TMDET: Web server for detecting transmembrane regions of proteins by using their 3D coordinates

Article *in* Bioinformatics · April 2005

DOI: 10.1093/bioinformatics/bti121 · Source: PubMed

CITATIONS
96

READS
95

3 authors:

Gabor E Tusnady
Hungarian Academy of Sciences
74 PUBLICATIONS **5,070** CITATIONS
SEE PROFILE

Zsuzsanna Dosztányi
Eötvös Loránd University
98 PUBLICATIONS **4,873** CITATIONS
SEE PROFILE

István Simon
Hungarian Academy of Sciences
165 PUBLICATIONS **10,687** CITATIONS
SEE PROFILE

BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 7 2005, pages 1276–1277
doi:10.1093/bioinformatics/bti121

Structural bioinformatics

TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates

Gábor E. Tusnády, Zsuzsanna Dosztányi and István Simon*

Institute of Enzymology, BRC, Hungarian Academy of Sciences, H-1518 Budapest, PO Box 7, Hungary

Received on June 18, 2004; revised on September 27, 2004; accepted on October 22, 2004
Advance Access publication November 11, 2004

ABSTRACT

Summary: The structure of integral membrane proteins is determined in the absence of the lipid bilayer; consequently the membrane localization of the protein is usually not specified in the corresponding PDB file. Recently, we have developed a new method called TMDET which determines the most possible localization of the membrane relative to the protein structure, and gives the annotation of the membrane embedded parts of the sequence. The entire Protein Data Bank has been scanned by the new TMDET algorithm resulting in the database of structurally determined transmembrane proteins (PDB_TM). Here we present the web interface of the TMDET algorithm to allow scientists to determine the membrane localization of structural data prior to deposition or to analyze model structures.

Availability: The TMDET server is available for academic users on http://www.enzim.hu/TMDET

Contact: tusi@enzim.hu

Although structure determination of integral membrane proteins by X-ray diffraction methods or NMR is still regarded as a major achievement, one vital component, the membrane itself, is missing from the determined structures. During the preparation of crystal for X-ray study or dense aqueous solution for NMR study, these proteins are taken out from the lipid bilayer, and the membrane-exposed hydrophobic parts are covered by amphiphilic detergents, so that the protein–detergent complex can be treated similarly to soluble proteins (Ostermeier and Michel, 1997). The detergent molecules are highly unstructured and are usually not visible in the X-ray picture; moreover they do not replace the removed lipid molecules one-by-one. With the exception of a few tightly bound lipid molecules, the deposited experimental data have no direct indication that the protein is immersed into the membrane under native conditions, and do not contain information about the exact location of the lipid bilayer (Lee, 2003). Lipid molecules are also missing from theoretical structural models. Given the difficulties of experimental structure determination of membrane proteins, in many cases molecular simulations or homology modelling techniques are used to obtain initial structural information. In these cases it is also important to assign the membrane embedded region in order to test the effect of mutations, or to design potential drugs that bind to receptor molecules of pharmacuetical interest.

*To whom correspondence should be addressed.

Currently, we have developed a new method to determine the most possible membrane localization of transmembrane proteins by utilizing their atomic coordinates as an input (Tusnády *et al*., 2004). The developed new method, called TMDET, can distinguish transmembrane and globular proteins by their atomic coordinates as well as identify the transmembrane segments of membrane proteins. Therefore, TMDET results provide an assignment of the membrane localization on amino acid sequence, similar to the DSSP algorithm (Kabsch and Sander, 1983), which defines the secondary structure on the amino acid sequences. TMDET has been already used to scan the entire Protein Data Bank (PDB) (Berman *et al*., 2000) to select all transmembrane proteins, and the result is collected into the PDB_TM database (Tusnády *et al*., 2005).

The detailed description of the TMDET algorithm can be found in our previous article (Tusnády *et al*., 2004). Briefly, the input of the method are the atomic coordinates of the query protein. After checking the chain type and length of each chain the first step is the construction of the possible biological oligomer structure by building up the biomolecule, using the information of the BIOMOLECULE record and by eliminating chains which form non-biological contacts as a result of the crystallographic process. Next, the lipid exposed solvent accessible area is calculated (Lee and Richards, 1971). The core of the algorithm is a direct search for the most probable position of membrane planes relative to the given coordinates, by measuring the fitness of membrane localization via an objective function. The objective function measures the hydrophobicity combined with certain structural features of the peptide segment in a given slice of the protein. Different membrane orientations are generated by rotating a vector around the 4π direction and cutting the protein to 1 Å slices by planes perpendicular to the given vector.

On the web interface the main input data is the atomic coordinate file uploaded for the TMDET server by the user. The user can alter some parameters such as requesting the results by e-mail and switch for generating a graphical output showing the determined membrane localization of the user's protein. Processing the input data takes from a few seconds to a few minutes depending on the number of the atoms submitted. The most time-consuming step is the calculation of the membrane-exposed solvent accessible surface area. The progress of the calculation can be followed as the server generates output from each step during its run. After the calculation, all temporary files including the uploaded files are deleted. The results are presented

as an XML file, the document type definition of which is the same as the PDB_TM files (Tusnády *et al*., 2005). The output can be downloaded in raw XML format or by processing the data through a nice style sheet. The output contains the final value of the objective function, the transformation matrix, that rotates the molecule so that the normal vector coincides with the Z-axis and the new origin is along this axis at the membrane half width; moreover the output contains the sequential position of each identified transmembrane region, inside and outside loops. Because the inside and outside part of the protein cannot be distinguished using only its coordinates, the server uses, side1 and side2 notation in the output. If it is requested, some pictures are also generated in portable network graphics (png) format, using the Pymol molecular visualization software (DeLano, 2003).

The TMDET server targets users who wish to determine the transmembrane regions of protein structures, which have not been submitted to the PDB. This can aid the annotation of structural data prior to deposition. Furthermore, as the server is not limited to experimentally determined structures, mutant and/or theoretically designed protein models can be investigated as well, for example to test the possible effects of various mutations to the membrane localization of a protein.

ACKNOWLEDGEMENT

This work has been sponsored by grants OTKA T34131, D42207 and F043609. Zs.D. and G.E.T. were supported by the Bolyai Janos Scholarship.

REFERENCES

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

DeLano,W.L. (2003) *The PyMOL Molecular Graphich System*. DeLano Scientific LLC San Carlos, CA, USA. http://pymol.sourceforge.net.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

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

Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

Ostermeier,C. and Michel,H. (1997) Crystallization of membrane proteins. *Curr. Opin. Struct. Biol.*, **7**, 697–701.

Tusnády,G.E., Dosztányi, Zs. and Simon,I. (2004) Transmembrane proteins in Protein Data Bank: identification and classification. *Bioinformatics*, **20**, 2964–2972.

Tusnády,G.E., Dosztányi,Zs. and Simon,I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the Protein Data Bank. *Nucleic Acids Res.*, **33**, D275–D278.