

# Modeling, Docking, and Simulation of the Major Facilitator Superfamily

John Holyoake,\* Victoria Caulfeild,\* Stephen A. Baldwin,<sup>†</sup> and Mark S. P. Sansom\*

\*Department of Biochemistry, University of Oxford, Oxford, United Kingdom; and <sup>†</sup>Astbury Centre for Structural Molecular Biology, Institute of Membrane and Systems Biology, University of Leeds, Leeds, United Kingdom

**ABSTRACT** X-ray structures are known for three members of the Major Facilitator Superfamily (MFS) of membrane transporter proteins, thus enabling the use of homology modeling to extrapolate to other MFS members. However, before employing such models for, e.g., mutational or docking studies, it is essential to develop a measure of their quality. To aid development of such metrics, two disparate MFS members (NupG and GLUT1) have been modeled. In addition, control models were created with shuffled sequences, to mimic poor quality homology models. These models and the template crystal structures have been examined in terms of both static and dynamic indicators of structural quality. Comparison of the behavior of modeled structures with the crystal structures in molecular dynamics simulations provided a metric for model quality. Docking of the inhibitor forskolin to GLUT1 and to a control model revealed significant differences, indicating that we may identify accurate models despite low sequence identity between target sequences and templates.

Received for publication 27 July 2006 and in final form 6 September 2006.

Address reprint requests and inquiries to Mark Sansom, E-mail: mark.sansom@bioch.ox.ac.uk.

The Major Facilitator Superfamily (MFS) is a large family of membrane transporter proteins present in bacteria, archaea, and eukarya (1). Sequence-based predictions indicate that a 12- or 14-transmembrane (TM) helix topology is shared by all MFS members. MFSs transport a wide range of solutes by diverse mechanisms (uniport, symport, and antiport). Problems associated with overexpression of membrane proteins mean that only three distinct x-ray structures are available for MFSs, namely: LacY (2); the glycerol-3-phosphate transporter (GlpT) (3); and EmrD, a multi-drug transporter (4).

Despite relatively low sequence identities (~15%) between LacY, GlpT, and EmrD, all share a similar fold and arrangement of TM helices. LacY and GlpT are resolved in an inward-facing open conformation, allowing intracellular access to the central binding site. EmrD is in a closed conformation (similar to that seen in the electron microscopy images of OxlT (5)). These structures offer the possibility of homology modeling of other MFSs (6,7), despite very low sequence identities. However, it is important to assess the quality of such models (8).

We have used a combined simulation and docking approach to assess MFS homology models. Two MFS members were modeled: GLUT1, a human facilitative glucose transporter using GlpT as a template; and NupG (a bacterial nucleoside transporter), using LacY as a template. Initial sequence alignments were adjusted manually to optimize agreement with experimental data.

In addition, two “control” models were created: LacY-Con and GLUT1Con (Table 1). In these, the amino-acid sequences of LacY and GLUT1, respectively, were subject to thorough pairwise shuffling (gaps in the GLUT1 alignment were not subject to shuffling) immediately before homology modeling (i.e., a shuffled alignment was used as the input to modeling). Note that this approach leaves the amino-acid

composition of the GLUT1Con model the same as that of the “true” GLUT1 model.

Structures and models were also used as starting structures for 15-ns molecular dynamics (MD) simulations using GROMACS ([www.gromacs.org](http://www.gromacs.org)) in solvated dimyristoyl phosphatidylcholine bilayers (system size ~65,000 atoms). Repeat simulations of the GLUT1 and GLUT1Con models were performed to provide an estimate of the variability in conformational sampling between simulations. Docking of the potent GLUT1 inhibitor forskolin (9) into the GLUT1 and GLUT1Con models was performed using Autodock 3 (10).

Previous modeling studies (6) have used static indicators of model stereochemical quality, e.g., Ramachandran analysis, reinforced by evaluation of the model against available experimental data. The latter approach is clearly difficult for high throughput modeling a wide range of MFS proteins (as is obtaining a high-quality sequence alignment). In this study, we employ a metric for model quality based on dynamic behavior in simulations. Inclusion of the LacY and GlpT crystal structures and the sequence-shuffled controls enables us to evaluate dynamic indicators of model quality for the GLUT1 and NupG models.

