Structure, Volume *23*

Supplemental Information

MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes

Phillip J. Stansfeld, Joseph E. Goose, Martin Caffrey, Elisabeth P. Carpenter, Joanne L. Parker, Simon Newstead, and Mark S.P. Sansom

Supplemental Information for:

MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes

Phillip J. Stansfeld, Joseph E. Goose, Martin Caffrey, Elisabeth P. Carpenter, Joanne L. Parker, Simon Newstead, and Mark S.P. Sansom

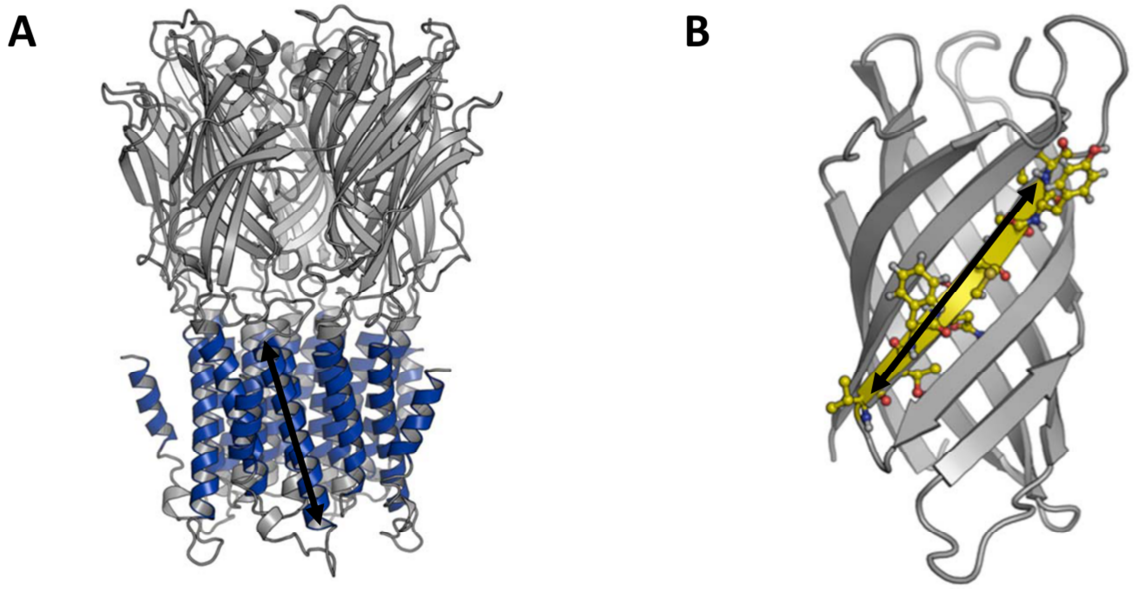

Figure S1, related to Figure 1. Detection of membrane protein structures. Identification of membrane protein structures is achieved through two strategies for membrane proteins of different secondary structures. **A** For α-helical membrane proteins, transmembrane helices are initially identified using Octopus, based on the amino acid sequence of the protein. DSSP is then used to define the secondary structure for these amino acids from the PDB file to see if these residues indeed form an α-helix that is longer than 20 Å, i.e. sufficiently long to span the membrane. The helix is then checked for surface accessibility therefore whether it would actually make contact with the membrane. If at least one transmembrane helix meets all of these criteria then the protein is classified as an integral membrane protein. **B** β-Barrel proteins are initially identified based on their secondary structure. They must contain a β-strand of at least 8 residues, that is at least 20 Å in length. The surface accessibility of the residues in this strand are then assessed using DSSP, with a per-residue hydrophobicity scale (based on OMPLA) applied to residues in the outer face of the strand to calculate the likelihood of membrane insertion. If at least 5 neighbouring strands share these attributes then the barrel is classified as transmembrane.

