ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 3, pp. 293-300. © Pleiades Publishing, Ltd., 2007.
Published in Russian in Biokhimiya, 2007, Vol. 72, No. 3, pp. 358-367.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM06-279, January 21, 2007.

A Novel Method for Packing Quality Assessment of Transmembrane α-Helical Domains in Proteins

A. O. Chugunov[1,2]*, V. N. Novoseletsky[1,3], A. S. Arseniev[1], and R. G. Efremov[1]

[1]*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (495) 336-2000; E-mail: batch2k@yandex.ru*
[2]*Department of Bioengineering, Biological Faculty, Lomonosov Moscow State University, 119899 Moscow, Russia*
[3]*Moscow Institute of Physics and Technology (State University), Institutskii Pereulok 9, 141700 Dolgoprudnyi, Moscow Region, Russia*

Received November 2, 2006
Revision received November 10, 2006

Abstract—Here we present a novel method for assessment of packing quality for transmembrane (TM) domains of α-helical membrane proteins (MPs), based on analysis of available high-resolution experimental structures of MPs. The presented concept of protein-membrane environment classes permits quantitative description of packing characteristics in terms of membrane accessibility and polarity of the nearest protein groups. We demonstrate that the method allows identification of native-like conformations among the large set of theoretical MP models. The developed "membrane scoring function" will be of use for optimization of TM domain packing in theoretical models of MPs, first of all G-protein coupled receptors.

DOI: 10.1134/S0006297907030066

Key words: integral membrane proteins, spatial structure prediction, protein environment, GPCR, visual rhodopsin

Integral membrane proteins (MPs) are of exceptional biological importance; they are known to mediate transmembrane (TM) signal transduction, light absorption, generation of TM potential, etc. For example, G-protein coupled receptors (GPCR), the broadest and most important class of MPs, are targets for more than 50% of all currently marketed drugs [1] since their malfunction is connected to a number of diseases [2]. The functioning of MPs depends primarily on the TM domain, which often binds a ligand and accommodates conformational reorganization initiating intracellular response. Information on the structure and functioning of TM domains is urgently required for pharmaceutical applications, such as structure-based drug design. Modern experimental techniques of three-dimensional (3D) structure determination, such as X-ray crystallography and NMR spectroscopy, on the other hand, often fail to solve the problem due to technical difficulties related to protein purification and crystallization [3]. Only a few tens of MP structures have been determined to this day, making up <1% of the total number of known structures in the Protein Data Bank (PDB) [4], although the genomes sequenced so far encode at least 15-30% of MPs [5].

Molecular modeling, however, could be useful for prediction of the 3D structure of membrane-spanning proteins, complementarily to experimental methods. Since most TM domains of MPs are formed of α-helices (they and β-barrels are the only two folds discovered in MPs to date), and therefore, given that exactly this class is pharmacologically important, we will further focus on α-helical TM proteins. Most of the modeling protocols include the following stages: (i) TM segments are identified in the protein's amino acid sequence; (ii) optimal mutual arrangement of the helices is predicted; (iii) specific structural/functional features, e.g. kinks and other deviations from the ideal α-helicity, are delineated; (iv) loop regions are reconstructed. In this protocol, stage (ii) poses the greatest challenge—even in the case of the simplest helix–helix systems the computational procedure is not straightforward and requires exhaustive sampling of the peptides' conformational space in a heterogeneous membrane-like environment [6].

Modeling approaches used to build such models can be divided into two groups: those starting from the "first

Abbreviations: GPCR) G-protein coupled receptors; MP) membrane protein; PhP) photosynthetic protein; RMSD) root-mean square deviation; S^{mem}) "membrane score"; TM) transmembrane.

* To whom correspondence should be addressed.

principles" (so-called *ab initio* approaches) and those utilizing related proteins' structural information. The first group is based only on *a priori* knowledge of TM topology of a protein and predicts its structure taking into account the physicochemical nature of the membrane and various empiric potentials that describe its properties [7-10]. Methods of the latter group are based on homology with proteins with known experimental spatial structure [11], assuming that 3D structure of a protein is more conservative than its sequence [12]. In the case of GPCRs, however, the choice of a template for homology modeling is restricted to a single structure available to date, namely, that of the bovine visual rhodopsin [13]. Note that this protein reveals only limited homology level with other family members. Given the functional differences between modeling target and template proteins, resulting models should be carefully optimized taking into account their biological specificity. To have this done (for example, by delicate tuning of mutual arrangement of helices), all available experimental data (such as information on functionally important residues and distances between them [14], probable mechanism of protein function, and general packing principles observed in MPs [15, 16]) have to be considered.

