Automated Method for Modeling Seven-Helix Transmembrane Receptors from Experimental Data

Pawel Herzyk and Roderick E. Hubbard
Department of Chemistry, University of York, Heslington, York YO1 5DD, United Kingdom

ABSTRACT A rule-based automated method is presented for modeling the structures of the seven transmembrane helices of G-protein-coupled receptors. The structures are generated by using a simulated annealing Monte Carlo procedure that positions and orients rigid helices to satisfy structural restraints. The restraints are derived from analysis of experimental information from biophysical studies on native and mutant proteins, from analysis of the sequences of related proteins, and from theoretical considerations of protein structure. Calculations are presented for two systems. The method was validated through calculations using appropriate experimental information for bacteriorhodopsin, which produced a model structure with a root mean square (rms) deviation of 1.87 Å from the structure determined by electron microscopy. Calculations are also presented using experimental and theoretical information available for bovine rhodopsin to assign the helices to a projection density map and to produce a model of bovine rhodopsin that can be used as a template for modeling other G-protein-coupled receptors.

INTRODUCTION

G-protein-mediated transmembrane signaling modulates a great number of physiological events in a wide range of organisms by the transmission of a variety of external signals through G-protein-coupled receptors (GPCRs). The external signals can be as diverse as light, odorants, neurotransmitters, and peptide hormones, which affect a GPCR to cause activation of a guanine nucleotide-binding protein (G-protein) on the intracellular side of the receptor. The activated G-protein initiates an effector system that produces an intracellular signal (Strader et al., 1994; Watson and Arkinstall, 1994). For these reasons, understanding the structure and mechanism of GPCRs is central to understanding many aspects of cellular signaling and control, and these molecules are also important targets for therapeutic agents, attracting considerable pharmaceutical interest.

Currently more than 700 GPCRs have been cloned and sequenced (Oliveira et al., 1993; G. Vriend, personal communication). Multiple sequence alignment of the family of GPCRs shows seven regions of high homology that are also the regions of highest hydrophobicity. Detailed sequence analysis reveals that these regions contain motifs that can be used as fingerprints in identifying new GPCRs (Attwood and Findlay, 1993). Fourier transform analysis of conserved residues has revealed the presence of α-helical periodicity (Donnelly et al., 1989), which extends earlier experimental work on rhodopsin (Chabre, 1985), which demonstrated that the transmembrane domain was highly α-helical. In addition, the two-dimensional (2-D) electron density projection map of bovine rhodopsin at 9 Å resolution shows a pattern that can be readily interpreted if the transmembrane domain is a seven-helix bundle (Schertler et al., 1993).

A common picture that emerges from this evidence is that the GPCR polypeptide chain spans the lipid bilayer seven times so that amino and carboxy termini are on extra- and intracellular sides of the membrane, respectively, with the transmembrane part composed of seven α-helices corresponding to the seven regions of increased hydrophobicity. This pattern is similar to the electron microscopy (EM) structure of bacteriorhodopsin (bR) (Henderson et al., 1990), a light-driven proton pump in *Halobacterium halobium*, which explains why the bulk of the 3-D structure modeling of GPCRs uses the EM structure of bR as a template (Cronet et al., 1993; Dahl et al., 1991; Findlay and Eliopoulos, 1990; Grötzinger et al., 1991; Hibert et al., 1991; Ijzerman et al., 1992; Lewell, 1992; Nordvall and Hacksell, 1993; Rippmann and Bottcher; Sylte et al., 1993; Trumpp-Kallmeyer et al., 1992; Underwood et al., 1994; Vriend; Yamamoto et al., 1993; Zhang and Weinstein, 1993). These models have mainly been constructed by using manual manipulation of structures through interactive graphics as the modeling technique.

