ORIGINAL PAPER

Molecular dynamics simulations and membrane protein structure quality

Anthony Ivetac · Mark S. P. Sansom

Received: 10 May 2007 / Revised: 28 September 2007 / Accepted: 3 October 2007 / Published online: 25 October 2007 © EBSA 2007

Abstract Despite a growing repertoire of membrane protein structures (currently ~120 unique structures), considerations of low resolution and crystallization in the absence of a lipid bilayer require the development of techniques to assess the global quality of membrane protein folds. This is also the case for assessment of, e.g. homology models of human membrane proteins based on structures of (distant) bacterial homologues. Molecular dynamics (MD) simulations may be used to help evaluate the quality of a membrane protein structure or model. We have used a structure of the bacterial ABC transporter MsbA which has the correct transmembrane helices but an incorrect handedness and topology of their packing to test simulation methods of quality assessment. An MD simulation of the MsbA model in a lipid bilayer is compared to a simulation of another bacterial ABC transporter, BtuCD. The latter structure has demonstrated good conformational stability in the same bilayer environment and over the same timescale (20 ns) as for the MsbA model simulation. A number of comparative analyses of the two simulations were performed to assess changes in the structural integrity of each protein. The results show a significant difference between the two simulations, chiefly due to the dramatic structural deformations of MsbA. We therefore propose that MD could become a useful quality control tool for membrane protein structural biology. In particular, it provides a way in which to explore the global conformational stability of a model membrane protein fold.

A. Ivetac · M. S. P. Sansom (☒)
Department of Biochemistry, University of Oxford,
South Parks Road, Oxford OX1 3QU, UK
e-mail: mark.sansom@bioch.ox.ac.uk

Introduction

Advances in structural biology of membrane proteins are yielding an increasing number of structures, with \sim 120 unique structures in the current database (http://blanco. biomol.uci.edu/Membrane Proteins xtal.html). However, it should be recalled that the majority of membrane protein structures are determined in a non-membrane environment (e.g. crystallization in the presence of detergents or antibodies). Given the key roles of lipids in, e.g. ion channel function (Schmidt et al. 2006), this may be an important limitation. Furthermore, a significant number of the membrane protein structures determined are at relatively low resolutions. Thus, there are 60 unique structures at 3 Å resolution or worse, i.e. ~50% of the unique membrane protein structures. In contrast, for the RCSB (http:// www.rcsb.org) as a whole, ~4% of the entries are at 3 Å resolution or worse. Thus, computational tools to evaluate the "quality" of a membrane protein should play a key role in membrane protein structural biology. Standard methods for measuring protein quality [e.g. Procheck (Laskowski et al. 1993) and Whatif (Vriend 1990)] provide an assessment of local stereochemistry (e.g. for the backbone via Ramachandran plot analysis and for the sidechains via analysis of torsion angle distributions) but do not consider the global fold (i.e. α-helix packing) of a membrane protein model. Molecular dynamics (MD) simulations in a lipid bilayer environment provide a possible tool to address the latter aspect.

We have previously used MD simulations to evaluate homology models of membrane proteins, e.g. a potassium channel (Capener et al. 2000) or members of the major facilitator superfamily of transport proteins (Holyoake et al. 2006). The former study explored the ability of MD simulations to discriminate between small differences in sequence



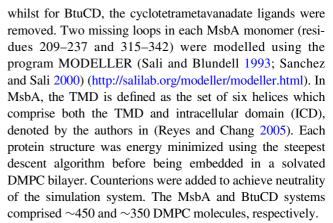
alignment. The MFS model simulations explored the use of MD simulations to detect poor homology models generated by randomization of the sequences of TM helices. MD simulations flagged an issue relating to resolution in an early structure of Aqp1 (Zhu et al. 2001) and have been used to explore the relationship between structure resolution and conformational drift in simulations of a number of membrane proteins (Law et al. 2005). Taken together, these studies have indicated that MD simulations may be able to highlight difficulties with low-resolution structures or imperfect sequence alignments in homology models. We now have an unusual opportunity to extend our range of test cases to the scenario where the TM helices have been correctly assigned, but where the packing together of the TM helices is incorrect. This has arisen because a recent 4.2 Å resolution structure for a bacterial ABC transporter [MsbA from S. typhimurium (Reves and Chang 2005)] has proved to be incorrect in both the handedness of the structure and the topology (Chang et al. 2006). Because the positions of the TM helices within the (incorrect) model agree well with a consensus of sequence based predictions (Cuthbertson et al. 2005), this system mimics a situation in which a membrane protein model accurately embodies the transbilayer topology, but has incorrect packing of the helices. Such a situation could conceivably also arise, e.g. either via homology modelling with a low sequence identity between target and template, via restraint-directed fitting of a TMH fold to low resolution electron density data, or via modelling of an α-helical membrane protein to NMR data with a relatively small number of paramagnetic relaxation enhancement restraints (Tamm and Liang 2006).

