Novel approach to computer modeling of seven-helical transmembrane proteins: Current progress in the test case of bacteriorhodopsin[★✪]

Gregory V. Nikiforovich[1✉], Stan Galaktionov[1], Juris Balodis[2] and Garland R. Marshall[1]

[1]*Department of Biochemistry and Molecular Biophysics, Washington University, Campus Box 8036, St. Louis, MO 63110, U.S.A.;* [2]*Latvian Institute of Organic Synthesis, Riga, LV-1006, Latvia*

G-protein coupled receptors (GPCRs) are thought to be proteins with 7-membered transmembrane helical bundles (7TM proteins). Recently, the X-ray structures have been solved for two such proteins, namely for bacteriorhodopsin (BR) and rhodopsin (Rh), the latter being a GPCR. Despite similarities, the structures are different enough to suggest that 3D models for different GPCRs cannot be obtained directly employing 3D structures of BR or Rh as a unique template. The approach to computer modeling of 7TM proteins developed in this work was capable of reproducing the experimental X-ray structure of BR with great accuracy. A combination of helical packing and low-energy conformers for loops most close to the X-ray structure possesses the r.m.s.d. value of 3.13 Å. Such a level of accuracy for the 3D-structure prediction for a 216-residue protein has not been achieved, so far, by any available *ab initio* procedure of protein folding. The approach may produce also other energetically consistent combinations of helical bundles and loop conformers, creating a variety of possible templates for 3D structures of 7TM proteins, including GPCRs. These templates may provide experimentalists with various plausible options for 3D structure of a given GPCR; in our view, only experiments will determine the final choice of the most reasonable 3D template.

G-protein coupled receptors (GPCRs) are transmembrane proteins with 7-membered transmembrane helical bundles (7TM proteins). Until recently, this view has been based on indirect evi-

---

Abbreviations: BR, bacteriorhodopsin; 3D, 3-dimensional; GPCR, G-protein-coupled receptor; OOC, cytochrom *c* oxidase; PCR, photosynthetic reaction center; Rh, rhodopsin; r.m.s., root mean square; r.m.s.d., root mean square displacement; TM proteins, transmembrane proteins; 7TM proteins, 7-membered transmembrane proteins.

dence, such as hydrophobicity plots for many GPCRs showing several fragments consisting of highly hydrophobic residues which are sufficiently long to form helical segments crossing the membrane (see, e.g., [1]). Direct structural evidence for 7TM-type 3D models of transmembrane proteins, including GPCRs, has been obtained recently. The X-ray structures have been solved for two 7TM proteins, bacteriorhodopsin (BR) [2] and rhodopsin (Rh) [3], the latter being a GPCR. It is known that most GPCRs display sequence homology with Rh [4].

Recent progress in the X-ray crystallography of 7TM proteins hopefully will reveal more 3D structures for other GPCRs. One can expect that those 3D structures will reflect differences between GPCRs that ensure the selectivity and specificity of receptor interaction with any given ligand. It is noteworthy that 3D structures revealed by the X-ray studies for BR and Rh, that both belong to the same class of the light-activated 7TM proteins, are similar, but by no means identical. On the other hand, the X-ray crystallography produces "snapshots" of 3D protein structures; it cannot elucidate structural details of highly flexible parts of proteins, such as loops (see, e.g., [5]). In the case of 7TM proteins, besides flexible interhelical loops and the N- and C-terminal fragments outside the membrane, one has to consider the possibility of movements within the helical bundle itself that appears to be associated with switching from the inactive to the active state of the GPCR (e.g., [6]).

It is, therefore, likely that while 3D models for different GPCRs obtained by X-ray crystallography may show overall similarities with BR or Rh, there will be also significant differences. For example, the N-terminus and extracellular loops of rhodopsin form a structured plug that caps the retinal binding site. This cannot be the case with GPCRs that bind ligands such as catecholamines or peptide hormones that require access to binding sites from the extracellular space. Thus, computational approaches to modeling 3D structures of 7TM proteins become even more important, since they are indispensable to complement new experimental structural data. In our view, this task can hardly be fulfilled by the wide-spread homology modeling that employs 3D structures of BR or Rh as unique templates (see, e.g., [7]). On the contrary, computer modeling for 7TM proteins would be much more productive if adequate computational tools and procedures to build a variety of energetically consistent 3D models for different GPCRs were available.

RESULTS

This study reports on a novel approach to molecular modeling of 7TM proteins that has been developed to build varieties of plausible 3D structures for any given GPCR. This *ab initio* approach starts from the amino-acid sequence, and is based on very general physical principles discussed below. As a validation test, we have used bacteriorhodopsin (BR), which until very recently (August 2000) was the only 7TM protein whose 3D structure has been determined by X-ray diffraction at high resolution (entry code 1AP9 in the Brookhaven Protein Data Bank [2]).

