Modeling the 3D Structure of Rhodopsin Using a De Novo Approach to Build G-protein–Coupled Receptors

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · FEBRUARY 1999

Impact Factor: 3.3 · DOI: 10.1021/jp9820471

3 AUTHORS, INCLUDING:

Marta Filizola
Icahn School of Medicine at Mount Sinai
128 PUBLICATIONS **2,819** CITATIONS
SEE PROFILE

Juan J Perez
Polytechnic University of Catalonia
161 PUBLICATIONS **2,013** CITATIONS
SEE PROFILE

Modeling the 3D Structure of Rhodopsin Using a De Novo Approach to Build G-protein−Coupled Receptors

Marta Filizola,[†,‡] Maria Cartenì-Farina,[†] and Juan J. Perez*[,‡]

Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche (CRISCEB), Seconda Università degli Studi di Napoli, Via Costantinopoli, 16, 80138 Napoli, Italy, and Department d'Enginyeria Quimica, UPC, ETS d'Enginyers Industrials, Av. Diagonal, 647, 08028 Barcelona, Spain

Received: April 29, 1998; In Final Form: November 4, 1998

The difficulties associated with the experimental determination of high-resolution, three-dimensional (3D) structures of G-protein−coupled receptors (GPCRs), make molecular modeling the only practicable technique to provide structural information at atomic resolution of these integral membrane proteins. However, these methods need to be critically evaluated to assess their predictive capabilities and their limitations. We have recently proposed a de novo method to construct 3D structures of the transmembrane domain of GPCRs. The procedure requires only the knowledge of the amino acid sequence and uses the helix spatial arrangement of rhodopsin deduced from its low-resolution electron density map. Models are refined using molecular dynamics. The aim of the present work is to apply this approach to construct and refine a 3D model of rhodopsin as a test case. Moreover, to assess the level of accuracy that can be expected of the models using our de novo approach, we constructed and refined a 3D model of bacteriorhodopsin using a low-resolution electron density map of this protein. After refinement the model of bacteriorhodopsin was compared with the experimental structure refined at 3.5 Å resolution.

Introduction

G-Protein−coupled receptors (GPCRs) are found in a wide range of organisms and are involved in the signal transduction mechanisms of diverse chemical messengers. At the present time there are more than 700 receptors sequenced, many of which have an unknown pharmacological function. Knowledge of their 3D structure would be a valuable aid in designing ligands targeted to bind to them. However, because of the difficulties associated with the isolation and purification of GPCRs in the amounts required to perform structural studies, modeling studies presently represent the only way to construct 3D models with which to perform docking experiments or to suggest receptor amino acid residues relevant for ligand binding.[1]

Rhodopsin is the pigment of retinal rod photoreceptor cells that mediates visual transduction in vertebrates. The system is composed of an integral protein, opsin, and a chromophore, 11-*cis*−retinal, covalently bonded to Lys^{296} via a protonated Schiff base linkage. Upon light absorption retinal reaches an electronic excited state and undergoes a *cis*−*trans* isomerization about its 11-*cis* double bond. This activates rhodopsin and triggers a subsequent signal transduction mediated through its G-protein transducin. Besides its intrinsic interest, rhodopsin is the only member of the GPCRs superfamily for which an electron density map has been recorded. Electron density maps of frog rhodopsin at 7.5 Å resolution[2,3] and of bovine rhodopsin at 9 Å resolution[4] provide direct evidence of the seven-helix bundle of the protein transmembrane domains. Thus, rhodopsin presently represents the only source of structural information to study the packing features of the seven transmembrane domains in GPCRs. Accordingly, any modeling approach aimed at constructing 3D structures of GPCRs should provide a model of this molecule as a test case.

Several modeling approaches have been used in the past to construct 3D models of GPCRs and applied to build 3D models of rhodopsin.[5] They can be classified broadly into two categories: those that use the 3D structure of bacteriorhodopsin as template and those that construct the 3D models de novo. Using bacteriorhodopsin as template to perform homology modeling studies at first was considered a reasonable choice, because similar to rhodopsin, bacteriorhodopsin is an integral protein with an all-*trans*−retinal prosthetic group covalently linked to Lys^{216}, and its transmembrane domains are also arranged to form a seven-helix bundle. However, there are many difficulties inherent in the use of this approach.[6] First, bacteriorhodopsin and rhodopsin do not share a common biological function. Thus, although bacteriorhodopsin as well as rhodopsin is activated by light, it functions as a proton pump and it is not coupled to a G-protein. Accordingly, bacteriorhodopsin plays the role of an energy transducer instead of a signal transducer as does rhodopsin. Second, sequence identity between bacteriorhodopsin and rhodopsin is low (<20%). Third, comparison of the electron density maps of bacteriorhodopsin and rhodopsin reveals different helix bundle arrangements. In rhodopsin, helix III is buried more deeply in the interior, giving rise to two differentiated pockets which are not observed in bacteriorhodopsin. Moreover, considerable differences exist in the helix packing of the two proteins at the intracellular surface[3] which may be associated with the requirement of rhodopsin to be coupled to a G-protein.

Model construction de novo using all the biophysical and chemical information available represents an alternative strategy. Different models have been constructed based on a detailed sequence analysis and multiple sequence alignments,[7,8] using molecular graphics to satisfy specific restrains[9−11] or using a

[†] Seconda Università degli Studi di Napoli.
[‡] ETS d'Enginyers Industrials.

