins is to determine the numbering in all the proteins investigated for contacts. Further to this, the application to a completely novel MFS protein would justify whether these key positions on each helix can provide an anchor point that describes the packing of the TM helices and hence can be used to align the target with a template MFS protein.

Due to the chronology of the research in this thesis, the numbering system for MFS transporters was not complete for the SV2A work. Therefore it would be of interest to use the numbering to refine the models of SV2A, to verify the binding site within the TM cavity. In addition, the extracellular loop domain could be added, using the X-ray crystal structure for that domain in the SV2A isoform, SV2C.

For FucP, the completion of the transport cycle and investigation of the occluded to inward open transition using steered MD is the first task. Once this is completed and unbiased MD simulations of the inward open state are investigated, then a free energy profile for FucP transport cycle would be useful. The steered MD simulations provide snapshots along the change in states from outward open to occluded and through to inward open, so these can be used to guide adaptive-bias simulations (Moradi & Tajkhorshid 2013a). Titrable residues such as E135, which mediates the ligand binding site as investigated by Liu et al. (Liu et al. 2015) should be further assessed along with the role of D46 in order to determine the effect of both on resulting energy landscapes and ligand binding to the central TM cavity. This is necessary particularly as the 'dry patch' in the TM cavity is hypothesized to be required for efficient movement between the open states.

8 Bibliography

Abraham, M.J. et al., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, pp.1–7.

Abramson, J. et al., 2003. The lactose permease of Escherichia coli: overall structure, the sugar-binding site and the alternating access model for transport. *FEBS letters*, 555(1), pp.96–101.

Almaula, N. et al., 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. *Molecular Pharmacology*, 50(1), pp.34–42.

Andersson, M. et al., 2012. Proton-coupled dynamics in lactose permease. *Structure*, 20(11), pp.1893–1904.

Andrec, M. et al., 2007. A large data set comparison of protein structures determined by crystallography and NMR: Statistical test for structural differences and the effect of crystal packing. *Proteins*, 69(3), pp.449–465.

Bajjalieh, S.M. et al., 1994. Differential expression of synaptic vesicle protein 2 (SV2) isoforms. *J. Neurosci.*, 14(9), pp.5223–5235.

Baker, D. & Sali, A., 2001. Protein structure prediction and structural genomics. *Science*, 294(5540), pp.93–96.

Ballesteros, J. et al., 1998. Functional microdomains in G-protein-coupled receptors. The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. *The Journal of Biological Chemistry*, 273(17), pp.10445–10453.

Ballesteros, J.A. & Weinstein, H., 1992a. Analysis and refinement of criteria for predicting the structure and relative orientations of transmembranal helical domains. *Biophysical Journal*, 62(1), pp.107–109.

Ballesteros, J.A. & Weinstein, H., 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In S. C. S. B. T.-M. in Neurosciences, ed. *Receptor Molecular Biology*. Academic Press, pp. 366–428.

Ballesteros, J.A. & Weinstein, H., 1992b. The role of Pro/Hyp-kinks in determining the transmembrane helix length and gating mechanism of a [Leu]zervamicin channel. *Biophysical Journal*, 62(1), pp.110–111.

Bambeke, F. Van, Balzi, E. & Tulkens, M., 2000. Antibiotic Efflux Pumps. *Biochemical Pharmacology*, 60, pp.457–470.

Becerra, J.E., Yebra, M.J. & Monedero, V., 2015. An l-Fucose Operon in the Probiotic Lactobacillus rhamnosus GG Is Involved in Adaptation to Gastrointestinal Conditions. *Applied and Environmental Microbiology*, 81(11), pp.3880–3888.

Becker, D.J. & Lowe, J.B., 2003. Fucose: biosynthesis and biological function in mammals. *Glycobiology*, 13(7), pp.41–53.

Beghi, E. & Hesdorffer, D., 2014. Prevalence of epilepsy-An unknown quantity. *Epilepsia*, 55(7), pp.963–7.

Benkert P. Biasini, M. & Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3), pp.343–350.

Benkert, P., Schwede, T. & Tosatto, S.C.E., 2009. QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Structural Biology*, 9, p.35.

Benoit, R.M. et al., 2014. Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. *Nature*, 505(7481), pp.108–11.

Berger, O., Edholm, O. & Jahnig, F., 1997. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophysical Journal*, 72, pp.2002–2013.

Bonomi, M. et al., 2009. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10), pp.1961–1972.

Burckhardt, G. & Wolff, N.A., 2000. Structure of renal organic anion and cation transporters. *Renal Physiology*, 278(6), pp.F853–66.

Bussi, G., Donadio, D. & Parrinello, M., 2007. Canonical sampling through velocity rescaling. *J Chem Phys*, 126(1), p.014101.

Cao, B. et al., 2006. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. *Bioinformatics*, 22(3), pp.303–309.

Chang, W. & Sudhof, T.C., 2009. SV2 renders primed synaptic vesicles competent for Ca2+-induced exocytosis. , 29(4), pp.883–897.

Cole, C., Barber, J.D. & Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, 36, pp.197–201.

Coyne, M.J. et al., 2005. Human symbionts use a host-like pathway for surface fucosylation. *Science*, 307(5716), pp.1778–81.

Crevecoeur, J. et al., 2013. Expression of SV2 isoforms during rodent brain development. *BMC Neurosci*, 14, p.87.

Dang, S. et al., 2010. Structure of a fucose transporter in an outward-open conformation. *Nature*, 467(7316), pp.734–738.

Daniels, V. et al., 2013. Modulation of the conformational state of the SV2A protein by an allosteric mechanism as evidenced by ligand binding assays. *Br. J. Pharmacol.*, 169(5), pp.1091–1101.

Deng, D. et al., 2014. Crystal structure of the human glucose transporter GLUT1. *Nature*, 510(7503), pp.121–5.

Doki, S. et al., 2013. Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), pp.11343–11348.

Doolittle, R.F., 1986. *Of URFs and ORFs: a primer on how to analyze derived amino acid sequences*, University Science Books.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5, p.113.

Edgar, R.C. & Batzoglou, S., 2006. Multiple sequence alignment. *Current Opinion in Structural Biology*, 16(3), pp.368–73.

Eilers, M. et al., 2000. Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), pp.5796–5801.

Elofsson, A. & von Heijne, G., 2007. Membrane protein structure: prediction versus reality. *Annual Review of Biochemistry*, 76, pp.125–140.

Eraly, S. a, 2008. Implications of the alternating access model for organic anion transporter kinetics. *The Journal of membrane biology*, 226(1-3), pp.35–42.

Essmann, U. et al., 1995. A smooth particle mesh Ewald method. *J Chem Phys*, 103, pp.8577–8593.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., et al., 2007. Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Protein Science*, 5, pp.2.9.1–31.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., et al., 2007. Comparative protein structure modeling using Modeller. *Current Protocols in Protein Science*, 5, pp.2.9.1–31.

Eyre, T. a, Partridge, L. & Thornton, J.M., 2004. Computational analysis of alpha-helical membrane protein structure: implications for the prediction of 3D structural models. *Protein Engineering Design and Selection*, 17(8), pp.613–624.

Felts, A.K. et al., 2002. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the Surface Generalized Born solvent model. *Proteins*, 48(2), pp.404–422.

Feng, G. et al., 2009. Down-regulation synaptic vesicle protein 2A in the anterior temporal neocortex of patients with intractable epilepsy. *J. Mol. Neurosci.*, 39(3), pp.354–359.

Finn, R.D. et al., 2014. Pfam: The protein families database. *Nucleic Acids Research*, 42, pp.222–230.

Fischbarg, J. et al., 1990. Glucose transporters serve as water channels. *Proceedings of the National Academy of Sciences*, 87, pp.3244–3247.

Fisher, R.S. et al., 2005. Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, 46(4), pp.470–472.

Forrest, L.R., Tang, C.L. & Honig, B., 2006. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophysical Journal*, 91(2), pp.508–517.

Fowler, P.W. et al., 2015. Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Structure*, 23(2), pp.290–301.

Gambardella, A. et al., 2008. Monotherapy for partial epilepsy: Focus on levetiracetam. *Neuropsychiatric Disease and Treatment*, 4(1), pp.33–38.

Guan, L. & Kaback, H.R., 2006. Lessons from Lactose Permease. *Annual Review of Biomolecular Strucure*, 35, pp.67–91.

Guan, L. & Kaback, H.R., 2009. Properties of a LacY efflux mutant. *Biochemistry*, 48(39), pp.9250–9255.

Van Gunsteren, W.F. & Berendsen, H.J.C., 1988. A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation*, 1(3), pp.173–185.

Harden, C., 2001. Safety profile of levetiracetam. *Epilepsia*, 42(Suppl. 4), pp.36–39.

Von Heijne, G., 1992. Membrane Hydrophobicity Protein Structure Prediction Analysis and the Positive-inside.

Heijne, G. Von, 1994. Membrane Proteins : From Sequence to Structure. *Annual Review of Biomolecular Strucure*, 23, pp.167–192.

Hess, B. et al., 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4, pp.435–447.

Hess, B. et al., 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12), pp.1463–1472.

Hinsen, K., 2008. Structural flexibility in proteins: Impact of the crystal environment. *Bioinformatics*, 24(4), pp.521–528.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S., 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14(4), pp.378–379.

Holyoake, J. & Sansom, M.S.P., 2007. Conformational change in an MFS protein: MD simulations of LacY. *Structure*, 15(7), pp.873–884.

Hopf, T.A. et al., 2012. Theory Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell*, 149(7), pp.1607–1621.

Humphrey, W., Dalke, A. & Schulten, 1996. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics & Modelling*, 14, pp.33–38.

Iantorno, S. et al., 2014. Who watches the watchmen? An appraisal of benchmarks for multiple sequence alignment. *Methods in Molecular Biology*, 1079, pp.59–73.

Isberg, V. et al., 2015. Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends in Pharmacological Sciences*, 36(1), pp.22–31.

Izrailev, S. et al., 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal*, 72(4), pp.1568–1581.

Jardetzky, O., 1966. Simple Allosteric Model for Membrane Pumps. *Nature*, 211, pp.969–970.

Javitch, J.A. et al., 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(4), pp.998–1006.

Jiang, D. et al., 2013. Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. *Proceedings of the National Academy of Sciences*, 110(36), pp.14664–14669.

Jones, D.T., 1999a. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2), pp.195–202.

Jones, D.T., 1999b. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, 292(2), pp.195–202.

Jorgensen, W.L. et al., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), p.926.

Jorgensen, W.L., 1981. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *Journal of the American Chemical Society*, 103, pp.335–340.

Jorgensen, W.L., Maxwell, D.S. & Tirado-rives, J., 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 7863, pp.11225–11236.

Jorgensen, W.L. & Tirado-rives, J., 1988. The OPLS Pontential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *Journal of the American Chemical Society*, 110(6), pp.1657–1666.

Kabsch, W. & Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, pp.2577–2637.

Kaminski, G. et al., 2001a. Comparison with Accurate Quantum Chemical Calculations on Peptides. *Journal of Physical Chemistry B*, 105, pp.6474–6487.

Kaminski, G. et al., 2001b. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *Journal of Physical Chemistry B*, 105, pp.6474–6487.

Kaminski, R.M. et al., 2008. SV2A protein is a broad-spectrum anticonvulsant target: functional correlation between protein binding and seizure protection in models of both partial and generalized epilepsy. *Neuropharmacology*, 54(4), pp.715–720.

Kasho, V.N., Smirnova, I.N. & Kaback, H.R., 2006. Sequence alignment and homology threading reveals prokaryotic and eukaryotic proteins similar to lactose permease. *Journal of Membrane Biology*, 358(4), pp.1060–1070.

Kitchen, D.B. et al., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug discovery*, 3(11), pp.935–949.

Klitgaard, H. et al., 1998. Evidence for a unique profile of levetiracetam in rodent models of seizures and epilepsy. *Eur. J. Pharmacol.*, 353(2-3), pp.191–206.

Klitgaard, H., 2001. Levetiracetam: the preclinical profile of a new class of antiepileptic drugs? *Epilepsia*, 42 Suppl 4, pp.13–18.

Koehler Leman, J., Ulmschneider, M.B. & Gray, J.J., 2015. Computational modeling of membrane proteins. *Proteins: Structure, Function, and Bioinformatics*, 83(1), pp.1–24.

Konvicka, K. et al., 1998. A proposed structure for transmembrane segment 7 of G protein-coupled receptors incorporating an asn-Pro/Asp-Pro motif. *Biophysical Journal*, 75(2), pp.601–611.

Kottra, G., Stamfort, A. & Daniel, H., 2002. PEPT1 as a paradigm for membrane carriers that mediate electrogenic bidirectional transport of anionic, cationic, and neutral substrates. *Journal of Biological Chemistry*, 277(36), pp.32683–32691.

Krogh, a et al., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Membrane Biology*, 305(3), pp.567–580.

Kumar, H. et al., 2014. Structure of sugar-bound LacY. *Proceedings of the National Academy of Sciences*, 111(5), pp.1784–8.

Kumar, S. et al., 2013. Modulation of Bacterial Multidrug Resistance Efflux Pumps of the Major Facilitator Superfamily. *International Journal of Bacteriology*, p.204141.

Lee, J. et al., 2015. Exploring the Interaction of SV2A with Racetams Using Homology Modelling, Molecular Dynamics and Site-Directed Mutagenesis. *PloS one*, 10(2), p.e0116589.

Lemieux, M.J., Huang, Y. & Wang, D.N., 2005. Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from E. coli. *Journal of electron microscopy*, 54(Suppl 1), pp.43–46.

Li, J. et al., 2015. Computational characterization of structural dynamics underlying function in active membrane transporters. *Current Opinion in Structural Biology*, 31, pp.96–105.

Li, J. et al., 2013. Transient formation of water-conducting states in membrane transporters. *Proceedings of the National Academy of Sciences*, 110(19), pp.7696–7701.

Li, W. & Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), pp.1658–1659.

Liu, Y., Ke, M. & Gong, H., 2015. Protonation of Glu135 Facilitates the Outward-to-Inward Structural Transition of Fucose Transporter. *Biophysical Journal*, 109(3), pp.542–551.

Lomize, M. a. et al., 2006. OPM: Orientations of proteins in membranes database. *Bioinformatics*, 22(5), pp.623–625.

Löscher, W. et al., 2013. New avenues for anti-epileptic drug discovery and development. *Nature reviews. Drug discovery*, 12(10), pp.757–76.

Löscher, W., Hönack, D. & Rundfeldt, C., 1998. Antiepileptogenic effects of the novel anticonvulsant levetiracetam (ucb L059) in the kindling model of temporal lobe epilepsy. *The Journal of pharmacology and experimental therapeutics*, 284(2), pp.474–479.

Lynch, B.A. et al., 2004. The synaptic vesicle protein SV2A is the binding site for the antiepileptic drug levetiracetam. *Proceedings of the National Academy of Sciences*, 101(26), pp.9861–9866.

Lynch, B.A. et al., 2008. Visualization of SV2A conformations in situ by the use of Protein Tomography. *Biochemical and biophysical research communications*, 375, pp.391–395.

Lyons, J.A. et al., 2014. Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO reports*, 15(8), pp.886–893.

Madej, M.G. et al., 2013. Evolutionary mix-and-match with MFS transporters. *Proceedings of the National Academy of Sciences*, 110(15), pp.5870–5874.

Madej, M.G., 2014. Function, Structure and Evolution of the Major Facilitator Superfamily: The LacY Manifesto. *Advances in Biology*, p.e523592.

Madej, M.G. & Kaback, H.R., 2013. Evolutionary mix-and-match with MFS transporters II. *Proceedings of the National Academy of Sciences*, 110(50), pp.E4831–8.

Madeo, M., Kovacs, A.D. & Pearces, D.A., 2014. The human synaptic vesicle protein, SV2A functions as a galactose transporter in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, 289(48), pp.33066–71.

Marger, M.D. & Saier, M.H., 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends in biochemical sciences*, 18(1), pp.13–20.

Margineanu, D.G. & Klitgaard, H., 2009. Mechanisms of drug resistance in epilepsy: relevance for antiepileptic drug discovery. *Expert Opin Drug Discov*, 4(1), pp.23–32.

Marks, D.S. et al., 2011. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PloS one*, 6(12), p.e28766.

Marti-Remon, M.A. et al., 2000. Comparative protein structure modeling of genes and genomes. *Annual Review of Biomolecular Strucure*, (29), pp.291–325.

Masureel, M. et al., 2014. Protonation drives the conformational switch in the multidrug transporter LmrP. *Nature chemical biology*, 10(2), pp.149–55.