For an already built model, one should have a possibility to assess its packing "quality" — compliance with similar characteristics known for high-resolution MPs' structures. Such assessment may guide further optimization of the models and be of use for elimination of gross errors made during modeling. There have been a number of attempts made to optimize theoretical models with utilization of such characteristic features of MPs as distributions of hydrophobic [17-19] and variability [17, 20-22] properties of TM segments, as well as the preferences of residues for membrane-interacting [23] or interhelical interface-forming [24-28] positions. These regularities are undoubtedly important for understanding MP organization principles, though availability of experimental high-resolution spatial structures of MPs permits more delicate structural examination of features by exploration of microscopic environments of individual residues. Solitary studies of this organization have already been described [29-31]; however there is still no packing quality assessment method for MPs, e.g. analogous to well-known ones designed for globular proteins [32, 33], and based on the analysis of individual microenvironments of residues in high-resolution structures. The necessity of separate methods, optimized or constructed especially for MPs, is imposed by major differences in physicochemical properties of surrounding medium for globular and membrane proteins, and therefore considerable dissimilarities in their spatial organization [17, 34, 35].

In this work we present a novel method for assessment of packing quality of TM domains of α-helical MPs based on the analysis of packing properties of residues observed in a non-redundant "training" set of 26 MP structures that have been determined to a high experimental resolution. This analysis allowed computation of preferences of residues to five predefined classes of membrane-protein environment, characterizing the residue by the degree of membrane exposure and the polarity of proximal protein groups. No proteins with TM domains abounded by a high number of cofactors, such as photosynthetic proteins, were used in parameterization of the method. We discovered a linear dependence of the "membrane score" function on TM domain sequence length, which permits comparison of the "quality" of theoretical models with those of experimental structures of comparable size. Moreover, the method allows identification of "close-to-native" structures among sets of models including those with serious structural errors. The presented approach will be applied to optimization of theoretical models of TM domains, first of all for GPCR modeling.

METHODS OF INVESTIGATION

MP structures and rhodopsin theoretical models. The initial list of MPs with known structure was obtained from site "Membrane proteins with known structure" (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Low-resolution (>3.5 Å) and NMR-derived structures, as well as β-barrel and closely related proteins were removed from it. Only the structure determined at highest resolution in a family of homologous proteins was taken for further inquiry. In total, a non-redundant training set of 26 α-helical MPs was established, and corresponding structures were taken from the PDB database [36]. Six structures of photosynthetic proteins comprised a separate set—couples of structures of light-harvesting complexes (PDB IDs: 1NKZ and 1LGH), photosynthetic reaction centers (PDB IDs: 1PRC and 1OGV), and photosystems (PDB IDs: 1JBO and 2AXT).

All proteins were spatially aligned in such a way that the principal moments of inertia of their TM domains were collinear to the *Z*-axis (corresponding to the membrane normal). Negative *Z*-values corresponded to the cytoplasmic side of the cell membrane or matrix side of the inner mitochondria membrane. Then, the structures were "scanned" along *Z*-axis with a virtual "hydrophobic slab", mimicking the membrane, and the "membrane"–protein interaction energy (E_{solv}) was calculated at every step. The "hydrophobic slab" was defined by two *XY* planes, separated with a distance *D* (varying in the range 20-40 Å), corresponding to the membrane thickness. Hydrophobic properties of the membrane were described by an empirical function E_{solv}, in the framework of the atomic solvation parameters (ASP) formalism:

$$E_{solv} = \sum_{i=1}^{N} \Delta\sigma_i \mathrm{ASA}_i, \quad (1)$$

where ASA_i and $\Delta\sigma_i$ are the solvent accessible surface area and ASP for atom i, respectively, and N is the number of atoms. Hereinafter, ASA means the area exposed to surroundings. Depending on the Z coordinate of atom i, the values of ASPs were taken to approximate cyclohexane and water inside ($|Z| \le D/2$) and outside the slab, respectively [37, 38]. Z and D values, corresponding to minimum of E_{solv}, were defined as "optimal" position and membrane thickness, respectively. Then, the PDB file coordinates origin was transferred into the middle of the "membrane". The database for parameterization of the "membrane score" method was constructed from residues in α-helical segments (as determined by the DSSP software [39]) inside the "optimal" hydrophobic layer, constricted by 5 Å at each side in order to avoid boundary effects ($|Z| \le D/2 -$ 5 Å). The final database (the "training set") consisted of 4991 residues in 217 TM α-helical segments.

Theoretical models of visual rhodopsin, which were used for validation of the "membrane score" method, were either collected from the public domain (from the GPCRDB server [40] or other sites) or specially built for this purpose using common homology modeling tools. All theoretical models were constructed *prior to* release of the crystal structure of rhodopsin [13] and contained inaccuracies or errors. In the case of specially built models, such errors as misalignments and others were advisedly introduced or not corrected.