10.1021/jp9820471 CCC: $18.00
Published on Web 02/24/1999

	1	I	50
BR_HALHA	MLELLPTAVE GVSQAQITGR **PEWIWLALGT ALMGLGTLYF LVKG**MGVSDP		
	51 II	**III**	100
BR_HALHA	**DAKKFYAITT LVPAIAFTMY LSMLL**GYGLT MVPFGGEQN**P IYWARYADWL**		
	101	**IV**	150
BR_HALHA	**FTTPLLLLDL ALL**VDAD**QGT ILALVGADGI MIGTGLVGAL** TKVYSY**RFVW**		
	151 **V**	**VI**	200
BR_HALHA	**WAISTAAMLY ILYVLFFGFT** SKAESMRPEV **ASTFKVLRNV TVVLWSAYPV**		
	201	**VII**	250
BR_HALHA	**VWLI**GSEGAG **IVPLNIETLL FMVLDVSAKV GFGLILLRS**R AIFGEAEAPE		
	251 262		
BR_HALHA	PSAGDGAAAT SD		

Figure 1. Amino acid sequence of bacteriorhodopsin retrieved from the SwissProt database with the putative seven transmembrane helices highlighted.

simulated annealing protocol coupled to a Monte Carlo procedure that places and orients rigid helices to satisfy specific structural restraints.[12] We recently have developed a general and simple de novo procedure to construct 3D models of GPCRs based on the fitting of rigid helices to the low-resolution electron density map of rhodopsin.[13,14] The procedure only requires knowledge of the sequences that define the transmembrane regions. Helix orientation is performed automatically based on the hydrophobic properties of the residues involved. The purpose of the present work is to explore the characteristics of this approach by applying the procedure to the construction of 3D models of bacteriorhodopsin and of rhodopsin. In each case, the corresponding electron density maps were used. The 3D models were refined subsequently by energy minimization. For bacteriorhodopsin, direct comparison of the minimized structure with the available crystal structure was performed. In rhodopsin, in the absence of an atomic resolution crystal structure, assessment of the geometrical features of the complex was performed by a molecular dynamics (MD) simulation, in which relevant distances between the ligand and the receptor in the complex were monitored.

Methods

1. Choice of the Putative Transmembrane Regions of Bacteriorhodopsin and Rhodopsin. Amino acid sequences of *Halobacterium halobium* bacteriorhodopsin and bovine rhodopsin were taken from the SWISS−PROT sequence data bank.[15] They comprise 262 and 348 residues, respectively, and share a sequence identity lower than 20% between them. In bacteriorhodopsin the putative transmembrane segments were assessed from its known 3D structure[16] (entry 2BRD of the Brookhaven data bank), whereas in rhodopsin they were assigned with use of the variability index.[7] Figures 1 and 2 show the sequences of the two different membrane proteins with their seven putative transmembrane regions highlighted.

2. Construction of Bacteriorhodopsin and Rhodopsin. First step was the construction of the transmembrane domains of the two proteins. Ideal α-helices were generated for each of the putative transmembrane regions by imposing values of $\phi = -57°$ and $\psi = -47°$ to the dihedral angles of each amino acid residue, except for prolines, where $\phi = -65°$. Side chains were generated in an extended conformation. Atomic coordinates were computed with the help of the AMBER4.0 program.[17] Helices

TABLE 1: Bundle Parameters Deduced from the 2D Projection of Bacteriorhodopsin Assuming the Coordinate Origin Placed at the Center of Helix II

helix	helix center (x,y)/Å	tilt angle (θ)/degrees	tilt orientation (χ)/degrees	bisector vectors (x,y)/Å
helix I	−3.05, −8.13	−20.04	−56.31	−9.30, −5.82
helix II	0.00, 0.00	−8.63	38.22	−5.61, 6.63
helix III	9.14, 1.78	12.83	−84.29	−3.72, 8.54
helix IV	17.53, 7.87	−10.71	0.0	1.22, 10.29
helix V	23.37, 1.02	15.85	29.74	8.99, 0.53
helix VI	17.27, −8.13	14.64	50.19	5.70, −9.40
helix VII	7.87, −9.14	14.23	71.56	0.08, −9.45

TABLE 2: Bundle Parameters Deduced from the 2D Electron Density Map of Rhodopsin Assuming the Coordinate Origin Placed at the Center of Helix VI

helix	helix center (x,y)/Å	tilt angle (θ)/degrees	tilt orientation (χ)/degrees	bisector vectors (x,y)/Å
helix I	−9.89, −15.71	10.63	−59.74	−9.43, −0.84
helix II	0.58, −20.56	−25.89	75.96	2.02, −11.37
helix III	8.15, −13.00	25.92	0.0	5.91, −8.92
helix IV	17.27, −9.31	21.54	0.0	9.25, −3.35
helix V	11.25, 1.94	16.76	−15.94	5.60, 11.46
helix VI	0.00, 0.00	13.94	33.69	−7.94, 8.21
helix VII	−2.13, −10.28	14.88	5.71	−8.77, 5.77

were arranged automatically to generate a bundle, using the in-house program BUNDLE.[14] This program treats each helix as a rigid body and arranges the helices to form a bundle with use of information derived from the electron density two-dimensional (2D) projection map of a given protein. Accordingly, helix centers, helix tilt orientation, and vectors pointing in the direction of the bisector of the angles defined by the centers of three consecutive helices in the helix bundle reported in Tables 1 and 2 were computed from the electron density maps of *H. halobium* bacteriorhodopsin[18] and bovine rhodopsin,[4] respectively, following the procedure described previously.[13,14] This information was provided to BUNDLE to construct the helix bundle in a completely automatic fashion. The main steps performed by the program on each helix to construct the unrefined models are: (i) transformation of the atomic coordinates of a helix to a local, right-handed orthonormal coordinate system, centered at the mean position of the helix and with one of the coordinate axes pointing in the direction of the helical one; (ii) rotation of the helix around one of the nonhelical axes to generate the adequate antiparallel orientation of the helices in the bundle; (iii)

```
            1                                                  I          50
opsd_bovin  MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML

            51                          II                             100
opsd_bovin  GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH

            101          III                                           150
opsd_bovin  GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE

            151          IV                                            200
opsd_bovin  NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN

            201          V                                             250
opsd_bovin  ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV

            251          VI                                 VII        300
opsd_bovin  TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV

            301                                                        348
opsd_bovin  YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
```

Figure 2. Amino acid sequence of bovine rhodopsin retrieved from the SwissProt database with the putative seven transmembrane helices highlighted.

computation of its hydrophobic moment according to the method described by Eisenberg et al.[19]; (iv) rotation of the helix around its helical axis to force its hydrophobic moment to point outward from the helix bundle and in the direction of the bisector of the angle between the centers of the previous helix, the present helix, and the next one; (v) tilting of the helix and translation of the helix centers, according to the available information deduced from the 2D projection of the two proteins, respectively.