In this paper we compare two MD simulations: one of the MsbA model structure and one of another ABC protein, the vitamin B_{12} transporter BtuCD (Ivetac et al. 2007; Locher et al. 2002), each protein embedded in a dimyristoylphosphatidlycholine (DMPC) bilayer. Comparative analyses of structural integrity of the protein in the two simulations demonstrate that MD simulations may be used as a measure of quality of a membrane protein structure.

Methods

Simulation systems

The starting structures for the two simulations were MsbA from *S. typhimurium* (Reyes and Chang 2005) (pdb code 1Z2R, obsoleted on 2006-11-21; resolution 4.2 Å) and BtuCD from *E. coli* (Locher et al. 2002) (pdb code 1L7V; resolution 3.2 Å). Non-protein atoms were removed from the structures, such that the proteins were simulated in a ligand-free (ie. apo) state. Thus, for MsbA, the ADP-vanadate ligand and lipopolysaccharide molecules were removed,



In each case the protein was embedded in a pre-equilibrated bilayer following the protocol described in detail in (Ivetac et al. 2007). Thus, a pre-equilibrated DMPC bilayer of 576 molecules formed the basis of subsequent membrane-inserted simulations. The protein was inserted in this bilayer. DMPC molecules overlapping the protein were removed, and the resultant protein/bilayer system was solvated, energy minimized, and equilibrated, with the non-H atoms of the protein restrained in order to allow a relaxation of the packing of lipids around the protein. In the case of BtuCD the position of the protein in the bilayer was determined visually using bands of charged residues either side of the transbilayer region. The MsbA structure was positioned using a preliminary coarse-grained self-assembly simulation (Bond et al. 2007; Bond and Sansom 2006; Sansom et al. 2007; Scott et al. 2007). The outcome of these two "docking" procedures is shown in Fig. 1. We note that more lipid molecules were used in the MsbA simulation (463) than in the BtuCD simulation (356). This was in order to accommodate the larger protein without any potential periodic boundary artifacts.

Simulation methods

Molecular dynamics simulations were carried out with GROMACS v3.2.1 (MsbA) and GROMACS v.3.1.4 (BtuCD) (http://www.gromacs.org) (Berendsen et al. 1995; Lindahl et al. 2001; van der Spoel et al. 2005), both using the GROMOS96 force-field (Scott et al. 1999; van Gunsteren et al. 1996). Simulations were run in the NPT ensemble at 310 K and 1 atmosphere. Each system was equilibrated (for 0.5 ns for MsbA and for 0.2 ns for the intact BtuCD structure), with the non-H atoms of the protein harmonically restrained (force constant 1,000 kJ mol⁻¹ nm⁻²). During equilibration the Berendsen thermostat and Berendsen barostat algorithms (Berendsen et al. 1984) were employed, while the 20 ns production runs employed the Nosé-Hoover thermostat (Hoover 1985; Nose 1984) and Parrinello-Rahman barostat (Parrinello and Rahman 1981) algorithms. Long-range electrostatic interactions were calculated with the particle



mesh Ewald (PME) method (Darden et al. 1993). All bonds were constrained using the LINCS algorithm (Hess et al. 1997), with the use of a 2 fs timestep.

Analysis

Analysis was carried out using the GROMACS suite of analysis programs. Secondary structure assignment was performed by DSSP (Kabsch and Sander 1983). Molecular graphics images were produced using PYMOL (http://www.pymol.org).