General considerations

Our approach to 3D modeling of 7TM proteins utilizes some 3D information on 7TM proteins that is known *a priori*. For instance, we know that TM helices are immobilized in the membrane, so their mobility is certainly lower than that of the interhelical loops. Also, protein fragments protruding from the outer surface of the membrane (the N-terminus and the extracellular loops) would certainly not interact with fragments inside the cell (the intracellular loops and the C-terminus). TM fragments are connected by relatively short loops (up to 20–25 residues), so it is reasonable to suggest that they pack together as antiparallel helices; the short loops also ensure interactions between TM helices neighboring in a sequence. Therefore, 3D models for 7TM proteins can be built in a "block-by-block" manner, which is certainly not the case for soluble globular proteins.

Figure 1 illustrates our general view of the problem. Generally, we assume that the 3D structures of TM proteins may be described in terms of local

internal coordinates, i.e., dihedral angles. Accordingly, we use the ECEPP force field [8, 9] in our energy calculations assuming rigid valence geometry. For assembling helical bundles, one needs to include also the "global" coordinates, three variables defining the spatial position of the center of mass, X, Y and Z, and three more variables of general rotations, Tx, Ty and Tz, for each TM helix (see Fig. 1). To simplify the problem further, we
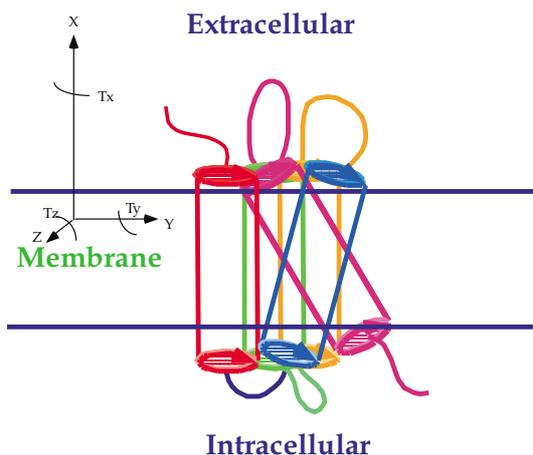
Figure 1. The working model for TM protein.

apply an assumption of "hard helical cores" (backbones) and "soft shells" (side chains) for each helix. Practically, this means that the dihedral angles for backbones of helices are "frozen" at the values obtained by energy minimization for each separate helix. For BR, energy calculations preserve the 3D structures of individual helices with great accuracy, the r.m.s. values calculated for C^α atoms differing from 0.42 Å to 1.61 Å ("coordinate" r.m.s. values).

We divide our approach into several distinct steps: (i) finding possible TM helices in a sequence; (ii) assembling helical bundle(s); and (iii) restoring interhelical loops.

Determining possible TM helices in a sequence

Since establishing of the rules for predicting helices requires substantial experimentally defined training sets, generally not available for TM proteins, we have developed a novel procedure that predicts TM helical fragments in the sequence by analyzing possible breaks in TM helices [10]. The procedure consists of stepwise elongation of the "core" helical fragment determined by the consensus results of several available statistical predictions to increase reliability. At each step, we calculate conformational energies corresponding to the helical conformer of the "core" fragment elongated by two flanking residues, $E(\alpha)$, as well as to several options for the fragment to continue or to exit the helix by changing conformations of the flanking residues, E_i. (We have used the options to continue or to exit the helix suggested in [11, 12].) The minimal values out of E_i–$E(\alpha)$, Δ_k, obtained at the each elongation step, can be viewed as a profile of relative energies, where each minimum of Δ_k is a signal to start or to stop the TM helix. We assumed that the boundaries of the TM helix would be determined by the largest break signals that are the closest to the "core" sequence in the Δ_k profiles.

Our procedure has been applied to the prediction of the N- and C-termini for 45 transmembrane helices from the photosynthetic reaction center (PRC) from *Rhodopseudomonas viridis* [13], bacteriorhodopsin (BR) [2] and cytochrome *c* oxidase (OCC) from *Paracoccus denitrificans* [14]. We have predicted helical endpoints within an error of ±2 residues with an accuracy of about 55%. For comparison, three different statistical approaches, namely the TMPRED program that is basically a neural network application [15], the SOSUI program based mainly on the spectral density of hydrophobicity plots [16], and the DAS method that is a generalization of sequence-alignment techniques [17] made the same predictions with an accuracy not greater than 40%. (Interestingly, a very recent paper on determining TM helices within a sequence specifically pointed out that the accuracy of our procedure remains unsurpassed [18].) Specifically for BR, our predictions have been as follows: (first to last residue in the helix; experimental data in parentheses): 10–32 for BR1 (10–32), 39–68 for BR2 (38–61), 83–101 for BR3 (80–100), 108–127 for BR4 (105–126), 138–157 for BR5 (134–157), 178–198 for BR6 (166–189), and 205–224 for BR7 (202–225). It is noteworthy that, in any particular case, the exact length of TM residues obtained by our predictions can be slightly altered in the process of restoring the interhelical loops.