Once the transmembrane domains were constructed, the amino acid segments corresponding to the loops were added to be bonded to the corresponding helical termini and in an extended conformation.

3. Ligand Conformations and Construction of the Starting Structures of the Complexes. Molecular mechanics calculations were carried out with AMBER 4.0[17] using the all-atom parm91 parameter set.[20] The structures of the all-*trans*−retinal and 11-*cis*−retinal molecules, the chromophores of bacteriorhodopsin and rhodopsin, respectively, were built using the PREP module of AMBER 4.0[17] guided by the corresponding structures taken from the Cambridge data bank. Atomic partial charges for the ligands were generated by fitting the molecular electrostatic potential computed with a STO-3G basis set using the GAUSSIAN94 suite of programs.[21] Figure 3 shows the structure and labeling for the all-*trans*−retinal and 11-*cis*−retinal. The two molecules were energy minimized and subsequently, manually docked inside the helix bundle of bacteriorhodopsin and rhodopsin, respectively. The proper orientation of bacteriorhodopsin was guided by its 3D structure.[16] Because there was no direct evidence about its orientation in the binding site, rhodopsin was placed manually inside the helix bundle and rotated to bring residues known to be involved in ligand−receptor interactions close to the corresponding ligand moieties, according to data.[22−35]

4. Model Refinement. The unrefined 3D models of bacteriorhodopsin and rhodopsin first were subjected to energy minimization. The minimization procedure was carried out with a distance-dependent dielectric constant of 4r according to the following protocol: The systems were first subjected to 1000 cycles of steepest descent, keeping all the backbone atoms frozen

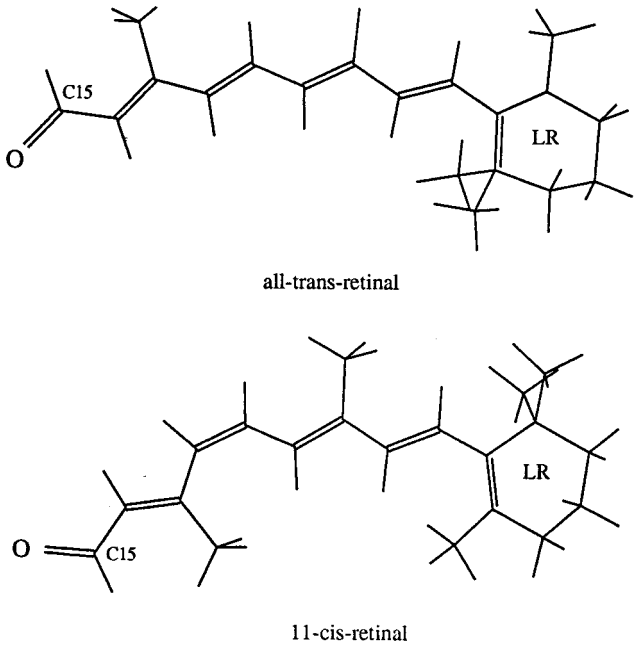

Figure 3. (top) Labeling scheme for the all-*trans*−retinal chromophore; (bottom) same as the top, but for 11-*cis*−retinal. LR denotes a dummy atom located on the centroid of the β-ionone ring.

in their initial positions to eliminate the worst steric interactions between atoms of the side chains. In a subsequent step, a conjugate gradient minimization was carried out with all the backbone atoms restrained by harmonic potentials and with the helix lengths and the distances between consecutive helices fixed. Convergence on the root mean square (rms) distance was set to 0.001 Å for the structures of two successive iterations. At this point, the harmonic potential restraints on all the backbone atoms were lifted and minimization was carried out until reaching a rms distance of 0.001 Å between the structures of two successive iterations. Finally, the restraints obtained by fixing of the helix lengths and the distances between consecutive helices were lifted. First distance constraints between consecu-

tive helices were slowly reduced until the rms distance between the structures of two consecutive iterations was less than 0.001 Å. Similarly, restraints fixing helix length were suppressed, carrying out a last minimization of the complexes completely relaxed.

MD simulations on rhodopsin were initiated by heating the minimized structure with all the restraints activated up to 300 K during 100 ps at a constant rate of 30 K/10 ps. After heating, a 100 ps of equilibration followed, keeping all the restraints at a constant temperature of 300 K. Further, 100 ps of equilibration followed with ligand coordinates and helix length and distance between neighbor helices constrained. Finally, a production run of 250 ps was carried out with the complexes completely relaxed. The time step was set to 2 fs using the SHAKE algorithm. A cutoff of 12 Å was used to compute nonbonded interactions, with the nearest neighbor list that was used updated every 10 steps.