Results

Visual inspection of the MsbA simulation reveals significant structural deformations in both the transmembrane domains (TMDs) and nucleotide-binding domains (NBDs) of the model (Fig. 1). The protein appears to "collapse" and the TMDs exhibit a striking loss of symmetry. The transmembrane helices (TMHs) undergo significant movements from their starting positions, with one TMD appearing to "peel" away from the other. The NBDs appear to retain their fold better, however, there are shifts away from the initial model coordinates. In contrast, the BtuCD structure exhibits little change during the course of the 20 ns simulation, retaining overall structural integrity. Comparison of the radius of

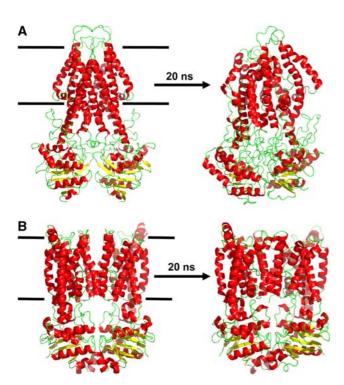
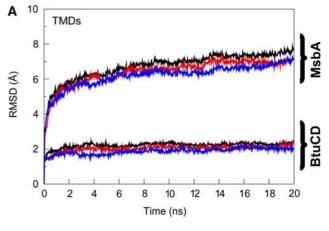


Fig. 1 Initial (t = 0 ns) and final (t = 20 ns) structures from simulations of **a** MsbA and **b** BtuCD. The approximate bilayer location is indicated by the *horizontal black lines*

gyration ($R_{\rm G}$) of each protein in the simulations shows a steady reduction for MsbA, with overall a $\sim 10\%$ decrease in $R_{\rm G}$ over 20 ns, consistent with the "collapse" of the structure. In contrast, BtuCD shows a $\sim 4\%$ decrease in $R_{\rm G}$ in the first 50 ps of the simulation and then remains stable.

To assess the conformational drift of each protein from its initial structure, we have calculated the root mean-square deviation (RMSD) of all $C\alpha$ atoms over 20 ns. In the MsbA simulation, there is an initial (over the first 1 ns) steep increase in the RMSD for all residues of the protein to \sim 6 Å, which is followed by a steady increase to a value of \sim 9 Å at 20 ns. It is also evident that the RMSD would continue to rise if the simulation was extended. In contrast, for the BtuCD simulation there is an initial (over the first 0.1 ns) rise to \sim 2 Å, and then the value plateaus, with a final $C\alpha$ RMSD of \sim 2.7 Å after 20 ns. In order to test whether the high RMSD of MsbA was due to a rigid-body repacking of the two subunits of the protein, the RMSDs for the individual subunits were compared with those of the whole (i.e. dimeric) protein (Fig. 2). For the TMDs



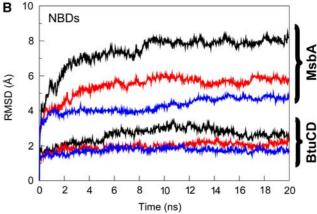
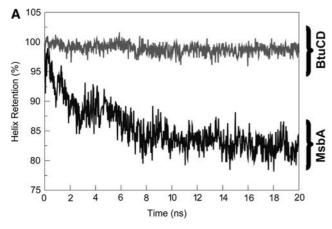


Fig. 2 Conformational drift of the **a** TMDs and **b** NBDs, measured as the RMSDs of the $C\alpha$ atoms from the initial (t = 0 ns) structures. *Black lines* represent RMSD for both subunits, *red* and *blue lines* represent RMSDs for the first subunit and the second subunit, respectively, fitted on just those subunits





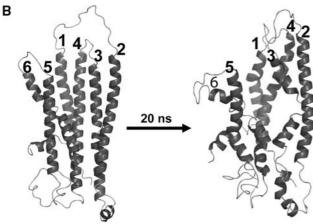


Fig. 3 a Helix content (as a percentage retention of the initial helix content) in the TMDs as a function of time. **b** Initial (0 ns) and final (20 ns) structure of TMD1 from the MsbA simulation, illustrating the distortion and loss of α -helicity

(Fig. 2a), it is clear that almost all of the conformational drift is the result of intra-TMD deformations. For the NBDs (Fig. 2b), the situation is a little more complex. Although the intra-NBD RMSDs are high (\sim 5 Å, compared to \sim 2 Å for BtuCD), it is evident that inter-NBD motions also make a contribution to the total conformational drift of this region of the protein. Thus, the major "instability" in the MsbA

fold would seem to be in the transmembrane domain, with a distortion of the NBDs consequential upon this instability.