For multiple structures/models/simulations of the same protein, analysis of Ramachandran plots of backbone dihedrals has proved useful (11). However, the percentage of residues in the “Core + Allowed” regions (as defined by Procheck) of the Ramachandran plot is equally high for x-ray structures, for the “true” models, and for the control models. Thus, although a necessary criterion for a high quality model, this measure is not sufficient to discriminate between good and poor models.

TABLE 1 Summary of models and simulations

Name	Sequence identity (%)	Core + Allowed (%)	RMSD (Å)	$\alpha$ -Helix loss (%)
LacY	—	99	$3.2 \pm 0.07$	$7 \pm 0.01$
LacYCon	8	99	$4.3 \pm 0.07$	$23 \pm 0.01$
NupG	10	99	$3.8 \pm 0.06$	$2 \pm 0.01$
GlpT	—	96	$2.4 \pm 0.05$	$4 \pm 0.01$
GLUT1	12	98	$3.6 \pm 0.07$	$7 \pm 0.01$
		98	$3.9 \pm 0.08$	$6 \pm 0.01$
GLUT1Con	6	98	$4.5 \pm 0.09$	$19 \pm 0.02$
		98	$4.5 \pm 0.06$	$20 \pm 0.01$

Sequence identity is with the template; Core + Allowed refers to the percentage of residues in the corresponding region of the Ramachandran plot. RMSDs are for core-domain C $\alpha$  atoms relative to the corresponding initial structures and are evaluated over the period 14–15 ns. The  $\alpha$ -helix loss is for core-domain residues. For GLUT1 and GLUT1Con, the second set of figures refer to the repeat simulations.

A simple measure of the conformational stability of an MFS fold is provided by the root mean-square deviation (RMSD) of the C $\alpha$  atoms of the core helical domains from the corresponding starting structure, thus excluding the flexible termini and interdomain linker regions (Table 1). It is evident that RMSDs for the crystal structures are lower than for each of their respective models, as might be expected. However, encouragingly the RMSDs of the models are significantly lower than that of both controls, indicating that in even the relatively short timescales accessible to MD simulations differential behavior can be observed. Furthermore, the repeat simulations yield similar RMSD values, lending confidence to our observations.

These differences in conformational stability are also apparent in the end structures of the simulations (Fig. 1). The structure of GlpT can be seen to have changed little during the course of the simulation, and the structure of the GLUT1 model in the two simulations remains close to that of the template.

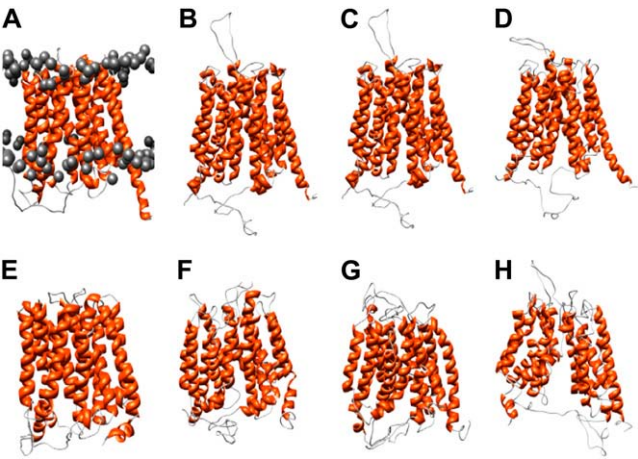


FIGURE 1 Initial (0 ns) and final (15 ns) structures (helices in orange, loops in blue/gray) from simulations: GlpT (A,E); GLUT1 (B,F); GLUT1Rep (C,G); and GLUT1Con (D,H). The phosphorus atoms of the dimyristoyl phosphatidylcholine bilayer are shown for the GlpT simulation as gray spheres.