Environmental parameters and "membrane score" calculation. Environments of residues in the database are characterized by two parameters, F_p^1 and F_{np}^1, corresponding to fractions of the side chain surface area covered by polar (nitrogen or oxygen) and non-polar atoms of neighboring helices in the TM domain, respectively:

$$F_p^1 = \frac{S_p - S_p^0}{S^0}, \qquad F_{np}^1 = \frac{S_{np} - S_{np}^0}{S^0}, \tag{2}$$

where S_p and S_{np} are the total areas of polar and non-polar contacts, respectively; S_p^0 and S_{np}^0 are the areas of polar and non-polar contacts within the same helical segment, respectively, and S^0 is the surface accessible area of a given residue (X) in Gly-X-Gly tripeptide [41] (Scheme).

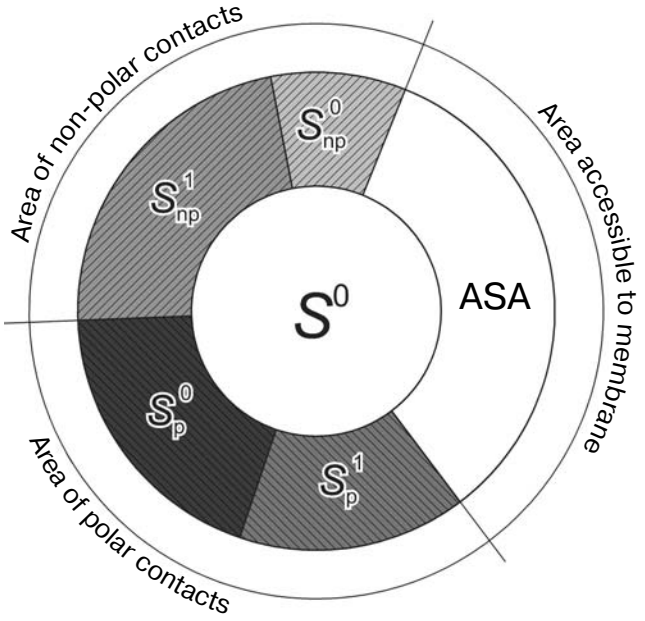
According to F_p^1 and F_{np}^1 values every residue was attributed to one of five predefined membrane-protein environment classes defined by three parameters: a, b, and tanα (Fig. 1b). "Compatibility" of each of 20 residue types with every environment class was calculated based upon frequencies of occurrence of residues in these classes:

$$S_{ij}^{mem} = \ln(P_{ij} / P_j), \tag{3}$$

where P_{ij} is the probability of finding residue i in class j, and P_j is the probability of finding *any* residue in class j ("occupancy" of the class). Boundaries of the classes (Fig. 1b) were systematically varied with simultaneous recalculation of S_{ij}^{mem} at each step in order to maximize the total membrane score value over the whole database (the "quality" of the training set should be maximal):

$$S_{total}^{mem} = \sum_{ij} N_{ij} S_{ij}^{mem}, \tag{4}$$

where N_{ij} is the number of residues of type i in class j. If there were no i residues in class j, N_{ij} in Eq. (3) was set as unity to avoid critical points in the logarithmic function, but the whole term in Eq. (4) was not accounted in the total sum. Final S_{ij}^{mem} values for "optimal" border parameters $a = 0.7$, $b = 0.05$, and $\tan\alpha = 2$ are given in the table.



Scheme of surface area division of residues. The total residue surface area (S^0) is subdivided into the following parts: ASA is the surface area accessible to membrane; S_p and S_{np} are the areas covered by polar and non-polar protein atoms, respectively. The two latter are further subdivided into S_p^0 and S_{np}^0 (the contact areas with residues of the same TM segment) and S_p^1 and S_{np}^1 (the contact areas with other helices)

"Rotameric test". Model "rotamers" were created by systematic independent rotation of all seven TM α-helices as rigid bodies around their axes with 90° increment. If a chemical bond of either residue penetrated cyclic structure of another residue, the χ_1 dihedral angle of the latter was varied until this error is gone. Then the structure was subjected to energy minimization in GROMACS 3.14 [42] to avoid rough atomic clashes. In total, $(360°/90°)^7 = 16{,}384$ rotameric conformations were constructed, and only one of them ("zero rotamer") corresponded to the crystal structure (nevertheless, it was energy-minimized in the same manner).

Preferences of residues for environment classes (for definition of classes, see Fig. 1b)

Residue	Environment class				
	1	*2*	*3*	*4*	*5*
Ala	0.09	−0.33	−0.10	0.29	0.64
Arg	−0.04	−0.94	0.51	−1.20	0.82
Asn	0.21	−1.05	0.56	−1.13	0.60
Asp	−0.04	−1.12	0.86	−1.67	−0.09
Cys	−0.42	−0.23	0.25	0.71	−0.41
Gln	−0.22	−1.02	0.66	−0.18	0.38
Glu	−0.56	−0.77	0.57	−0.34	0.63
Gly	−0.16	−0.32	−0.15	0.36	0.79
His	−0.72	−0.55	0.90	−0.97	−1.87
Ile	0.27	0.40	−0.65	−0.19	−2.29
Leu	0.10	0.34	−0.53	0.01	−0.97
Lys	0.48	−0.78	0.64	−2.02	−0.15
Met	−0.21	0.44	−0.74	0.19	−1.27
Phe	−0.20	0.25	−0.28	0.26	−0.64
Pro	−0.36	−0.38	0.50	−0.87	0.35
Ser	−0.25	−0.51	0.54	−0.37	0.23
Thr	−0.19	−0.19	0.22	−0.11	0.30
Trp	−0.31	0.11	0.10	−0.11	−0.32
Tyr	−0.58	−0.40	0.44	−0.21	0.46
Val	0.28	0.10	−0.12	0.03	−0.81