A fundamental measure of the structural integrity of the transmembrane domain fold of a membrane protein is provided by monitoring the secondary structure of this domain during the course of the simulation (Holyoake et al. 2006). For example, we have measured the extent to which those TMD residues present in an α -helix retain this secondary structure during a simulation. In the MsbA simulation, 15% of these residues lose their α -helicity, whereas in the BtuCD simulation the comparable loss is only 1% (Fig. 3a). Visualization of the TMDs from MsbA at the end of the simulation reveals both truncation of and breaks in the α -helices (Fig. 3b). In contrast, for the TMDs of BtuCD, all α -helices remain intact at the end of the simulation. One may assign a "dominant secondary structure" to each residue over the course of a simulation, i.e. the secondary structure adopted for $\geq 80\%$ of the simulation. Comparison for the two simulations (Fig. 4) reveals a marked loss of α -helicity for most of the 12 TMHs of MsbA, most dramatically so in TM3, whereas for BtuCD the TMHs show almost no loss of secondary structure. These changes are significant given that the location of the TMHs within the initial MsbA model is in good agreement with the consensus predicted locations of TMHs (Cuthbertson et al. 2005) in the MsbA sequence, which suggests that the secondary structure assignment of the initial model is broadly correct. We also analyzed the number of inter-helix H-bonds during the MsbA simulations, but no significant trends were observed (data not shown).

It is also informative to compare the orientation and interaction of the two proteins with respect to the lipid bilayer during the simulations. For the orientation of each protein relative to the DMPC bilayer, the tilt angle of the long axis of the protein relative to the bilayer normal was calculated via a vector from the centre-of-mass of the NBDs to the centre-of-mass of the TMDs. For MsbA there was a significant change over 20 ns, with the tilt angle initially \sim 0° but progressively shifting to \sim 22° over the course of the simulation (Fig. 5). In contrast that for BtuCD fluctuates

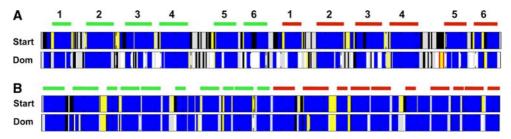


Fig. 4 Secondary structure analysis (using DSSP, Kabsch and Sander (1983)) of TMDs of a MsbA and b BtuCD. In each case, the secondary structure at start of each simulation (*Start*) is compared with the dominant secondary structure (*Dom*). Position of the transmembrane heli-

ces of TMD1 and of TMD2 is indicated by *green* and *red bars*, respectively (*blue* α -helix; *yellow* turn; *black* β -bridge; *purple* 5-helix; *dark grey* 3_{10} -helix; *light grey* coil)



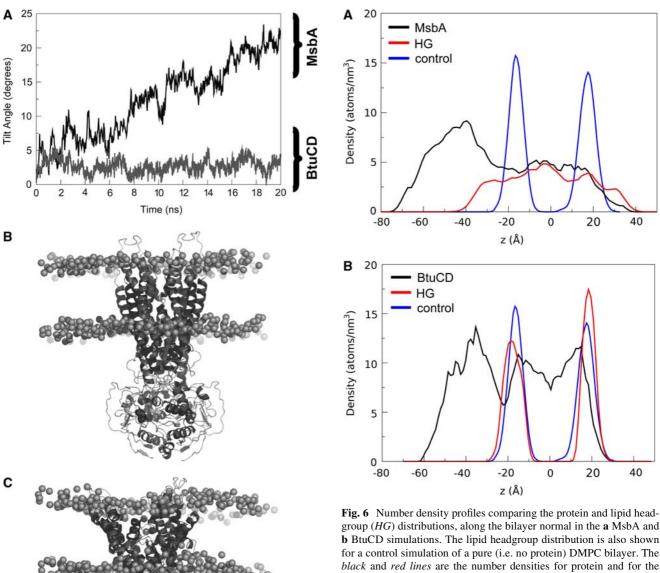


Fig. 5 a Tilt angle of the protein long axis relative to the bilayer normal as a function of time for the two simulations. Initial (b 0 ns) and final (c 20 ns) structures of MsbA are shown relative to the bilayer as defined by the lipid headgroup phosphate atoms (grey spheres)

within 0° to 5° over the course of the simulation. Analysis of the tilt angles of just the TMDs shows that for MsbA there is considerable tilt, with final angles of $\sim 25^{\circ}$ for the TMD of one subunit and of $\sim 34^{\circ}$ for the other TMD.