Matagne, a et al., 2008. Anti-convulsive and anti-epileptic properties of brivaracetam (ucb 34714), a high-affinity ligand for the synaptic vesicle protein, SV2A. *British journal of pharmacology*, 154(8), pp.1662–1671.

Matveeva, E.A. et al., 2007. Asymmetric accumulation of hippocampal 7S SNARE complexes occurs regardless of kindling paradigm. *Epilepsy Res.*, 73(3), pp.266–274.

Michino, M. et al., 2009. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nature reviews. Drug discovery*, 8(6), pp.455–463.

Moradi, M. & Tajkhorshid, E., 2013a. Driven Metadynamics: Reconstructing Equilibrium Free Energies from Driven Adaptive-Bias Simulations. *Journal of Physical Chemistry Letters*, 4, pp.1882–1887.

Moradi, M. & Tajkhorshid, E., 2013b. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proceedings of the National Academy of Sciences*, 110(47), pp.18916–21.

Moradi, M. & Tajkhorshid, E., per commun.

Müller, T., Spang, R. & Vingron, M., 2002. Estimating amino acid substitution models: a comparison of Dayhoff's estimator, the resolvent approach and a maximum likelihood method. *Molecular biology and evolution*, 19(1), pp.8–13.

Newstead, S. et al., 2011. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO journal*, 30, pp.1–10.

Newstead, S., 2011. Towards a structural understanding of drug and peptide transport within the proton-dependent oligopeptide transporter (POT) family. *Biochemical Society transactions*, 39(5), pp.1353–1358.

Nosé, S. & Klein, M.L., 1983. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5), pp.1055–1076.

Nowack, A. et al., 2010. SV2 regulates neurotransmitter release via multiple mechanisms. *Am. J. Physiol., Cell Physiol.*, 299(5), pp.C960–967.

Noyer, M. et al., 1995. The novel antiepileptic drug levetiracetam (ucb L059) appears to act via a specific binding site in CNS membranes. *European journal of pharmacology*, 286(2), pp.137–146.

Nugent, T. & Jones, D.T., 2012. Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. *Proceedings of the National Academy of Sciences*, 109(24), pp.E1540–7.

Otoul, C., De Smedt, H. & Stockis, A., 2007. Lack of pharmacokinetic interaction of levetiracetam on carbamazepine, valproic acid, topiramate, and lamotrigine in children with epilepsy. *Epilepsia*, 48(11), pp.2111–2115.

Pack, A.M., 2014. Brivaracetam, a Novel Antiepileptic Drug : Is it Effective and Safe? Results from One Phase III Randomized Trial. *Epilepsy Curr*, (4), pp.196–198.

Padan, E., Zilberstein, D. & Schuldiner, S., 1981. pH homeostasis in bacteria. *Biochimica et biophysica acta*, 650(2-3), pp.151–166.

Pao, S.S., Paulsen, I.A.N.T. & Saier, M.H., 1998. Major Facilitator Superfamily. *Microbiology*, 62(1), pp.1–34.

Pardo, L. et al., 1992. On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 89(9), pp.4009–4012.

Parker, J.L., Mindell, J.A. & Newstead, S., 2014. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLIFE*, (3), p.e04273.

Parrinello, M. & Rahman, a., 1981. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *Journal of Applied Physics*, 52(12), pp.7182–7190.

Pascual, J.M. et al., 2008. Structural Signatures and Membrane Helix 4 in GLUT1: Inferences from human blood-brain glucose transport mutants. *Journal of Biological Chemistry*, 283, pp.16732–16742.

Paulsen, I.T., Brown, M.H. & Skurray, R. a, 1996. Proton-dependent multidrug efflux systems. *Microbiological reviews*, 60(4), pp.575–608.

Pazdernik, N.J., Cain, S.M. & Brooker, R.J., 1997. An analysis of suppressor mutations suggests that the two halves of the lactose permease function in a symmetrical manner. *Journal of Biological Chemistry*, 272(42), pp.26110–26116.

Pedersen, B.P. et al., 2013. Crystal structure of a eukaryotic phosphate transporter. *Nature*, 496(7446), pp.533–536.

Pendse, P.Y., Brooks, B.R. & Klauda, J.B., 2010. Probing the Periplasmic-Open State of Lactose Permease in Response to Sugar Binding and Proton Translocation. *Journal of Molecular Biology*, 404(3), pp.506–521.

Phillips, A.J., 2006. Homology assessment and molecular sequence alignment. *Journal of Biomedical Informatics*, 39, pp.18–33.

Pronk, S. et al., 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), pp.845–54.

Quistgaard, E.M. et al., 2013. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature structural & molecular biology*, 20(6), pp.766–8.

Radestock, S. & Forrest, L.R., 2011. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of molecular biology*, 407(5), pp.698–715.

Reddy, V.S. et al., 2012. The major facilitator superfamily (MFS) revisited. *The FEBS journal*, 279(11), pp.2022–2035.

Ri, Y. et al., 1999. The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. *Biophysical journal*, 76(6), pp.2887–2898.

Rost, B., 1999. Twilight zone of protein sequence alignments. *Protein Engineering*, 12(2), pp.85–94.

Saier, M.H. et al., 2014. The transporter classification database. *Nucleic Acids Research*, 42, pp.251–258.

Saier, M.H. & Paulsen, I.T., 2001. Phylogeny of multidrug transporters. *Seminars in cell & developmental biology*, 12(3), pp.205–213.

Sankoff, D. & J., Cedergren, R., 1983. Time Warps, String Edits and Macromolecules. The Theory and Practice of Sequence Comparison. In Addison-Wesley, pp. 253–263.

Sargentini-Maier, M.L. et al., 2008. Pharmacokinetics and metabolism of 14C-brivaracetam, a novel SV2A ligand, in healthy subjects. *Drug Metab. Dispos.*, 36(1), pp.36–45.

Schlitter, J. et al., 1993. Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T ↔ R Transition in Insulin. *Molecular Simulation*, 10(2-6), pp.291–308.

Semah, F. et al., 1998. Is the underlying cause of epilepsy a major prognostic factor for recurrence? *Neurology*, 51, pp.1256–1262.

Shaheen, A. et al., 2015. Characterization of putative multidrug resistance transporters of the major facilitator-superfamily expressed in Salmonella Typhi. *Journal of Infection and Chemotherapy*, 21(5), pp.357–362.

Shayeghi, M. et al., 2005. Identification of an intestinal heme transporter. *Cell*, 122(5), pp.789–801.

Shen, M.-Y. & Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11), pp.2507–2524.

Shi, J. et al., 2011. Combining modelling and mutagenesis studies of synaptic vesicle protein 2A to identify a series of residues involved in racetam binding. *Biochemical Society transactions*, 39(5), pp.1341–1347.

Shi, J., Blundell, T.L. & Mizuguchi, K., 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of molecular biology*, 310(1), pp.243–257.

Shi, Y., 2013. Common folds and transport mechanisms of secondary active transporters. *Annual review of biophysics*, 42, pp.51–72.

Shokri, A. & Larsson, G., 2004. Characterisation of the Escherichia coli membrane structure and function during fedbatch cultivation. *Microbial cell factories*, 3, p.9.

Sills, G.J. et al., 1997. Neurochemical studies with the novel anticonvulsant levetiracetam in mouse brain. *European Journal of Pharmacology*, 325(1), pp.35–40.

Smirnova, I. et al., 2012. Role of protons in sugar binding to LacY. *Proceedings of the National Academy of Sciences*, 109(42), pp.16835–16840.

Smirnova, I., Kasho, V. & Kaback, H.R., 2011. Lactose permease and the alternating access mechanism. *Biochemistry*, 50(45), pp.9684–9693.

Smirnova, I.N., Kasho, V. & Kaback, H.R., 2008. Protonation and sugar binding to LacY. *Proceedings of the National Academy of Sciences*, 105(26), pp.8896–8901.

Søndergaard, C.R. et al., 2011. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *Journal of Chemical Theory and Computation*, 7(7), pp.2284–2295.

Stahl, M. et al., 2011. L-fucose utilization provides Campylobacter jejuni with a competitive advantage. *Proceedings of the National Academy of Sciences*, 108, pp.7194–7199.

Stelzl, L.S. et al., 2014. Flexible gates generate occluded intermediates in the transport cycle of LacY. *Journal of molecular biology*, 426(3), pp.735–51.

Suzek, B.E. et al., 2007. UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10), pp.1282–1288.

Tanner, V.A., Ploug, T. & Tao-Cheng, J.H., 1996. Subcellular localization of SV2 and other secretory vesicle components in PC12 cells by an efficient method of preembedding EM immunocytochemistry for cell cultures. *J. Histochem. Cytochem.*, 44(12), pp.1481–1488.

Tieleman, D.P. et al., 2006. Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. *Journal of physics. Condensed matter*, 18(28), pp.S1221–S1234.

Tramontano, A., 1998. Homology modeling with low sequence identity. *Methods*, 14(3), pp.293–300.

Trott, O. & Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput Chem.*, 31(2), pp.455–461.

Tusnády, G.E. & Simon, I., 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17(9), pp.849–850.

Del Val, C., White, S.H. & Bondar, A.N., 2012. Ser/Thr motifs in transmembrane proteins: Conservation patterns and effects on local protein structure and dynamics. *Journal of Membrane Biology*, 245(11), pp.717–730.

Vastermark, A. et al., 2015. Conserved movement of TMS11 of LacY and XylE Keywords: TMS11, LacY, XylE, MFS, contact maps, occluded conformational states. *Proteins: Struct. Funct. Bioinf.*, 4(83), pp.735–45.

Västermark, A. & Saier, M.H., 2014. Major Facilitator Superfamily (MFS) evolved without 3-transmembrane segment unit rearrangements. *Proceedings of the National Academy of Sciences*, 111(13), pp.E1162–3.

Venkatakrishnan, a J. et al., 2013. Molecular signatures of G-protein-coupled receptors. *Nature*, 494(7436), pp.185–94.

Van Vliet, E. a et al., 2009. Decreased expression of synaptic vesicle protein 2A, the binding site for levetiracetam, during epileptogenesis and chronic epilepsy. *Epilepsia*, 50(3), pp.422–433.

Wan, Q.-F. et al., 2010. SV2 acts via presynaptic calcium to regulate neurotransmitter release. *Neuron*, 66(6), pp.884–895.

Wang, L. & Jiang, T., 1994. On the complexity of multiple sequence alignment. *Journal of computational biology*, 1(4), pp.337–348.

Wisedchaisri, G. et al., 2014. Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nature communications*, 5, p.4521.

Wolf, M.G. et al., 2010. g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *Journal of computational chemistry*, 31, pp.2175–2174.

Xu, T. & Bajjalieh, S.M., 2001. SV2 modulates the size of the readily releasable pool of secretory vesicles. *Nature cell biology*, 3(8), pp.691–698.

Yaffe, D. et al., 2013. Identification of molecular hinge points mediating alternating access in the vesicular monoamine transporter VMAT2. *Proceedings of the National Academy of Sciences*, pp.E1332–E1341.

Yao, J. & Bajjalieh, S.M., 2009. SVOP is a nucleotide binding protein. *PloS one*, 4(4), p.e5315.

Zeidel, M.L. et al., 1992. Role of glucose carrier in human erythrocyte water permeability. *Biochemistry*, 31(2), pp.589–596.

Zeuthen, T., Zeuthen, E. & Macaulay, N., 2007. Water transport by GLUT2 expressed in Xenopus laevis oocytes. *The Journal of physiology*, 579(Pt 2), pp.345–361.

Zhou, W. et al., 1994. A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Molecular Pharmacology*, 45(2), pp.165–170.

9 Appendices

9.1 Script for sequence analysis: msa-analysis.py

```
#!/usr/bin/python
import sys
from get_uniref_id import get_uniref_id
from return_unique import return_parsed
import os
import glob
import numpy as np
import re
from operator import itemgetter
from itertools import tee, izip

#############Usage: python msa-analysis.py <uniprotID>###################


########################Get Alignment############################

#output put into this directory
os.system('mkdir output')

#print sys.argv[1]
#defined uniprot reference - user input
cl = get_uniref_id( sys.argv[1] )
#print("%s" % cl)

#edit the cluster fasta and then return unique organisms with alignment
without newlines into unique-query.aln
fa = return_parsed( cl )
#print("%s" % fa)

#align the unique organism files
os.system('muscle -in ./output/all.fa -out ./output/align.afa')


######################Analyse Alignment#########################

#get MSA to count number of alignments
seq = []
len_no = ""
f_aln = open('./output/align.afa', 'r')
f_aln_content = f_aln.read()
spl = re.compile('>seq\w+\n').split(f_aln_content)
for k in spl:
      j = k.split('\n')
      seq.append(''.join(j[0:-1]))
f_aln.close
len_no = len(seq)

#add MSA to file with spaces
aln_len = ""
f_alnnh = open('./output/unique-align-gap.afa', 'w+')
for k in seq:
      n = 1
```

```
      l = [k[i:i+n] for i in range(0, len(k), n)]
      aln_len=str(len(l))
      for m in l:
            f_alnnh.write("%s " % m)
      f_alnnh.write("\n")
f_alnnh.close

print("%s = len_no \t %s = aln_len\n\Next Step\nopen R script\nR\nlen <-
%s\naln_len <- %s" % (len_no, aln_len, aln_len, len_no))
```

And the 2 dependency scripts for the above script:

get_uniref_id.py

```
import urllib,urllib2

def get_uniref_id( n ):
      url = 'http://www.uniprot.org/mapping/'
      cluster = ''
      url2 = 'http://www.uniprot.org/uniprot/?query=cluster:(' + n +
')&format=fasta'
      #print("%s" % url2)
      response = urllib.urlopen(url2)
      url2_data = response.read()
      #print("%s" % url2_data)

      file_n = './output/' + n
      #print("%s" % file_n)
      with open(file_n, 'w') as f:
            f.write(url2_data)
      f.close

      with open(file_n, 'r') as f:
            for line in f:
                  cluster+=str(line)
      f.close

      return cluster
```

return_unique.py

```
import re

def return_unique( n ):
      lines = n.split("\n>")
      single = []
      head = []

      #get unique organisms in header
      for i in lines:
            spl = i.split(' ')
            #print("%s" % spl)
            for j in spl:
                  if "OS=" in j and (j in head) == False:
```

```
                        single.append(i)
                        head.append(j)

        #get unique fasta sequences
        fw = open('./output/unique.fa', 'w')
        num = "1"
        for j in single:
              name = '>seq' + str(num)
              k = ''
              spl = j.split('\n')
              del spl[0]
              #print("%s" % spl)
              l = "\n".join(spl)
              #print("%s" % l)
              fw.write('%s\n%s\n' % (name, l))
              num = int(num) + 1
        fw.close

        return single

def return_parsed( n ):
        lines = n.split("\n>")
        all_seq = []

        #remove header
        for i in lines:
              spl = i.split('\n')
              spl.pop(0)
              #print("%s" % spl)
              spl_jn = "".join(spl)
              all_seq.append(spl_jn)

        #get all fasta sequences
        fw = open('./output/all.fa', 'w')
        num = "1"
        for j in all_seq:
              name = '>seq' + str(num)
              fw.write('%s\n%s\n' % (name, j))
              num = int(num) + 1
        fw.close
```

9.2 Heat maps of conserved residues for YajR, FucP, EmrD and GlpT

Figure 9.1 The conservation of residues in YajR. Each heat map shows the percentage conservation of each residue type (hydrophobic, small hydrophobic, aromatic, polar, positive, negative, glycine and proline) is given as 0 – 100 % with white – dark colour.