RESULTS AND DISCUSSION

Development of the method using structural data. There are now more than 200 experimentally-solved spatial structures of MPs. However, this list is highly redundant, and after omitting low-resolution and NMR-derived structures, β-barrel proteins, and closely homologous proteins, the training set of 26 α-helical membrane proteins was established. A separate set was composed of six structures of photosynthetic proteins (see "Methods of Investigation"). Development of a scoring function for TM α-helical domains requires accurate delineation of the "optimal" boundaries of membrane-spanning segments in each protein structure. These boundaries were determined by "scanning" of a protein model, aligned along the membrane normal, by a "virtual hydrophobic slab" of varying thickness. The latter imitates the hydrophobic membrane environment, specific for MPs (see "Methods of Investigation"). A single "solvation energy" minimum was detected for every structure. It corresponds to an optimal position of the protein relative to the "membrane" center and an optimal "membrane" thickness. A database used for parameterization of the method was comprised of residues that belong to α-helical segments in the "hydrophobic layer" (in total, 4991 residues in 217 TM helices). Spatial distributions of residues along the membrane normal, hydrophobic properties of their microenvironments, and preferences for membrane-exposed and buried positions turned out to be in a very good agreement with data reported earlier [29, 30, 43] (data not shown).

To characterize environments of residues and to define environment classes, we chose parameters F_p^1 and F_{np}^1, which represent the fractions of side chain area forming, respectively, polar and non-polar contacts with *other* TM helices in MPs (see "Methods of Investigation"). Examples of distributions $F_p^1 \times F_{np}^1$ for leucine and arginine residues in the training set, as well as the scheme defining environment classes, are given in Fig. 1.

As seen in Fig. 1a, distributions for leucines and arginines differ significantly. Arginines are concentrated near the membrane boundaries, while almost all of the leucines are found in the hydrophobic membrane core. Moreover, leucines mainly occur in either membrane-exposed (class *1*; see definitions of environment classes in Fig. 1b) or partly buried non-polar surroundings (class *2*), while arginines prefer more polar positions (classes *3* and *5*). Arginines situated in the central membrane regions are usually buried to avoid unfavorable contacts with hydrophobic acyl chains of lipids. Finally, interfacial