Results

Structure of the Membrane-Spanning Regions of Bacteriorhodopsin. A 3D model of the transmembrane segments of bacteriorhodopsin was constructed using the program BUNDLE with the structural parameters listed in Table 1, deduced from its 2D electron density map projection.[18] In addition to these parameters, the program requires as input the sets of atomic Cartesian coordinates of each of the transmembrane helices whose sequences are highlighted in Figure 1. The coordinate sets subsequently were manipulated automatically by BUNDLE, following the steps described in Methods to generate the helix bundle. The ligand structure was first energy minimized and manually docked inside the seven-helix bundle pocket as explained in Methods. The initial structure of the ligand-opsin complex was energy minimized using the procedure described in the methods section.

The model of bacteriorhodopsin was constructed to assess the level of accuracy that should be expected from these models. For this purpose the coordinates of the energy-minimized model were compared with those of the experimental structure[16] solved at 3.5 Å resolution. To make a direct comparison, the experimental structure was subjected to energy minimization completely relaxed, in the same conditions as the 3D model constructed in this work. Coordinates of the C^{α} atoms of the 3D model of bacteriorhodopsin were superimposed with the corresponding coordinates of the energy-minimized structure of 2BRD. The rms distance after optimal superimposition of the Cα atoms of the transmembrane domains was 3.8 Å.

Structure of the Membrane-Spanning Regions of Rhodopsin. An unrefined structure of the membrane-spanning segments of bovine rhodopsin was constructed using the BUNDLE program with the parameters listed in Table 2 and consistent with the 2D electron density projection map[4] of the newly reported structure of frog rhodopsin.[3] In addition, the program requires as input the sets of atomic Cartesian coordinates of each of the transmembrane helices relative to an arbitrary origin. Coordinates were transformed by BUNDLE following the steps described in Methods to generate the helix bundle. The pre−energy-minimized model of rhodopsin presented the following residues facing the interior of the bundle. Specifically, residues Asp^{83} and Gly^{90} in helix II; Glu^{113}, Gly^{114}, Phe^{115}, Ala^{117}, Gly^{121}, Glu^{122}, Leu^{125}, and Trp^{126} in helix III; Ala^{164} in helix IV; Phe^{208} and His^{211} in helix V; Phe^{261}, Trp^{265}, and Tyr^{268} in helix VI; Ala^{292}, Lys^{296}, Asn^{302}, Ile^{305}, and Tyr^{306} in helix VII. These residues have been proposed by different experimental studies to be involved in ligand−receptor interactions, and are depicted explicitly in Figure 4.

TABLE 3: Selected Inter- and Intramolecular Distances Measured Between the Corresponding Atoms Listed in Columns 1 and 2 on the 11-*cis*−retinal Rhodopsin Complex

atom-residue	atom-residue	d/Å	ref
Cα-Gly^{90}	Nζ-Lys^{296}	13.8	30
Cδ-Glu^{113}	Nζ-Lys^{296}	6.9	43
Cβ-Ala^{117}	Nζ-Lys^{296}	9.0	35
Cδ-Glu^{122}	LR−His^{211}	8.5	24, 25, 31
atom-residue	atom-rhodopsin		
Cα-Gly^{90}	C_{15}	13.0	30
Cδ-Glu^{113}	O	5.1	43
Cβ-Ala^{117}	C_{15}	6.3	35
Cδ-Glu^{122}	LR	4.5	24, 25, 31
Cβ-Ala^{164}	LR	9.6	27
LR-Phe^{208}	LR	11.2	28
LR-His^{211}	LR	10.6	28
LR-Phe^{261}	LR	14.5	27
LR-Trp^{265}	LR	14.4	24, 25, 31
LR-Tyr^{268}	LR	6.4	24, 25, 31
Cβ-Ala^{269}	LR	9.9	29, 33
Cβ-Ala^{292}	C_{15}	5.2	23
Nζ-Lys^{296}	O	6.6	29, 33

[a] LR refers to a dummy atom located on the centroid of the corresponding rings. Relevant experimental results suggesting the neighborhood between the different pairs of atoms are provided in column 4. Upper limits of these distances, as deduced from the experimental studies, are reported in Table 3 of ref 12.

The ligand structure was first energy minimized and manually docked into the seven-helix bundle pocket, guided by the results of different experimental studies. The oxygen O1 of 11-*cis*−retinal was placed at a distance of about 2 Å from the nitrogen of Lys^{296} side chain located in helix VII, according to the results gathered from Raman spectroscopy[29] and solid-state NMR[33] studies, which suggest a covalent link between Lys^{296} and 11-*cis*−retinal.

The initial structure of ligand-opsin complex was energy minimized and subjected to a MD simulation at 300 K as explained in Methods. Distance fluctuations between the interacting side chains of rhodopsin and 11-*cis*−retinal were monitored during the trajectory. Average distances along the 250-ps production run between a subset of the rhodopsin side chains highlighted in Figure 4 and the ligand as well as a few intermolecular distances are reported in Table 3. Fluctuations of these distances are shown in Figure 5.

Discussion

The model obtained for bacteriorhodopsin provides a test of this new methodology. Rms distance obtained from the comparison of the Cα atoms of the transmembrane domains of the refined model and the minimized crystal structure is 3.8 Å. This value is within the resolution of the crystal structure (3.5 Å) and validates the de novo approach to build the transmembrane domains of GPCRs used in the present work.[36]

In contrast to bacteriorhodopsin, there are no experimental results for rhodopsin or any GPCR that can be used as guidance to perform a fully reliable assessment of the residues involved in each of the helical transmembrane segments. There are several procedures described in the literature to assess the sequences of the transmembrane domains,[37,38] including the hydrophobicity profile,[39] variability index,[7] multiple alignments,[40] or combined methods.[41] The use of reliable methods to predict the sequence of the transmembrane domains is critical for the quality of the 3D models. Differences between alternative methods of predicting the transmembrane segments affect (i) protrusion from the membrane core, (ii) the relative axial displacement between