There are, however, a number of arguments against using bR as a template for GPCRs. First, bR is not a member of the GPCR family. bR carries out light-dependent proton translocation from the inside to the outside of the *H. halobium* cell and its response to light is not coupled to G-protein activation. Secondly, there is no significant homology between the bR sequence and any of the GPCRs, suggesting there is no evolutionary link, and detailed sequence analysis (Attwood and Findlay, 1993) revealed that the characteristic GPCR fingerprint is absent in bR. Two hypotheses have been formulated to show that this link may exist. The first one implies that the homologous helices in bR and GPCRs are not colinear, which may have been caused by an exon shuffling event that occurred during the

Received for publication 9 June 1995 and in final form 15 August 1995.
Address reprint requests to Dr. Pawel Herzyk, Department of Chemistry, University of York, Heslington, York YO1 5DD, UK. Tel.: 011-44-1904-43-2599; Fax: 011-44-1904-41-0519; E-mail: pavel@yorvic.york.ac.uk.

0006-3495/95/12/2419/24 $2.00

evolution of the GPCRs and bR from a common ancestor (Pardo et al., 1992). The second suggests that helices 5–7 originated from helices 1–3 as a result of ancestral gene duplication, which may have led to the situation that helices 1–3 in bR are homologous to 5–7 in the GPCRs (Taylor and Agarwal, 1993). It was, however, recently demonstrated that sequence analysis supports neither of these hypotheses (Soppa, 1994). Third, although both bR and rhodopsin respond to light through isomerization of a retinal molecule that in both cases is covalently bound to lysine on helix G, the retinal molecules differ in both configuration and conformation. The isomerization is all-*trans* → 13-*cis* in bR and 11-*cis* → all-*trans* in rhodopsin (Wald, 1968). Furthermore, the conformation of the C_6-C_7 bond in retinal are different, being 6-*s*-*trans* in bR (Harbison et al., 1985) and 6-*s*-*cis* in rhodopsin (Smith et al., 1987). Fourth, the low resolution projection map of bovine rhodopsin reveals that, although the overall topology of rhodopsin is identical to that of bR, the actual arrangement shows significant differences (Schertler et al., 1993).

The idea of 3-D modeling of GPCRs de novo is not new. The existing models, however, either use automated techniques based only on detailed sequence analysis and multiple sequence alignment (Alkorta and Du, 1994; Taylor et al., 1994) or use interactive graphics to visually satisfy selected restraints (Donnelly et al., 1994; Kontoyianni and Lybrand, 1993; MaloneyHuss and Lybrand, 1992). The existing modeling procedures do not use the full range of available experimental and theoretical data to direct the modeling process. This will become increasingly important as the number of new GPCR sequences grows rapidly and the range and quality of experimental and theoretical data is extended.

The aim of this work is to propose a rule-based automated technique for aggregating helices in seven-helix membrane proteins using restraints generated from experimental and theoretical data. The resulting low resolution structure provides a template that can then be used for further modeling. Any 3-D structure modeling method used for the GPCRs needs first to be tested on bR for which a high resolution EM 3-D structure is known (Henderson et al., 1990). In this paper, we apply our method to bR and use this test case protein to explore the characteristics of the method. Finally, we describe the generation of a template for rhodopsin and discuss the way the methodology presented can be used for other GPCR modeling.

METHODS

The four main steps in our method are to (1) analyze the available experimental and theoretical data to derive geometrical restraints, (2) use a protein representation adequate for applying the restraints, (3) construct a penalty function that efficiently penalizes violations of the restraints, and (4) optimize the penalty function to find a family of structures that best satisfy the restraints. Each of these steps is described separately below.

Protein representation

Much of the structural information that will be derived as restraints in our modeling relates to helices and residues rather than atoms. Therefore we have decided to reduce the protein representation so that each residue is represented by one virtual C_α atom with a virtual C_β side-chain atom positioned in the centroid of a side chain with a radius dependent on the size of the amino acid side chain. This is a simplified version of our earlier protein-reduced representation (Herzyk and Hubbard, 1993). Details are presented in Appendix A.

Transmembrane helices are currently considered as rigid with rigidly assigned C_α and C_β virtual atoms. The centroid position and the extracellular and intracellular end of the helix are represented by three dummy atoms labeled DC, DE, and DI. These dummy atoms can be used as anchor points for certain restraints but do not take part in the steric overlap calculations. The dummy atoms for particular helices are designated as, for example, DE-A for the extracellular end of helix A.