In terms of lipid/protein interactions, the total number of H-bonds between the protein and the lipid headgroups was analysed (data not shown). There was monotonic drift in the number of such H-bonds for the MsbA/DMPC

group (HG) distributions, along the bilayer normal in the a MsbA and **b** BtuCD simulations. The lipid headgroup distribution is also shown for a control simulation of a pure (i.e. no protein) DMPC bilayer. The black and red lines are the number densities for protein and for the DMPC headgroup atoms, respectively, for the last 4 ns of each protein/ bilayer simulation. The blue lines are the corresponding number densities for the headgroup atoms of a 2.5 ns long control DMPC bilayer simulation

system (from \sim 10 at the start of the simulation to \sim 60 at the end) whereas for the BtuCD/DMPC system there was little change (mean number \sim 55) over the course of the simulation. This difference is consistent with the observed "collapse" of the NBDs (and some of the TMHs) towards the lipid/water interface in the MsbA simulation (Fig. 5c). Of course, this is rather specific to MsbA, and would not provide a quality control measure for all membrane proteins.

We also monitored the behaviour of the lipid bilayer in the two simulations. Thus, as can be seen from Fig. 6a, the bilayer in the MsbA simulation shows considerable local deformation relative to a control simulation of a DMPC bilayer without an inserted protein. In contrast, for BtuCD



(Fig. 6b) the thickness of the bilayer remains identical to that in the control DMPC simulation.

Discussion

In summary, these simulations suggest that MD can be used to detect an incorrect membrane protein structure or model as one which exhibits considerable conformation instability during such simulations. We therefore propose that MD simulation could be a valuable quality control tool for membrane protein structural biology. We note that a single 20 ns simulation of a protein of comparable size to MsbA takes ~480 CPU-days, i.e. ~1 month on 16 CPUs. Furthermore, the "pathological" simulation behaviour of the MsbA simulation was evident after <5 ns of simulation time. Combining this with the observation that many researchers have access to larger clusters either locally or via supercomputing centres, a quality control simulation of an average sized membrane protein could be completed in a matter of a few days.

In previous studies, MD simulations have previously been performed on the $E.\ coli$ MsbA structure (4.5 Å resolution), which only contained an incomplete $C\alpha$ trace (Chang and Roth 2001). Both studies highlighted the relatively high structural drift of the "repaired" protein structure, along with some loss of secondary structure (Campbell et al. 2003; Haubertin et al. 2006). However, at the time of these simulations it was not appreciated that this model was also erroneous.

It would be of interest to extend this study to a wider range of membrane proteins, and to more subtle differences/ errors in structure. One way in which this could be done would be to consider those membrane proteins for which multiple structures at a range of resolutions are available. These include bacteriorhodopsin (resolutions ranging from 3.5 to 1.55 Å), rhodopsin (4.1 to 2.2 Å), KcsA (3.2 to 2.0 Å), Aqp0 (3.0 to 1.9 Å), Aqp1 (3.8 to 2.2 Å), the Rhodobacter sphaeroides photosynthetic reaction centre (3.1 to 2.35 Å), and photosystems I (4.0 to 2.5 Å) and II (3.8 to 3.0 Å). There are also a number of outer membrane proteins for which both X-ray and NMR structures have been determined, and simulation studies have been used to explore the differences in conformational stability of these (Cox et al. 2007). It would also be of interest to compare the same protein(s) in simulations in different lipid bilayers, differing both in tail and in headgroup species.

In more general terms, the simulations described above suggest that one may use MD simulations to detect incorrect packing of otherwise correctly predicted TM α -helices in a membrane protein model. A previous study (Holyoake et al. 2006) suggested simulations may be used to identify a model with poor assignment of TM helices. It remains to explore less extreme cases of low quality models, e.g. a

model with correctly packed helices (in xy plane) but with errors of ± 2 turns in assignment of TMs, such as one might envisage to occur in models based on cryoelectron microscopy maps with anisotropic resolution. However, we are encouraged by studies which show convergence of subtly different models of, e.g. the GpA helix dimer (Cuthbertson et al. 2006), that if lower resolution models "pass" the MD simulation test, refinements in their structure may also be possible via simulation.

Acknowledgments Research in MSPS's laboratory is supported by the BBSRC and the Wellcome Trust. AI was supported by the MRC. Our thanks to all our colleagues, and especially to John Holyoake for insightful comments on this manuscript.

