BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 5 2006, pages 623–625
doi:10.1093/bioinformatics/btk023

Structural bioinformatics

OPM: Orientations of Proteins in Membranes database

Mikhail A. Lomize[1], Andrei L. Lomize[2,*], Irina D. Pogozheva[2] and Henry I. Mosberg[2]

[1]College of Literature, Science and the Arts and [2]College of Pharmacy, University of Michigan, Ann Arbor, MI 48109-1065, USA

Received on November 28, 2005; revised on December 20, 2005; accepted on December 21, 2005
Advance Access publication January 5, 2006
Associate Editor: Keith A Crandall

ABSTRACT

Summary: The Orientations of Proteins in Membranes (OPM) database provides a collection of transmembrane, monotopic and peripheral proteins from the Protein Data Bank whose spatial arrangements in the lipid bilayer have been calculated theoretically and compared with experimental data. The database allows analysis, sorting and searching of membrane proteins based on their structural classification, species, destination membrane, numbers of transmembrane segments and subunits, numbers of secondary structures and the calculated hydrophobic thickness or tilt angle with respect to the bilayer normal. All coordinate files with the calculated membrane boundaries are available for downloading.

Availabililty: http://opm.phar.umich.edu

Contact: almz@umich.edu

INTRODUCTION

There are hundreds of integral and peripheral membrane proteins with known three-dimensional (3D) structure deposited in the Protein Data Bank (PDB, Berman *et al*., 2000), but their precise positioning in the lipid bilayer is missing. This positioning is essential for biological activity, intermolecular interactions, stability and folding of membrane protein complexes, and it has been studied by a variety of experimental methods, including chemical modification, fluorescence, spin-labeling, X-ray scattering, neutron diffraction, electron cryo-microscopy and NMR or infrared spectroscopy for several dozen cases, such as rhodopsin, lactose permease, mechanosensitive and potassium channels or C2 domains (Frilingos *et al*., 1998; Lee, 2003; Hubbell *et al*., 2003; Malmberg and Falke, 2005). However, since the amount of such experimental data is still very limited, this problem should be addressed computationally to keep up with the expanding flow of structures in the PDB.

Orientations of proteins in membranes may be theoretically calculated by minimizing a protein's transfer energy from water to a planar slab that serves as a crude approximation of the membrane hydrocarbon core (Rees *et al*., 1989). The transfer energy can be estimated in different ways, including whole-residue hydrophobicity scales ('Garlic', Zucic and Juretic, 2004), the normalized accessible surface area of non-polar residues ('TMDET', Tusnady *et al*., 2004) or atomic solvation parameters ('IMPALA', Basyn *et al*., 2003). One of these methods, TMDET, has been applied recently to detect all transmembrane (TM), but not monotopic or peripheral, proteins in the PDB, which were deposited in the PDB_TM database (Tusnady *et al*., 2005). However, the calculated orientations of TM proteins in PDB_TM were not verified through experimental data.

In an attempt to develop a method that better agrees with experimental studies, we designed a more elaborate computational approach for optimizing the spatial arrangement of proteins in membranes. This method combines atomic solvation parameters for the water–decadiene system, interfacial polarity profiles in membranes determined in EPR studies, ionization energies of charged residues and elimination of energetic contributions from any atoms situated in the polar pores or channels of TM proteins that do not interact with lipids (A. L. Lomize, I. D. Pogozheva, M. A. Lomize and H. I. Mosberg, manusript submitted). The developed theoretical approach discriminates between TM and water-soluble proteins and determines the positions of TM proteins with a precision of $\sim$1 Å for the hydrophobic thickness and $\sim 2^\circ$ for the tilt angle relative to the membrane normal. Most importantly, our results are in good agreement with experimental studies of 24 TM proteins, though they are less consistent with results of other computational methods, such as TMDET (Tusnady *et al*., 2004) or IMPALA (Basyn *et al*., 2003) (see more details at OPM website).