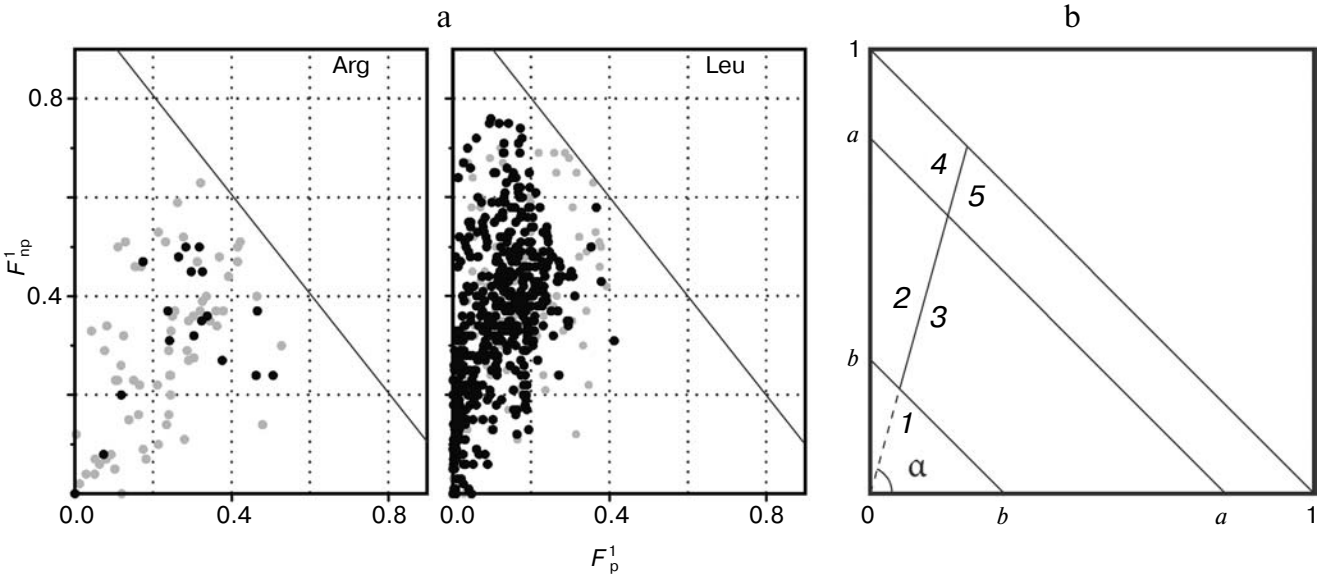


Fig. 1. Classes of membrane-protein environments for MP residues. a) Distributions of parameters characterizing arginine and leucine residues from the database of TM α-helical domains (F^1_p and F^1_{np} definitions are given in the text and in "Methods of Investigation"). Every data point corresponds to a single residue in the database. Gray and black dots denote residues on the interface ($|Z| \geq 15$ Å) and in the membrane core ($|Z| < 15$ Å), respectively. b) Scheme for environment classes based upon parameter distribution from (a). Class *1* corresponds to residues preferentially exposed to the membrane (proximity to the zero of both F^1_p and F^1_{np} parameters means that almost all area is accessible to the membrane; see Scheme), classes *2*, *3* and *4*, *5* are composed of residues partly and completely buried inside the bundle, respectively. Classes *2* and *4* denote non-polar, while classes *3* and *5* denote polar environment.

leucine residues have a higher affinity to polar environments than the "central" ones. To quantitatively characterize these preferences, the "membrane score" function was introduced, which describes propensities of a residue to occur in a certain environment class in the training data set (see "Methods of Investigation"). Positive membrane score values indicate favorable environment for this residue, and negative a disadvantageous one (to be more precise, that this residue is placed to the current environment rarer than any residue on average). Membrane scores for each pair of residues and environment classes are listed in the table.

The proposed method is conceptually close to the well-known 3D-profiles approach developed by Eisenberg et al. [32], but there are important differences. First, apart from the 3D-profiles method that was parameterized on a set of 16 structures of *globular* proteins, the "membrane score" approach is designed specially for α-helical MPs and considers their structural peculiarities. Second, it takes into account only one type of secondary structure, i.e. α-helices, since we concentrated on this MP subclass, and irregular structure is very unfavorable in the membrane environment. And, finally, the proposed scheme of environment classes significantly differs from that used in [32].

Validation of the method. Assuming that almost all residues in proteins are located in favorable environments, one could expect that the total membrane score value (that equals a sum of all residual values (S^{mem})) will linearly depend on protein sequence length (L). Figure 2a shows this dependence for proteins of the training set, six structures of photosynthetic proteins (PhPs) (that were not used upon parameterization), and a set of theoretical models of rhodopsin, built *prior to* release of its crystal structure (here they are used to assess the method's ability to distinguish between "well-packed" from misfolded structures). It is seen that the data points corresponding to the training set proteins fit well a regression line, confirming the assumption. If the method of parameterization was done using only some of the proteins from the training set, utilizing another part as a test set gave a very similar membrane score dependence on protein length, membrane score values, and classes borders (Fig. 1b). Data points, corresponding to test set proteins fitted well a regression line for the training set (data not shown). All of the preceding confirms that the proposed method is robust and reliable, and the data set used is rather representative and does not lead to over-determination of the method.

At the same time, as seen from Fig. 2a, there is a distinct dependence of S^{mem} on L for photosynthetic proteins (PhPs). This effect does not depend on training set (data not shown). Evidently, this is caused by considerable distinctions in PhPs and other MP packing characteristics due to high abundance of their (the PhPs') TM domains in light-accepting photosynthetic pigments. Probably the latter change inter-helical interactions characteristic for other MPs and/or partially substitute them with protein–pigment interactions. Probably, individual