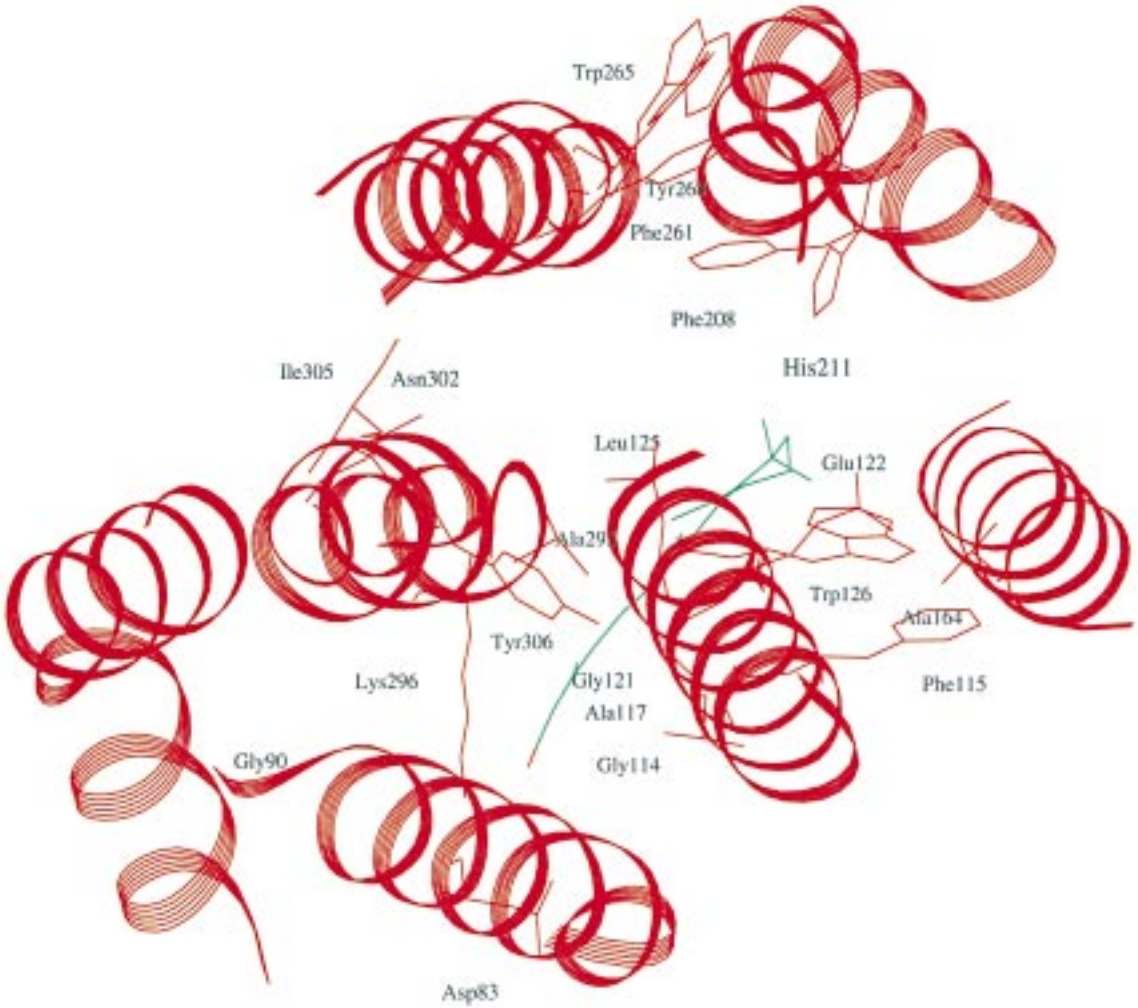

Figure 4. Model of rhodopsin constructed by BUNDLE before any energy minimization. Residues proposed by different experimental studies to be orientated toward the inside of the bundle are depicted explicitly: Asp83 and Gly90 for helix II; Gly114, Phe115, Ala117, Gly121, Glu122, Leu125, and Trp126 for helix III; Ala164 for helix IV; Phe208 and His211 for helix V; Phe261, Trp265, and Tyr268 for helix VI; Ala292, Lys296, Asn302, Ile305, and Tyr306 for the helix VII.

neighbor helices, and (iii) eventually helix orientation. Obviously, in the present work, the model built for bacteriorhodopsin avoids this source of inaccuracies because transmembrane regions were taken directly from the experimental structure. However, rhodopsin, inaccuracies in the prediction of the transmembrane segments may induce errors in the axial displacement between neighbor helices. If inaccuracies in the assessment of helix beginning are similar to those regarding the assessment of helix ending, the present approach will properly calculate the *z*-coordinates of the helices because of a compensation of errors. However, if one helix end is properly assessed but not the other, the error in computing its *z*-coordinates could be of the order of a turn of a helix, i.e., ≈5 Å. Accordingly distances measured in the complexes need to be interpreted within a certain tolerance.

As expected, initial models of the helix bundles exhibit numerous atomic contacts between the side chains, because they are in extended conformations. For this reason energy minimization is performed in several steps. First, only side chains are relaxed avoiding the worst steric contacts. Subsequent minimization is performed with helix lengths and distances between neighboring helices constrained. After this step is completed, the whole structure is relaxed without any constraint. This brings the helix bundle to a local energy minimum close to the initial structure, where all steric effects are removed. MD calculations are carried out from this initial structure, although a similar protocol needs to be considered during the heating and equilibration period to avoid the unpacking of the helices.

