MAPS: An interactive web server for membrane annotation of transmembrane protein structures

Jitender Cheema* and Gautam Basu*
Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700 054, India.

Received 07 October 2010; revised 11 March 2011

The exact positioning of the membrane in transmembrane (TM) proteins plays important functional roles. Yet, the structures of TM proteins in protein data bank (pdb) have no information about the explicit position of the membrane. Using a simple hydrophobic lipid-protein mismatch energy function and a flexible lipid/water boundary, the position of lipid bilayer for representative TM proteins in pdb have been annotated. A web server called MAPS (Membrane Annotation of Protein Structures; available at: http://www.boseinst.ernet.in/gautam/maps) has been set up that allows the user to interactively analyze membrane-protein orientations of any uploaded pdb structure with user-defined membrane flexibility parameters.

Keywords: Web server MAPS, Transmembrane proteins, Membrane flexibility

Transmembrane (TM) protein structures in the pdb have no explicit information about the position of the membrane. It is not very difficult to come up with a physically reasonable orientation of the protein with respect to the membrane using residue-based hydrophobicities. A simple score based on hydrophobic mismatch energy can be made more complex by adding lipid perturbation contribution or a structure factor. In addition, the simple description of the membrane as a planar slab with a constant thickness can be changed, especially with respect to the water lipid boundary. A target function more sophisticated than simple hydrophobic mismatch energy, a non-rigid lipid bilayer model and an exhaustive search for the protein-membrane orientation can yield more reliable membrane-protein orientations.

Even with such exhaustive searches using sophisticated objective functions, the final membrane-protein boundary is no more than just a prediction. Results from two different prediction methods differ in their physical meaning, which depends on the algorithm and the models used. In other words, given a number of such predictions, it is up to the user to choose the prediction that uses the most appropriate model and parameters for the specific problem. Thus, the availability of more than one publicly accessible web servers employing different methodologies for membrane annotation of TM proteins is to the advantage of the user. Two web servers currently available towards this end are PDBTM and OPM. However, they do not: i) provide options for membrane flexibility, and ii) allow interactive sessions in where the user can change parameters and redo the calculations.

We have set up a freely available web server MAPS (Membrane Annotated Protein Structures), where a flexible membrane boundary algorithm is implemented to annotate TM proteins and which allows a user to interactively change parameters and obtain different membrane annotations for the same protein. In addition to being complementary to the existing servers, the interactive nature of MAPS can provide additional benefit to students and new learners of structural bioinformatics.

Results and Discussion

Formulation of the problem

Given a protein structure, all residues are first assigned a fraction solvent accessible surface area (ASA) weighted hydrophobicity \( H(i) \)

\[
H(i) = f_{ASA} h(i) \quad \ldots (1)
\]
The hydrophobicity of residues $h(i)$ is user defined and current choices are WW$^{6,7}$, KD$^8$ and GES$^9$ scales. KD and GES scales are adjusted by 1.0 kcal mol$^{-1}$ as compensation for per-residue backbone dehydration cost as suggested by Jayasinghe et al$^{10}$. The fraction ASA values $f_{ASA}(i)$ are calculated by dividing the DSSP-generated$^{11}$ ASA by the maximum ASA (estimated from DSSP-generated ASA of the residue when the residue is present with extended conformation in the tripeptide Gly-Xxx-Gly).

The protein-bilayer orientation is described schematically in Fig. 1a. The lipid bilayer is considered to be a rectangular slab of equilibrium half width $w_h$ and the protein is considered as a rigid body. During the calculation, the bilayer is fixed and is centered at the origin with its normal parallel to the z-axis. The protein is placed in the bilayer such that the z-component of its geometric center is 0. During the calculation the protein is translated along the z-axis (by $z_0$) and tilted from the z-axis ($\theta$) and rotated around the z-axis ($\phi$) to produce new bilayer-protein orientations. Each new orientation is assigned a score as discussed below.

The lipid bilayer with equilibrium half width $w_h^0$ can fluctuate by $\pm w_f$ without any energy penalty (Fig. 1b). However, if the boundary stretches such that $w_h > (w_h^0 + w_f)$ or it shrinks such that $w_h < (w_h^0 - w_f)$, then there is an energy penalty for stretching ($E_+$) or shrinking ($E_-$):

$$E_+ = k_m [w_h - (w_h^0 + w_f)]^2$$

$$E_- = k_m [(w_h^0 - w_f) - w_h]^2$$

$E_+$ and $E_-$ are shown as a function of $(w_h^0 + w_f)$ and $(w_h^0 - w_f)$ respectively in Fig. 2.