A ligand is treated as a rigid molecule represented by all non-hydrogen atoms. In this study, the ligand is retinal, which is designated as RET; particular atoms are labeled according to convention as shown in Fig. 1 and dummy atoms are defined as LS, LC, and LR representing the Schiff base, C11 atom, and β-ionone ring (Heyn et al., 1988), respectively. The van der Waals radii are defined by PARAM19 parameter set (Brünger, 1990) and presented in Appendix A.

Generation of restraints

Table 1 lists the types of experimental and theoretical studies that we have used or can potentially be used in the derivation of the structural restraints. Each of these can produce some kind of geometrical relationship between residues or portions of the protein molecule. In addition, the physical principles of protein structure and the presence of the membrane bilayer can be used to further constrain the model.

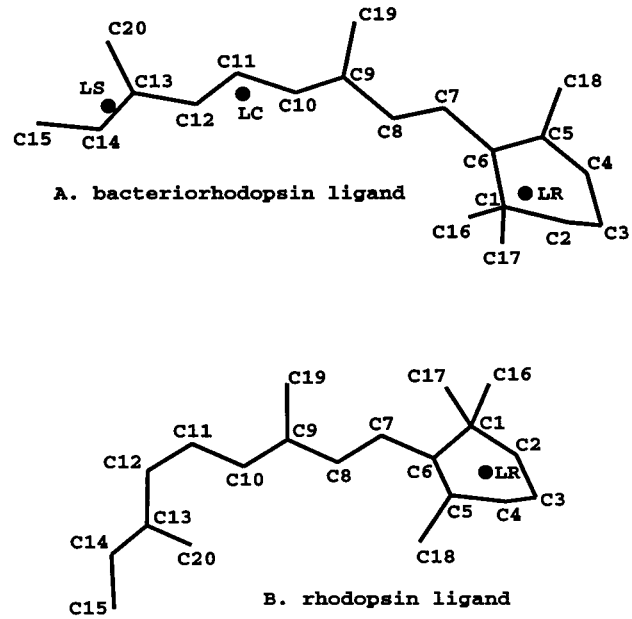

FIGURE 1 Labeling scheme for the all-*trans* retinal and 11-*cis* retinal, which are the ligands for bR and rhodopsin, respectively. LS, LC, and LR of all-*trans* retinal denote dummy atoms associated with deuteration centers of bR representing Schiff base, C11, and β-ionone ring, respectively (Heyn et al., 1988). LR in 11-*cis* retinal denotes the dummy atom for the centroid of the β-ionone ring.

TABLE 1 Structural information available for seven-helix membrane proteins that can be used for deriving structural restraints