Atom−atom distances can be easily measured with use of the present model. A set of the most significant inter- and intramolecular distances are listed in Table 3. References of the experimental studies supporting the involvement of the specific residues monitored are also provided. First, the chromophore is adequately oriented in the neighborhood of Lys^{296}, consistent with the covalent linkage actually formed between the chromophore and the opsin.[42] The distance between Lys^{296} and the carbonyl oxygen of the chromophore in the MD simulation fluctuates between 4 and 8 Å (Figure 5a). This indicates that the ligand is well accommodated inside the receptor and close to Lys^{296}. Similarly, Glu^{113} remains close to both, the ligand and residue Lys^{296} consistently with its putative role as counterion for the Schiff base nitrogen formed between Lys^{296} and the chromophore,[34,43] as shown in Figures 5b and 5c. Specifically, Figure 5b confirms the proximity of Glu^{113} to the ligand during the 250 ps of the simulation, whereas Figure 5c shows the fluctuations of the distance Glu^{113}−Lys^{296} that correlate well with those shown in Figure 5a. Ala^{292} and Gly^{90} are also known to be in the neighborhood of the Schiff base nitrogen, because mutations $Ala^{292}Glu^{23}$ and $Gly^{90}Asp^{44}$ cause congenital night blindness because of a competitive effect of

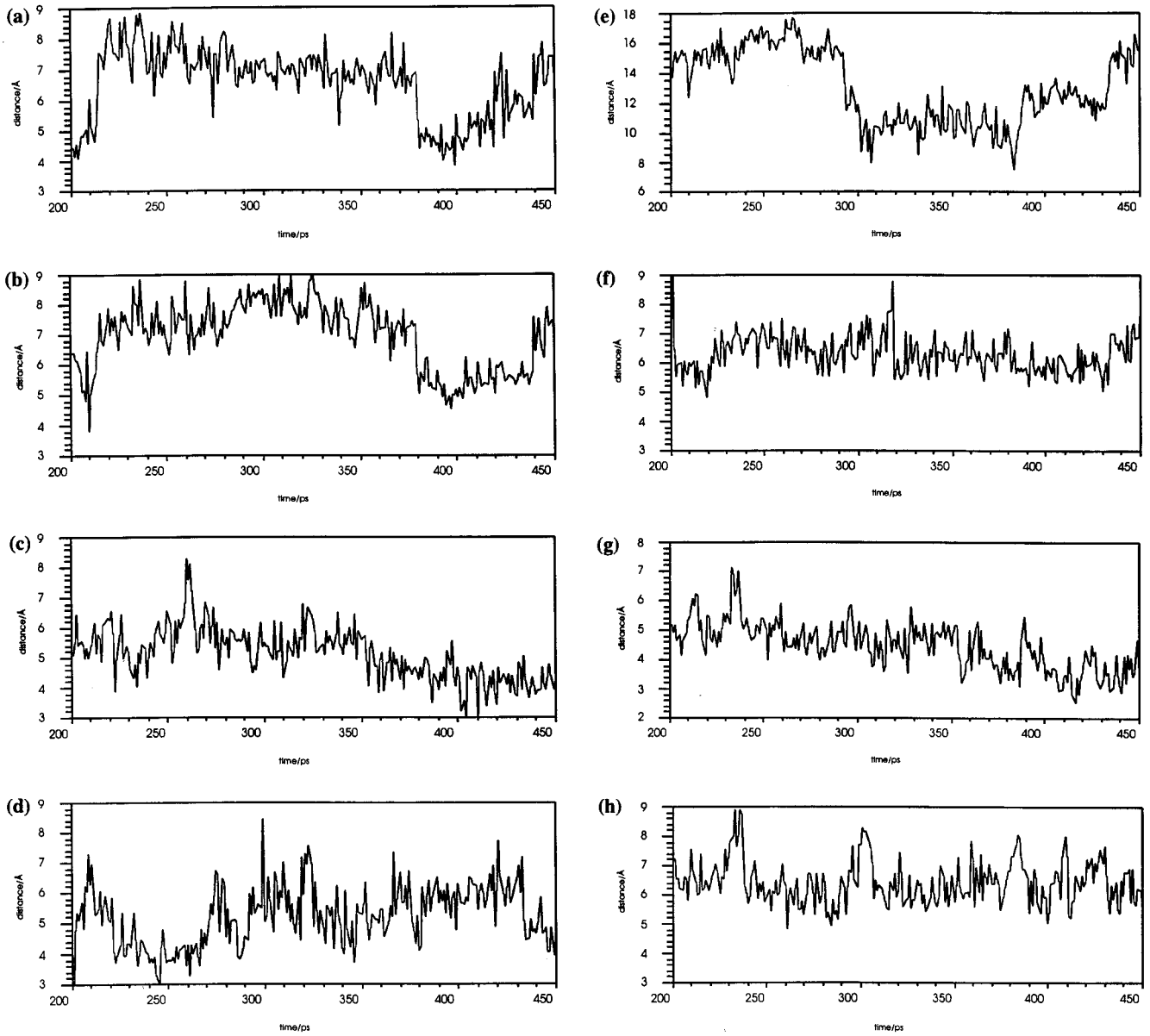

Figure 5. Fluctuations of selected distances reported in Table 3 during the production run of the opsin−chromophore complex: (a) Lys296-O; (b) Glu113-O; (c) Glu113-Lys296; (d) Ala292-C15; (e) Gly90-C15; (f) Ala117-C15; (g) Glu122-LR; (h) Tyr268-LR.

these residues with Glu^{113} for the positively charge nitrogen on Lys^{296}. In our model, Ala^{292} is at a short distance of the ligand in accordance with experimental observations. Figure 3d shows the fluctuations of this distance along the MD trajectory suggesting that the residue is in the neighborhood of the ligand during the simulation. In contrast, the distance ligand−Gly^{90} is too long in the present model. Although the model shows that Gly^{90} is adequately oriented toward the interior of the bundle, the residue was assigned to belong to the loop connecting helices II and III (Figure 2). Accordingly, it can be deduced that the axial displacement of helix II is the origin of the long distance observed between this residue and the chromophore. Figure 5e shows that the residue exhibits two different orientations: one at 9 Å and the other at 15 Å of the ligand, suggesting that the residue is properly oriented to interact with the chromophore. Experimental evidence suggests that Ala^{117} is also close to the ligand, because experiments with the double-replacement mutant $Glu^{113}Ala/Ala^{117}Glu^{35}$ indicate that the position of the Schiff base counterion is displaced one helix turn. This result is consistent with the predictions based on the present model as can be seen by the small fluctuations of Ala^{117} ligand distance (Figure 5f).