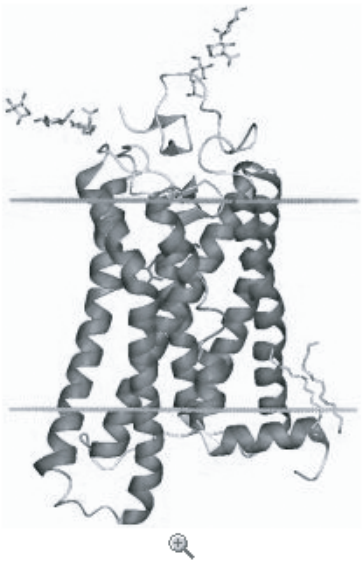
OPM database

The OPM database has several important features. First, the calculated spatial arrangements of TM proteins in membranes were verified by a large sample of published experimental data. Second, protein orientations were calculated for quaternary complexes (biological units) rather than individual subunits or domains. Complexes were generated by the PQS server (Henrick and Thornton, 1998) and verified through the literature to exclude functionally irrelevant oligomers found in crystals, as in the PDB_TM database (Tuesday *et al*., 2004). Third, we included a small initial set of 33 integral monotopic and peripheral membrane proteins, which will be significantly expanded in the future. Fourth, all protein complexes were classified based on the structure of their main membrane-associated domains. Our classification has four hierarchal levels: type (TM or peripheral/monotopic protein and peptides), class (all-α, all-β, α+β, α/β), superfamily (evolutionarily related proteins) and family (proteins with clear sequence homology). The superfamilies and families were taken from SCOP (Andreeva *et al*., 2004), with some corrections, and included proteins not present in the latest release of SCOP. TM superfamilies and families (except enzymes and structural proteins) were ordered by their biological function, as in TCDB (Busch and Saier, 2004), starting from complexes involved in photosynthesis, respiration and

*To whom correspondence should be addressed.

1gzm » Rhodopsin

- *Type:* 1. Transmembrane (2 classes)
- *Class:* 1.1. Alpha-helical transmembrane (31 superfamilies)
- *Superfamily:* 1.1.01. Rhodopsin-like proteins (2 families)
- *Family:* 1.1.01.02. G-protein coupled receptors (1 protein)
- *Species:* *Bos taurus* (8 proteins)
- *Localization:* Eukaryotic plasma membrane (25 proteins)



1gzm » Rhodopsin	
Hydrophobic Thickness	32.4 ± 1.7 Å
Tilt Angle	8 ± 3°
$\Delta G_{transfer}$	-91.3 kcal/mol
Links to 1gzm	PDB Sum, SCOP, MSD, OCA, MMDB, FSSP, HSSP
Topology	subunit A (N-terminus extracellular)
Resolution	2.70 Å
Related PDB Sum entries	1f88, 1hzx, 1l9h, 1u19
Number of TM Secondary Structures	7

3D view in Chime or Webmol

Download Coordinates

Topology in Eukaryotic plasma membrane

1 transmembrane subunit
A - Tilt: 8° - Segments: 1(38-63), 2(72-96), 3(109-133), 4(153-172), 5(202-224), 6(253-274), 7(286-309)

Comments on 1gzm » Rhodopsin
Hydrophobic boundaries expand when calculated with detergent parameters. Monomer and dimer are functional. Structure of functional dimer is unknown. Localization in disc membrane derived from plasma membrane.

Fig. 1. A page from OPM displaying the characteristics of bovine rhodopsin (1gzm).

other primary active transport processes. Finally, we included the tilt angle of the protein relative to the membrane normal, the hydrophobic thickness, the transfer energy of the protein from water to the membrane, the topology, the type of destination membrane and other parameters, which are not provided together in any other resource.

