MAPAS: a tool for predicting membrane-contacting protein surfaces

To the editor: Many important biological processes, from serum phospholipid metabolism to amyloid disease, involve formation of protein-membrane complexes. Thus, tools for identifying membrane-contacting features in a protein structure are very important. However, few algorithmic approaches for membrane-contacting surface prediction have yet been reported[1,2].

We developed a program and web-based tool called MAPAS, or membrane-associated-proteins assessment (http://cancer-tools.sdsc.edu/MAPAS/pro2.html). MAPAS uses a set of algorithmic scoring functions to predict whether a given protein structure can form strong membrane contacts and to define the regions of the protein surface that most likely form such contacts (**Supplementary Methods** online). The MAPAS input window (**Supplementary Fig. 1** online) accepts Protein Data Bank (PDB) protein identifiers or a pasted file in pdb format.

The MAPAS algorithm is based on the assumption that membrane-contacting protein surfaces have a specific distribution of membranephilic surface residues in a plane. This planar region would contact the membrane (the explicit assumption is that, on the scale of proteins, the cell membrane can be considered as a plane). These residues must provide the necessary binding energy to keep the protein at the membrane surface. MAPAS (i) identifies the planar surfaces that encompass a given protein, and (ii) scores them according to their membranephilic properties. To provide a measure of membranephilicity, we estimated the relative tendency of individual residues to bind to a phospholipid bilayer. We calculated scoring functions using a semi-empiric approach based on steered molecular dynamics (**Supplementary Figs. 2–4** and **Supplementary Table 1** online) and Poisson-Boltzmann calculations (**Supplementary Methods**). MAPAS accepts a protein's three-dimensional structure as input and identifies all planes encompassing the protein structure (**Fig. 1a**) then calculates all residues that lie in the layer of a given thickness (**Supplementary Fig. 5** online). Then MAPAS sorts the planar protein surfaces based on their membranephilic character. The output window displays rotatable three-dimensional presentations of submitted proteins with their possible membrane–contacting surfaces indicated (see for example, **Supplementary Figs. 6** and **7** online).

We validated the performance of MAPAS with several known membrane-contacting proteins (**Fig. 1b** and **Supplementary Tables 2** and **3** online). MAPAS can predict membrane-contacting proteins, membrane-associated proteins and the membrane-contacting surfaces of proteins including transmembrane proteins (**Supplementary Discussion** online).

Nevertheless, as with all prediction programs, MAPAS can yield false positive and false negative predictions. One possible source of error is the fact that coordinates of proteins listed in PDB as membrane-contacting do not include the membrane–contacting regions, either because they are disordered or because they are engineered out of the protein to permit crystallization. Another problem is the relatively small area of membrane contact found in some proteins. Our tests show that MAPAS is reliable when the number of membrane-contacting residues is at least 5 (data not shown). With fewer residues in the membrane-contacting zone the statistical error increases.
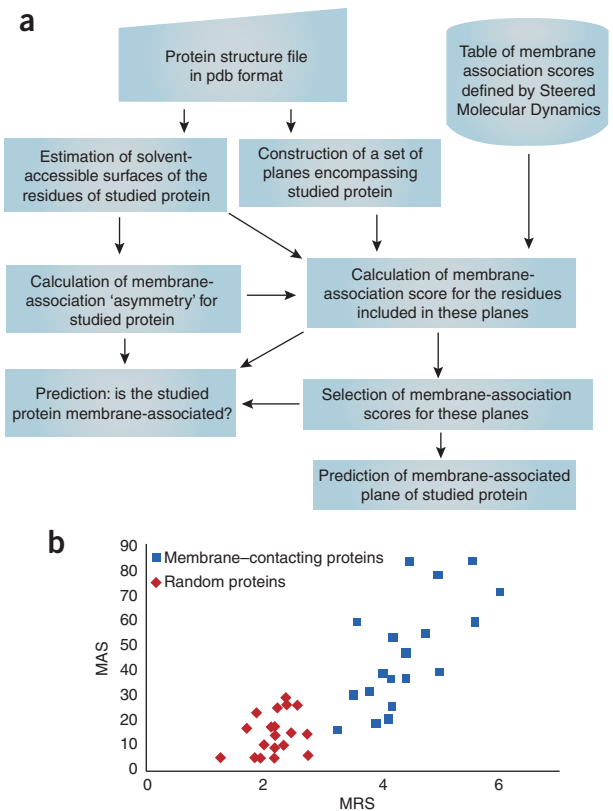
Figure 1 | The MAPAS algorithm workflow and performance. (**a**) The solvent-accessible surface of each residue of the entrance protein is calculated, a set of planes encompassing the entire protein is constructed, and then membrane-association asymmetry scores and membrane-contact scores for these planes are calculated using the table of membrane-association scores defined with semi-empirical method. Finally, membrane-associated proteins and the membrane-associated surfaces of these proteins are predicted. (**b**) Membranephilic area scores (MAS) and membranephilic residues scores (MRS) define membrane-contacting proteins. The majority of membrane-related proteins cluster differently than random non-membrane-contacting proteins. If a protein has MRS > 3 or MAS > 60%, there is a high probability that the protein has a true membrane-contacting region. Given the limitations of each scoring method, we suggest that users select proteins with high MAS and MRS, and then refine the predictions by considering K_{mpha}, the coefficient of 'membranephilic asymmetry' (see **Supplementary Methods** for definitions).

Note: Supplementary information is available on the Nature Methods website.

Yuriy Sharikov[1], Ross C Walker[1], Jerry Greenberg[1], Valentina Kouznetsova[2], Sanjay K Nigam[2,3], Mark A Miller[1], Eliezer Masliah[4,5] & Igor F Tsigelny[1,6]

[1]San Diego Supercomputer Center, [2]Department of Medicine, [3]Department of Pediatrics, [4]Department of Neuroscience, [5]Department of Pathology, [6]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.
e-mail: itsigeln@ucsd.edu

1. Lomise, A., Pogozheva, I., Lomize, M. & Mosberg, H. *Protein Sci.* **15**, 1318–1333 (2006).
2. Scott, D., Diez, G. & Goldmann, W. *Theor. Biol. Med. Model.* **3**, 17–20 (2006).