References

- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
- Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 95:43–56
- Bond PJ, Sansom MSP (2006) Insertion and assembly of membrane proteins via simulation. J Am Chem Soc 128:2697–2704
- Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP (2007) Coarsegrained molecular dynamics simulations of membrane proteins and peptides. J Struct Biol 157:593–605
- Campbell JD, Biggin PC, Baaden M, Sansom MSP (2003) Extending the structure of an ABC transporter to atomic resolution: modelling and simulation studies of MsbA. Biochemistry 42:3666–3673
- Capener CE, Shrivastava IH, Ranatunga KM, Forrest LR, Smith GR, Sansom MSP (2000) Homology modelling and molecular dynamics simulation studies of an inward rectifier potassium channel. Biophys J 78:2929–2942
- Chang G, Roth CB (2001) Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. Science 293:1793–1800
- Chang G, Roth CB, Reyes CL, Pornillos O, Chen YJ, Chen AP (2006) Retraction. Science 314:1875
- Cox K, Bond PJ, Grottesi A, Baaden M, Sansom MSP (2007) Outer membrane proteins: comparing X-ray and NMR structures by MD simulations in lipid bilayers. Eur Biophys J. doi:10.1007/ s00249-007-0185-8
- Cuthbertson JM, Doyle DA, Sansom MSP (2005) Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng Des Sel 18:295–308
- Cuthbertson JM, Bond PJ, Sansom MSP (2006) Transmembrane helix-helix interactions: comparative simulations of the glycophorin A dimer. Biochemistry 45:14298–14310
- Darden T, York D, Pedersen L (1993) Particle mesh Ewald—an N.log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
- Haubertin DY, Madaoui H, Sanson A, Guérois R, Orlowski S (2006) Molecular dynamics simulations of E. coli MsbA transmembrane domain: formation of a semipore structure. Biophys J 91:2517–2531
- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comp Chem 18:1463–1472
- Holyoake J, Caulfeild V, Baldwin SA, Sansom MSP (2006) Modelling, docking and simulation of the major facilitator superfamily. Biophys J 91:L84–L86



- Hoover WG (1985) Canonical dynamics: equilibrium phase–space distributions. Phys Rev A 31:1695–1697
- Ivetac A, Campbell JD, Sansom MSP (2007) Dynamics and function in a bacterial ABC transporter: simulation studies of the BtuCDF system and its components. Biochemistry 46:2767–2778
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck—a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
- Law RJ, Capener C, Baaden M, Bond PJ, Campbell J, Patargias G, Arinaminpathy Y, Sansom MSP (2005) Membrane protein structure quality in molecular dynamics simulation. J Mol Graph Mod 24:157–165
- Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7:306–317
- Locher KP, Lee AT, Rees DC (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 296:1091–1098
- Nose S (1984) A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 52:255–268
- Parrinello M, Rahman A (1981) Polymorphic transitions in singlecrystals—a new molecular-dynamics method. J Appl Phys 52:7182–7190
- Reyes CL, Chang G (2005) Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. Science 308:1028–1031
- Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815

- Sanchez R, Sali A (2000) Comparative protein structure modeling. Introduction and practical examples with modeller. Methods Mol Biol 143:97–129
- Sansom MSP, Scott KA, Bond PJ (2007) Coarse grained simulation: a high throughput computational approach to membrane proteins. Biochem Soc Trans (in press)
- Schmidt D, Jiang QX, MacKinnon R (2006) Phospholipids and the origin of cationic gating charges in voltage sensors. Nature 444:775–779
- Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem A 103:3596–3607
- Scott KA, Bond PJ, Ivetec A, Chetwynd AP, Khalid S, Sansom MSP (2007) Membrane protein/bilayer interactions: structural bioinformatics via coarse-grained MD simulations. Structure (submitted)
- Tamm LK, Liang B (2006) NMR of membrane proteins in solution. Prog Nucl Magn Res Spect 48:201–210
- van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718
- van Gunsteren WF, Kruger P, Billeter SR, Mark AE, Eising AA, Scott WRP, Huneberger PH, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Biomos/Hochschulverlag AG an der ETH Zurich, Groningen/Zurich
- Vriend G (1990) WhatIf—a molecular modeling and drug design program. J Mol Graph 8:52–56
- Zhu FQ, Tajkhorshid E, Schulten K (2001) Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. FEBS Lett 504:212–218