Similar to PDB_TM, OPM provides an up-to-date list of TM proteins with their hydrophobic boundaries. Our current release includes 126 unique 3D structures that represent 506 PDB entries. Typically, a complex with the most complete quaternary structure or one determined with the highest resolution is selected as a representative model. Other structures, such as mutants or conformational states of the same protein, are included in OPM as 'related PDB entries.' However, six types of structures are temporarily excluded: (1) complexes with many unassigned residues and sets of backbone coordinates; (2) low-resolution electron microscopy-based models; (3) incomplete or non-functional assemblies, such as peptide fragments, monomeric units of TM channels (protegrin, mellitin, alamethicin, zervamicin, etc.) or double helices of gramicidin A; (4) NMR models derived from orientational rather than distance constraints; (5) ionophores (valionomycin, monesin, etc.) and (6) theoretical models.

Data access and visualization

OPM allows either searching of proteins by their name or PDB ID or sorting of proteins in tables for many specific categories (type, class, superfamily, family, destination membrane or biological source). Sorting within tables may be executed on the proteins' name, PDB ID, number of TM helices or subunits, number of secondary structures, hydrophobic thickness (membrane penetration depth for peripheral proteins), tilt angle, transfer energy, and structural family, biological source or destination membrane.

An individual web page is generated for each membrane protein complex, as shown in Figure 1, with pictures prepared with QUANTA (Accelrys Inc.). 3D visualization is available through Jmol, MDL Chime and WebMol (Walther, 1997). Coordinate files of all proteins with calculated membrane boundary planes are available for downloading separately for each protein or as a whole dataset. Membrane hydrophobic core planes are marked by dummy atoms of nitrogen for the inner and oxygen atoms for the outer membrane sides according to topology definitions in MPtopo (Jayasinghe *et al*., 2001). OPM is made with PHP, MySQL and uses the Smarty template framework, which separates the program logic (PHP, MySQL) and presentation (XHTML, CSS, JavaScript), and enables caching. OPM is manually curated and will be updated regularly.

ACKNOWLEDGEMENTS

We thank Jim Zajkowski for technical expertise. This work was supported by NIH grant DA003910 and the Upjohn Research Award from the College of Pharmacy, University of Michigan.

Conflict of Interest: none declared.

REFERENCES

Andreeva,A. *et al*. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res*., **32**, D226–D229.

Basyn,F. *et al*. (2003) Insertion of X-ray structures of proteins in membranes. *J. Mol. Graph. Model*., **22**, 11–21.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Busch,W. *et al*. (2004) The IUBMB-endorsed transporter classification system. *Mol. Biotechnol*., **27**, 253–262.

Frillingos,S. *et al*. (1998) Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins. *FASEB J*., **12**, 1281–1299.

Henrick,K. and Thornton,J.M. (1998) PQS: a protein quaternary structure file server. *Trends Biochem. Sci*., **23**, 358–361.

Hubbell,W.L. *et al*. (2003) Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Adv. Protein Chem*., **63**, 243–290.

Jayasinghe,S. *et al*. (2001) MPtopo: A database of membrane protein topology. *Protein Sci*., **10**, 455–458.

Lee,A.G. (2003) Lipid–protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta*., **1612**, 1–40.

Malmberg,N.J. and Falke,J.J. (2005) Use of EPR power saturation to analyze the membrane-docking geometries of peripheral proteins: applications to C2 domains. *Ann. Rev. Biophys. Biomol. Struct*., **34**, 71–90.

Rees,D.C. *et al*. (1989) The bacterial photosynthetic reaction center as a model for membrane proteins. *Annu. Rev. Biochem*., **58**, 607–633.

Tusnady,G.E. *et al*. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics*, **20**, 2964–2972.

Tusnady,G.E. *et al*. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res*., **33**, D275–D278.

Walther,D. (1997) WebMol—a Java-based PDB viewer. *Trends Biochem. Sci*., **22**, 274–275.

Zucic,D. and Juretic,D. (2004) Precise annotation of transmembrane segments with garlic—a free molecular visualization program. *Croat. Chem. Acta*, **77**, 397–401.