Assembling of helical bundle(s)

Any significant similarity between the X-ray structures of BR and Rh, noted above, is most pronounced in the type of helical packing in the plane of the membrane (see [2] and [3]). Our analysis of possible reasons for such packing has been associated with finding the rough values for the "transitional" global parameters X, Y and Z.

Y and Z coordinates

Packing of TM helices in the membrane plane YZ may be considered as packing of the cylinders with the identical diameter parallel to the X axis (see Fig. 1). Obviously, according to the principle of dense packing, the densest packing of these cylinders can be achieved in a hexagonal arrangement where the central cylinder is in contact ("interacts") with six neighbors. However, the most dense packing of "amphiphilic cylinders" that avoids unfavorable contacts between "hydrophilic" and "hydrophobic" cylinder surfaces results in two other types of packing each featuring 5 contacts of the non-neighboring cylinders (see Fig. 2). One of these types is very close to the helical packing observed in BR and Rh (Fig. 2). Assuming that the diameter of the cylinders is 10 Å, which corresponds to the average diameter of TM helices [19], we immediately obtain Y and Z coordinates for the roughly packed BR helical bundle that are in good agreement with those from the X-ray structure (the largest individual difference is 3.45 Å, see Table 1, the Z coordinate for helix

Figure 2. Dense packing of "amphiphilic" cylinders.

Numbering corresponds to that in BR and Rh.

BR5). Note that the numbering of cylinders (helices) in the two possible types of packing in Fig. 2 may be different; in that case, new options for packing helices may be created.

X coordinates

To solve this problem, we have applied a previously developed computational procedure for arranging isolated TM helices across a membrane [20]. This simple procedure considers the "membrane" as a space with a given "membrane thickness" between two infinite parallel planes filled inside with a homogeneous octanol environment. Similarly, the homogeneous water environment occupies the space outside the planes on both sides. Octanol and water environments differ in their contributions to the solvation energy for each particular atom transferred from one environment to another. The procedure pulls the helical fragment through the membrane along the X axis with step increments of 1.5 Å. At each step, the helical fragment is tilted in the direction normal to the axis (tilt angle Ty) up to $\pm 40°$ in $20°$ increments, and is rotated around the X axis (rotation angle Tx) from $0°$ to $150°$ in $30°$ increments. This corresponds to the total circle of rotation, since both positive and negative tilts are considered. (Note that a recent analysis of helical packing for TM proteins based on X-ray data [19] showed that the values of the tilt angles (Ty and Tz in our notation) are generally small, not exceeding $\pm 40°$.) Totally, at each vertical step, 25 positions of the helical fragment are examined. For each of these 25 global spatial positions, the optimal solvation energy of the helical fragment, E_{sol}, is calculated. Energy values themselves are the sum of two parts. The first one is the regular intramolecular atom–atomic interactions inside the helical fragment calculated by the ECEPP/2 force field. The second one is a sum of solvation energies for each atom calculated by the "excluded volume" procedure [21]. As a general result, this procedure calculates a profile of the energetic cost for transferring the helical fragment from one side of the membrane to another. The lowest point(s) in this profile may be used for indicating the optimal position(s) of the helical frag-

ment in the membrane. (A similar idea has been implemented earlier in [22], where differences in

demonstrates helical packing obtained by using only the values of X, Y and Z described in Table 1

Table 1. Global coordinates X, Y, Z (Å) and Ty, Tz (degrees) for the X-ray structure of BR (upper row), and those suggested for rough packing of helices (lower row).