Source	Information	Derived restraints	Illustrative references
Electron microscopy	2-D electron density projection map displaying arrangements of helices in the membrane plane	Positions of helix centroids and ends projected onto the membrane plane to match their 2-D image	Baldwin et al., 1988; Henderson and Unwin, 1975; Schertler et al., 1993
Neutron diffraction coupled with ^2H labeling of certain residues	2-D projection map of marked residues	Orientations of helices about their axes to match the image of the labeled residue	Popot et al., 1989
Neutron diffraction coupled with ^2H labeling of different parts of retinal	2-D projection map of labeled centers in a ligand with respect to helix positions	Positions of labeled centers with respect to the projections of helix centers and/or ends	Heyn et al., 1988; Seiff et al., 1985, 1986
Site-directed mutagenesis	Detection of residues that are functionally important (ligand binding, signal transduction)	Orientation of the helix to position the mutated functional residue toward inside of the seven-helix bundle. Position of the residues important for ligand binding to be close to ligand	Khorana, 1988; Savarese and Fraser, 1992
Data on naturally occurring mutations	Detection of functionally important residues	Same as above	Rao et al., 1994
Cross-linking	Detection of residues bound to the photoactivatable ligand analogue	Positions of the residues close to the ligand	Huang et al., 1982; Nakayama and Khorana, 1990
Site-directed spin labeling	Positions of labeled residues with respect to the aqueous and lipid phases	Orientation of the helix to position labeled residues toward inside or outside of the seven helix bundle	Altenbach et al., 1991; Greenhalgh et al., 1991
FTIR difference spectroscopy combined with site-directed mutagenesis	Detection of functionally active residues	Orientation of the helix to position functionally active residues inside the seven-helix bundle	Rath et al., 1993; Rothschild, 1992
Solid state ^{13}C NMR	Data on ligand conformation or ligand interaction with charged residues	Ligand conformation or position of the ligand with respect to certain charged residues	deGroot et al., 1989; Smith et al., 1987, 1990
Semiempirical calculations on ligand-protein interactions	Same as above	Same as above	Honig et al., 1979; Kakitani et al., 1985
Disulfide bridge detection	Detection of Cys residues participating in a disulfide bond	Position of the identified Cys residues next to each other	Curtis et al., 1989
Multiple sequence alignment	Positions of conserved, variable, and polar residues in receptor families	Orientations of helices to position the conserved and polar residues toward inside and variable residues toward outside of the seven-helix bundle	Baldwin, 1993
Calculations of periodicity in hydropathy, sequence variability, and substitution pattern	Detection of hydrophobic versus hydrophilic, variable versus conserved, and lipid-exposed versus buried faces of transmembrane helices	Orientations of the helices with the hydrophobic/variable/lipid-exposed face outside and hydrophilic/conserved/buried face inside the seven-helix bundle	Donnelly et al., 1993

Fig. 2 shows the types of restraints currently considered by our method. These can be summarized as (1) distance restraints between atoms in 3-D space, i.e., *xyz*-distance restraints; (2) distance restraints between the projections of atoms on the membrane plane (*xy* plane), i.e., *xy*-distance restraints; (3) position restraints imposed on *x* and *y* coordinates of atoms, restraining the positions of the projections of atoms on the membrane plane, i.e., *xy*-position restraints; (4) position restraints imposed on the *z* coordinate of certain atoms, i.e., *z*-position restraints; and (5) orientation restraints for residues facing inside or outside of a protein, i.e., ORI inside/outside restraints (not shown in Fig. 2).

Tables 2 and 3 list the experimentally derived restraints used in the modeling of bR and rhodopsin, respectively. The table footnotes describe the source of the restraints. The following describes general features of the different classes of restraints.

Choice of transmembrane helical segments for bR

For the test calculations, the helix-forming residues were assigned on the basis of the known structure of bR (Henderson et al., 1990) with helices C and E modified according to the results of spin-labeling experiments (Altenbach et al., 1991; Greenhalgh et al., 1991). The final assignment is helix A, residues 10–32; helix B, 38–62; helix C, 80–100; helix D, 108–127; helix E, 132–157; helix F, 167–191; and helix G, 203–225. The