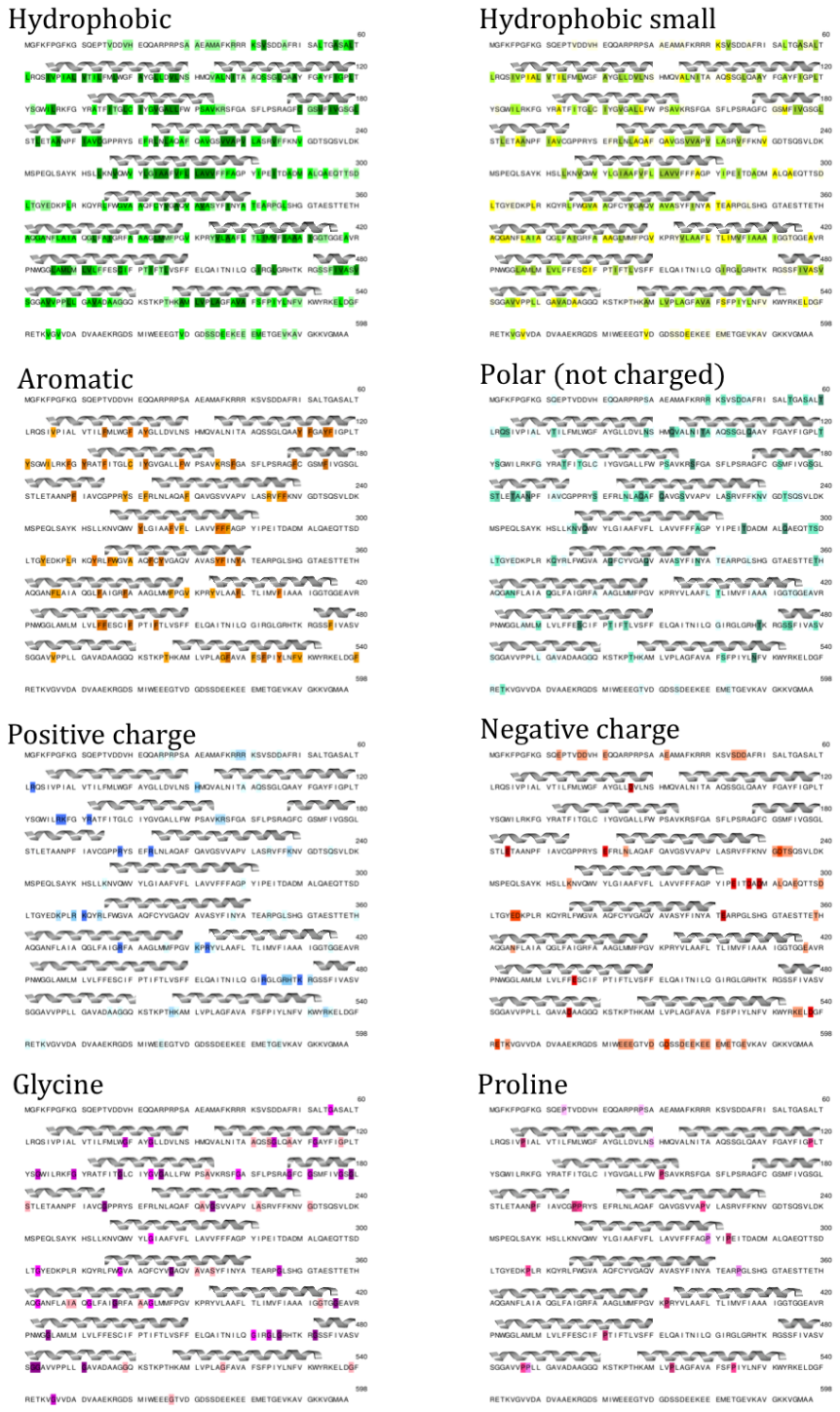


Figure 9.2 The conservation of residues in FucP. Each heat map shows the percentage conservation of each residue type (hydrophobic, small hydrophobic, aromatic, polar, positive, negative, glycine and proline) is given as 0 – 100 % with white – dark colour.

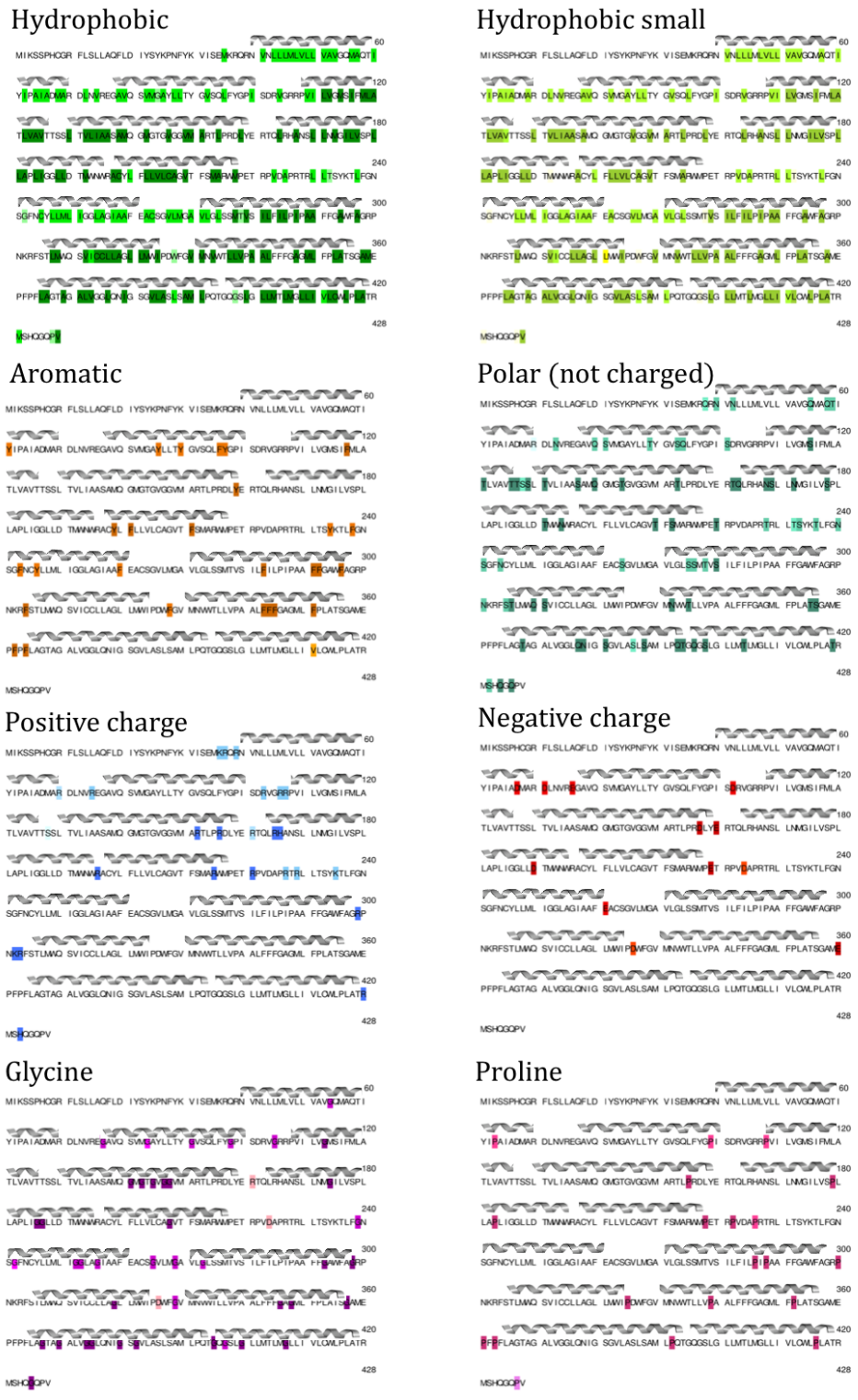


Figure 9.3 The conservation of residues in EmrD. Each heat map shows the percentage conservation of each residue type (hydrophobic, small hydrophobic, aromatic, polar, positive, negative, glycine and proline) is given as 0 – 100 % with white – dark colour.

Figure 9.4 The conservation of residues in GlpT. Each heat map shows the percentage conservation of each residue type (hydrophobic, small hydrophobic, aromatic, polar, positive, negative, glycine and proline) is given as 0 – 100 % with white – dark colour.

9.3 Heat maps for conserved contacts

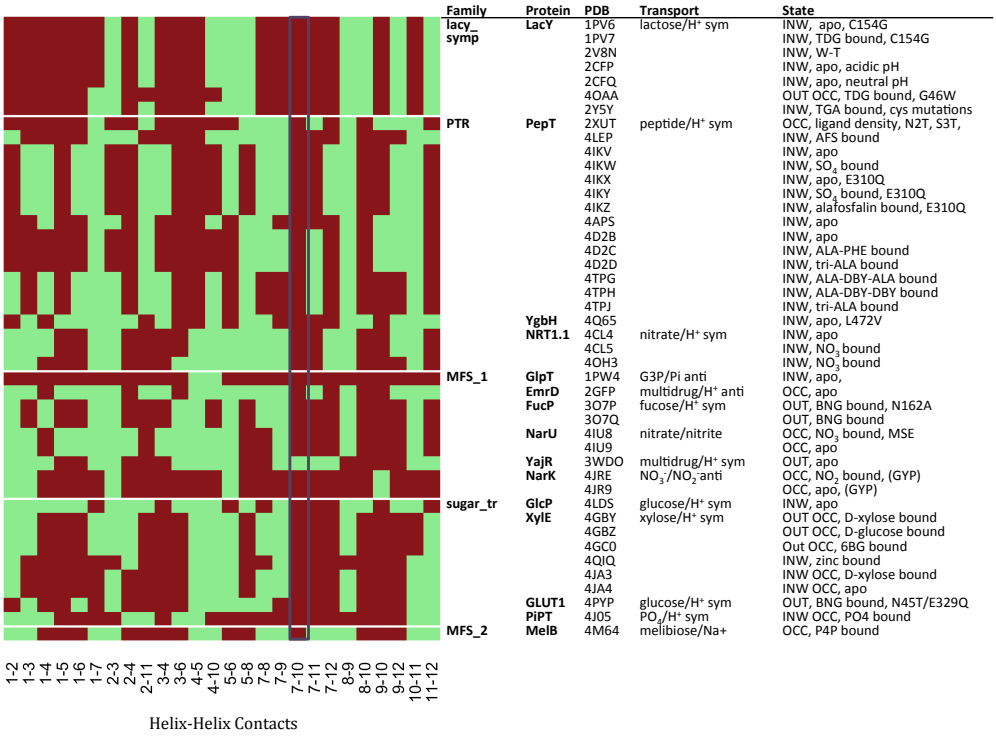
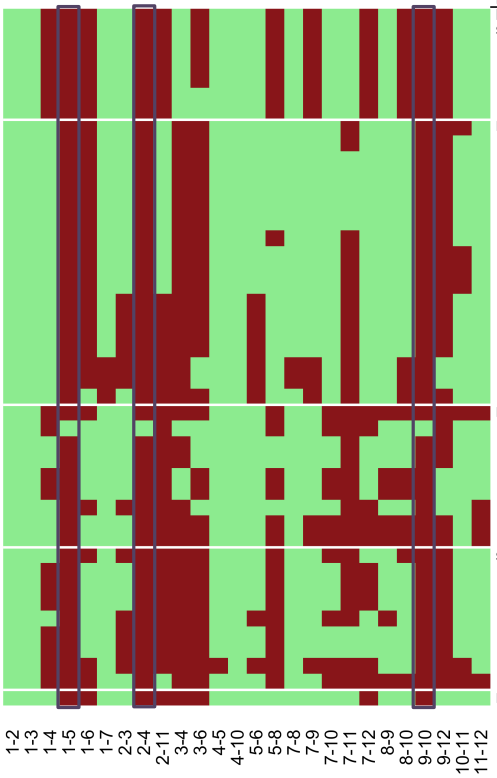


Figure 9.5 The contact map for aromatic residues (W, F, Y) in MFS X-ray crystal structures. The black boxes show the most conserved contacts. This is an example of the heat maps used to determine the conserved contacts in MFS proteins.



Family	Protein	PDB	Transport	State
lacy_ symp	LacY	1PV6	lactose/H^+ sym	INW, apo, C154G
		1PV7		INW, TDG bound, C154G
		2V8N		INW, W-T
		2CFP		INW, apo, acidic pH
		2CFQ		INW, apo, neutral pH
		4OAA		OUT OCC, TDG bound, G46W
		2Y5Y		INW, TGA bound, cys mutations
PTR	PepT	2XUT	peptide/H^+ sym	OCC, ligand density, N2T, S3T,
		4LEP		INW, AFS bound
		4IKV		INW, apo
		4IKW		INW, SO_4 bound
		4IKX		INW, apo, E310Q
		4IKY		INW, SO_4 bound, E310Q
		4IKZ		INW, alafosfalin bound, E310Q
		4APS		INW, apo
		4D2B		INW, apo
		4D2C		INW, ALA-PHE bound
		4D2D		INW, tri-ALA bound
		4TPG		INW, ALA-DBY-ALA bound
		4TPH		INW, ALA-DBY-DBY bound
		4TPJ		INW, tri-ALA bound
	YgbH	4Q65		INW, apo, L472V
	NRT1.1	4CL4	nitrate/H^+ sym	INW, apo
		4CL5		INW, NO_3 bound
		4OH3		INW, NO_3 bound
MFS_1	GlpT	1PW4	G3P/Pi anti	INW, apo,
	EmrD	2GFP	multidrug/H^+ anti	OCC, apo
	FucP	3O7P	fucose/H^+ sym	OUT, BNG bound, N162A
		3O7Q		OUT, BNG bound
	NarU	4IU8	nitrate/nitrite	OCC, NO_3 bound, MSE
		4IU9		OCC, apo
	YajR	3WDO	multidrug/H^+ sym	OUT, apo
	NarK	4JRE	NO_3^-/NO_2^- anti	OCC, NO_2 bound, (GYP)
		4JR9		OCC, apo, (GYP)
sugar_tr	GlcP	4LDS	glucose/H^+ sym	INW, apo
	XylE	4GBY	xylose/H^+ sym	OUT OCC, D-xylose bound
		4GBZ		OUT OCC, D-glucose bound
		4GC0		Out OCC, 6BG bound
		4QIQ		INW, zinc bound
		4JA3		INW OCC, D-xylose bound
		4JA4		INW OCC, apo
	GLUT1	4PYP	glucose/H^+ sym	OUT, BNG bound, N45T/E329Q
	PiPT	4J05	PO_4/H^+ sym	INW OCC, PO4 bound
MFS_2	MelB	4M64	melibiose/Na+	OCC, P4P bound

Figure 9.6 The contact map for glycine residues in MFS X-ray crystal structures. The black boxes show the most conserved contacts. This is an example of the heat maps used to determine the conserved contacts in MFS proteins.

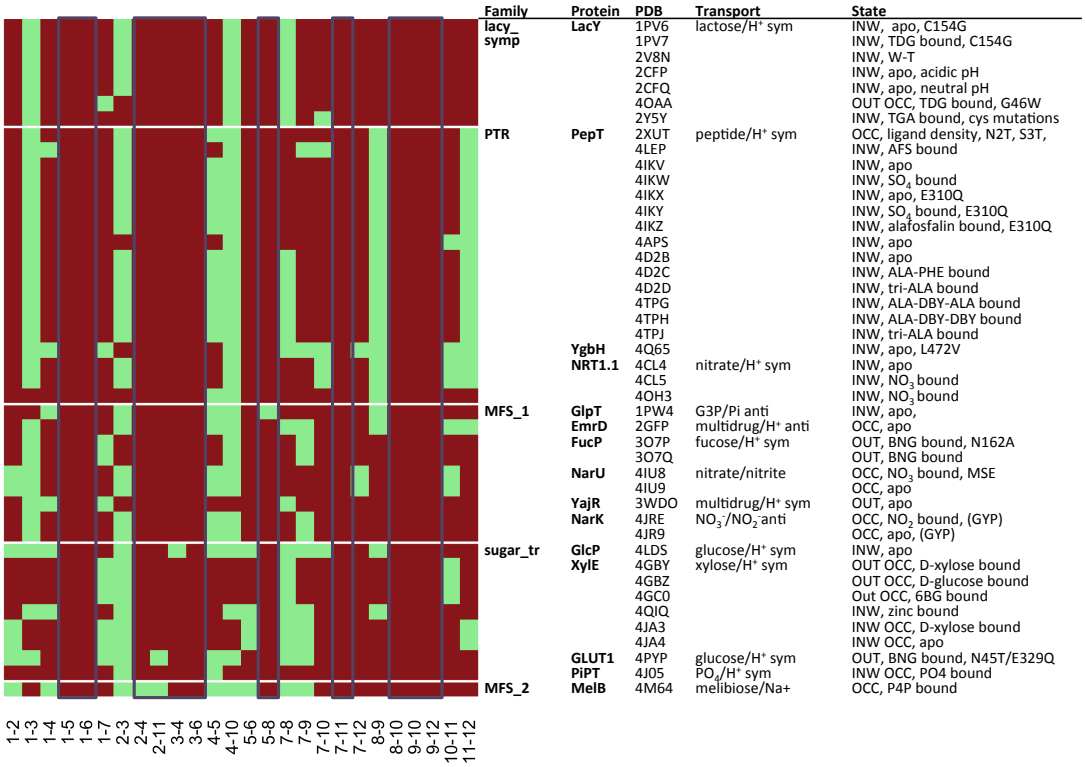


Figure 9.7 The contact map for small hydrophobic residues (A, V, I, L) in MFS X-ray crystal structures. The black boxes show the most conserved contacts. This is an example of the heat maps used to determine the conserved contacts in MFS proteins.