Interestingly, in both the GlpT and GLUT1 simulations there is a degree of kinking of the C-terminal helix, enabling the intracellular segment to interact with the lipid headgroups. This indicates that changes in MFS structure can occur on an  $\sim$ 15-ns timescale. In contrast, the GLUT1Con model exhibits substantial helix loss, and dissociation of the two six TM helix domains. Indeed, loss of  $\alpha$ -helicity in the two six TM helix bundle domains is the clearest indicator of a difference in structural stability (Table 1). Both control models show over 20% loss in  $\alpha$ -helix content in their core domains, while the crystal structures show only 7% and 4% for LacY and GlpT, respectively. The models show a maximum loss of only 7% for GLUT1 and as low as 2% for NupG. Such low levels, comparable to the two crystal structures, lend confidence to the quality of these models. Taken together, these analyses suggest that one may discriminate between good and poor models of MFS proteins using dynamic structural properties more readily than via static stereochemical analyses.

One use of homology models is in the study of protein/inhibitor interactions. For example, a number of authors have used docking to explore interactions of inhibitors with GLUT1 models (6,12). We have docked the high affinity inhibitor forskolin to the GLUT1 and GLUT1Con models (using Autodock3 (10)). Comparing ensembles of 1000 docks (from the Lamarckian genetic algorithm) reveals a clear difference in behavior between the two models (Fig. 2). For GLUT1, the 1000 docks converge to only a few consistent

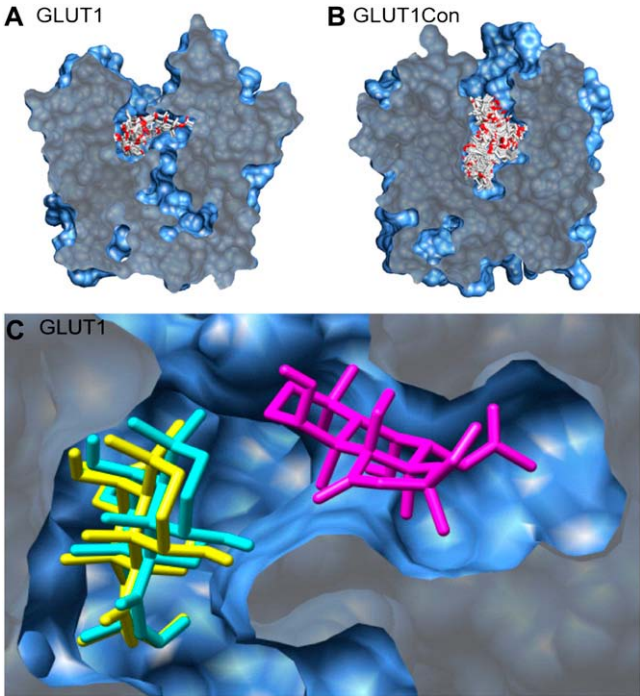


FIGURE 2 Docking of forskolin (shown in bonds format in gray/red/white) into the central cavity of models: (A) GLUT1; and (B) GLUT1Con. Panel C shows the three lowest interaction energy clusters for GLUT1 in purple, cyan, and yellow, respectively.

binding modes. The lowest energy docking mode (seen for 31% of docks) corresponds to a pocket formed by the packing of TM helices 7, 10, and 11. Forskolin forms H-bonds to Trp<sup>388</sup>, and its concatenated ring stacks on top of Trp<sup>412</sup>. Both these residues have been implicated in forskolin binding. Clusters 2 and 3 represent a different binding mode, differing from each other only by a small translation. Interactions in this mode are formed by TM helices 1, 4, and 5, with few specific side chain interactions.

In marked contrast, the output of docking for GLUT1Con failed to reveal a consistent binding mode. Instead, the 1000 docks simply filled the available volume of the central cavity. Thus, for GLUT1 >90% of the 1000 docking attempts were within the top two docking clusters. This contrasts with only 56% for docking into GLUT1Con, increasing to only 66% if the top five clusters are considered. In total there are eight docking clusters for GLUT1 and 42 for GLUT1Con. This difference in docking behavior suggests that the environment within the central cavity of the GLUT1 model is at least a reasonable approximation to that of the true structure.

Taken together, our results indicate that ~10-ns MD simulations in a simple lipid bilayer environment can distinguish the conformational stability of a crystal structure and a control model, or of a plausible homology model and a control model. The latter is encouraging, given the low percentage identity of the model and template sequences. Furthermore, the conformational stability (measured in terms of C $\alpha$  RMSDs and especially in terms of loss of  $\alpha$ -helicity of the core fold) of the (plausible) models is comparable to that of the x-ray structures and consistent between repeat simulations. This degree of discrimination is possible despite all of the starting models ("true" and controls) having comparable stereochemistry as judged by, e.g., Ramachandran plots.