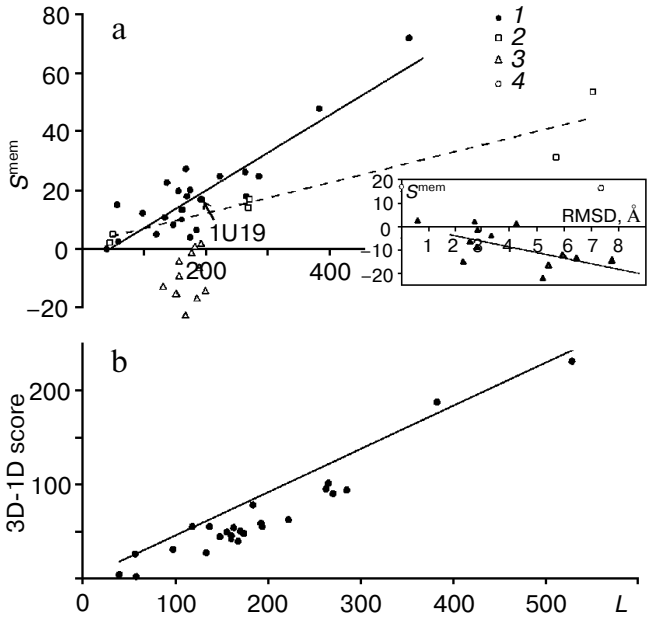


Fig. 2. Validation of the "membrane score" (S^{mem}) method. a) S^{mem} values vs. TM domain length (L) for the training set (*1*), photosynthetic proteins (*2*), and computer models of rhodopsin (*3*). The solid and dashed lines represent the least squares fits obtained for the training set and photosynthetic proteins, respectively. The position of rhodopsin crystal structure (PDB ID: 1U19) is shown with an arrow. Inset: dependence of the values S^{mem} on the root-mean-square deviation from the crystal structure (*4*) for theoretical models of rhodopsin. The regression trend-line is given. b) Values of 3D-1D score [32] for the final set of membrane proteins. The solid line shows the "ideal" score value for a protein with a given length (L) [46]: $\text{3D-Score}_{ideal} = \exp(-0.83 + 1.008 \cdot \ln(L))$.

parameterization of the "membrane score" method for PhPs is needed.

The "membrane score" method was also applied to assess a set of theoretical models of visual rhodopsin (Fig. 2a). These models were either constructed before the crystallographic structure of rhodopsin became available or built using common homology modeling software. Such estimations were specially designed to validate the "membrane score" approach. The proposed method permits unambiguous identification of the crystallographic ("correct") rhodopsin structure—its score value is significantly higher than for other models. Moreover, there is a prominent relation between closeness of a model to the experimental structure (in RMSD (root-mean square deviation) terms; see Fig. 2a, inset) and the membrane score. Generally, the models, carefully optimized by their authors, score better, and *post factum* (after release of the rhodopsin crystal structure) reveal lower RMSD values. On the contrary, the models that were built in an automated manner (*via* available web-servers Swiss-Model [44], ModBase [45], or GPCRDB [40]) and often containing improper sequence-alignment or other modeling errors, have high RMSD values and poor scores.

Identification of "native" (or "close-to-native") structures in voluminous sets of theoretical models is one of the crucial tasks in molecular modeling. In fact, having this task solved and being able to effectively sample conformational space of a protein of interest, one could resolve the whole problem of structure prediction. Unlike the "membrane score" method, the 3D-1D algorithm mishandles this problem in the case of MPs. For instance, it does not distinguish crystal structure from computer models—3D-1D Score in the former case was either not maximal, or the gap with the next model was small and not steady. More importantly, since the 3D-1D method is not optimized for MPs, the correlation between RMSD and 3D-Score value gets lost. As seen in Fig. 2b, MPs systematically score worse than globular proteins of the same sequence length. Although they do not fall into "misfolded structures" zone (which is characterized by 3D-Score values lower than $0.45 \cdot \text{3D-Score}_{ideal}$ [46]; see Fig. 2b legend), this difference gives evidence for distinctions in organization of membrane and globular proteins.

This systematic impairment of 3D-1D Score is not the only consequence of its application to MP packing quality assessment. As has been already said, the main limitation in this case is the loss of ability to discriminate between "correct" and misfolded structures. In order to compare in more detail such "differentiation" capabilities of "classic" assessment method (3D-1D Score [32]) and the novel approach, designed specially for MPs, the following test was proposed. The crystal structure of rhodopsin was modified in such a way that model structures were, given the fine MP organization, most probably misfolded and therefore unstable. These modifications consisted in systematic rotation of all TM α-helices as rigid bodies around their axes and energy relaxation of resulting "rotamer" models for elimination of rough atomic bumps. With account of independent rotation of all seven helices with 90° increment, this procedure yielded more than 16,000 conformations, only one of which ("zero-rotamer") corresponded to the crystal structure (see "Methods of Investigation"). Results of application of both methods to this data set are given in Fig. 3.

The 3D-profiles method declared itself incapable of identifying a "correct" structure among a generated decoy set, placing it near the mean value for the distribution, shown in Fig. 3a. The "membrane score" method appeared to be much more sensitive (Fig. 3b), placing the "zero-rotamer" at the very top of the list, outside of the 3σ range from the mean value of the membrane scores distribution. Moreover, only ~4% of structures received positive scores, eliminating the necessity of considering the majority of the generated conformations.

Practical optimization of MP packing is unlikely to require such thorough exploration of "rotational space" of an α-helical bundle, since even 90° rotations (not to mention 180° ones) will conflict with known regularities in orientations of helices with the most hydrophobic [17]