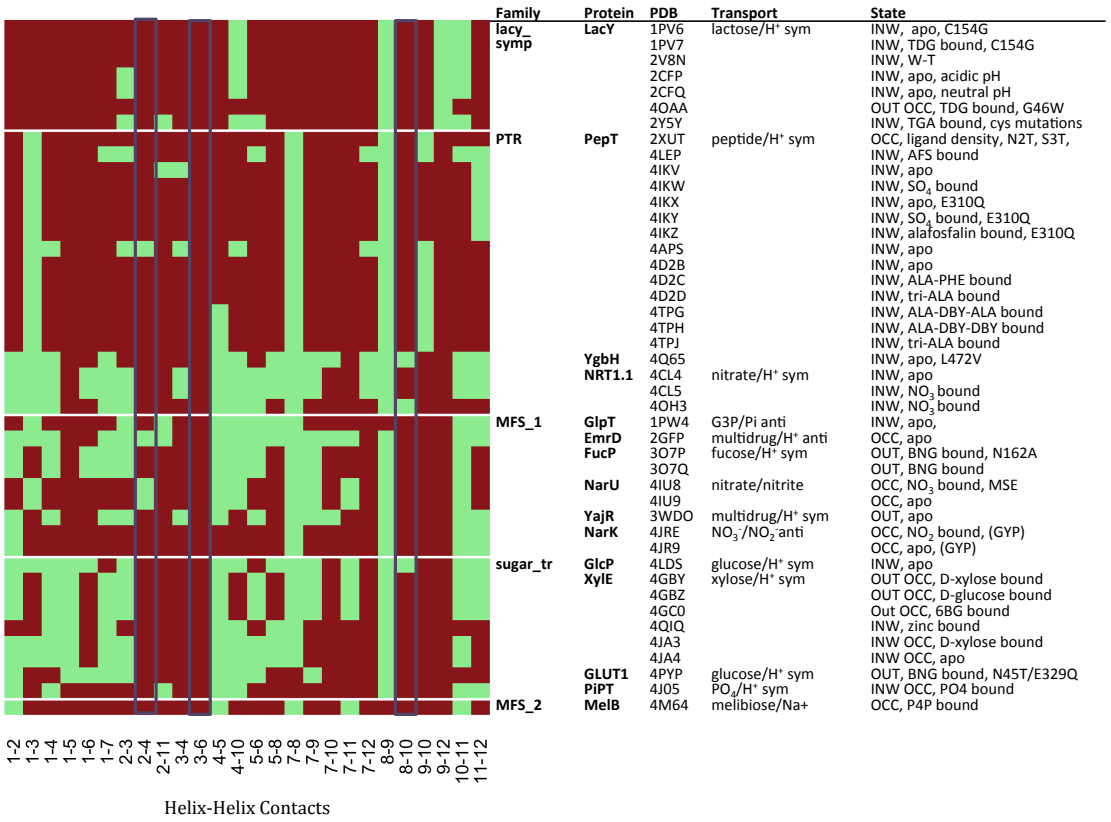


Figure 9.8 The contact map for large hydrophobic residues (C, M, W, Y, F) in MFS X-ray crystal structures. The black boxes show the most conserved contacts. This is an example of the heat maps used to determine the conserved contacts in MFS proteins.

9.4 Numbering in XylE, PepT and LacY

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
1	-	1-5 Gly	CA - CA: 4.9	G25 and G174	G25
		1-2 STNQ			
		1-5 small			
		1-6 small			
2	-	2-4 Gly	CA-CA: 4.6	G71 and G141	G71
		2-11 Gly			
		2-4 small			
		2-4 large			

Table 9.1 The residues on each helix of XylE that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
3	-	3-4 small			
		3-6 small	CA - CA: 5.6	A92 and A210	A92
		3-6 large			
4	Gly	2-4 Gly	CA-CA: 4.6	G71 and G141	G141
		2-4 small			
		3-4 small			
		2-4 large			
5	Gly	1-5 Gly	CA - CA: 4.9	G25 and G174	G174
		5-8 STNQ			
		1-5 small			
		5-8 small			
6	-	1-6 small			
		3-6 small	CA - CA: 5.6	A92 and A210	A210
		3-6 large			
7	aromatic	7-11 STNQ			Y298
		7-10 FY			
		7-11 small			
8	-	5-8 STNQ			
		5-8 small	CD1 – O: 5.1	Q168 – L326	L326
		8-10 small			
		8-10 large			
9	-	9-10 Gly	CA - CA: 5.8	G348 - A382	G348
		9-10 small			
		9-12 small			
10	Gly	9-10 Gly	-	-	G388
		7-10 FY			
		8-10 small			
		9-10 small			
		8-10 large			
11	Gly	2-11 Gly			
		7-11 STNQ	OG - CA: 3.8	S285 - Q415	Q415
		7-11 small			
12	-	9-12 small	CA - CA: 4.7	I346 - A456	A456

Figure 9.1 Continued. The residues on each helix of XylE that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
1	-	1-5 Gly	C - CA: 6.9	G160 - A/G22	A22
		1-2 STNQ			
		1-5 small			
		1-6 small			
2	-	2-4 Gly	CA - CA: 6.7	G124 - G66	G66
		2-11 Gly			
		2-4 small			
		2-4 large			
3	-	3-4 small			
		3-6 small			
		3-6 large	CA - CA: 8.3	L93 - F189	L93
4	Gly	2-4 Gly	CA - CA: 6.7	G124 - G66	G124
		2-4 small			
		3-4 small			
		2-4 large			
5	Gly	1-5 Gly	C - CA: 6.9	G160 - A/G22	G160
		5-8 STNQ			
		1-5 small			
		5-8 small			
6	-	1-6 small			
		3-6 small			
		3-6 large	CA - CA: 8.3	L93 - F189	F189
7	aromatic	7-11 STNQ			F315
		7-10 FY			
		7-11 small			
8	-	5-8 STNQ	CA - O: 4.5	S165 - A342	A342
		5-8 small			
		8-10 small			
		8-10 large			
9	-	9-10 Gly	CA - CA: 4.9	G379 - G418	G379
		9-10 small			
		9-12 small			

Table 9.2 The residues on each helix of PepT that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
10	Gly	9-10 Gly			G418
		7-10 FY			
		8-10 small			
		9-10 small			
		8-10 large			
11	Gly	2-11 Gly			
		7-11 STNQ	CA - O: 4.2	T309 - S449	S449
		7-11 small			
12	-	9-12 small	CA - CA: 5.2	I388 - A489	A489

Table 9.2 Continued. The residues on each helix of PepT that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
1	-	1-5 Gly	CA - CA: 6.0	G147 - G13	G13
		1-2 STNQ			
		1-5 small			
		1-6 small			
2	-	2-4 Gly	-	-	-
		2-4 small	-	-	-
		2-4 large	CA – CG: 8.0	G115 - F55	F55
		2-11 Gly			
3	-	3-4 small			
		3-6 small	C - C: 6.7	L84 - A177	L84
		3-6 large			
4	Gly	2-4 Gly	-	-	-
		2-4 small	-	-	-
		3-4 small			
		2-4 large	CA - CG: 8.0	G115 - F55	G115
5	Gly	1-5 Gly	CA - CA: 6.0	G147 - G13	G147
		5-8 STNQ			
		1-5 small			
		5-8 small			

Table 9.3 The residues on each helix of LacY that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

Helix	Conserved Residue	Conserved Contact	Contact Distance (Å)	Residues involved in Contact	Residue to Number from
6	-	1-6 small			
		3-6 small	C - C: 6.7	L84 - A177	A177
		3-6 large			
7	aromatic	7-11 STNQ			Y236
		7-10 FY			
		7-11 small			
8	-	5-8 STNQ	CA - CA: 6.2	S156 - T266	T266
		5-8 small			
		8-10 small			
		8-10 large			
9	-	9-10 Gly	CA - CA: 7.9	G296 - L329	G296
		9-10 small			
		9-12 small			
10	Gly	9-10 Gly			G332
		7-10 FY			
		8-10 small			
		9-10 small			
		8-10 large			
11	Gly	2-11 Gly			
		7-11 STNQ	SG - CA: 7.0	S234 - Q359	Q359
		7-11 small			
12	-	9-12 small	CA - CD1: 5.6	I304 - L390	L390

Table 9.3 Continued. The residues on each helix of LacY that are numbered x.0, where x is the helix index. The table lists the conserved contacts between each helix and then notes the contact used in the numbering, with the distance between atoms of the residues in the contact.

9.5 Position of numbered sites in XylE, PepT and LacY

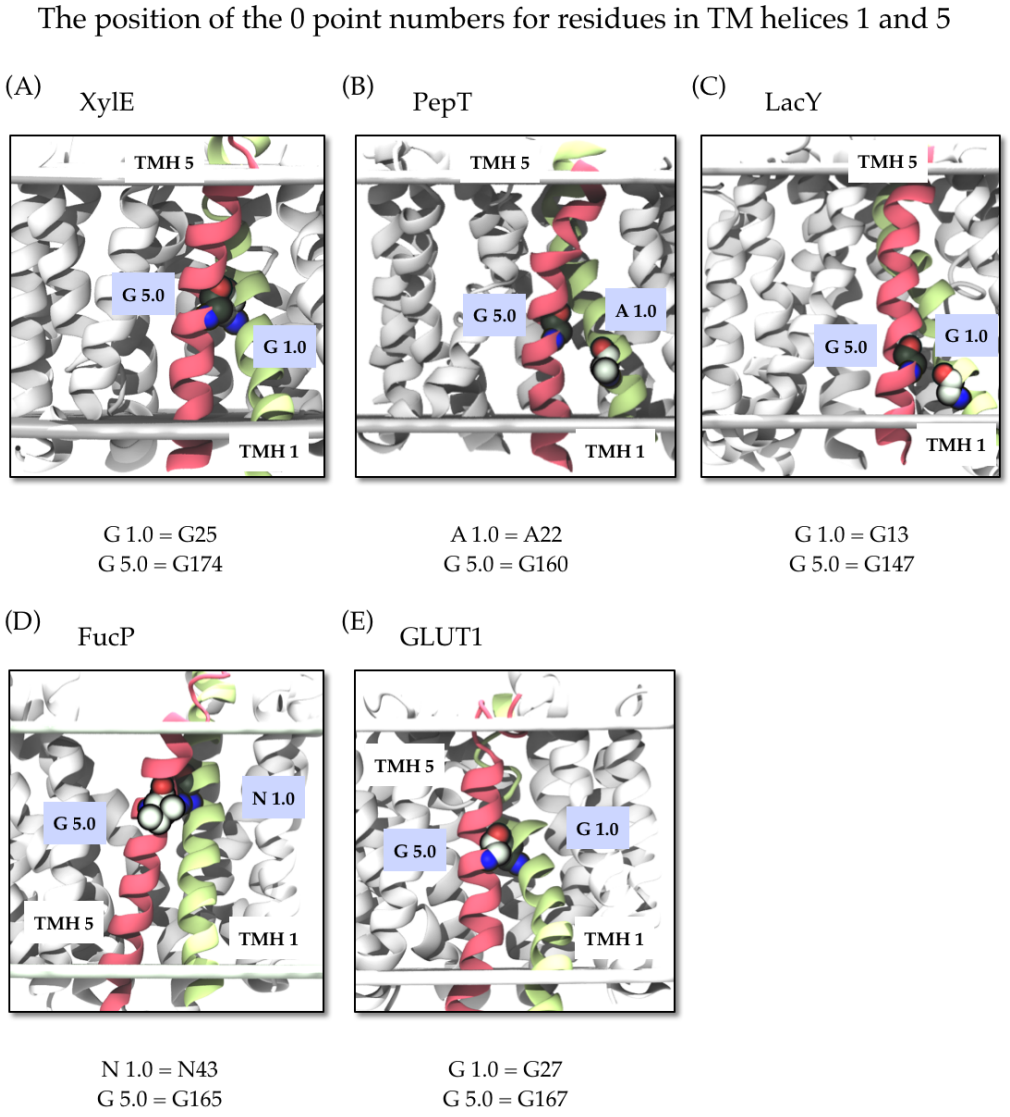


Figure 9.9 The position of numbered site 0 in TMHs 1 and 5.

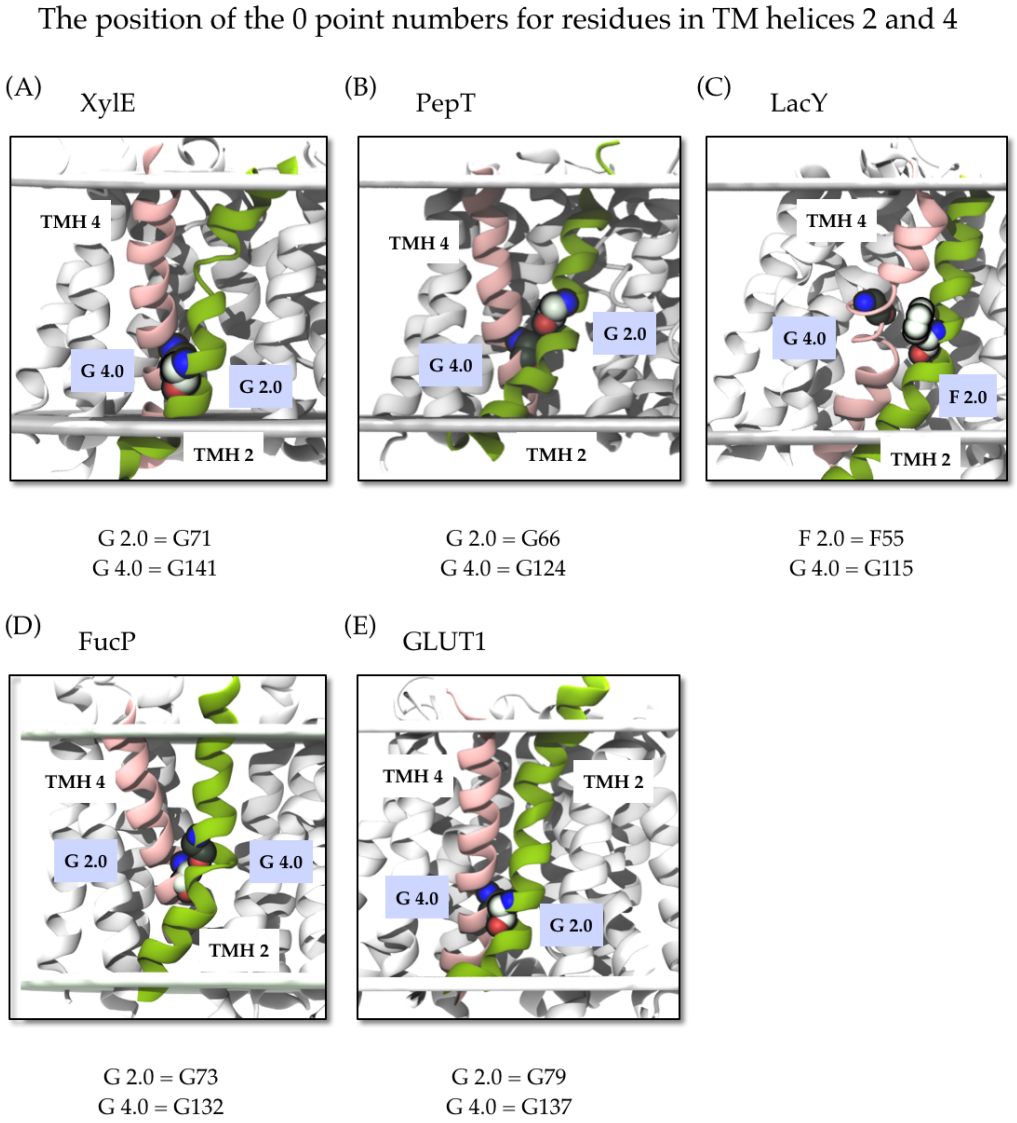


Figure 9.10 The position of numbered site 0 in TMHs 2 and 4.