Several residues have been implicated to be interacting with the β-ionone ring of the 11-*cis*−retinal,[12] based on mutagenesis experiments that alter the spectral characteristics of the chromophore or affect the activation of the G-protein. Distances measured to the center of the β-ionone ring to several residues are within van der Waals range of distances for most of the residues. To illustrate, Figures 3g and 3h show the fluctuation of the distance between the chromophore ring and residues Glu^{122} and Trp^{265}, respectively. In contrast distances between the ligand and Trp^{265} and Phe^{261} appear long. In our model these two residues are not oriented toward the interior of the bundle, but involved in intramolecular interactions with other hydrophobic residues. Accordingly, these residues may be important for the stability of the receptor, rather than for direct interaction with the receptor.

The present model compares well with many of the features of other de novo models of rhodopsin reported in the literature.

Thus, distances reported in the present study fit well with those used by[12] with the exception of that of residue Gly^{90}. However, this inconsistency is caused by the erroneous assignment of Gly^{90} as part of a loop. Also distances between retinal and residues Trp^{265} and Phe^{261} are longer in the present model. This is because these residues are not adequately oriented toward the interior of the receptor. However, they exhibit hydrophobic intramolecular interactions with neighbor residues that may be important for the structure of the receptor and explain the experimental observations in their corresponding mutants. Baldwin et al.[9] recently published a model of rhodopsin. The model was constructed using the electron density map of a refined structure of frog rhodopsin.[3] In a way similar to our procedure Baldwin's model was constructed from the computation of straight axes for helices I−IV and VII and two axes for each of helices V and VI enabling a bent structure. These axes were determined by a trial-and-error procedure guided by the electron density map. The authors reported values of the (x,y) coordinates of the helical axes as well as their tilt and azimuthal angles, in a fashion similar to that reported in Table 2. Regarding the positions of the axes on the $Z = 0$ plane, it should be considered a certain uncertainty in the figures reported, because in simultaneous publications by the same authors,[3,9] differences can be found in the values of the coordinates of up to 2 Å. To compare these coordinates with those of the present work reported in Table 2, it is necessary to carry out the transformation between the two coordinate frameworks. Although both sets of coordinates exhibit the same trends there are differences of 5 Å in some cases. Another differential feature between the two models that should be considered concerns the distances between pairs of neighbor helix centers. With use of the coordinates reported in Table 2, distances are ≈10 Å, corresponding to the distance between two packed helices, whereas in the model proposed by Baldwin et al.,[9] distances are shorter than 8 Å in some cases. Tilt angles are comparable, although in some cases there are differences of 15°. However, these differences should not be considered significant because this information is used only in the present work for the generation of an unrefined starting model and consequently, a subsequent refinement procedure is required to produce a more reliable model.

Extension of the Present Procedure to Other GPCRs. Most of the models of GPCRs published in the literature are constructed by homology modeling using rhodopsin or bacteriorhodopsin as a template.[5] This is unfortunate because GPCRs exhibit low sequence identity with any of these templates, and furthermore, the latter is not even a GPCR. Indeed, the paradigm on which homology modeling studies are based states that two proteins sharing a high sequence identity and carrying out similar functions must exhibit a similar 3D structure. Thus, models based on bacteriorhodopsin may contain errors from two sources. First, those associated with a low sequence identity between the target protein and the template and second, those that occur because the template does not carry out the same function as the protein modeled. On the other hand, protocols to model GPCRs using rhodopsin as a template do not exhibit errors associated with this latter source. However, because the atomic coordinates of rhodopsin are not available, it is necessary to construct a 3D structure of this protein first. Models of this protein have been constructed by manual manipulation using interactive graphics as the modeling technique and imposing restraints generated from experimental and theoretical data. The resulting low-resolution structure is then used as template for modeling the GPCR of interest. To generate a molecular model of any GPCR, the putative transmembrane regions are first aligned with those of rhodopsin to optimize the alignment of hydrophobic and hydrophilic residues within each of the transmembrane helices. Many times the hydrophobic segments do not have the same length as those in the template consequently adding an additional uncertainty to the models.

One of the few exceptions to this general procedure is based on sequence divergence analysis and provides an independent outcome to be contrasted with other models.[7] However the use of this procedure is hampered by the limited number of sequences with a high degree of identity available in a family.

The procedure presented in this work represents a totally general, automated technique for modeling membrane-embedded bundles of seven-helix membrane proteins, assuming that they adopt a similar topology to the one observed in rhodopsin. The major strength of the present approach lies in that it is not a homology modeling based procedure.

Conclusions

The general automated technique proposed in this work allows the construction of seven helix bundles of any GPCR with low homology with rhodopsin. No structure is used as a template for model building. The method can be considered as completely ab initio because no experimental or theoretical information is used, excluding the quantified values of tilting and helices positions deduced from projection map of rhodopsin. Taking the coordinates of ideal α-helices for the predicted transmembrane region as the only input, the BUNDLE program permits us to calculate hydrophobic moments for each helix, to rotate them so that hydrophobic moment vectors point out of the bundle, and to locate and tilt helices according to the values deduced from the projection map of rhodopsin. Because of the general approach of this method, the developed program can provide a good approximation for modeling of new GPCR with a low homology with rhodopsin and for which few experimental data are known. In this case, the homology modeling suggests useful experiments that can be used to understand the function of the particular GPCR studied.