For a given bilayer model ($w_h^0$, $w_f$ and $k_m$), the energy $E(z0, \theta, \phi)$ of a given lipid-protein orientation (Fig. 1a) is given by a sum of residue-based (residue index $i$) energy contributions, $R_i[z(i), H(i), z0, \theta, \phi]$, where $z(i)$ is the value of the $z$ Cartesian coordinate of the i-th residue, as:

$$E(z0, \theta, \phi, w_h, w_f, k_m) = \sum_i R_i[z(i), H(i), z0, \theta, \phi, w_h, w_f, k_m]$$

Contributions from individual residues, $R(i)$ in Eq. 3 are given by:

$$w_h < (w_h^0 - w_f), \text{ then there is an energy penalty for stretching (} E_+) \text{ or shrinking (} E_-):$$

$$E_+ = k_m [w_h - (w_h^0 + w_f)]^2$$

$$E_- = k_m [(w_h^0 - w_f) - w_h]^2$$

$E_+$ and $E_-$ are shown as a function of $(w_h^0 + w_f)$ and $(w_h^0 - w_f)$ respectively in Fig. 2.

For a given bilayer model ($w_h^0$, $w_f$ and $k_m$), the energy $E(z0, \theta, \phi)$ of a given lipid-protein orientation (Fig. 1a) is given by a sum of residue-based (residue index $i$) energy contributions, $R_i[z(i), H(i), z0, \theta, \phi]$, where $z(i)$ is the value of the $z$ Cartesian coordinate of the i-th residue, as:

$$E(z0, \theta, \phi, w_h, w_f, k_m) = \sum_i R_i[z(i), H(i), z0, \theta, \phi, w_h, w_f, k_m]$$

Contributions from individual residues, $R(i)$ in

$$w_h < (w_h^0 - w_f), \text{ then there is an energy penalty for stretching (} E_+) \text{ or shrinking (} E_-):$$

$$E_+ = k_m [w_h - (w_h^0 + w_f)]^2$$

$$E_- = k_m [(w_h^0 - w_f) - w_h]^2$$

$E_+$ and $E_-$ are shown as a function of $(w_h^0 + w_f)$ and $(w_h^0 - w_f)$ respectively in Fig. 2.

For a given bilayer model ($w_h^0$, $w_f$ and $k_m$), the energy $E(z0, \theta, \phi)$ of a given lipid-protein orientation (Fig. 1a) is given by a sum of residue-based (residue index $i$) energy contributions, $R_i[z(i), H(i), z0, \theta, \phi]$, where $z(i)$ is the value of the $z$ Cartesian coordinate of the i-th residue, as:

$$E(z0, \theta, \phi, w_h, w_f, k_m) = \sum_i R_i[z(i), H(i), z0, \theta, \phi, w_h, w_f, k_m]$$

Contributions from individual residues, $R(i)$ in

$$w_h < (w_h^0 - w_f), \text{ then there is an energy penalty for stretching (} E_+) \text{ or shrinking (} E_-):$$

$$E_+ = k_m [w_h - (w_h^0 + w_f)]^2$$

$$E_- = k_m [(w_h^0 - w_f) - w_h]^2$$

$E_+$ and $E_-$ are shown as a function of $(w_h^0 + w_f)$ and $(w_h^0 - w_f)$ respectively in Fig. 2.
For a hydrophobic residue \((H(i) < 0)\) when \(|z(i)| \leq (w^0_h + w_f)\), the residue is considered to always be inside the membrane with an energy contribution \(H(i)\) (Type I in Eq. 4 and Fig. 1b). When \(|z(i)| > (w^0_h + w_f)\), the residue is within the membrane with an energy contribution \(E_+ + H(i)\), only if the unfavorable membrane stretching energy \(E_+\) is less than the favorable hydrophobic mismatch energy \(H(i)\); (Type II in Eq. 4 and Fig. 1b). On the other hand, if the unfavorable membrane stretching energy exceeds the favorable hydrophobic mismatch energy, the residue is considered to be outside the membrane with no energy contribution (Type IIa in Fig. 1b).