The position of the 0 point numbers for residues in TM helices 3 and 6

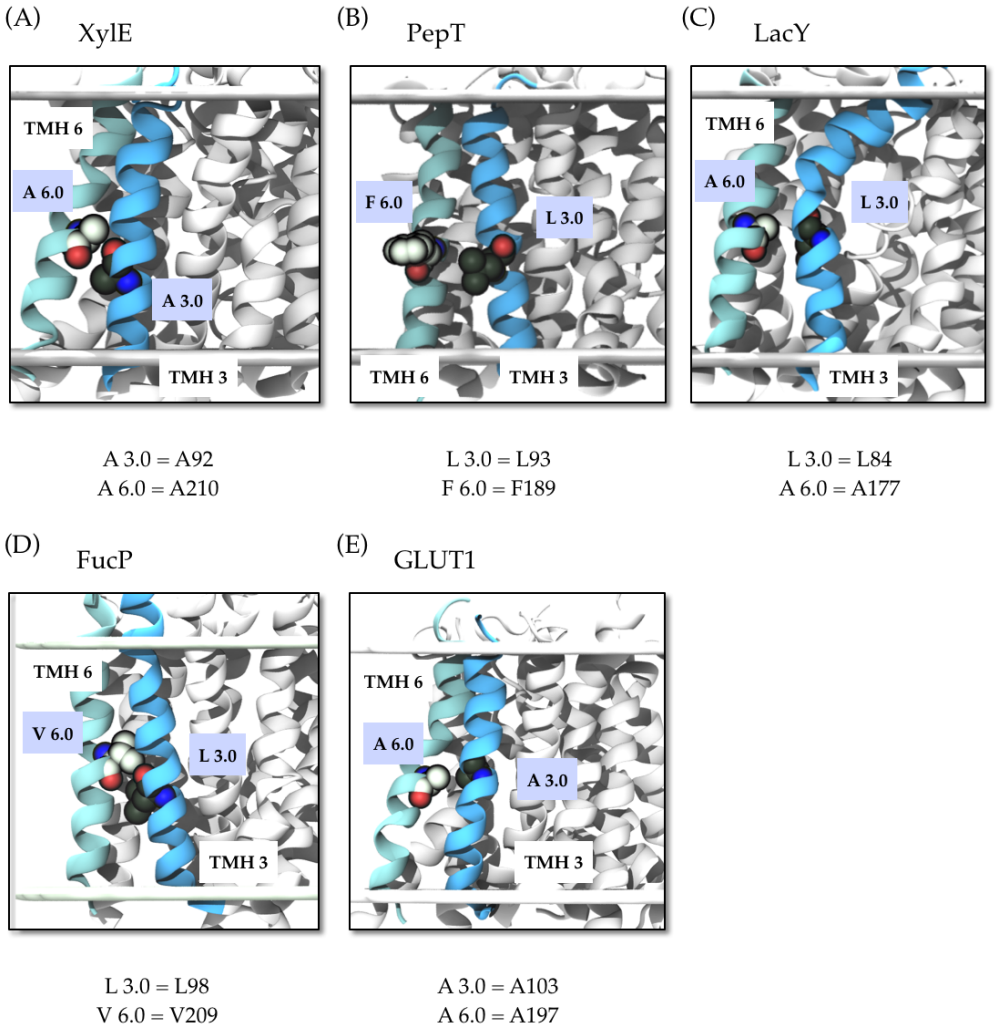


Figure 9.11 The position of numbered site 0 in TMHs 3 and 6.

The position of the 0 point numbers for residues in TM helices 7 and 8

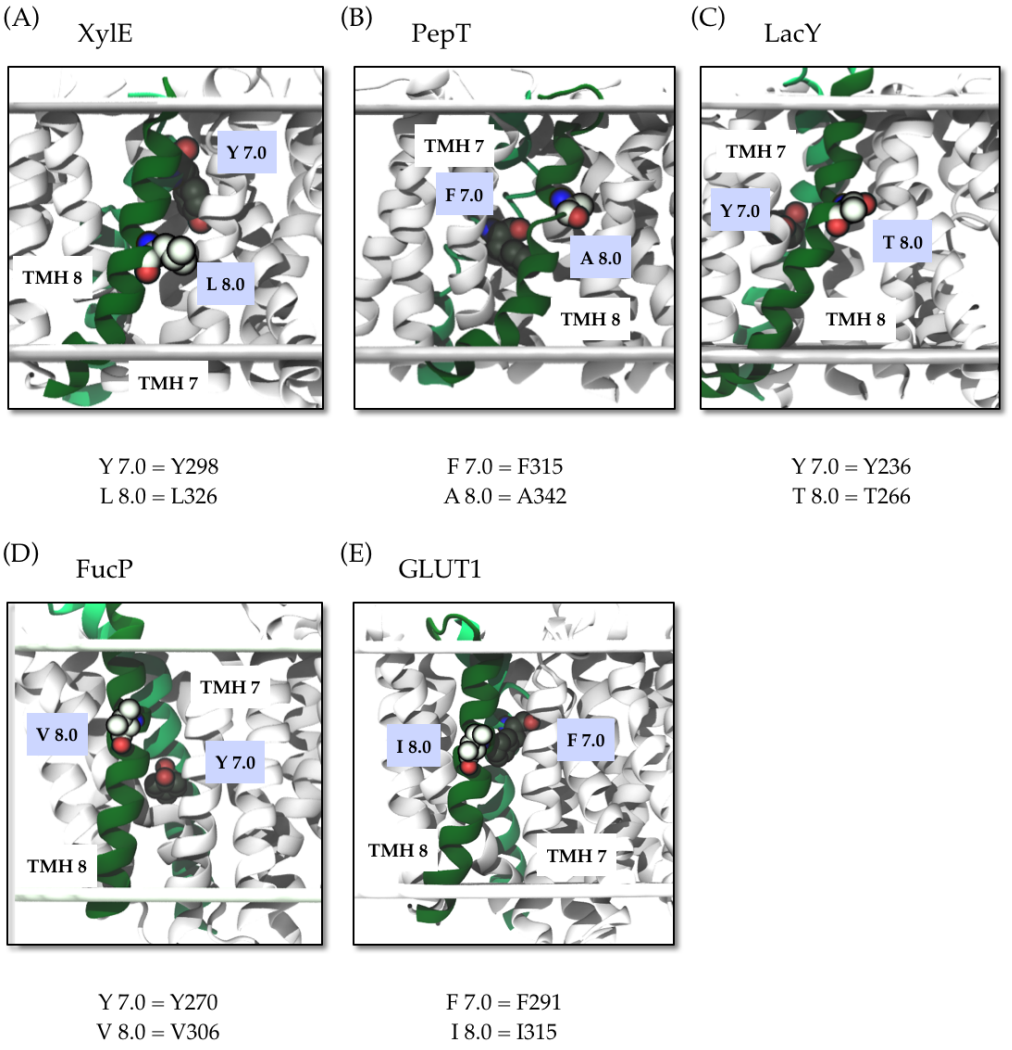


Figure 9.12 The position of numbered site 0 in TMHs 7 and 8.

The position of the 0 point numbers for residues in TM helices 9 and 10

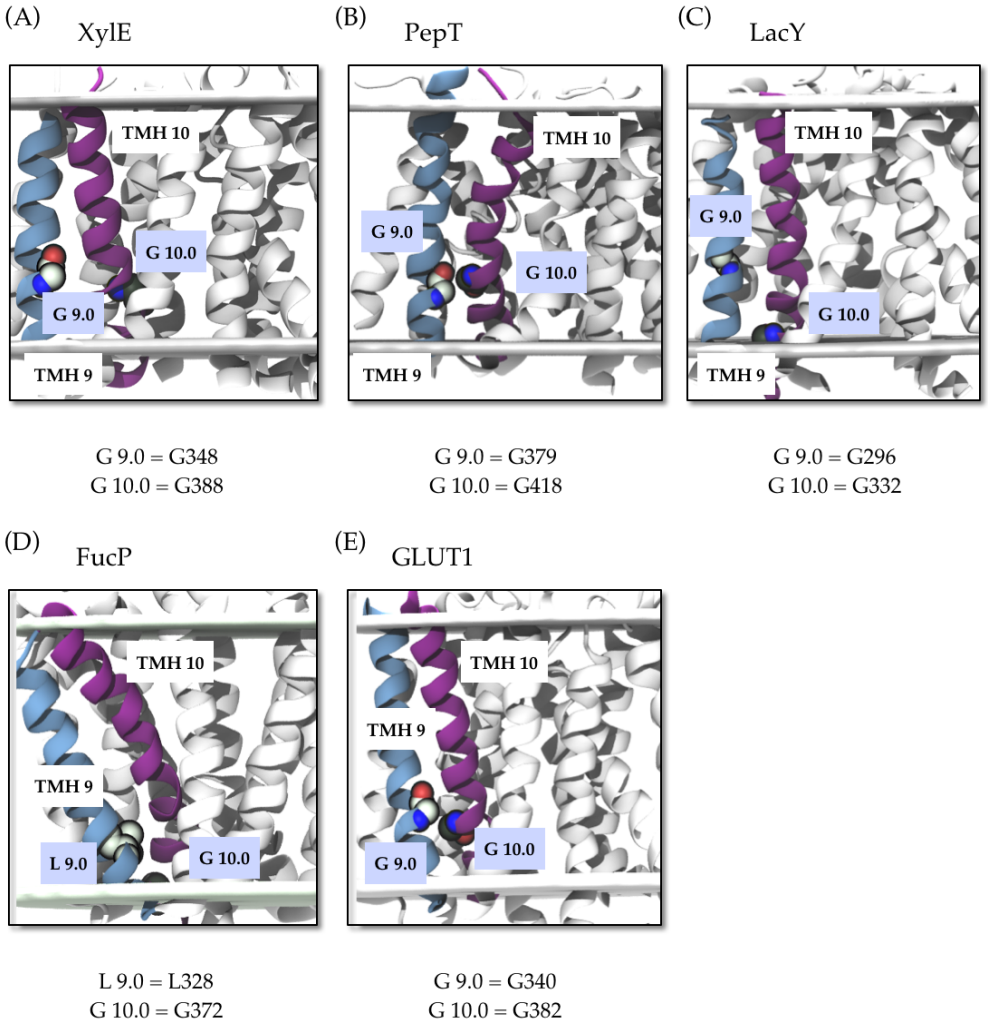


Figure 9.13 The position of numbered site 0 in TMHs 9 and 10.

The position of the 0 point numbers for residues in TM helices 11 and 12

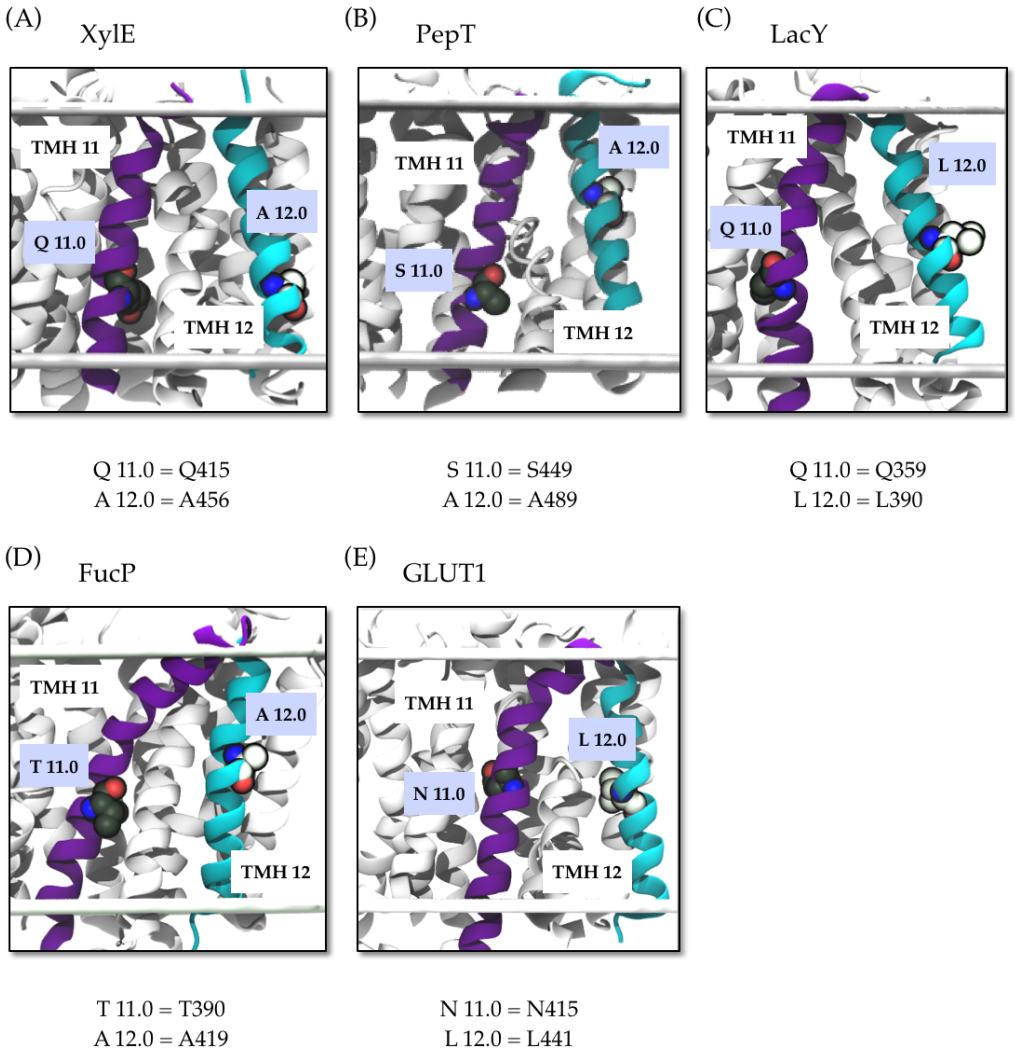


Figure 9.14 The position of numbered site 0 in TMHs 11 and 12.

9.6 Exploring the interaction of SV2A with racetams using homology modelling, molecular dynamics and site directed mutagenesis

PLOS | ONE

RESEARCH ARTICLE

Exploring the Interaction of SV2A with Racetams Using Homology Modelling, Molecular Dynamics and Site-Directed Mutagenesis

Joanna Lee[1], Veronique Daniels[2], Zara A. Sands[2], Florence Lebon[2], Jiye Shi[2], Philip C. Biggin[1]*

1 Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom, **2** UCB Pharma S.A., Chemin du Foriest, B-1420 Braine-l'Alleud, Belgium

* philip.biggin@bioch.ox.ac.uk

OPEN ACCESS

Citation: Lee J, Daniels V, Sands ZA, Lebon F, Shi J, Biggin PC (2015) Exploring the Interaction of SV2A with Racetams Using Homology Modelling, Molecular Dynamics and Site-Directed Mutagenesis. PLoS ONE 10(2): e0116589. doi:10.1371/journal.pone.0116589

Academic Editor: Chandra Verma, Bioinformatics Institute, SINGAPORE

Received: August 13, 2014

Accepted: December 11, 2014

Published: February 18, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Joanna Lee is a BBSRC-funded student in receipt of additional financial support from UCB BioPharma SPRL. Zara Sands, Florence Lebon and Jiye Shi are all employees of UCB BioPharma SPRL. During the course of the research reported herein, Veronique Daniels was an employee of UCB BioPharma SPRL and her research was supported by a grant from the Walloon Region (Belgium) – DGO6 (Convention 6536). There are no patents or products in development or marketed products to

Abstract

The putative Major Facilitator Superfamily (MFS) transporter, SV2A, is the target for levetiracetam (LEV), which is a successful anti-epileptic drug. Furthermore, SV2A knock out mice display a severe seizure phenotype and die after a few weeks. Despite this, the mode of action of LEV is not known at the molecular level. It would be extremely desirable to understand this more fully in order to aid the design of improved anti-epileptic compounds. Since there is no structure for SV2A, homology modelling can provide insight into the ligand-binding site. However, it is not a trivial process to build such models, since SV2A has low sequence identity to those MFS transporters whose structures are known. A further level of complexity is added by the fact that it is not known which conformational state of the receptor LEV binds to, as multiple conformational states have been inferred by tomography and ligand binding assays or indeed, if binding is exclusive to a single state. Here, we explore models of both the inward and outward facing conformational states of SV2A (according to the alternating access mechanism for MFS transporters). We use a sequence conservation analysis to help guide the homology modelling process and generate the models, which we assess further with Molecular Dynamics (MD). By comparing the MD results in conjunction with docking and simulation of a LEV-analogue used in radioligand binding assays, we were able to suggest further residues that line the binding pocket. These were confirmed experimentally. In particular, mutation of D670 leads to a complete loss of binding. The results shed light on the way LEV analogues may interact with SV2A and may help with the ongoing design of improved anti-epileptic compounds.

declare that arise from this work. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Competing Interests: Joanna Lee is a BBSRC-funded student in receipt of additional financial support from UCB BioPharma SPRL. Zara Sands, Florence Lebon and Jiye Shi are all employees of UCB BioPharma SPRL. During the course of the research reported herein, Veronique Daniels was an employee of UCB BioPharma SPRL and her research was supported by a grant from the Walloon Region (Belgium) – DGO6 (Convention 6536). There are no patents or products in development or marketed products to declare that arise from this work. This does not alter the authors' adherence to PLoS ONE policies on sharing data and materials.