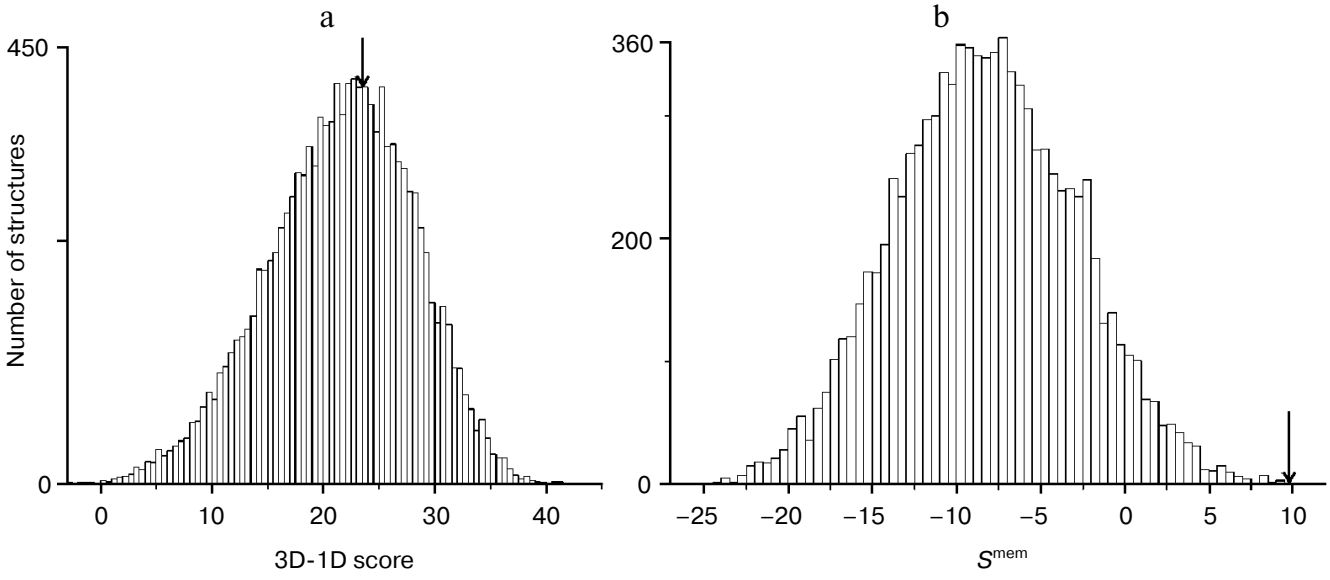


Fig. 3. Distributions of 3D-1D [32] (a) and membrane (b) scores for the ensemble of "rotameric" conformations of visual rhodopsin. The position of the crystallographic structure ("zero rotamer") is shown with an arrow.

and most variable [17, 20-22] sides to the membrane. However, while exploring the position of helices within a single 90° sector, the "membrane score" method might be of use and supplement existing methods for optimization of MP models.

The proposed "membrane score" method for assessment of packing quality of TM segments in membrane proteins permits determination of "compatibility" of amino acid residues with their environments in the structure, based on the data derived from the analysis of experimental high-resolution MP structures. The method was tested on a number of theoretical models of MPs, and its ability to identify "close-to-native" structures was demonstrated. This approach may be used for optimization of α-helical TM domain packing of models. One of the promising computational protocols looks as follows: starting mutual orientations of helices in TM bundle (obtained *via* homology modeling or by *ab initio* predictions) are varied in a systematic manner (e.g. by translations and rotations of helices as rigid bodies), and the values of S^{mem} are calculated for each resulting conformation. Those obtaining the highest S^{mem} values are considered as the most probable candidates for the "native-like" model. One of the shortcomings of the method is that the membrane score is quite sensitive to positions of the side chain rather than to overall protein fold. However, being combined with effective procedures for sampling of side chains, this shortcoming may turn into an advantage and give a possibility of optimizing "low-resolution" theoretical models, i.e. containing only C_α atoms or protein backbone. We believe that one of the most practically important applications of the "membrane score" method is optimization of GPCR models, either built on a rhodopsin template or generated *ab initio*.

Supplementary materials about the method are available on the website: http://model.nmr.ru.

We thank Drs. P. E. Volynsky and D. E. Nolde for their assistance with the implicit membrane model and software for generation of "rotamer" models, respectively.

This work was supported by the Russian Foundation for Basic Research (grants 04-04-48875-a, 05-04-49283-a, and 06-04-49194-a), by the Program "Physicochemical Biology" of the Russian Academy of Sciences, and by the Russian Federation Federal Agency for Science and Innovations (State contract 02.467.11.3003 of 20.04.2005, grant SS-4728.2006.4 "Leading Scientific Schools").

REFERENCES

1. Klabunde, T., and Hessler, G. (2002) *Chembiochem.*, **10**, 928-944.
2. Schoneberg, T., Schulza, A., Biebermann, H., Hermsdorf, T., Romplera, H., and Sangkuhl, K. (2004) *Pharmacol. Ther.*, **104**, 173-206.
3. Torres, J., Stevens, T. J., and Samso, M. (2003) *Trends Biochem. Sci.*, **28**, 174.
4. Fleishman, S. J., Unger, V. M., and Ben-Tal, N. (2006) *Trends Biochem. Sci.*, **31**, 106-113.
5. Wallin, E., and von Heijne, G. (1998) *Protein Sci.*, **7**, 1029-1038.
6. Vereshaga, Y. A., Volynsky, P. E., Nolde, D. E., Arseniev, A. S., and Efremov, R. G. (2005) *J. Chem. Theory Comput.*, **1**, 1252-1264.

7. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D. (2005) *Science*, **310**, 638-642.
8. Yarov-Yarovoy, V., Schonbrun, J., and Baker, D. (2006) *Proteins*, **62**, 1010-1025.
9. Zhang, Y., Devries, M. E., and Skolnick, J. (2006) *PLoS Comput. Biol.*, **2**, e13.
10. Becker, O. M., Shacham, S., Marantz, Y., and Noiman, S. (2003) *Curr. Opin. Drug. Discov. Devel.*, **6**, 353-361.
11. Hillisch, A., Pineda, L. F., and Hilgenfeld, R. (2004) *Drug Discov. Today*, **9**, 659-669.
12. Lesk, A. M., and Chothia, C. (1986) *Philos. Trans. R. Soc. Lond. Biol. Sci.*, **317**, 345-356.
13. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) *Science*, **289**, 739-745.
14. Sorgen, P. L., Hu, Y., Guan, L., Kaback, H. R., and Girvin, M. E. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 14037-14040.
15. Bowie, J. U. (2005) *Nature*, **438**, 581-589.
16. White, S. H., and von Heijne, G. (2005) *Curr. Opin. Struct. Biol.*, **15**, 378-386.
17. Rees, D. C., DeAntonio, L., and Eisenberg, D. (1989) *Science*, **245**, 510-513.
18. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982) *Nature*, **299**, 371-374.
19. Efremov, R. G., and Vergoten, G. (1996) *J. Prot. Chem.*, **15**, 63-76.
20. Donnelly, D., Overington, J. P., Ruffle, S. V., Nugent, J. H., and Blundell, T. L. (1993) *Protein Sci.*, **2**, 55-70.
21. Stevens, T. J., and Arkin, I. T. (2001) *Protein Sci.*, **10**, 2507-2517.
22. Beuming, T., and Weinstein, H. (2004) *Bioinformatics*, **20**, 1822-1835.
23. Pilpel, Y., Ben-Tal, N., and Lancet, D. (1999) *J. Mol. Biol.*, **294**, 921-935.
24. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O., and Fleming, P. J. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 5796-5801.
25. Eilers, M., Patel, A. B., Liu, W., and Smith, S. O. (2002) *Biophys. J.*, **5**, 2720-2736.
26. Fleishman, S. J., and Ben-Tal, N. (2002) *J. Mol. Biol.*, **321**, 363-378.
27. Park, Y., Elsner, M., Staritzbichler, R., and Helms, V. (2004) *Proteins*, **57**, 577-585.
28. Liu, W., Eilers, M., Patel, A. B., and Smith, S. O. (2004) *J. Mol. Biol.*, **337**, 713-729.
29. Ulmschneider, M. B., Sansom, M. S., and Di Nola, A. (2005) *Proteins*, **59**, 252-265.
30. Eyre, T. A., Partridge, L., and Thornton, J. M. (2004) *Protein Eng. Des. Sel.*, **17**, 613-624.
31. Efremov, R. G., and Vergoten, G. (1996) *Protein Eng.*, **9**, 253-263.
32. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) *Science*, **253**, 164-170.
33. Delarue, M., and Koehl, P. (1995) *J. Mol. Biol.*, **249**, 675-690.
34. Stevens, T. J., and Arkin, I. T. (1999) *Proteins*, **36**, 135-143.
35. Chothia, C. (1976) *J. Mol. Biol.*, **105**, 1-12.
36. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.*, **28**, 235-242.
37. Efremov, R. G., Nolde, D. E., Vergoten, G., and Arseniev, A. S. (1999) *Theor. Chem. Acc.*, **101**, 170-174.
38. Efremov, R. G., Volynsky, P. E., Nolde, D. E., and Arseniev, A. S. (2001) *Theor. Chem. Acc.*, **106**, 48-54.
39. Kabsch, W., and Sander, C. (1983) *Biopolymers*, **22**, 2577-2637.
40. Horn, F., Weare, J., Beukers, M. W., Horsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F., and Vriend, G. (1998) *Nucleic Acids Res.*, **26**, 277-281.
41. Eisenberg, D., Wesson, M., and Yamashita, M. (1989) *Chem. Scr.*, **29A**, 217-221.
42. Lindahl, E., Hess, B., and van der Spoel, D. (2001) *J. Mol. Mod.*, **7**, 306-317.
43. Von Heijne, G. (1986) *EMBO J.*, **5**, 3021-3027.
44. Kopp, J., and Schwede, T. (2006) *Nucleic Acids Res.*, **34**, D315-D318.
45. Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B. M., Eramian, D., Shen, M. Y., Kelly, L., Melo, F., and Sali, A. (2006) *Nucleic Acids Res.*, **34**, D291-295.
46. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) *Nature*, **356**, 83-85.