For a hydrophilic residue \((H(i) > 0)\), when \(|z(i)| \leq (w^0_h - w_f)\), the residue is considered to be inside the membrane only if the unfavorable membrane contracting energy \(E_-\) is less than the unfavorable mismatch energy \(H(i)\). Such a residue will contribute \(E_-\) to the total energy and is covered in Type III in Eq. 1 and Fig. 1b. On the other hand, when the unfavorable membrane contracting energy exceeds the unfavorable hydrophobic mismatch energy \(H(i)\), the residue is considered inside the membrane with an energy contribution \(H(i)\) (type IV in Fig. 1b and Eq. 4). When \(|z(i)| > (w^0_h - w_f)\), the residue is always outside the membrane with no energy contribution (shown as type V in Fig. 1b). In addition, all residues for which \(H(i) = 0\) are considered to be inside the membrane, if they lie within \(w^0_h\) Å from the membrane center, although they do not contribute to \(E(z0)\).

For a given set of \(w^0_h, w_f\) and \(k_m\), the accepted membrane-protein orientation corresponds to that \(\{z0, \Theta, \phi\}\) for which \(E(z0, \Theta, \phi)\) (Eq. 4) is the minimum.

**The MAPS server**

**Options and features**

Java, JavaScript, C and Python languages have been used to implement the program and setting up the web server MAPS (www.boseinst.ernet.in/gautam/maps). From the homepage, the option 'Interactive' allows the user to start the program (Fig. 3). The first step is choosing a PDB file, which can be chosen from the existing searchable database (currently 980 pdb files) or uploaded by the user. Visualization (3D) of protein structures in MAPS is implemented using Jmol (http://jmol.sourceforge.net/) Applets that can be viewed in any Java-enabled browser with Java runtime environment support. The membrane-annotated sequence database, generated using the default parameters on a select set of pdb structures can be downloaded. Membrane-annotated pdb files (membrane in/out assignments are shown in the Cα B-factor column: 96 (in), 3 (out)), sequences in fasta format (where upper case and lower case letters signify amino acids inside and outside the membrane, respectively) and rasmol figures (gif) can also be downloaded for any pdb structure in the results section in the interactive mode (Fig. 4).

**Program parameters**

**Hydrophobicity scale**

The choices for \(h(i)\) are: WW\(^6-7\), ii) KD\(^8\) (default) and GES\(^9\) scales. KD and GES scales are adjusted by 1.0 kcal mol\(^{-1}\) for compensation for per-residue backbone dehydration cost\(^10\). The program also allows a user defined hydrophobic scale.

**Residue centering**

The position of each residue is assigned either by: i) CA (The position of its Cα atoms), or ii) SC
(geometric center of its side chain). Option i) is the default and the only option, if the pdb contains only Cα atoms.

ASA weighing

Two options are available: i) ON and ii) OFF. Option i) is the default. Option ii) cannot be used, if the pdb contains only Cα atoms.

Membrane flexibility

The membrane can be rigid or flexible (default). Default Parameters: i) \( w_h = 26.0 \) Å (DMPC bilayer width at 30°C); ii) \( w_f \) and \( k_m \) are 0.25 Å and 0.1 kcal.mol\(^{-1}\).Å\(^{-1}\), the latter derived by fitting a Gaussian distribution to the equilibrium DMPC carbonyl density along the bilayer normal generated from a long MD simulation of DMPC bilayer\(^{13}\).

Initial orientation

Options for initial protein orientation are: i) pdb (Cartesian z-axis of the input pdb), and ii) MAPS (this option automatically assigns a suitable protein axis similar to that described in TMDET\(^{3}\) with minor modifications).

Utility and uniqueness

The primary aim of MAPS is to provide the user with an interactive environment for membrane-annotation of TM proteins based on a simple model that includes membrane flexibility. All TM proteins are associated with “natural” membrane partners (endoplasmic reticulum, plasma membrane etc.) with unique width and flexibility. The web server MAPS allows the user to adjust membrane parameters that best fits the problem in an interactive fashion. The uniqueness of MAPS is its interactive nature and the inclusion of membrane flexibility. We hope MAPS will not only be beneficial to active scientists in the field, but also be used by students to appreciate the subtleties of the problem by interactively using the server and changing the parameters that suit their purpose.

---

Fig. 4—MAPS membrane annotation page for human aquaporin 1 (pdb code: 1H6I) [The color-annotated structure is shown on the left, the color-annotated sequence is shown on the right and parameters used for the calculation are shown below the sequence. The user can change parameters and redo calculations and download the figure, sequence and annotated pdb structure from this page]
Acknowledgements
Funding for this work was provided by MEXT, Japan, and DBT, India. We thank Sanjib K Gupta and Madhurima Das for their assistance. GB likes to acknowledge critical comments by Nobuhiro Go and Takeshi Kawabata.

References
4 Tusnady G E, Dosztanyi Z & Simon I (2005) Bioinformatics 21, 1276-7
11 Kabsch W & Sander C (1983) Biopolymers 22, 2577-2637