Introduction

Epilepsy remains one of the most debilitating neurological disorders and is characterized by recurrent spontaneous seizures [1]. Furthermore, about 30% of patients with epilepsy cannot be adequately treated due to poor efficacy or undesirable side-effects [2]. One of the more successful treatment strategies has revolved around the administration of a compound from a class of compounds collectively known as racetams, which are distinguished by a central pyrrolidine ring system. The only example currently on the market is levetiracetam (LEV; (*S*)-α-ethyl-2-oxo-pyrrolidine acetamide; KEPPRA) which is a second-generation anti-epileptic drug [3]. Studies in rats in the 1990s led to the identification of the target for the action of LEV being the synaptic vesicle protein, SV2A [4]. It has since been confirmed as the target for related racetam compounds in both rat and human brains [5]. Although the mode of action of LEV is not known at the molecular level, it appears to have a different activity profile compared to many anti-epileptic drugs and thus may act via a different pathway, which in turn may account for fewer adverse side-effects when compared to many existing anti-epileptic compounds [6]. Thus there is considerable interest in understanding its interaction with SV2A.

SV2A is an integral membrane protein found in both synaptic and dense-core vesicles [7]. It is one of three SV2 isoforms that are distributed widely across the brain and has been found to be necessary for Ca^{2+}-dependent vesicular neurotransmitter release [8,9]. It has been implicated in the formation of the SNARE complex [9], with a proposed binding site for synaptotagmin in the amino terminal domain of SV2A. There is also a putative ATP binding site [10], though SV2A has not been proven to affect ATP uptake in synaptic vesicles [11]. Despite intense effort, the actual physiological function of SV2A remains unclear, though recently it has been shown to be capable of transporting galactose at least when expressed in hexose transport-deficient EBY.VW4000 yeast cells [12].

Knock out mice die within 3 weeks of birth, suffering seizures within 7 days [13], which correlates with the first expression of SV2A [14]. Furthermore, in knock-out models, neurons exhibited sustained increase in Ca^{2+} dependent synaptic transmission when two or more action potentials were triggered in quick succession [15]. Although Wan *et al.* [16] have shown that retinal neurons from SV2B knockout mice exhibit apparent changes in cytoplasmic calcium at the presynaptic terminal, the precise role or relevance of SV2 proteins in calcium homeostasis remains rather unclear at the present time. Kindling experiments, where repeated stimulation induces seizures that are proposed to mimic partial onset epilepsy, have indicated an up-regulation of SV2A in rat models [17]. However, other experiments have shown that SV2A expression decreases in chronically epileptic animals [18] and in patients with temporal lobe epilepsy [19]. This experimental data points towards a key role in synaptic maturation, though the mechanism for that remains unclear.

By sequence homology, SV2A has been described as a member of the Major Facilitator Superfamily (MFS) of transporters [20–22]. Structural evidence for this came from protein tomography experiments where two major conformations could be inferred [23]. Furthermore, LEV binding did not cause large-scale conformation changes or appear to lock the protein in a specific conformational state. Ligand-binding assays have also suggested that the conformational state of SV2A can be modulated in an allosteric fashion [24]. Taken together, these data demonstrate that the SV2A protein is likely to be highly dynamic and adopt many conformational states.

Although there is no structure for SV2A, it has been postulated, through remote sequence relationships, to resemble the architecture found in the MFS clan of transporters. However, the sequence identity to known examples of this family is very low (16% and 15% for GlpT and FucP respectively). If indeed it is a true member of the MFS family then one might expect it to

be involved in the uptake of a key metabolite, which would be transported via the alternating access mechanism [25]. Work by Shi et al. [26] determined 13 residues important for binding the racetam, ucb 30889, which were chosen according to their alignment to functionally relevant residues in LacY. Here we extend this work further by considering two alternative conformational states based on two additional templates: FucP as a template for the outward-open state and GlpT as an additional (to LacY) model of the inward-open state. In this terminology, inward refers to the cytosol, thus as SV2A is a synaptic vesicle membrane protein, an outward-facing state means the binding site would be exposed to the inside of the vesicle. On the basis of *in-silico* modelling and molecular dynamics simulations, we were able to suggest additional residues that line the binding pocket for ucb 30889. These predictions were subsequently confirmed by site-directed mutagenesis in conjunction with binding affinity assays. Taken in conjunction with previous results, the outward-open models (i.e. with the putative binding site exposed to the interior of the vesicle) tend to give better agreement with site-directed mutagenesis experiments.

Materials and Methods

Sequence Analysis and Model Building

Transmembrane (TM) predictions for the 12 potential TM helices of SV2A were made with SOSUI [27], HMMTop [28], JPRED [29] and PSIPRED [30]. A final consensus TM prediction was made manually (see S1 Fig.). As we do not know which conformational state of SV2A compounds like levetiracetam (LEV) bind to, we explored two distinct models corresponding to an inward-open and outward-open state. The FucP structure (PDB: 3O7P [31]) was used as the template for the outward-open state and GlpT (PDB: 1PW4 [32]) was used as the template for the inward-open state. Initial structural alignments based on the consensus predictions were manually adjusted. These alignments were used to generate homology models with Modeller 9.10 [33]. In the case where the model was predicted to be helical but the template did not have helical structure, restraints were applied within Modeller to ensure helicity. As the focus here was on the racetam-binding site within the TM region, we did not attempt to make structural predictions for TMH-connecting regions and modeled them as loops within Modeller. 100 models were generated in each case and the model with the best DOPE score [34] was selected.

To refine these models further we used conservation data from a sequence alignment. A BLAST [35] search with SV2A_RAT (Uniprot: Q02563) as the query sequence returned 758 sequences with $E < 10^{-30}$ and an alignment was constructed using CLUSTAL Omega [36]. The conserved sites were determined using in-house R scripts to produce a heat map: blue—red corresponding to 0–100% conservation. More usefully, for the initial refinement of the transmembrane helices, we generated heat-maps for the degree of hydrophobic conservation which we determined by computing the occurrence of residues M, A, V, I, L, C, Y, F and W (Fig. 1), as has been previously demonstrated for class A G-protein Coupled Receptors [37].

We then examined the initial models and used the conservation analysis to refine the structural alignment. The final alignment is given in Supporting Information (S2 Fig.). Models were then constructed using Modeller as before and additionally assessed in terms of their QMEAN score [38] to provide an absolute measure of the structural quality that can be compared to known structures.

Molecular Dynamics. The final models were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer using the *g_membed* [39] feature of GROMACS and an energy minimization with a steepest descent algorithm until convergence with a force tolerance of 0.239 kcal mol^{-1} $Å^{-1}$ was performed. Sodium and chloride ions were then added to the

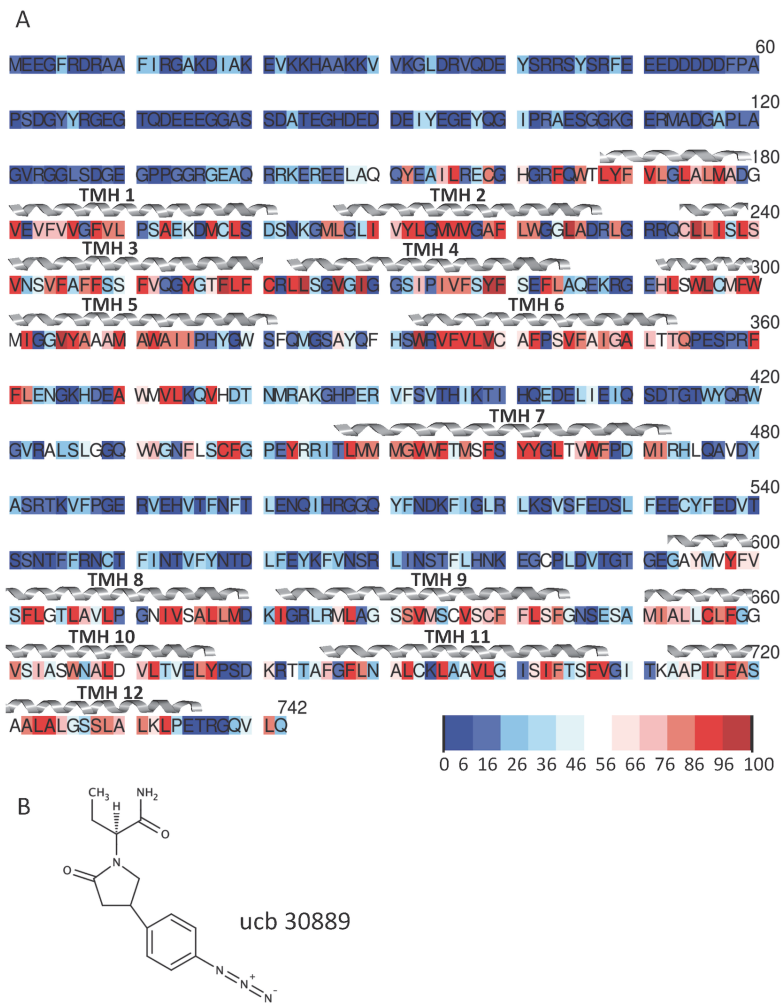


Fig 1. (A) Conservation pattern of residues (M, A, V, I, L, C, Y, W and F) as ascertained by an alignment of 758 sequences form a BLAST search against rat SV2A . The degree of conservation is coloured from blue to red as a function of percentage. The position of the TM helices as predicted from the consensus prediction (S1 Fig.) are indicated. (B) Chemical structure of ucb 30889, a commonly used radio-ligand that is an analogue of LEV.

doi:10.1371/journal.pone.0116589.g001

systems to a concentration of 150 mM followed by as restrained MD run whereby all heavy atoms were restrained by a harmonic potential of 2.39 kcal mol^{-1} $Å^{-2}$ for 1 ns. Finally, 80 ns of production runs were performed on three repeats that differed in their initial velocities only. As an additional check of the simulations we analyzed the area per lipid for all simulations and

Table 1. Summary of Simulations.

Name	Description
Inward-apo	Model based on GlpT template
Outward-apo	Model based on FucP template
Inward-ucb 30889	Inward apo model with ucb 30889 docked into the binding pocket
Outward-ucb 30889	Outward apo model with ucb 30889 docked into the binding pocket

Simulations were 80ns in duration and were repeated with 3 different random seeds for the initial velocity assignment.

doi:10.1371/journal.pone.0116589.t001

observed that there was no significant difference compared to t = 0 or between runs. The areas per lipid (in $Å^2$) for the Inward-apo, Inward-ucb 30889, Outward-apo and Outward-ucb 30889 simulations were 65.58 ± 0.82, 68.63 ± 0.74, 68.21 ± 0.71 and 67.49 ± 0.71 respectively. Simulations were performed for inward and outward models both in the apo state and in complex with the radioligand (used in previous functional assays) ucb 30889 as summarized in Table 1. MD simulations were carried out with GROMACS v4.5.4 [40] using the OPLS-AA [41,42] force-field and TIP3P water molecules [43]. Production simulations were performed in an NPT ensemble maintained at 323 K and 1 bar pressure. The integration time step was set as 2 fs and a stochastic dynamics integrator [44] was used. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method [45] with a 14 Å cut-off and 1 Å space grid. The Lennard-Jones potential used a cut-off of 9 Å, with a switch at 8 Å. The LINCS algorithm [46] was used in order to constrain bond lengths in both the lipid molecules and the protein. For the simulations with ucb 30889 bound, the protein coordinates were taken from the apo simulation at 80 ns timestep and Autodock Vina [47] was used to dock the ligand into the cavity to provide the initial starting coordinates for the protein-ligand complex. The docking grid was defined as a rectangle that encompassed the TM cavity between the lipid head groups of the bilayer. The conformation of ligand chosen was that with the best score and energy minimization of the complex was conducted as above, before MD simulation.

Radioligands, drugs and chemicals. ucb 30889 ((2S)-2-[4-(3-azidophenyl)-2-oxopyrrolidin-1-yl]butanamide) was synthesized at UCB (Braine-l'Alleud, Belgium). [^{3}H]ucb 30889 (47 Ci $mmol^{-1}$) was custom labeled by Amersham Biosciences (Roosendaal, The Netherlands). HEK cells and zeocin were purchased from Life Technologies (Merelbeke, Belgium). Phosphate buffered saline (PBS), Dulbecco's Modified Eagle Medium (DMEM), L-glutamine, trypsin and fetal bovine serum were purchased from Lonza (Verviers, Belgium). Complete protease inhibitor cocktail was purchased from Roche (Vilvoorde, Belgium) and DNAse (Deoxyribonuclease I, Type II from Bovine Pancreas) from Sigma Aldrich (Bornem, Belgium). GF/B filters and GF/C filters were from Brandel (Alpha Biotech Ltd., London, UK) and Pall (Zaventem, Belgium) respectively. All other reagents were of analytical grade and obtained from conventional commercial sources.

Preparation of membrane proteins from HEK cells. Human SV2A was cloned from a fetal brain cDNA library as described in Lynch *et al.* [4] SV2A point mutants were synthesized at Genscript and cloned into pcDNA3.1(+) (Life Technologies, Gent, Belgium). Final constructs were verified by sequencing and were expressed using transient transfections in FreeStyleTM 293-F suspension cells using 293fectinTM following the manufacturers instruction (all from Life Technologies). Cells were subcultured in DMEM containing 200 mM of L-glutamine and 100 μg ml^{-1} zeocin, supplemented with 10% fetal bovine serum. The cells were grown in a humidified atmosphere of 5% CO_2 at 37°C. Confluent cells were pelleted by centrifugation at

1,500 g for 10 min at 4°C. The pellet was washed once with ice cold PBS using the same centrifugation protocol. The resulting pellet was homogenized in a buffer containing 15 mM Tris-HCl, 1 mM EGTA, 0.3 mM EDTA and 2 mM MgCl2 (pH 7.5) supplemented with complete protease inhibitor cocktail Roche. The homogenate was freeze-thawed twice and equilibrated at 25°C followed by a 10 min DNAse (10 U ml^{-1}) treatment. Subsequently, the solution was centrifuged for 25 min at 40,000 g and 4°C. Finally, the pellet was resuspended in a buffer containing 20 mM Tris-HCl (pH 7.4) and 250 mM of sucrose at a protein concentration of 5 to 10 mg ml^{-1} and stored in liquid nitrogen.

Competition binding experiments. Experiments were performed essentially as described before [5]. For all assays, membrane proteins (10 μg per assay) were incubated for 120 min at 4°C in 0.2 ml of a 50 mM Tris-HCl buffer (pH 7.4) containing 2 mM $MgCl_2$. Increasing concentrations of compounds were added in the presence of 5 nM of [^{3}H]ucb 30889. At the end of the incubation period, the membrane-bound radioligand was recovered by rapid filtration through GF/B glass fiber filter plates pre-soaked in 0.1% polyethyleneimine (PEI). Plates were washed rapidly with 0.8 ml of ice-cold Tris buffer (pH 7.4); the total washing procedure did not exceed 10 sec. Scintillation cocktail (Ultima Gold MV, Perkin Elmer, Zaventem, Belgium) was added to the filter plates and the radioactivity trapped on the filters was measured using a β-counter.

Data analysis. IC_{50} values of competition binding experiments were calculated using computerized nonlinear curve fitting methods (Graphpad Prism 6 software, San Diego, CA), according to the equation of a sigmoidal dose response curve with variable slope.

Results and Discussion

Given that SV2A has a low sequence identity to any of the known MFS structures (GlpT has the highest at 16% when aligned to SV2A TM domains: residues 170–375, 445–480, 590–742), and consequently the accuracy of any structural predictions is likely to be low, we attempted to supplement this with multiple sequence comparisons. An alignment of 758 sequences (see Methods) was used to investigate to what extent hydrophobic conservation could be used to suggest TM helix positions. When the conservation of hydrophobic residues is analysed (as shown in Fig. 1), and compared to the consensus TM predictions (S1 Fig.), there is good agreement, thus giving us confidence in the TM predictions. We used this information to structurally align the TM helices of SV2A to FucP and GlpT to provide models of the Outward and Inward (with respect to the cytosol) facing conformations respectively. The alignment was further refined (S2 Fig.) using the hydrophobic conservation patterns, which correspond to buried faces of the helices. The quality of the resulting models (Fig. 2A, B) was assessed with QMEAN where the scores indicated that they sit within the expected range for membrane proteins. The inward-facing model had a score of 0.301 (the range for all 100 models was between 0.330 and 0.401) which when compared to the GlpT template, which has a score of 0.525, was considered reasonable. The outward-facing model had a score of 0.381 (the range for all 100 models was between 0.317–0.388) which when compared to the FucP template with a QMEAN score of 0.512 was also considered reasonable. We also used QMEANclust [48] to assess the confidence of model quality in both models. Unsurprisingly, the loop regions had the highest estimate error (see S3 Fig.).

