Bioinformatics, 32(16), 2016, 2548–2550
doi: 10.1093/bioinformatics/btw208
Advance Access Publication Date: 19 April 2016
Applications Note

OXFORD

Structural bioinformatics

OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins

Guillaume Postic[1,2,3,4,*], Yassine Ghouzam[1,2,3,4] and Jean-Christophe Gelly[1,2,3,4,*]

[1]Inserm U1134, Paris, France, [2]Univ. Paris Diderot, Sorbonne Paris Cité, UMR_S 1134, Paris, France, [3]Institut National De La Transfusion Sanguine, Paris, France and [4]Laboratory of Excellence GR-Ex, Paris, France

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on January 27, 2016; revised on March 22, 2016; accepted on April 9, 2016

Abstract

Summary: The experimental determination of membrane protein orientation within the lipid bilayer is extremely challenging, such that computational methods are most often the only solution. Moreover, obtaining all-atom 3D structures of membrane proteins is also technically difficult, and many of the available data are either experimental low-resolution structures or theoretical models, whose structural quality needs to be evaluated. Here, to address these two crucial problems, we propose OREMPRO, a web server capable of both (i) positioning α-helical and β-sheet transmembrane domains in the lipid bilayer and (ii) assessing their structural quality. Most importantly, OREMPRO uses the sole alpha carbon coordinates, which makes it the only web server compatible with both high and low structural resolutions. Finally, OREMPRO is also interesting in its ability to process coarse-grained protein models, by using coordinates of backbone beads in place of alpha carbons.

Availability and Implementation: http://www.dsimb.inserm.fr/OREMPRO/

Contact: guillaume.postic@univ-paris-diderot.fr or jean-christophe.gelly@univ-paris-diderot.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Determining the orientation of membrane protein structures in the lipid bilayer is necessary to understand their function—with numerous medical and industrial applications. Membrane proteins are difficult to study, due to the laborious and expensive solubilization, purification and reconstitution methods they require. This is worsened by the need of working on multimeric complexes, since membrane proteins often function as such. Actually, the orientational disorder that exists in the lipid bilayer makes membrane positioning so difficult that data are available for only ∼20 proteins (Nugent and Jones, 2013), which most often makes computational methods the last resort (besides visual analysis). Moreover, the determination of an all-atom structure of membrane protein is already a challenge, and the relevance of positioning side-chains in environments that strongly differ from a biological lipid bilayer (i.e. detergents and solvents) is questionable. These difficulties result in an important number of 'low-resolution' structures (N-Cα-CO backbones and Cα-only), the quality of which needs to be evaluated. The same goes for the modelling of membrane protein structure, which is less advanced than for globular proteins, due to the complexity of the lipid environment, therefore making low-resolution and coarse-grained approaches widely used. Here, we present OREMPRO, the first web server aimed at orienting and evaluating membrane protein structures (alpha and beta) from the sole Cα coordinates, and thus compatible with both high and low structural resolutions. Our web server is also able to process coarse-grained structures, such as those used with the MARTINI force field (Marrink *et al.*, 2007), by using coordinates of backbone beads ('BB') in place of Cα.

2 Methods

2.1 Web interface

Our web server integrates the two MAIDEN (Postic *et al.*, 2015) and ANVIL (Postic *et al.*, 2016) algorithms, which are aimed at orienting and evaluating membrane protein structures, respectively, and have shown high accuracies in extensive and rigorous benchmarks. Through the user-friendly interface, which does not require any registration, users can upload a protein coordinate file in the PDB format, select options, and start the calculations. If no PDB file is submitted, OREMPRO can still be tested with an example. The resulting membrane assignment of the protein 3D structure is interactively displayed with the PV molecular viewer (Biasini, 2014). The server automatically adapts the visual representation of the structure, depending on the type of data submitted—e.g., 'spheres' for Cα-only and coarse-grained models. The output structure with its membrane can be downloaded as a PDB file. The results page also contains bilayer thickness, sequence positions, and tilt angles for each transmembrane segment assigned, and the MAIDEN quality score. The results are kept on the server and can be accessed later at the same address.