Helix	X	Y	Z	Ty	Tz
BR1	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00
BR2	6.73	7.28	3.17	−13.56	−162.47
	4.50	8.66	5.00	−20.00	−160.00
BR3	3.61	16.97	3.41	25.47	18.34
	5.00	17.32	0.00	10.00	20.00
BR4	1.13	26.30	6.71	−9.13	−163.00
	2.00	25.98	5.00	−20.00	180.00
BR5	−2.74	29.59	−1.55	26.90	3.60
	−1.50	25.98	−5.00	0.00	0.00
BR6	2.40	20.29	−10.24	−17.05	177.82
	3.50	17.32	−10.00	−20.00	180.00
BR7	−2.82	9.15	−5.27	9.23	1.00
	0.50	8.66	−5.00	0.00	0.00

experimentally estimated energies for "residue–water" *vs* "residue–cyclohexane" interactions were used for establishing possible arrangements of TM helices across the membrane in the process of a stepwise residue- by-residue immersion.)

The procedure has been applied to a validation set of the same 45 transmembrane helices mentioned above. In two thirds of the helical fragments considered, the procedure has predicted the vertical shifts of the fragments across the membrane with an accuracy of (−0.15 ± 3.12) residues compared to the experimental data. The accuracy for the remaining 15 fragments was (2.17 ± 3.07) residues, i.e. about one-half of a helix turn. The procedure has been applied for finding the relative positions for all BR helices along the X coordinate. Four options of such relative arrangements have been found that correspond to two equivalently optimal positions for helices BR3 and BR6. Table 1 contains the X coordinates most close to those revealed by the X-ray studies. The largest individual difference between the calculated X coordinates and those from the X-ray structure was 3.32 Å (helix BR7 in Table 1).

The efficiency of employing the procedures developed for finding the X, Y and Z coordinates

(i.e., with the Tx, Ty and Tz values equal to zero). Even this rough option of helical packing possessed an r.m.s.d. value of 3.37 Å relative to the X-ray structure (C^α-atoms, "distance" r.m.s. value). Other options differing in the X coordinates (see above) possessed r.m.s.d. values ranging from 3.33 Å to 4.47 Å.

Ty and Tz coordinates

The types of helical packing shown in Fig. 2 suggest a high degree of cooperativity in the values for the tilt angles (Ty and Tz) for each individual helix included into a helical bundle. For instance, if the Ty and Tz angles for BR3 helix are defined as equal to zero, all other helices will try to adopt the values of Ty's and Tz's to maintain dense packing between helices that exists in the "initial" point when all Ty's and Tz's are equal to zero. As a result, we can suggest only two most probable sets of Ty's and Tz's for BR1, BR2, BR4, BR5, BR6 and BR7, corresponding to the "clockwise" or "anti-clockwise" tilt sense of these helices around BR3. The absolute values of those angles have been chosen as 20°, in agreement with literature data [19]. The set of Ty's and Tz's closest to

the values in the X-ray structure of BR is described in Table 1; the values are recalculated to the reference point where Ty and Tz for BR1 are equal to zero. One can see that the suggested Tz values are very close to those in the X-ray structure; as to the Ty values, the differences are more pronounced. Interestingly, the set of Ty and Tz values most close to the X-ray structure of Rh [3] possesses the same absolute values, but with the opposite tilt sense.

Tx coordinates

Energy calculations with optimization of the side chain dihedral angles have been performed for all neighboring pairs of BR helices. The standard ECEPP/2 force field has been used; however, electrostatic interactions have been ignored. The X, Y, Z, Ty and Tz starting points for energy calculations were those listed in Table 1, except for small variations (Tz = $-160°$ for BR4, Ty = $20°$ for BR3, Ty = $0°$ for BR4, and Ty = $20°$ for BR5). These values, as well as Tx's and the dihedral angles of side chains were allowed to vary. One hundred and fourty-four combinations of Tx's (12 by 12, i.e. a grid of $30°$) were considered for each pair of neighboring helices. Energy calculations found possible low-energy configurations ($\Delta E < 10$ kcal/mol) for each neighboring pair (2 configurations for pair of helices BR1–BR2, 13 for BR2–BR3, 25 for BR3–BR4, 35 for BR4–BR5, 18 for BR5–BR6, 23 for BR6–BR7, and 22 for BR7–BR1).

Assembling seven-helical bundle

Obviously, not all of the low-energy pairwise configurations are geometrically compatible with each other. Only 43 possible configurations of the seven-helical bundle were obtained as a result of combining the pairwise configurations. The r.m.s.d. values for these configurations in comparison with the X-ray structure of BR fall into the range from 2.93 Å to 3.52 Å, which means that all of them are close to the X-ray helical bundle. However, it is possible to rank the possible configurations further by applying a scoring function based on the frequencies of observed residue–residue contacts between TM helices in TM proteins within the known X-ray structures.

