A METHOD TO ASSESS CORRECT/MISFOLDED STRUCTURES OF TRANSMEMBRANE DOMAINS OF MEMBRANE PROTEINS

Chugunov A.O.1, 2, Novoseletsky V.N.1, 3, Efremov R.G.1

1 Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, GSP Moscow, 117997, Russia; 2 Department of Bioengineering, Biological Faculty, M.V. Lomonosov Moscow State University, 119899, Moscow; 3 Moscow Institute of Physics and Technology (State University), 141700, Russia

* Corresponding author: e-mail: volster@nmr.ru

Key words: membrane proteins, transmembrane domain, scoring function, residue environment, G-protein coupled receptors, rhodopsin

SUMMARY

Motivation: Integral membrane proteins (MP) are pharmaceutical targets of exceptional importance since more than 50 % of currently marketed drugs target these objects. Due to technical difficulties, modern experimental methods often fail to determine 3D structure of MPs. Computational methods for modeling MPs structure and assessment of these models’ quality may be very helpful in this case.

Results: We propose a novel method for quantitative estimation of the transmembrane (TM) domains models’ quality. The approach is based on the concept of environmental profile. A non-redundant set of 26 high-resolution X-ray structures of α-helical TM domains is used to define five classes of residues’ environment, considering polarity of nearest protein surrounding and accessibility for a given residue. Residues’ preferences for each environment class are calculated. The main results are: (1) The proteins length correlates with the proposed scoring function values, defining a way to differentiate “well-folded” structures from misfolded ones; (2) The method efficiently delineates crystallographic structure of visual rhodopsin both in a set of twelve its computer models, containing certain errors and ensemble of artificially generated misfolded structures of rhodopsin; (3) Photosynthetic MPs demonstrate different score-length dependency, suggesting distinct packing characteristics for these proteins.

INTRODUCTION

Integral membrane proteins (MP) are objects of special biological and pharmaceutical importance, establishing every cell’s communication with the rest of the world, including signal transduction, light absorption and formation of TM potential. A very large and important class of MPs, G-protein coupled receptors (GPCRs), is a target for > 50 % of currently marketed drugs. Unfortunately, possibilities of modern experimental techniques for MPs 3D structure determination are far under pharmaceutical industry (e.g., structure-based drug design) requirements. Current proportion of MPs in Protein Data Bank (PDB) is less than 1 %, whereas every sequenced to date genome encodes 15–30 % of membrane proteins. To overcome this discrepancy, development of computational methods for modeling MPs structure and assessment of these models’ quality is believed to be very helpful.
Many efforts have been made to understand the principles of structural organization of MPs, but the problem is yet to be solved. What are the differences in their packing and structure as compared to soluble proteins? Some methods designed for MPs structure prediction utilize sequence statistics or more general characteristics (e.g. protein packing density), but only very few of them use high-resolution structural data on residues’ environments.

**METHODS AND ALGORITHMS**

**Creation of membrane proteins database.** We used MPs structures with primarily \( \alpha \)-helical TM domains, determined by high-resolution (< 3.5 Å) X-ray crystallography. The training set contains 26 structures of proteins that have no sequence homology to each other. A separate set contains 6 structures of photosynthetic proteins. All protein structures were aligned along the membrane normal (hereinafter, \( Z \) axis) in order to place cytoplasmic sides of plasma MPs and matrix sides of inner mitochondria MPs to \( Z < 0 \) area, and vice versa. Optimal \( Z \) position of the entire structure and the thickness of “TM” hydrophobic layer were determined by finding solvation energy minimum using implicit membrane-water environment model (Efremov et al., 2000). Only \( \alpha \)-helical residues (as determined by DSSP (Kabsch, Sander, 1983)) within “optimal” hydrophobic layer (27 Å in average) plus 5 Å at each side were selected for the study.

**“Membrane Score” calculations.** In order to characterize the environment of a particular residue, we used fractions of full residue’s surface that face polar and non-polar atoms of other TM \( \alpha \)-helices, respectively:

\[
F_{\text{p}} = F_{\text{p}} - F_{\text{p}}^0 = \frac{S_{\text{p}}}{S^0} - \frac{S_{\text{p}}^0}{S^0}, \quad F_{\text{n}} = F_{\text{n}} - F_{\text{n}}^0 = \frac{S_{\text{n}}}{S^0} - \frac{S_{\text{n}}^0}{S^0},
\]

where \( F_{\text{p}} \) and \( F_{\text{n}} \) are areas of polar (Sp) or non-polar (Sn) contacts, divided by the residue “self” area in Gly-Res-Gly motif. \( F_{\text{p}}^0 \) and \( F_{\text{n}}^0 \) are the corresponding values for isolated TM helices. The difference between them is a measure of interhelical effects. Given the \( F_{\text{p}} \times F_{\text{n}} \) distributions for \( i \) residue (Fig. 1a) and membrane environments scheme with parameters \( a, b \) and \( tga \) (Fig. 1b), we define a membrane scoring function:

\[
\text{MemScore}_i = \ln\left(\frac{P_i}{P_j}\right),
\]

\[
\text{TotalMemScore} = \sum_j N_j \ln\left(\frac{P_i}{P_j}\right),
\]

where \( P_{ij} \) is a probability to find residue \( i \) in environment \( j \), and \( P_j \) is a probability to find any residue in environment \( j \) (j occupancy). The total score was calculated as in (3), where \( N_j \) is a number of residues \( i \) in class \( j \). Corresponding value in (2) was never zero and was set to unity if there were no \( i \) residues in \( j \) class, but if it was zero in (3), the entire term was not considered. Exact \( a, b \) and \( tga \) values were determined in order to maximize the total score value (3) for the whole training set.

**Rotameric test.** We generated an ensemble of conformations of visual rhodopsin, where every TM \( \alpha \)-helix was rotated around its axis with increment 90°, resulting in 16384 \((4^4)\) rotameric conformations. A simple energy minimization was applied to avoid sterical clashes.
RESULTS AND DISCUSSION

To establish a method for assessment of MPs packing quality, we introduced two residues’ environment characteristics, namely fractions of full residue surface that are in contact with polar and non-polar atoms of other TM α-helices, Fp\(^1\) and Fnp\(^1\), respectively (see Methods for details). For the whole training database (see Methods) we obtained distributions of these parameters for each residue type, as shown in Fig. 1a.

![Figure 1](image)

**Figure 1.** Residues’ environmental characteristics distributions derived from “training” database for Arginine and Leucine (a). Black circles correspond to the “central” residue location (|Z| < 15 Å), and gray – to the interfacial. Proposed scheme for membrane environment classes (b). Class 1 corresponds to exposed one, classes 2 and 3 – to intermediately buried, 4 and 5 – to buried. Classes 2 and 4 correspond to non-polar environment, whereas 3 and 5 – to polar.

As one can see, there are much more leucine residues in TM domains than arginines. This demonstrates strong preference of these residues to “central” and to interfacial locations, respectively. Also, most of leucines are situated in non-polar environment (close to “non-polar” Fnp\(^1\) axis) and arginines – in polar. In Fp\(^1\) × Fnp\(^1\) coordinates the proximity to zero means high accessibility for membrane milieu (high values of accessible solvent area, ASA), whereas location near the Fp\(^1\) = 1 – Fnp\(^1\) line means considerable burial (ASA ≈ 0). Based on these observations, we propose the following scheme for definition of environmental classes for TM α-helical domains of proteins (Fig. 1b). The scores were calculated for each combination of residue type and membrane class (not shown), enabling assessment of the quality of the whole structures.

In Fig. 2 the membrane scores for proteins from the training set are plotted against TM domain length. It is seen that, there is a good correlation between them. This enables differentiation between correct (e.g. crystal) and misfolded structures. To test the possibility, we chose from public domain 12 computer models of bovine visual rhodopsin, built prior to release of its crystal structure (Palczewski et al., 2000), and compared them in terms of membrane score values. As seen in Fig. 2, all of them lie below the crystal structure, and those that have been built in a fully automatic manner (e.g., at Swiss-Prot, MODBASE, GPCRDB servers), score much lower than the carefully optimized ones (data not shown). Also, a notable correlation exists between model’s deviation from the crystal structure (in terms of r.m.s.d.) and score impairment (not shown). It was noticed, that photosynthetic proteins demonstrate different score-length dependency, suggesting distinct packing characteristics for them.
In order to further validate our method’s possibility to distinguish correct and misfolded structures, we generated more than 16000 rhodopsin’s rotameric conformations (see Methods), vast majority of which are believed to be misfolded. For this ensemble, we compared the ability of our method to rank the crystal structure among it, with the results obtained using the well-known Eisenberg’s method (Bowie et al., 1991), conceptually close to ours, but parameterized for globular proteins. As seen from Fig. 3, “classical” method, very good for soluble proteins, is unable to mark out the crystal structure, whereas the present membrane-tuned method performs rather well.

To conclude, we have developed a novel method to estimate the packing quality of TM α-helical domains in proteins. We suppose that this method will be especially useful for GPCRs’ models construction and optimization.
ACKNOWLEDGEMENTS

This work was supported by the Programme RAS MCB, by the Russian Foundation for Basic Research (grants 02-04-48882-a, 05-04-49346-a, 06-04-49194-a), by the Russian Federation Federal Agency for Science and Innovations (The State contract 02.467.11.3003 of 20.04.2005, grant SS-4728.2006.4).

REFERENCES