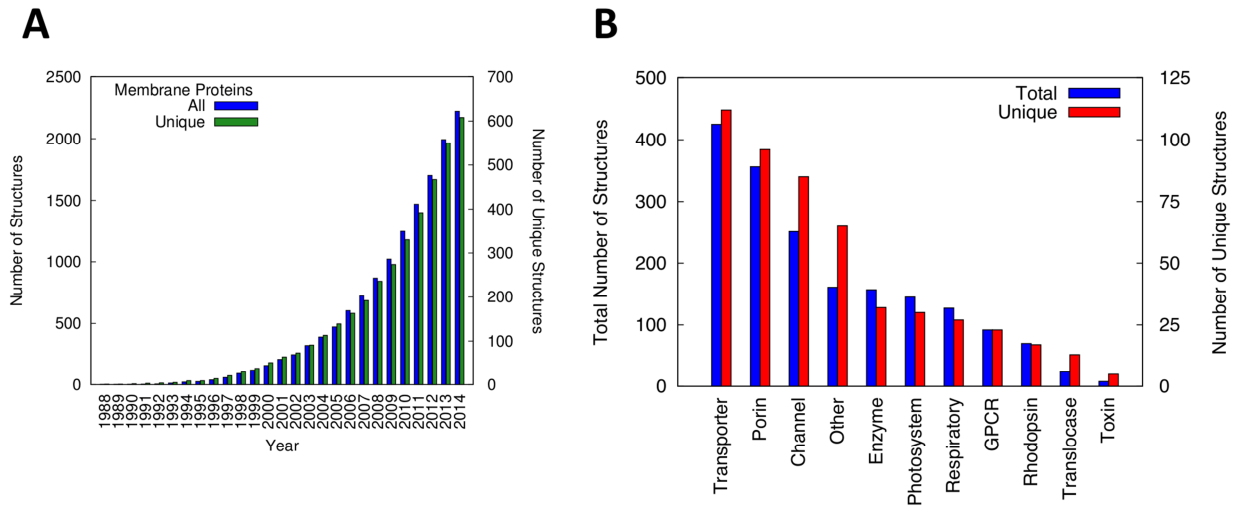

Figure S2, related to Figure 2. Progress in membrane protein structural biology. A The total and unique numbers of membrane protein structures deposited in the PDB since 1988. **B** A classification of membrane protein structures.

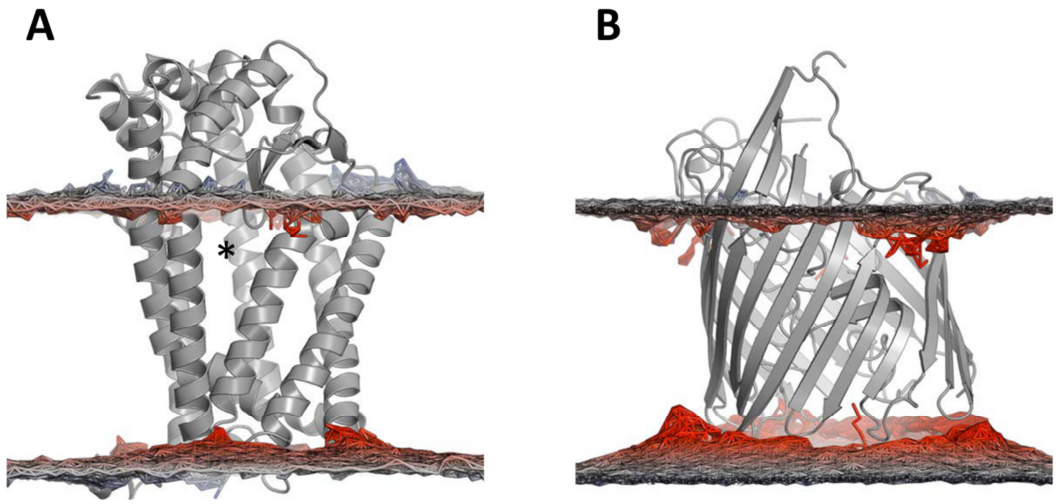

Figure S3, related to Figure 5. Local distortions of the lipid bilayer. **A** Many proteins show relatively small local deformations of the bilayer within the annular shell, as shown here for ZMPSTE24 (PDB id: 4AW6). This enzyme is responsible for the cleavage of a farnesylated peptide, which is believed to enter the barrel at the position marked by an asterisk. **B** Outer Membrane Proteins (OMPs), such as AlgE (PDB id: 4AZL) usually sit in a membrane that is thinner than that formed by a DPPC bilayer and therefore local deformations of the bilayer occur to accommodate the protein. In this instance the periplasmic leaflet this due to a high composition of acidic amino acids on the extracellular side of AlgE, locking it in place on the outer leaflet.