These results have important consequences for attempts to apply high throughput modeling (13) to transporters. There are estimated to be >1000 members of the MFS (www.tcd.org) in 54 different families. Assuming accurate sequence alignments to be achievable, to generate a good homology model of each member would take ~1 h of cpu. To run an ~10-ns simulation of a best homology model and a decoy (i.e., control) for each member would require ~3000 cpu hours. This is not an unreasonable challenge.

In summary, it appears that combining homology modeling with MD simulation can be used to extrapolate from a few x-ray structures to a complete set of plausible homology models, annotated with comparative metrics for their stability. Such models may then be further evaluated by, e.g., cysteine-scanning mutagenesis (14). As x-ray structures of further MFS members emerge, a more fine-grained analysis approach may be possible, e.g., comparing the conformational stability of a model of GlpT based on the structure of EmrD. In this manner it will be possible to cautiously progress to high throughput modeling from all available structures.

## ACKNOWLEDGMENTS

Our thanks to all our colleagues, and especially to Peter Henderson for insightful comments on this manuscript.

J.H. was supported by a Medical Research Council Studentship. This research was supported by grants from Biotechnology and Biological Sciences Research Council (Membrane Protein Structure Initiative), MRC, and the Wellcome Trust.

## REFERENCES and FOOTNOTES

1. Saier, M. H., Jr., J. T. Beatty, A. Goffeau, K. T. Harley, W. H. Heijne, S. C. Huang, D. L. Jack, P. S. Jahn, K. Lew, J. Liu, S. S. Pao, I. T. Paulsen, T. T. Tseng, and P. S. Virk. 1999. The major facilitator superfamily. *J. Mol. Microbiol. Biotechnol.* 1:257–279.
2. Abramson, J., I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata. 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*. 301:610–615.
3. Huang, Y., M. J. Lemieux, J. Song, M. Auer, and D. N. Wang. 2003. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science*. 301:616–620.
4. Yin, Y., X. He, P. Szewczyk, T. Nguyen, and G. Chang. 2006. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science*. 312:741–744.
5. Hirai, T., J. A. Heymann, D. Shi, R. Sarker, P. C. Maloney, and S. Subramaniam. 2002. Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* 9:597–600.
6. Salas-Burgos, A., P. Iserovich, F. Zuniga, J. C. Vera, and J. Fischbarg. 2004. Predicting the three-dimensional structure of the human facilitative glucose transporter GLUT1 by a novel evolutionary homology strategy: insights on the molecular mechanism of substrate migration, and binding sites for glucose and inhibitory molecules. *Biophys. J.* 87: 2990–2999.
7. Vardy, E., I. T. Arkin, K. E. Gottschalk, H. R. Kaback, and S. Schuldiner. 2004. Structural conservation in the major facilitator superfamily as revealed by comparative modeling. *Protein Sci.* 13:1832–1840.
8. Forrest, L. R., C. L. Tang, and B. Honig. 2006. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.* 91:508–517.
9. Lu, L., A. Lundqvist, C. M. Zeng, C. Lagerquist, and P. Lundahl. 1997. D-Glucose, forskolin and cytochalasin B affinities for the glucose transporter GLUT1. Study of pH and reconstitution effects by biomembrane affinity chromatography. *J. Chromatogr. A*. 776:81–86.
10. Morris, G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19:1639–1662.
11. Law, R. J., C. Capener, M. Baaden, P. J. Bond, J. Campbell, G. Patargias, Y. Arinaminpathy, and M. S. Sansom. 2005. Membrane protein structure quality in molecular dynamics simulation. *J. Mol. Graph. Model.* 24:157–165.
12. Cunningham, P., I. Afzal-Ahmed, and R. J. Naftalin. 2006. Docking studies show that D-glucose and quercetin slide through the transporter GLUT1. *J. Biol. Chem.* 281:5797–5803.
13. Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29:291–325.
14. Mueckler, M., and C. Makepeace. 2005. Cysteine-scanning mutagenesis and substituted cysteine accessibility analysis of transmembrane segment 4 of the GLUT1 glucose transporter. *J. Biol. Chem.* 280: 39562–39568.