Acknowledgment. This research was partially supported by the Human Capital and Mobility program "Access to Large Installations", under contract CHGE-CT92-0009 between The European Community and CESCA/CEPBA. Financial support from the "Programma di Intervents per la promozione della Ricerca Scientifica in Campania" LR no. 41-31/12/94 is also acknowledged.

References and Notes

(1) (a) Beck-Sckinger, A. G. *Drug Discov. Today* **1996**, *1*, 502. (b) Bikker, J. A.; Trumpp-Kallmeyer, S.; Humblet, C. *J. Med. Chem.* **1998**, *41*, 2911.

(2) Schertler, G. F. X.; Hargrave, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11578.

(3) Unger, V. M.; Hargrave, P. A.; Baldwin, J. M.; Schertler, G. F. X. *Nature* **1997**, *389*, 203.

(4) Schertler, G. F. X.; Villa, C.; Henderson, R. *Nature* **1993**, *362*, 770.

(5) Donnelly, D.; Findlay, J. B. C. *Curr. Opin. Struct. Biol.* **1994**, *4*, 582.

(6) Hoflack, J; Trumpp-Kallmeyer, K. S.; Hilbert, M. *Trends Pharmacol. Sci.* **1994**, *15*, 7.

(7) Alkorta, I.; Du, P. *Protein Eng.* **1994**, *7*, 1231.

(8) Taylor, W. R.; Jones, D. T.; Green, N. M. *Proteins Struct. Funct. Genet.* **1994**, *18*, 281.

(9) Baldwin, J. M.; Schertler, G. F. X.; Unger, V. M. *J. Mol. Biol.* **1997**, *272*, 144.

(10) Donnelly, D.; Findlay, J. B. C.; Blundell, T. L. *Receptor Channels* **1994**, *2*, 61.

(11) Kontoyianni, M.; Lybrand, T. P. *Drug Design Discovered* **1993**, *9*, 29.
(12) Herzyk P.; Hubbard, R. E. *Biophys. J.* **1995**, *69*, 2419.
(13) Perez, J. J.; Filizola, M.; Carteni-Farina, M. *J. Math. Chem.* **1998**, *23*, 229.
(14) Filizola, M.; Perez, J. J.; Carteni-Farina, M. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 111.
(15) Bairoch, A.; Bieckmann, B. *Nucleic Acids Res.* **1994**, *22*, 3578.
(16) Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, *259*, 393.
(17) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 4.0; University of California: San Francisco, 1991.
(18) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. *J. Mol. Biol.* **1990**, *213*, 899.
(19) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. *Nature* (*Lond.*) **1982**, *299*, 371.
(20) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J. A. Gaussian94, Revision B.2; Gaussian Inc.: Pittsburgh, PA, 1995.
(22) Cohen, G. B.; Oprian, D. D.; Robinson, P. R. *Biochemistry* **1993**, *32*, 6111.
(23) Dryja, T. P.; Berson, E. L.; Rao, V. R.; Oprian, D. D. *Nat. Genet.* **1993**, *4*, 280.
(24) Nakayama, T. A.; Khorana, H. G. *J. Biol. Chem.* **1990**, *265*, 15762.
(25) Nakayama, T. A.; Khorana, H. G. *J. Biol. Chem.* **1991**, *266*, 4269.
(26) Nathans, J. *Biochemistry* **1990**, *29*, 9746.
(27) Neitz, M.; Neitz, J.; Jacobs, G. H. *Science* **1991**, *252*, 971.
(28) Oprian, D. D. *J. Bioenerg. Biomembr.* **1992**, *24*, 211.
(29) Palings, I.; Pardeon, J. A.; Vandenberg, E.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. *Biochemistry* **1987**, *26*, 2544.
(30) Rao, V. R.; Cohen, G. B.; Oprian, D. D. *Nature* **1994**, *367*, 639.
(31) Ridge, K. D.; Bhattacharya, S.; Nakayama, T. A.; Khorana, H. G. *J. Biol. Chem.* **1992**, *267*, 6770.
(32) Robinson, P. R.; Cohen, G. B.; Zhukovsky, E. A.; Oprian, D. D. *Neuron* **1992**, *9*, 719.
(33) Smith, S. O.; Palings, I.; Miley, M. E.; Courtin, J.; de Groot, H.; Lugtenburg, J.; Mathies, R. A.; Griffin, R. G. *Biochemistry* **1990**, *29*, 8158.
(34) Zhukovsky, E. A.; Oprian, D. D. *Science* **1989**, *246*, 928.
(35) Zvyaga, T. A.; Fahmy, K.; Sakmar, T. P. *Biochemistry* **1994**, *33*, 9753.
(36) Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.
(37) Fasman, G. D.; Gilbert, W. A. *Trends Biochem. Sci.* **1990**, *15*, 89.
(38) Jähning, F. *Trends Biochem. Sci.* **1990**, *15*, 93.
(39) Degli Esposti, M.; Crimi, M.; Venturoli, G. *Eur. J. Biochem.* **1990**, *190*, 207.
(40) Persson B.; Argos, P. *J. Mol. Biol.* **1994**, *237*, 182−192.
(41) Rost, B.; Casadio, R.; Fariselli, P.; Sander, C. *Protein Sci.* **1995**, *4*, 521.
(42) Dratz, E. A.; Hargrave, P. A. *Trends Biochem. Sci.* **1983**, *8*, 128.
(43) Sakmar, T. P.; Franke, R. R.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8309.
(44) Sieving, P. A.; Richards, J. E.; Naarendorp F.; Bingham, E. L.; Scott, K.; Alpern, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 880.