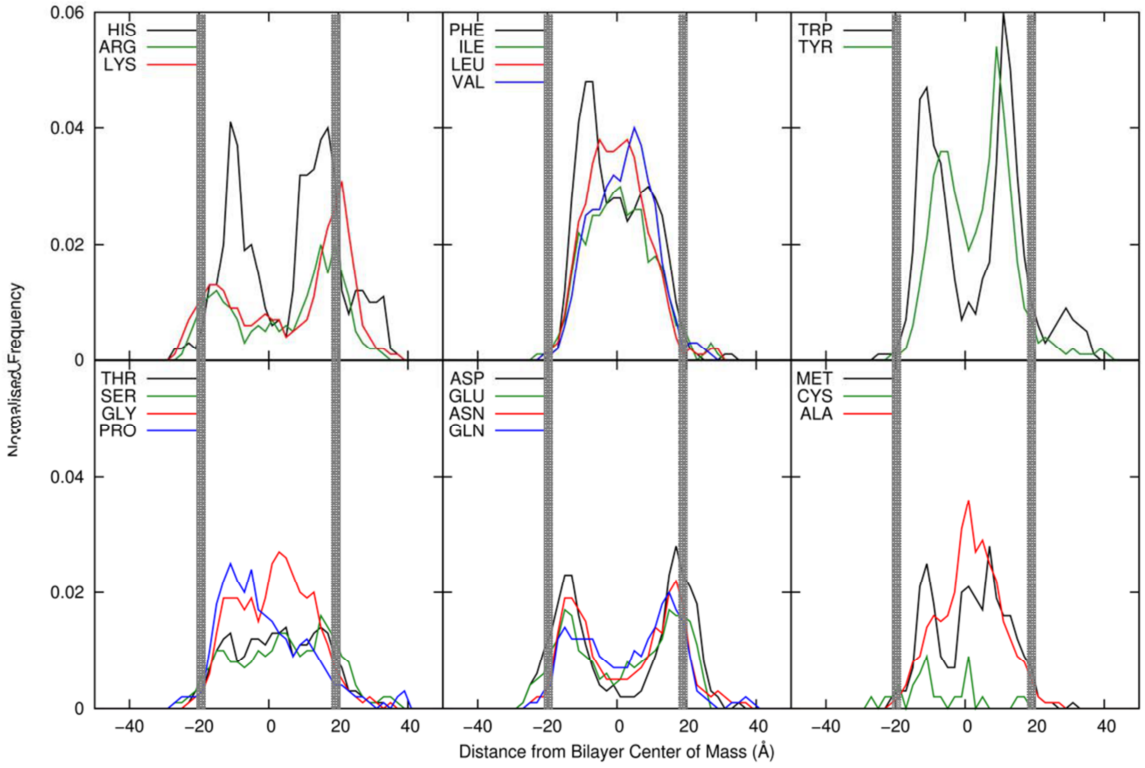

Figure S4, related to Figure 6. OMP amino acid distributions. The residue distributions in OMPs differ from those of the α-helical membrane proteins. Key differences are an apparent "positive-outside" rule corresponding to a higher frequency of lysine and arginine residues on the extracellular face of the (outer) membrane. There also appears to be a preference for proline in the periplasmic leaflet, as this residue is required to terminate the beta strands at the periplasmic interface. Phenylalanine is also found at a higher level at this interface, while tyrosine is found to a greater extent at the extracellular leaflet, likely to maintain H-bonds contacts with the sugar moieties of LPS.