The employed scoring function E was as follows:

$$E = \Sigma_{ij}\ E_{ij}, \qquad \text{where:}$$

$$E_{ij} = 0, \qquad \text{when } r_{ij} > 8 \text{ Å};$$

$$E_{ij} = \alpha f_{mn}, \qquad \text{when } R_0 < r_{ij} < 8 \text{ Å};$$

$$E_{ij} = -E_0, \qquad \text{when } r_{ij} < R_0.$$

Here i and j are indexes related to the C^{α}-atoms of interacting helices; r_{ij} is the distance between atomic centers i and j, and the values of parameters R_0, α and E_0 are 3.0 Å, 1.0 and 500, respectively, f_{mn} is a 20×20 matrix of normalized frequences of interhelical contacts between residues of the types n and m:

$$f_{nm} = N_{mn} / N_m * N_n,$$

where N_{mn} is the total number of found interhelical contacts between residues of the types m and n, and N_m, N_n are the numbers of contacts of any kind found for residues of the type m or n, respectively.

The f_{mn} matrix (Table 2) has been deduced by analyzing the X-ray structures of nine TM proteins (PDB access codes 1brd; 1occ; 1prc; 1be3; 1bl8; 1kzu; 1lgh; 1msl and 1lih) including a total of 2 337 residues in TM helices participating in 2 155 interhelical contacts. Figure 3 presents the plot of the E values for each of the 43 found configurations of the seven-helical bundle *vs.* the r.m.s.d. values relative to the X-ray structure of BR. Despite no really pronounced correlation between the E and r.m.s.d. values, it is clear that the higher values of scoring function correspond to the lower values of r.m.s.d.. The top seven configurations differ from the X-ray structure of BR by the r.m.s.d. values of 2.9–3.1 Å; the highest score TM bundle (r.m.s.d. = 3.045 Å) is depicted in Fig. 4. This r.m.s.d. value is comparable to that between the predicted 7TM helical bundle of Rh [23] and the actual X-ray structure of Rh [3], which is 3.085 Å (backbone atoms). However, in that case,

the electron microscopy maps for Rh [24], as well as the proposed assignment of the TM helices to the maps [25] have been employed as a starting point for refinement of the H-bond network within the bundle [23]; i.e., the rough values of the X, Y, Z, Tx, Ty and Tz coordinates have been taken directly from the experimental data on Rh.

Restoring interhelical loops

For *small loops* (4 to 10 residues), we have developed a rather straightforward approach to restoring their 3D structures. It consists of generating all possible combinations of the backbone conformations close to the local backbone energetic minima for the entire loop with the inherent distance limitations implied by the known positions of the adjacent helices. All generated conformations are then subjected to energy minimization, and then the "averaged" conformer representing the geometrical average for the entire set of low-energy conformers for a given small loop is selected.

For instance, for the seven-residue loop 30–38 in BR (the loop between helices BR1 and BR2), 1594 conformers of the peptide backbone have

been found to satisfy all distance constraints. For each of these conformers, we have performed energy calculations that included parabolic poten-

Figure 3. Scoring function calculated for 43 options of helical bundle compared to the r.m.s.d. values in relation to the X-ray structure of BR.

tials to comply with the distance constraints. (In these energy calculations, all dihedral angles of the backbone were not "frozen" as in the case of interhelical interactions, but were allowed to vary). One hundred and ten low-energy conformers ($\Delta E < 10$ kcal/mol) have been obtained that are all very close to the "averaged" 3D structure

Table 2. Normalized frequences of interhelical contacts between residues of various types in TM proteins.

T y p e s   o f   R e s i d u e s (one-letter code)

	G	A	V	L	I	F	P	W	S	T	M	N	Q	C	D	E	Y	H	K	R
G	3	11	12	11	5	7	3	5	4	4	3	2	1	1	0	5	3	4	1	3
A		15	19	24	15	14	3	6	9	8	6	1	2	0	5	0	4	4	4	2
V			10	19	12	8	3	3	4	9	2	2	1	0	0	5	4	2	4	1
L				16	14	15	4	4	14	10	6	3	1	2	3	5	8	4	4	4
I					6	6	2	1	6	8	4	1	1	1	1	3	3	2	1	1
F						6	3	3	6	5	5	1	1	2	1	1	3	2	1	3
P							0	2	1	2	1	1	0	0	2	3	1	1	2	2
W								1	2	3	1	1	2	0	1	0	0	1	1	5
S									3	5	2	1	1	1	1	1	2	1	0	6
T										7	4	3	1	1	2	1	2	2	0	1
M											1	0	1	1	1	0	1	1	1	1
N												0	1	0	0	0	0	0	2	0
Q													0	0	0	3	3	1	0	2
C														0	0	0	0	0	0	1
D															0	0	0	0	2	0
E																0	2	3	0	0
Y																	3	1	1	1
H																		1	2	0
K																			1	0
R																				0