To explore the conformational stability of both models, we performed MD simulations. The root means square deviation (RMSD) of the TM helix Cα, averaged over three runs for each model, were found to stabilize to 4.29 ± 0.07 Å and 3.57 ± 0.07 Å (Fig. 2C,D) for the Inward-apo and Outward-apo models respectively. A second factor for model validity is the packing of the helices, and to determine this the degree of deviation from an ideal α-helix was calculated

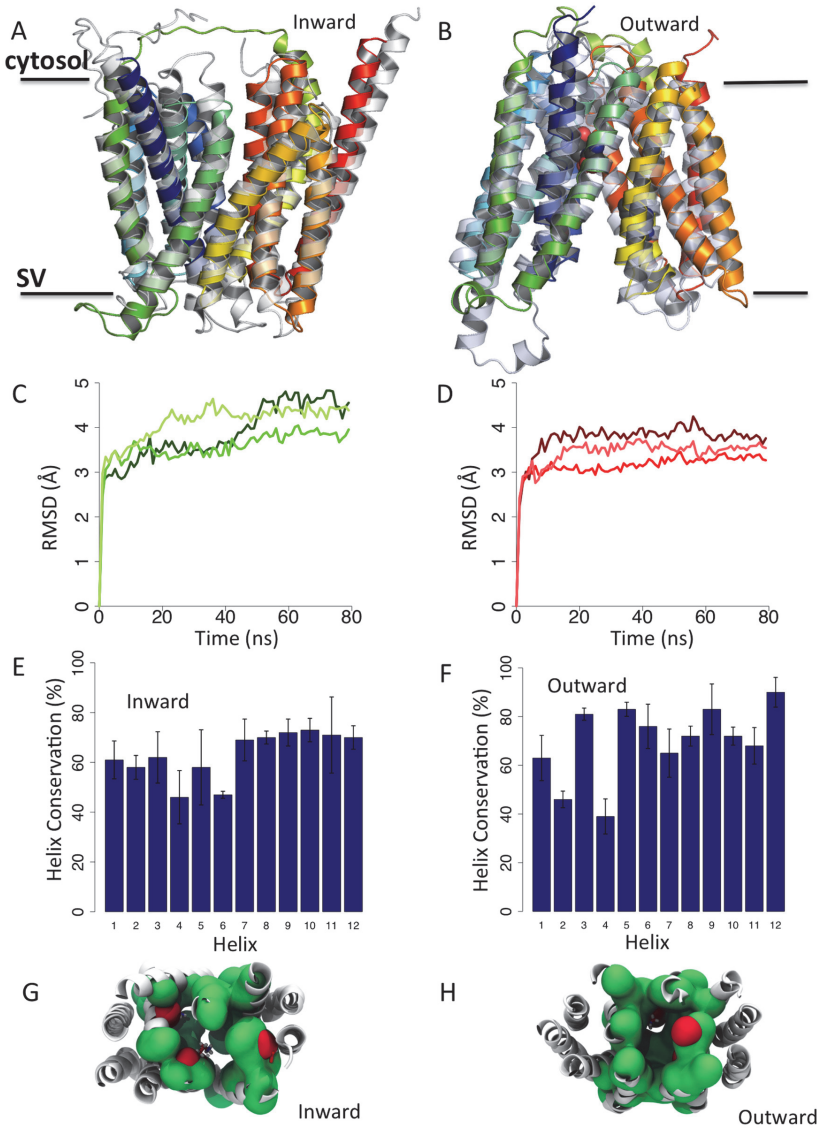


Fig 2. (A) Models of the Inward (based on GlpT template shown in grey) and (B) Outward (based on FucP template shown in grey) SV2A protein. (C) Root mean squared deviation (RMSD) of the Inward-apo (green lines) and (D) the Outward-apo (red lines) simulations (three independent simulations each) over 80 ns. The degree of helix conservation as described by DSSP for each residue in the (E) Inward-apo and (F) Outward-apo models of SV2A. Error bars are the standard deviation (n = 3). (G) A space fill view of the cavity for the Inward-apo and (H) Outward-apo models, with hydrophobic residues coloured green and polar residues shown in red.

doi:10.1371/journal.pone.0116589.g002

(Fig. 2E,F). Since SV2A is a putative transporter that may undergo conformational change as part of its function, a certain amount of structural fluctuation might be expected. In terms of helical character, the apo-systems have greater than 60% conservation of helicity (as defined by DSSP [49]) in all but 3 helices (TMHs 2, 4 and 6) for the Inward-apo and 2 helices (TMHs 2 and 4) for the Outward-apo model, which we take to indicate adequate TMH packing in the models, given that simulations of the templates, GlpT and FucP, maintained helicity in equivalent TM regions. We should note an important caveat at this point and that is that we have performed these simulations in a pure POPC bilayer, and thus at this stage we cannot rule out the specific effects of lipid and protein components that might be found *in vivo*. Nevertheless, these simulations should provide some reassurance that the model is reasonable and compatible with a membrane environment.

We then proceeded to analyze the cavity in the different models. The fluctuations in the volume throughout the simulations were smaller than the differences between models. For example the Inward-apo simulation had a volume of 3843 ± 158 $Å^3$ whilst the Inward-ubc 30889 simulation had a mean volume of 3263 ± 111 $Å^3$. The outward models had similarly low levels of fluctuation; 2929 ± 45 $Å^3$ and 3553 ± 103 $Å^3$ for the Outward-apo and Outward-ucb 30889 simulations respectively. These data indicate that on this timescale the models are conformationally stable. The residues lining the cavity are predominately hydrophobic in character (Fig. 2G, H). Further analysis of the conservation of residues within the proposed binding site indicates a conservation of hydrophobicity in this specific region of the cavity. In particular V276, F280, L284 and L296 have hydrophobic conservations between 76 and 96%, despite lower conservations of the particular residue found in each site of SV2A (Table 2) and all of which interact with the docked ligand in both the Inward and Outward models.. This conservation suggests a functional relevance in these positions, tentatively indicating that the endogenous ligand would have some hydrophobic character, especially considering the importance of W300, Y462 and W666 in racetam binding, as determined by Shi et al. [26] all of which display hydrophobic conservation in those sites of 93.9, 86.1 and 82.5% respectively.

Table 2. Conservation of residues in an alignment of 758 sequences from a BLAST search of SV2A.

Residue	Conservation (%)	Hydrophobic Conservation (%)
SDM residues		
C297	28.4	35.5
W300	83.4	93.9
Y462	83.0	86.1
W666	60.4	82.5
K694	66.4 (R), 23.9 (K)	
Charged residue in cavity		
E194	96.6	
Other residues in cavity		
W454	61.5	79.8
F688	24.9	88.1
Hydrophobic pocket		
V296	61.4	96.0
F280	30.7	80.2
L284	41.8	96.3
L296	40.8	76.3

doi:10.1371/journal.pone.0116589.t002

In addition to the hydrophobic character, there are 3 charged residues within the TM cavity in both models (Table 2), two of which (K694 and E194) are highly conserved. The final residue, D670, has 45% conservation (across the 758 sequences obtained from BLAST; see Methods), though an alignment of just the 24 SV2A sequences indicates that this aspartate is 100% conserved (S4 Fig.) suggesting a role specific to SV2A, rather than a broader function applicable to related proteins. D670 has not been postulated to be involved in ligand-binding before, however its position correlates well with the position of charged residues in other MFS transporters (for example the E135 residue in FucP). The models indicate that D670 is in close proximity to K694 and also to other residues already determined by Shi et al. [26] to impact racetam binding (W300 and Y462). E194 is located towards the cytosolic end of the TM region and so in these models is not expected to form part of the racetam binding site *per se* though its conservation of 97% implies an important but as of yet undefined role.

The ligand [^{3}H]ucb 30889 (Fig. 1B) has been used in assays on site-directed mutagenesis (SDM) studies reported previously. We docked this compound into the TM cavity of both the inward and outward SV2A models and ran further MD to assess the interaction of the compound with the protein and its influence on the dynamics. RMSD values of the TM Cα atoms show a plateauing (Fig. 3A, B) but at slightly higher values than the apo MD runs at 4.59 ± 0.08 Å (Inward-ucb 30889) and 4.14 ± 0.06 Å (Outward-ucb 30889), suggesting there is some

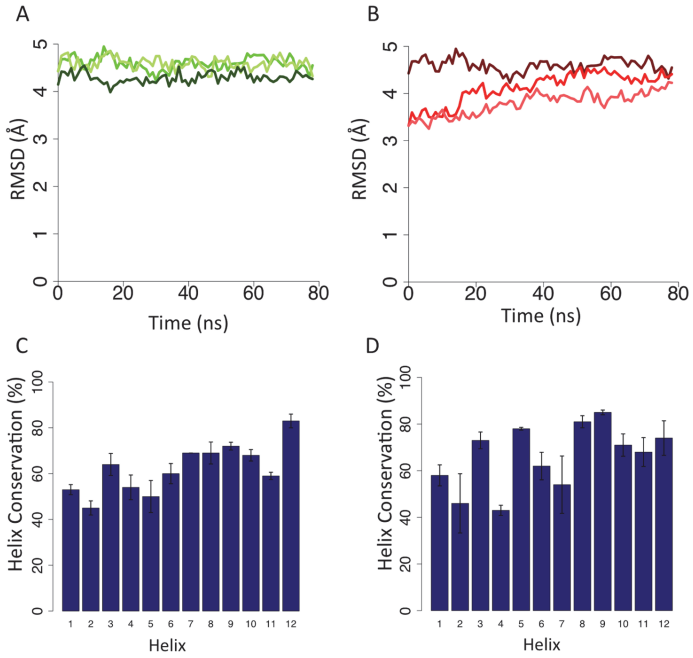


Fig 3. (A) Change in RMSDs from 80 ns of the Inward-ucb 30889 simulation (green) and (B) the Outward-ucb 30889 simulation (red lines) from the initial snapshot from the respective apo simulations . Helix conservation in the (C) Inward-ucb 30889 and (D) Outward-ucb 30889 simulations.

doi:10.1371/journal.pone.0116589.g003

fluctuation in these models after docking the ligand to the apo-system. This observation was supported by the amount of helicity that was preserved in the models. TMHs 1, 2, 4, 5 and 11 are below 60% conservation for the Inward-ucb 30889 run (Fig. 3C) and for the Outward-ucb 30889 run, TMHs 1, 2, 4, 6 and 7 are below 60% (Fig. 3D). All these helices, except TMH 6 line, the cavity, indicating the helices are changing conformation in order to accommodate the ligand. Though the timescale of the simulation is too short to see the large movements associated with the transition between distinct conformational states (expected to be of the order of a few Ångstroms on the basis of structures from other MFS proteins), these fluctuations in the helices could represent the commencement of the protein moving to an occluded state from the open states. However, in order to address that much longer simulation times would be required.

The binding cavity in the Inward and Outward models is pictured in Fig. 4A,B and schematically shown in Fig. 4C,D. Residues known to impact binding as reported previously by Shi et al. [26], were found to interact with the docked ligand: Y462, W666 and K694 for the Inward-ucb 30889 system and C297, W300, M301, Y462 and K694 for the Outward-ucb 30889 system. In the Inward model, C297 does appear to interact with the ligand, but is still in close proximity. Analysis of interaction time throughout the simulations between ligand and protein has further highlighted these residues and indicated those key in each system to be: Y462, W666, N667 and K694 in the Inward-ucb 30889 and C297, W300, Y462, N667 and K694 in the Outward-ucb 30889 system (Fig. 4A, B). In addition to these residues, the Outward-ucb 30889 showed a key interaction between D670 and ucb 30889, showing that 2 of the 3 charged residues in the binding cavity are interacting with the ligand. Thus overall, the ligand-binding cavity is predicted to have a predominantly hydrophobic character supplemented by two or three polar residues; N667, K694 and D670 in the Outward model.

The model allowed us to make predictions that could be tested experimentally—3 residues were explored; W454, F688 and D670. W454 is located near to the binding site, but in the Inward open model is pointing away from the binding cavity. In the Outward model, W454 does not appear to interact directly with ucb 30889 when docked to the last simulation frame, but it is however, pointing towards the cavity and potentially could interact with the ligand (Fig. 4F). Indeed, in MD simulations (Outward-ucb 30889), we found that the ligand interacts with W454 for 21% of the time (across all 3 repeats of Outward-ucb 30889). Thus we chose this residue to help delineate the two models better, and predicted that there would be a modest effect on ligand-binding for this residue. F688 is found at the cytosolic end of the TM cavity in the Inward open model and is buried in the Outward open model, and on this basis we predicted the mutation to have very little, if any, effect on the ligand binding site. D670, in the Inward-apo model, is located at the edge of the cavity, but in the Outward-apo model was located in a more central (in terms of the membrane normal) location and could potentially interact with K694. Indeed in the simulations, the distance between the carboxy oxygens of D670 and the amino hydrogens of K694 was less than 3.5 Å for 35% of the simulation time (with a mean of 4.5 ± 2.6 Å). Given the proposed transporter nature of SV2A, we hypothesized that this interaction may be necessary to help stabilize the Outward open conformation and thus replacing D670 with alanine should result in a decrease in binding ucb 30889. Thus, we predicted that mutating this residue would have a large impact on ligand-binding.

These predictions were borne out by experiments. As predicted, only a small effect on affinity was observed experimentally for W454A and there was almost no effect for F688A (Fig. 4F, Table 3). The position of the W454 is very different in the Inward open and Outward open models. In the Inward open model it is pointing away from the binding cavity, and although we cannot rule out indirect packing effects, we take this to suggest that the Outward open model accounts for this result better as in that model it does point into the cavity (Fig. 4F). For D670A the experiments again confirmed the prediction, with the binding being completely

fluctuation in these models after docking the ligand to the apo-system. This observation was supported by the amount of helicity that was preserved in the models. TMHs 1, 2, 4, 5 and 11 are below 60% conservation for the Inward-ucb 30889 run (Fig. 3C) and for the Outward-ucb 30889 run, TMHs 1, 2, 4, 6 and 7 are below 60% (Fig. 3D). All these helices, except TMH 6 line, the cavity, indicating the helices are changing conformation in order to accommodate the ligand. Though the timescale of the simulation is too short to see the large movements associated with the transition between distinct conformational states (expected to be of the order of a few Ångstroms on the basis of structures from other MFS proteins), these fluctuations in the helices could represent the commencement of the protein moving to an occluded state from the open states. However, in order to address that much longer simulation times would be required.

The binding cavity in the Inward and Outward models is pictured in Fig. 4A,B and schematically shown in Fig. 4C,D. Residues known to impact binding as reported previously by Shi et al. [26], were found to interact with the docked ligand: Y462, W666 and K694 for the Inward-ucb 30889 system and C297, W300, M301, Y462 and K694 for the Outward-ucb 30889 system. In the Inward model, C297 does appear to interact with the ligand, but is still in close proximity. Analysis of interaction time throughout the simulations between ligand and protein has further highlighted these residues and indicated those key in each system to be: Y462, W666, N667 and K694 in the Inward-ucb 30889 and C297, W300, Y462, N667 and K694 in the Outward-ucb 30889 system (Fig. 4A, B). In addition to these residues, the Outward-ucb 30889 showed a key interaction between D670 and ucb 30889, showing that 2 of the 3 charged residues in the binding cavity are interacting with the ligand. Thus overall, the ligand-binding cavity is predicted to have a predominantly hydrophobic character supplemented by two or three polar residues; N667, K694 and D670 in the Outward model.

The model allowed us to make predictions that could be tested experimentally—3 residues were explored; W454, F688 and D670. W454 is located near to the binding site, but in the Inward open model is pointing away from the binding cavity. In the Outward model, W454 does not appear to interact directly with ucb 30889 when docked to the last simulation frame, but it is however, pointing towards the cavity and potentially could interact with the ligand (Fig. 4F). Indeed, in MD simulations (Outward-ucb 30889), we found that the ligand interacts with W454 for 21% of the time (across all 3 repeats of Outward-ucb 30889). Thus we chose this residue to help delineate the two models better, and predicted that there would be a modest effect on ligand-binding for this residue. F688 is found at the cytosolic end of the TM cavity in the Inward open model and is buried in the Outward open model, and on this basis we predicted the mutation to have very little, if any, effect on the ligand binding site. D670, in the Inward-apo model, is located at the edge of the cavity, but in the Outward-apo model was located in a more central (in terms of the membrane normal) location and could potentially interact with K694. Indeed in the simulations, the distance between the carboxy oxygens of D670 and the amino hydrogens of K694 was less than 3.5 Å for 35% of the simulation time (with a mean of 4.5 ± 2.6 Å). Given the proposed transporter nature of SV2A, we hypothesized that this interaction may be necessary to help stabilize the Outward open conformation and thus replacing D670 with alanine should result in a decrease in binding ucb 30889. Thus, we predicted that mutating this residue would have a large impact on ligand-binding.