2.2 Membrane positioning

The positioning of the structure in the lipid bilayer is achieved by using the ANVIL algorithm, which follows an innovative approach whereby the membrane assignment problem is treated as a binary classification. The algorithm performs an orientational search of the best fitting membrane, trying to maximize a *C*-value that is similar to a Matthews Correlation Coefficient. In this web server version, two novel features are proposed, in addition to the options that were already present in the stand-alone version. First, we have implemented the possibility of using a geometrical criterion to improve the membrane assignment for certain types of protein structures (see Supplementary information). The other new option is the possibility to select the residue types which, in the binary classification, will be considered as having the tendency to be located either within the membrane (*M*) or at one of the two sides of the lipid bilayer (*S*). This new option has emerged as necessary, since all biological membranes do not have the same lipid composition and, therefore, might require different separations of the 20 residue types into the two *M* and *S* categories. Additionally, a weight can be defined for each residue type (positive defines *M*, negative defines *S*). For example, the Gly amino acids may be categorized as *M*, but with a weight of +0.1 rather than +1.0, in order to take into account the low hydrophobicity of this residue type. In this way, instead of using the default dichotomized membrane propensity scale of ANVIL—which is derived from the PM1D scale (Punta and Maritan, 2003)—users can test any of the existing hydrophobicity scales or even develop new ones. Finally, since the *C*-value does not have any cutoff to discriminate membrane proteins from globular ones, the algorithm will always try to assign a membrane, without regarding the type of protein submitted. This is particularly important, since it makes OREMPRO able to process structural models of medium and low qualities that may not be recognized as membrane proteins by other methods.

2.3 Structural assessment

Following the membrane assignment by ANVIL, a structural assessment is performed with MAIDEN, which is a distance-dependent statistical potential, analog of the DOPE score (Shen and Sali, 2006) but optimized on a dataset of membrane protein structures. For this web server version, new features that were not available on the stand-alone version of MAIDEN have been developed. For example, if one wants to calculate the MAIDEN pseudo-energy (or score) for another transmembrane domain than that delimited by ANVIL, then the web interface offers the possibility to manually define the lipid bilayer boundaries, by entering residue positions of each transmembrane segment. This new option allows calculating the MAIDEN score from membrane assignments produced by other positioning algorithms than ANVIL, such as OPM (Lomize *et al.*, 2012) or TMDET (Tusnády *et al.*, 2005). Another new feature is the possibility to download a detailed list (in text format) of the pseudo-energies for all the residue pairs in the query structure.

The MAIDEN score measures the relative quality of protein structures, which means that it can be used to rank several models by their quality: the lower the pseudo-energy, the better the model. However, to have an idea of the absolute quality of a single structure, we have added to the MAIDEN score a color code that is derived from a Z-score calculation of the pseudo-energy. This Z-score is calculated by $(\bar{u}_s - \mu_r)/\sigma_r$, where $\bar{u}_s$ is the pseudo-energy of the assessed structure, μ_r and σ_r are the mean and standard deviation of the pseudo-energy distribution of random sequence decoys. Each of these decoys is generated by randomly permuting the primary sequence of the protein (to <50% sequence identity) while keeping the coordinates of the side chain centers unchanged (Melo *et al.*, 2002), and therefore has a very high (positive) pseudo-energy. To ease interpretation, the Z-score is converted into a color code, with the red color corresponding to a Z-score ≥0, which means that the evaluated structure is not distinguishable from random decoys by the pseudo-energy.

3 Conclusion

Thanks to the use of a minimalistic representation of protein structures, OREMPRO is the first web server capable of processing low-resolution and coarse-grained models of membrane proteins, thus meeting the need for treating the growing number of these structures. Our web server offers a complementary combination of membrane assignment and structural quality assessment, by integrating the innovative ANVIL membrane positioning algorithm and the MAIDEN statistical potential into a user-friendly interface. Despite its dual function, OREMPRO is fast and processes most protein structures within a minute.

Funding

This work has been supported by grants from Inserm, Univ. Paris Diderot, Institut National de la Transfusion Sanguine and Laboratory of Excellence GR-Ex.

Conflict of Interest: none declared.

References

Biasini,M. (2014) PV—WebGL-based protein viewer. *Zenodo*, doi:10.5281/zenodo.12620.

Lomize,M.A. *et al.* (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, **40**, D370–D376.

Marrink,S.J. *et al.* (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, **111**, 7812–7824.

Melo,F. *et al.* (2002) Statistical potentials for fold assessment. *Protein Sci.*, **11**, 430–448.

Nugent,T. and Jones,D.T. (2013) Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics*, **14**, 276.

Postic,G. *et al.* (2015) An empirical energy function for structural assessment of protein transmembrane domains. *Biochimie*, **115**, 155–161.

Postic,G. *et al*. (2016) Membrane positioning for high- and low-resolution protein structures through a binary classification approach. *Protein Eng. Design Select*., **29**, 87–92.

Punta,M. and Maritan,A. (2003) A knowledge-based scale for amino acid membrane propensity. *Proteins*, **50**, 114–121.

Shen,M. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci*., **15**, 2507–2524.

Tusnády,G.E. *et al*. (2005) TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*, **21**, 1276–1277.