(the r.m.s. values (coordinates of C^{α} atoms) are 0.63–2.09 Å), and to the X-ray structure (the r.m.s. values of 1.40–1.73 Å) (see Fig. 5). The same approach has been used for loops between helices BR3 and BR4; BR4 and BR5; BR5 and BR6, and BR6 and BR7. A summary of the results is presented in Table 3. One can see that the "av-

tween BR2 and BR3) that has been subjected to this procedure. During the elongation steps, the number of generated conformers satisfying all limitations started with 326 259 at the 61–72 fragment, then reached the maximal value of 783 781 at the 61–75 fragment, and then finally achieved the value of 129 conformers for the entire 61–80
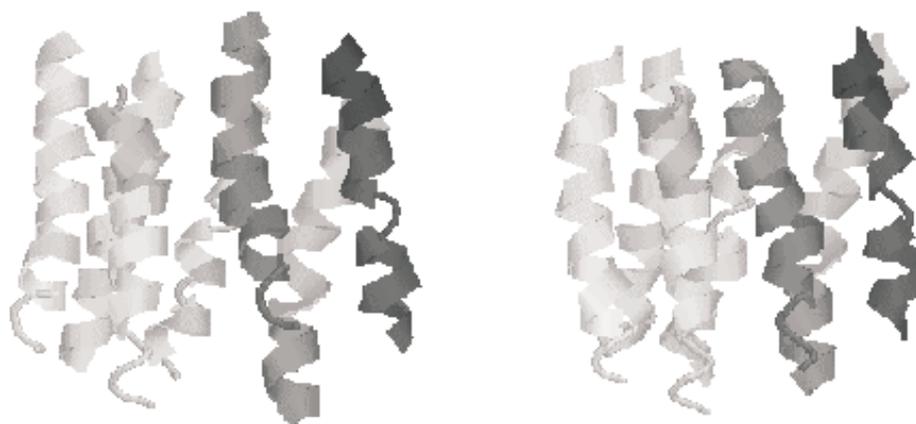
Figure 4. The highest scored helical bundle (right) compared to the X-ray structure of BR (left).

eraged" conformers indeed represent the available set of low-energy conformers even for the twelve-residue loop between helices BR6 and BR7; in this case, however, the geometrical variations between low-energy conformers are understandably the largest.

For ***medium loops*** (12 to 20 residues), the straight generation of all possible combinations of the backbone conformations is computationally limited. Therefore, we have developed a special build-up algorithm for a stepwise elongation of the peptide backbone for a loop connecting helices i and i +1 starting from the last residue of the i-th helix. First, all the "averaged" structures of the small loops are added to the already known helical bundles to create a united "framework" for the further building procedure. Then, at each elongation step, the system of limitations is imposed on the minimal and maximal C^{α}–C^{α} distances inside the loop, as well as between any C^{α} atom in the loop and any C^{α} atom within the "framework" consisting of the known helices and the smaller loops. After selection of the loop structures that satisfy these limitations, energy calculations are performed, and the resulting low-energy structures are clustered by their 3D shape. There was one eighteen-residue loop in BR (loop 61–80 be-

fragment. The subsequent energy calculations were performed in the same manner as described above for the small loops. They yielded 56 low-energy conformers (ΔE <10 kcal/mol), one of them depicted in Fig. 6. This particular structure
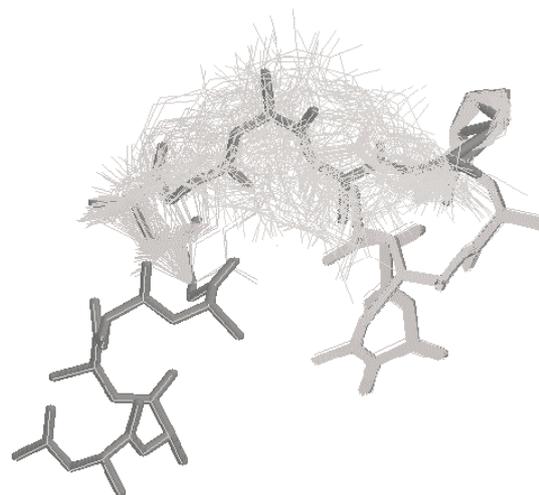
Figure 5. Variety of low-energy conformers for loop 30–38 (in light gray) in comparison with the "averaged" conformer (in dark gray).