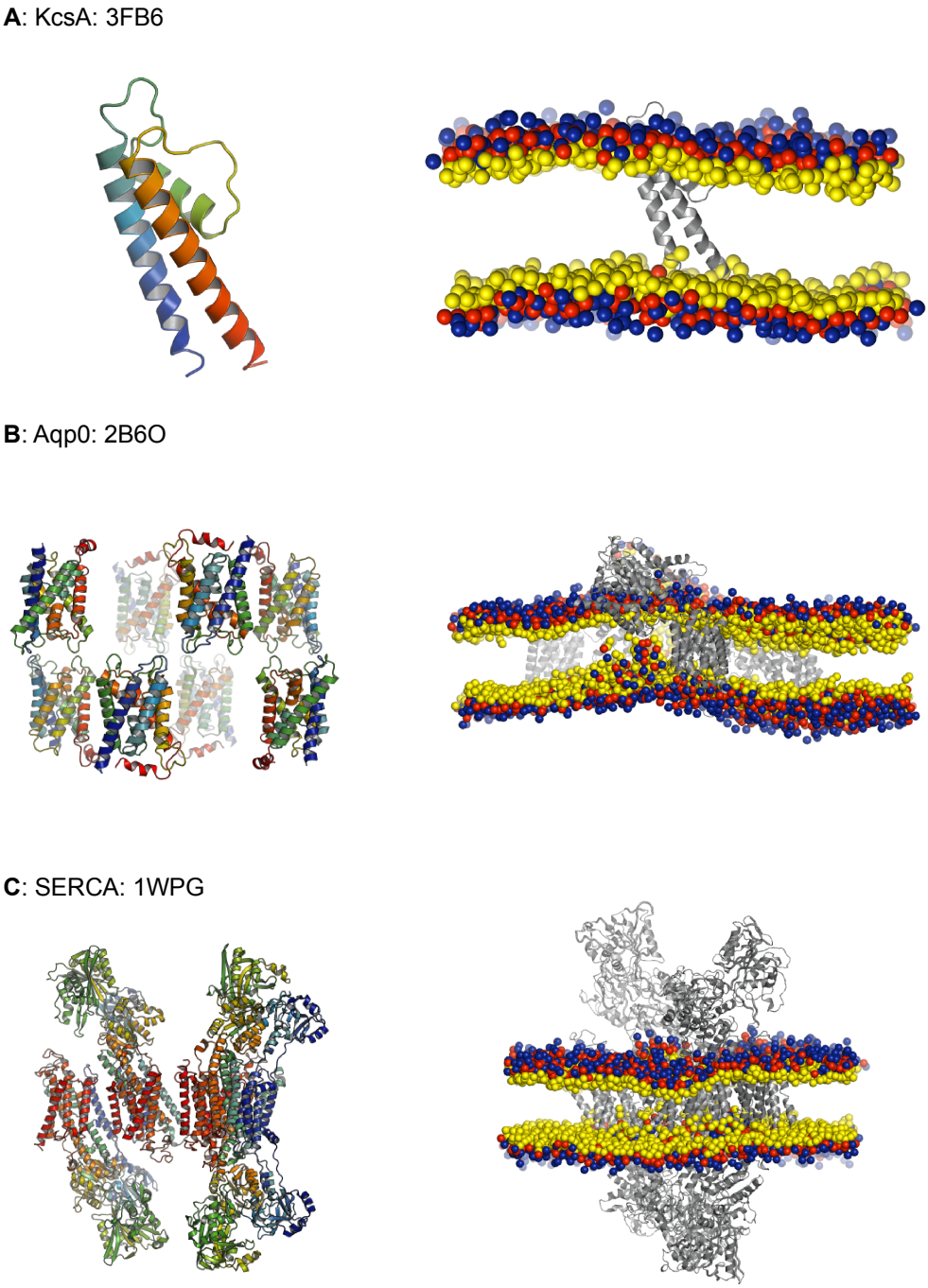

Figure S5, related to Figure 3. Issues with Biological Units. A A membrane protein monomer, which appears to be favourably inserted in the membrane by MemProtMD. This is exemplified by one of the KcsA structures (PDB id 3FB6) for which the 'biological assembly' in the PDB is a monomer rather than a tetramer. In this case the polar residue exposure to the membrane is able to identify oligomerisation interfaces. **B** A non-biological oligomer, which we could identify by MemProtMD using two criteria: (i) different monomers of the oliogomer adopt radically different orientations relative to the bilayer; and (ii) as a consequence the bilayer is seriously distorted. This is exemplified by Aqp0 (PDB id 2B6O) for which 'Biological Assembly 1' in the PDB is a non-biological octamer (whereas 'Biological Assembly 2' is the correct biological tetramer). **C** A non-biological oligomer, which we might flag by its behavior in MemProtMD but for which we would need additional (biochemical) information to decide on the correct oligomerization state. This is

exemplified by the 1WPG structure of SERCA, for which the 'Biological Assembly' in the PDB is the same as the asymmetric unit, namely an anti-parallel tetramer. This would be flagged by the previously mentioned test of "different monomers of the oligomer adopt radically different orientations relative to the bilayer". However, we note that the monomers would be antiparallel, which is not completely excluded biologically (for example, anti-parallel dimers may be formed by EmrE). Therefore curating this requires additional biochemical insight.