These predictions were borne out by experiments. As predicted, only a small effect on affinity was observed experimentally for W454A and there was almost no effect for F688A (Fig. 4F, Table 3). The position of the W454 is very different in the Inward open and Outward open models. In the Inward open model it is pointing away from the binding cavity, and although we cannot rule out indirect packing effects, we take this to suggest that the Outward open model accounts for this result better as in that model it does point into the cavity (Fig. 4F). For D670A the experiments again confirmed the prediction, with the binding being completely

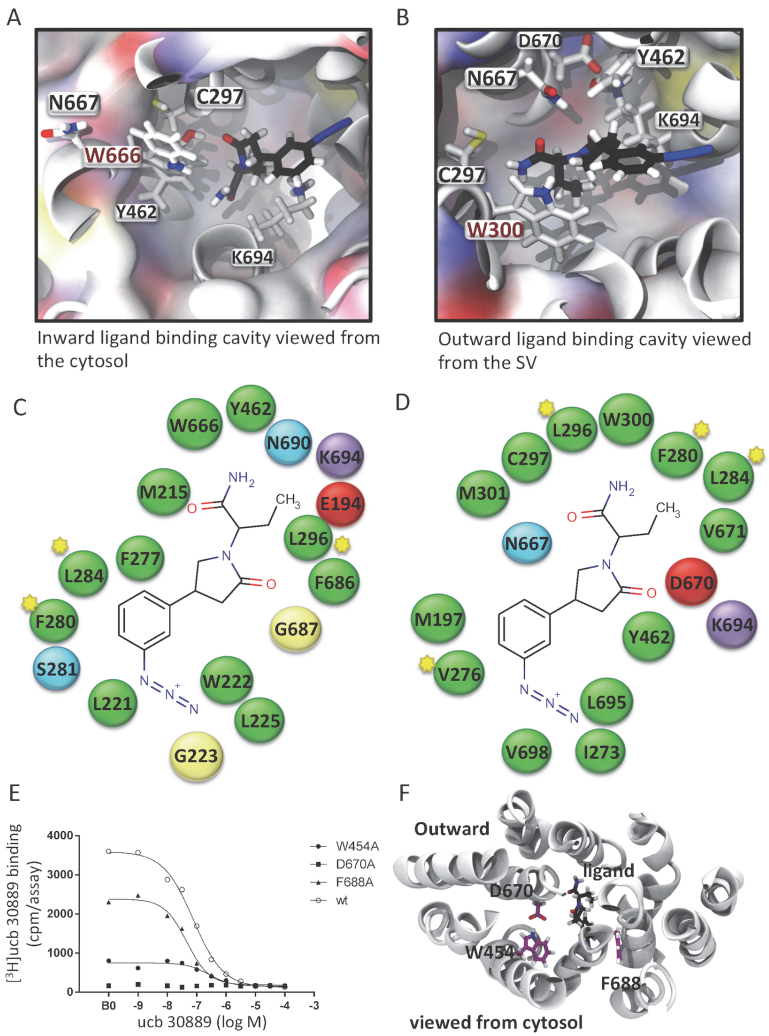


Fig 4. The ligand binding sites in the Inward-apo model of SV2A (A) and the Outward-apo model (B) from simulation (60 ns) . The ligand (black stick) was docked to a snapshot the apo-model after 80 ns simulation. Key residues identified by mutagenesis are highlighted as stick representations. Schematic interaction maps of the docked ligand, generated via MOE with an interaction cut-off of 6 Å are shown for the Inward (C) and Outward (D) models. Residues starred are conserved hydrophobic residues common to both the Inward and Outward ligand binding pockets. (E) Affinity of ucb 30889 for recombinant rat SV2A (wt and mutants). A concentration range of ucb 30889 was incubated with 5 nM of [^{3}H]ucb 30889 during 120 min at 4°C. B0 is the binding of [^{3}H]ucb 30889 in the absence of any competing compound. Data are representative of three independent experiments. pIC_{50} values were calculated from untransformed raw data by non-linear regression using a model describing a sigmoidal dose-response curve with variable slope and are reported in Table 3. The position of the mutants with respect to the ligand in the Outward model is shown in (F).

doi:10.1371/journal.pone.0116589.g004

Table 3. Affinity of ucb 30889 for rat recombinant SV2A (wildtype and mutants) labeled with [^{3}H]ucb 30889 (n = 3).

Rat recombinant SV2A	pIC_{50} (mean ± S.D.)
Wildtype	7.1 ± 0.05
W454A	6.7 ± 0.14
D670A	/
F688A	7.1 ± 0.19

doi:10.1371/journal.pone.0116589.t003

abolished in a radioligand binding assay (Fig. 4E). The position of the D670 in the Inward open model, although suggests it points towards the cavity, is nevertheless on the edge of the cavity. Given that the D670A mutant abolishes binding and the importance of interactions between ligand and TM helix 5 (W300A, C297alkylated are each reported to abolish binding by Shi et al. [26]), the position in the Outward open model is much more conducive to interaction with the ligand and thus we take this also to favour the Outward-apo model over the Inward-apo model. However, we must be extremely cautious, because mutations can manifest their influence on binding through indirect changes as well as direct changes. Indeed, this has been explicitly demonstrated for SV2 proteins where mutant proteins can for example become trapped in the endoplasmic reticulum, presumably reflecting a misfolded state [50]. Another caveat that we should raise at this point is that the mutations are performed in HEK cells and thus the influence of any vesicle proteins on drug binding will also be absent. We must also keep in mind that the sequence identity between the templates and SV2A is extremely low and the possibility of structural differences remains high at this level of similarity.

We should also be clear that we have generated two independent models here, rather than a single model that corresponds to two different states. Although the latter might ultimately be desirable to investigate state-dependent binding effects, we felt that generating the best model for each state independently was more useful at this stage. The development of a unified model is an ongoing area of research.

Conclusions

In this paper we have used homology modelling based on templates corresponding to two different possible states of SV2A. Analysis of the sequence conservation of hydrophobic residues in SV2A in conjunction with additional structural templates has allowed us to identify additional residues that play distinct roles in ucb 30889 (and by inference, LEV) binding. MD simulations of the apo-system confirmed that the model was stable in the timescale of 80 ns but with substantial flexibility within the TM regions. The results suggest that the Outward model is more consistent with the experimental data than the Inward model, though we should stress caution there because the sequence identity between the templates and SV2A is very low. Nevertheless, we were able to use the models in a predictive way to advance our understanding of small-molecule SV2A interactions.

Supporting Information

S1 Fig. The consensus agreement for the position of α-helices (red) and β-sheets (blue) in SV2A, using HMMTop, PSIPred, SOSUI and JPRED. The 12 predicted TM helices for SV2 are indicated by black bars across the top of the alignment.
(PDF)

S2 Fig. The final alignments for (A) GlpT and (B) FucP templates to SV2A, which were used to produce the model.
(PDF)

S3 Fig. The average confidence in model at each residue as given by QMEANlocalscore, as calculated by QMEANclust [48], for the (A) Inward and (B) Outward models respectively. The plots indicate high confidence in the TM helices in all models from MODELLER. Helices are indicated by black lines (standard deviation across the 100 models is shown in pale red)
(PDF)

S4 Fig. The alignment of 24 sequences identified as SV2A in the uniprotKB database—red indicates complete conservation, blue similarity and grey greater variability. These indicate complete conservation of D670 and K694, while Y462 is found in all but one sequence (where it is glutamine).
(PDF)

Acknowledgments

Joanna Lee is a BBSRC-funded student in receipt of additional financial support from UCB BioPharma SPRL. Zara Sands, Florence Lebon and Jiye Shi are all employees of UCB BioPharma SPRL. During the course of the research reported herein, Veronique Daniels was an employee of UCB BioPharma SPRL and her research was supported by a grant from the Walloon Region (Belgium)—DGO6 (Convention 6536. There are no patents or products in development or marketed products to declare that arise from this work. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Author Contributions

Conceived and designed the experiments: JL VD ZAS FL PCB. Performed the experiments: JL VD. Analyzed the data: JL VD ZAS FL JS PCB. Contributed reagents/materials/analysis tools: JL VD ZAS FL JS PCB. Wrote the paper: JL VD ZAS FL JS PCB.

References

1. Fisher RS, Boas WvE, Blume W, Elger C, Genton P, et al. (2005) Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). Epilepsia 46: 470–472. PMID: 15816939
2. Loscher W, Klitgaard H, Twyman RE, Schmidt D (2013) New avenues for anti-epileptic drug discovery and development. Nat Rev Drug Discov 12: 757–776. doi: 10.1038/nrd4126 PMID: 24052047
3. Gambardella A, Labate A, Colosimo E, Ambrosio R, Quattrone A (2008) Monotherapy for partial epilepsy: focus on levetiracetam. Neuropsychiatr Dis Treat 4: 33–38. PMID: 18728811
4. Lynch BA, Lambeng N, Nocka K, Kensel-Hammes P, Bajjalieh SM, et al. (2004) The synaptic vesicle protein SV2A is the binding site for the antiepileptic drug levetiracetam. Proceedings of the National Academy of Sciences of the United States of America 101: 9861–9866. PMID: 15210974
5. Fuks B, Gillard M, Michel P, Lynch B, Vertongen P, et al. (2003) Localization and photoaffinity labelling of the levetiracetam binding site in rat brain and certain cell lines. Eur J Pharm 478: 11–19. PMID: 14555179
6. Klitgaard H, Matagne A, Gobert J, Wülfert E (1998) Evidence for a unique profile of levetiracetam in rodent models of seizures and epilepsy. European Journal of Pharmacology 353: 191–206. PMID: 9726649
7. Tanner VA, Ploug T, Tao-Cheng JH (1996) Subcellular localization of SV2 and other secretory vesicle components in Pc12 cells by an efficient method of preembedding EM immunocytochemistry for cell cultures. J Histochem Cytochem 44: 1481–1489. PMID: 8985140
8. Chang W- P, Südhof TC (2009) SV2 renders primed synaptic vesicles competent for Ca2+-induced exocytosis. J Neurosci 29: 883–897. doi: 10.1523/JNEUROSCI.4521-08.2009 PMID: 19176798

9. Xu T, Bajjalieh SM (2001) SV2 modulates the size of the readily releasable pool of secretory vesicles. Nat Cell Biol 3: 691–698. PMID: 11483953
10. Yao J, Bajjalieh SM (2009) SVOP Is a nucleotide binding protein. PLoS ONE 4: e5315. doi: 10.1371/journal.pone.0005315 PMID: 19390693
11. Yao J, Bajjalieh SM (2008) Synaptic Vesicle Protein 2 binds adenine nucleotides. J Biol Chem 283: 20628–20634. doi: 10.1074/jbc.M800738200 PMID: 18524768
12. Madeo M, Kovács AD, Pearce DA (2014) The human synaptic vesicle protein, SV2A functions as a galactose transporter in Saccharomyces cerevisiae. J Biol Chem.
13. Crowder KM, Gunther JM, Jones TA, Hale BD, Zhang HZ, et al. (1999) Abnormal neurotransmission in mice lacking synaptic vesicle protein 2A (SV2A). Proc Natl Acad Sci USA 96: 15268–15273. PMID: 10611374
14. Crevecoeur J, Foerch P, Doupagne M, Thielen C, Vandenplas C, et al. (2013) Expression of SV2 isoforms during rodent brain development. BMC Neurosci 14: 87. doi: 10.1186/1471-2202-14-87 PMID: 23937191
15. Janz R, Goda Y, Geppert M, Missler M, Südhof TC (1999) SV2A and SV2B function as redundant Ca2 + regulators in neurotransmitter release. Neuron 24: 1003–1016. PMID: 10624962
16. Wan Q- F, Zhou Z- Y, Thakur P, Vila A, Sherry DM, et al. (2010) Sv2 acts via presynaptic calcium to regulate neurotransmitter release. Neuron 66: 884–895. doi: 10.1016/j.neuron.2010.05.010 PMID: 20620874
17. Matveeva EA, Vanaman TC, Whiteheart SW, Slevin JT (2007) Asymmetric accumulation of hippocampal 7S SNARE complexes occurs regardless of kindling paradigm. Epilepsy Res 73: 266–274. PMID: 17174072
18. Van Vliet EA, Aronica E, Redeker S, Boer K, Gorter JA (2009) Decreased expression of synaptic vesicle protein 2A, the binding site for levetiracetam, during epileptogenesis and chronic epilepsy. Epilepsia 50: 422–433. doi: 10.1111/j.1528-1167.2008.01727.x PMID: 18717715
19. Feng G, Xiao F, Lu Y, Huang Z, Yuan J, et al. (2009) Down-regulation synaptic vesicle protein 2A in the anterior temporal neocortex of patients with intractable epilepsy. J Mol Neurosci 39: 354–359. doi: 10.1007/s12031-009-9288-2 PMID: 19757204
20. Bajjalieh SM, Peterson K, Shinghal R, Scheller RH (1992) SV2, a brain synaptic vesicle protein homologous to bacterial transporters. Science 257: 1271–1273. PMID: 1519064
21. Feany MB, Lee S, Edwards RH, Buckley KM (1992) The synaptic vesicle protein SV2 is a novel type of transmembrane transporter. Cell 70: 861–867. PMID: 1355409
22. Gingrich JA, Andersen PH, Tiberi M, Mestikawy SE, Jorgensen PN, et al. (1992) Identification, characterization, and molecular cloning of a novel transporter-like protein localized to the central nervous system. FEBS Lett 312: 115–122. PMID: 1426240
23. Lynch BA, Matagne A, Brännström A, von Euler A, Jansson M, et al. (2008) Visualization of SV2A conformations in situ by the use of protein tomography. Biochem Biophys Res Comms 375: 491–495. doi: 10.1016/j.bbrc.2008.07.145 PMID: 18692481
24. Daniels V, Wood M, Leclercq K, Kaminski RM, Gillard M (2013) Modulation of the conformational state of the SV2A protein by an allosteric mechanism as evidenced by ligand binding assays. Br J Pharmacol 169: 1091–1101. doi: 10.1111/bph.12192 PMID: 23530581
25. Smirnova I, Kasho V, Kaback HR (2011) Lactose permease and the alternating access mechanism. Biochemistry 50: 9684–9693. doi: 10.1021/bi2014294 PMID: 21995338
26. Shi J, Anderson D, Lynch BA, Castaigne JG, Foerch P, et al. (2011) Combining modelling and mutagenesis studies of synaptic vesicle protein 2A to identify a series of residues involved in racetam binding. Biochem Soc Trans 39: 1341–1347. doi: 10.1042/BST0391341 PMID: 21936812
27. Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14: 378–379. PMID: 9632836
28. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17: 849–850. PMID: 11590105
29. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucl Acids Res 36: 197–201.
30. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195–202. PMID: 10493868
31. Dang S, Sun L, Huang Y, Lu F, Liu Y, et al. (2010) Structure of a fucose transporter in an outward-open conformation. Nature 467: 734–738. doi: 10.1038/nature09406 PMID: 20877283
32. Huang Y, Lemieux MJ, Song J, Auer M, Wang DN (2003) Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. Science 301: 616–620. PMID: 12893936

33. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling using Modeller. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.
34. Shen M-Y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15: 2507–2524. PMID: 17075131
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410. PMID: 2231712
36. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7.
37. Ballesteros JA, Weinstein H (1995) Integrated methods for construction three dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods Neurosci 25: 366–428.
38. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27: 343–350. doi: 10.1093/bioinformatics/btq662 PMID: 21134891
39. Wolf MG, Hoefling M, Aponte-Santamaría C, Grubmüller H, Groenhof G (2010) g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. J Comp Chem 31: 2169–2174.
40. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simuation. J Chem Theory Comput 4: 435–447.
41. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225–11236.
42. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105: 6474–6487.
43. Jorgensen WL (1981) Transferable intermolecular potential functions for water, alcohols, and ethers, application to liquid water. J Am Chem Soc 103: 335–340.
44. Van Gunsteren WF, Berendsen HJC (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1: 173–185.
45. Essman U, Perera L, Berkowitz ML, Darden T, Lee H, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103: 8577–8593.
46. Hess B, Bekker J, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comp Chem 18: 1463–1472.
47. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455–461. doi: 10.1002/jcc.21334 PMID: 19499576
48. Benkert P, Schwede T, Tosatto S (2009) QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. BMC Structural Biology 9: 35. doi: 10.1186/1472-6807-9-35 PMID: 19457232
49. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637. PMID: 6667333
50. Nowack A, Yao J, Custer KL, Bajjalieh SM (2010) SV2 regulates neurotransmitter release via multiple mechanisms. Am J Physiol Cell Physiol 299: C960–C967. doi: 10.1152/ajpcell.00259.2010 PMID: 20702688