possesses the r.m.s. value of 3.45 Å compared to the actual X-ray conformation, and it reproduces

a complicated twisted β-hairpin motif of the natural loop (see Fig. 6). Other low-energy conformers possess the r.m.s. values up to 8.87 Å. The quality of this prediction may be estimated in comparison with the results of predictions obtained by other authors. The best r.m.s.d. values of 0.93 Å [26], 1.79 Å [27] and 1.87 Å [28] have been reported for nine-residue loops; all attempts to predict conformations of loops larger than 10–11 residues resulted in the r.m.s.d. values greater than 3 Å [29]. No, at least somewhat realistic, predictions for loops as large as 18 residues were reported at the CASP-3 meeting [29]. It is noteworthy that energy calculations performed for loops have involved the backbone structures only; it should be expected that the r.m.s. values would further improve after addition of side chains. These could be added in a refinement procedure to discriminate between alternative loop conformations.

Figure 6. Best of low-energy conformers found for loop 61–80 (light gray) overlapped with the X-ray structure (dark gray); r.m.s. = 3.45 Å

DISCUSSION

The approach to computer modeling of 7TM proteins discussed in this work satisfies several requirements. It was capable of reproducing the experimental X-ray structure of BR with great accuracy. A combination of helical packing and low-energy conformers for loops most close to the X-ray structure is depicted in Fig. 7, and possesses the r.m.s.d. value of 3.13 Å. Such a level of accuracy for the 3D-structure prediction for a 216-residue protein has not been achieved, so far, by any available *ab initio* procedure of protein folding (see, e.g., the results of the CASP-3 meeting [30]). The approach may produce also other energetically consistent combinations of helical bundles and loop conformers, creating a variety of possible templates for 3D structures of 7TM proteins, including GPCRs. These templates may differ in the numbering of helices in helical bundles; in different optimal positions of individual helices arranged across the membrane; in different packing patterns; and in different low-energy conformers of interhelical loops. In fact, the main goal of this approach is to provide experimentalists with various plausible options for 3D struc-

Figure 7. The "best" combination of calculated helical packing and interhelical loops (in light gray) in comparison with the X-ray structure of BR (in dark gray).

ture of a given GPCR, not confining us with a single template option, as that of BR or Rh. In our

helical packing should be performed employing predicted TM helices, not the actual ones, as it

Table 3. Restored small loops in BR

Loop between helices	Loop	No. of residues	No. of low-energy conformers	r.m.s. values, Å		
				average *vs.* X-ray	*vs.* average	*vs.* X-ray
12	30–38	7	110	1.59	0.63–2.09	1.40–1.73
34	100–105	4	3	1.20	0.38–0.71	1.15–1.35
45	126–134	7	66	1.70	0.77–2.16	1.58–2.55
56	157–166	8	147	1.34	0.68–2.38	1.04–2.55
67	189–202	12	131	3.46	1.24–5.80	2.54–5.23

view, only experiments will determine the final choice of the most reasonable 3D template.

Also, an important feature of the approach is that it is implemented in an easy-to-understand stepwise procedure. Adequate computational tools have been developed for each step of this procedure. Employing these tools, we can find boundaries (breaks) of TM helices in a sequence. We can also arrange TM helices across membranes, and, using the principle of dense packing, perform a rough packing of helical bundles. We can combine compatible configurations found for neighboring pairs of helices together, forming various helical bundles. We can rank the bundles by a scoring function based on the frequencies of interhelical interactions between residues of various types. We also can restore the interhelical loops by different procedures depending on the loop size. It is noteworthy that each element of the entire procedure can be applied independently. For example, it would be very simple to restore interhelical loops attached not to the calculated option of helical packing, but to the experimentally determined helical bundle. We are currently using this procedure to build homology models of GPCRs that bind peptides, based on the crystal structure of rhodopsin.

So far, the approach cannot be regarded as fully developed and validated. Many tests still should be performed to validate this approach as a routine application for TM modeling. For instance,

was in this work. Various ways of numbering helices in the bundle should be checked, as well as various combinations of helical packing and loop conformers. The scoring function for interhelical interactions should be refined, as well as a final refinement of the entire 3D structure for the TM protein should be elaborated. And, of course, the X-ray structure of Rh has to be reproduced. However, even in the current state of development, the results obtained by our approach are encouraging indeed.

REFERENCES

1. Marshall, G.R. (1997) Therapeutic approaches to human immunodeficiency virus — structural studies on G-protein-coupled receptors. *Pharmacol. & Therap.* **76**, 135–139.

2. Pebay-Peyroula, E., Rummel, G., Rosenbush, J.P. & Landau, E.M. (1997) X-Ray structure of bacteriorhodopsin at 2.5 ångstroms from microcrystals grown in lipidic cubic phases. *Science* **277**, 1676–1681.

3. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. & Miyano, M. (2000) Crystal structure of rhodopsin, A G protein-coupled receptor. *Science* **289**, 739–745.

4. Soppa, J. (1994) Two hypotheses – one answer. Sequence comparison does not support an evolutionary link between halobacterial retinal proteins including bacteriorhodopsin and eukaryotic G-protein-coupled receptors. *FEBS Lett.* **342**, 7–11.

5. Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, R. & Hendrickson, W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659.

6. Chollet, A. & Turcatti, G. (1999) Biophysical approaches to G protein-coupled receptors, Structure, function and dynamics. *J. Comput.-Aided Mol. Des.* **13**, 209–219.

7. Alkorta, I. & Loew, G.H. (1996) A 3D model of the delta opioid receptor and ligand-receptor complexes. *Protein Eng.* **9**, 573–583.

8. Nemethy, G., Pottle, M.S. & Scheraga, H.A. (1983) Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occuring amino acids. *J. Phys. Chem.* **87**, 1883–1887.

9. Dunfield, L.G., Burgess, A.W. & Scheraga, H.A. (1978) Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. *J. Phys. Chem.* **82**, 2609–2616.

10. Nikiforovich, G.V. (1998) A novel, non-statistical method for predicting breaks in transmembrane helices. *Protein Eng.* **11**, 279–283.

11. Efimov, A.V. (1986) Standard conformations of a polypeptide chain in irregular protein regions. *Molek. Biol.* **20**, 250–260 (in Russian).

12. Efimov, A.V. (1991) Structure of α–α hairpins with short connections. *Protein Eng.* **4**, 245–250.

13. Deisenhofer, J., Epp, O., Sinning, I. & Michel, H. (1995) Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. *J. Mol. Biol.* **246**, 429–457.

14. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995) Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* **376**, 660–669.

15. Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995) Transmembrane helices predicted at 95% accuracy. *Protein Sci.* **4**, 521–533.

16. Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998) SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379.

17. Cserzö, M., Bernassau, J.-M., Simon, I. & Maigret, B. (1994) New alignment strategy for transmembrane proteins. *J. Mol. Biol.* **243**, 388–396.

18. Gromiha, M.M. (1999) A simple method for predicting transmembrane α-helices with better accuracy. *Protein Eng.* **12**, 557–561.

19. Bowie, J.U. (1997) Helix packing in membrane proteins. *J. Mol. Biol.* **272**, 780–789.

20. Tseitin, V.M. & Nikiforovich, G.V. (1999) Isolated transmembrane helices arranged across a membrane, computational studies. *Protein Eng.* **12**, 305–311.

21. Hopfinger, A.J. (1973) *Conformational Properties of Macromolecules*. Academic Press, N.Y.

22. Röper, D., Jacoby, E., Krüger, P., Engels, M., Grötzinger, J., Wollmer, A. & Strassburger, W. (1994) Modeling of G-protein coupled receptor with bacteriorhodopsin as a template. A novel approach based on interaction energy differences. *J. Receptor Res.* **14**, 167–186.

23. Pogozheva, I.D., Lomize, A.L. & Mosberg, H.I. (1997) The transmembrane α-bundle of rhodopsin: Distance geometry calculations with hydrogen bonding constraints. *Biophys. J.* **70**, 1963–1985.

24. Schertler, G.F.X., Villa, C. & Henderson, R. (1993) Projection structure of rhodopsin. *Nature* **362**, 770–772.

25. Baldwin, J.M. (1993) The probable arrangement of the helices in G-protein-coupled receptors. *EMBO J.* **12**, 1693–1703.

26. Zhang, H., Lai, L., Wang, L., Han, Y. & Tang, Y. (1997) A fast and efficient program for modeling protein loops. *Biopolymers* **41**, 61–72.

27. van Vlijmen, H.W.T. & Karplus, M. (1997) PDB-based protein loop prediction: Parameters for selection and methods for optimization. *J. Mol. Biol.* **267**, 975–1001.

28. Carlacci, L. & Englander, S.W. (1993) The loop problem in proteins: A Monte Carlo simulated annealing approach. *Biopolymers* **33**, 1271–1286.

29. Bates, P.A. & Sternberg, M.J.E. (1999) Model building by comparison at CASP3: Using expert knowledge and computer automation. *Proteins* (Suppl.) **3**, 47–54.

30. Orengo, C.A., Bray, J.E., Hubbard, T., LoConte, L. & Sillitoe, I. (1999) Analysis and assessment of *ab initio* three-dimensional prediction, secondary structure, and contacts prediction. *Proteins* (Suppl.) **3**, 149–170.