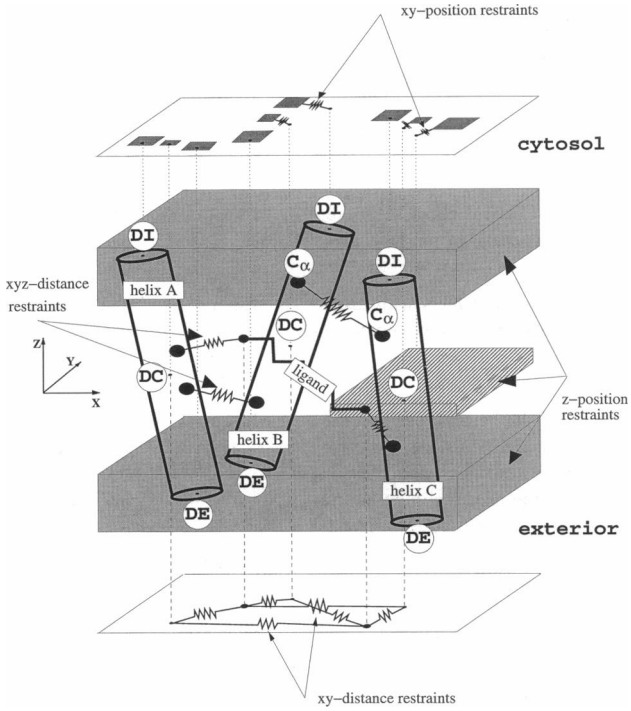

FIGURE 2 Graphical representation of distance and position restraints applied for helices. Three helices are represented as cylinders with dummy atoms DI, DC, and DE as shown. The dark shaded regions represent z-position restraints. The thick shaded region represents the regions close to the cytosolic and extracellular surfaces of the membrane where dummy atoms DI and DE are allowed to be. The thin shaded region represents the region where the ligand atom is allowed to be. All shaded regions are unlimited along x and y directions and have been cut only for the clarity of the picture. The xyz-distance restraints between C_α atoms and/or ligand atoms are shown as springs. The xy-distance and xy-position restraints applied to atom projections in the plane of the membrane are shown projected to other planes for clarity. The xy-distance restraints are displayed as springs in the lower plane whereas the xy-position restraints are displayed in the upper plane. The dark shaded squares in the upper plane represent the regions where the projections of dummy atoms DI, DC, and DE are allowed to be, whereas the springs show the direction of the restraints.

rigid structure of all-*trans* retinal was also extracted from the EM structure coded as 1brd in the Brookhaven databank (Bernstein et al., 1977). These are the only instances that the knowledge of the EM bR structure has been used in our calculations.

Choice of transmembrane helical segments for rhodopsin

The start and finish of each helical segment in rhodopsin were taken from the work of Baldwin (1993) in which the sequences of 105 GPCRs were analyzed. For bovine rhodopsin the transmembrane helical segments were proposed as follows: helix A, residues 38–63; helix B, 70–95; helix C, 111–136; helix D, 151–176; helix E, 202–227; helix F, 250–275; and helix G, 286–311.

xyz-distance restraints

The xyz-distance restraints correspond to 3-D distances between particular virtual atoms and/or helix end/centroid coordinates. They are derived from several experimental sources and, unless specified otherwise, the restraint upper limits for virtual atoms were derived by modeling a particular interaction in its extended form using an all-atom representation and then converting the model to virtual atoms, including an estimate of possible errors. Numbers of significant digits in the upper bounds estimated this way (Tables 2 *A* and 3 *A*) should not imply the accuracy of their estimation.

xy-distance restraints

This class of restraint corresponds to the distance apart of particular groups in the xy plane of the membrane. The information was derived only for bR from the 2-D projection map generated from neutron diffraction data in which different parts of the retinal were ^2H labeled (Heyn et al., 1988; Seiff et al., 1985, 1986). These distances can be used as xy-distance restraints once the assignment of bR helices onto their images has been made. The lower and upper limits of the restraints have been evaluated from the uncertainty associated with the experimental results.

xy-position restraints

These restraints are used to position and orient the helices relative to each other and, for both molecules, were derived from low resolution 2-D electron density projection maps as described in the table footnotes (see Table 2, footnote 8, and Table 3, footnote 9). Fig. 3 shows a schematic of the projection map used for bR (Henderson and Unwin, 1975) and Fig. 4 that used for rhodopsin (Schertler et al., 1993). The positions derived contain considerable error due to the resolution of the data and the uncertainty in assigning the tilt angle for the helices, and this is reflected in the size of tolerances assigned.

The images can provide xy-position restraints for particular helix centroids and ends if it is known which direction of the map corresponds to the extra- and intracellular surfaces of the membrane, how to assign helices to their images, and how the helix tilts in the projection map correspond to the intra- or extracellular end. For bR some of this information is available from specific labeling of the protein and lipid on the extracellular surface of purple membrane (Henderson et al., 1978), which demonstrated that the original projection map (Henderson and Unwin, 1975) presents a view from the intracellular side. Also, the assignment of helices A and B to images 1 and 2, respectively, was possible from neutron diffraction studies of ^2H labeling of Leu and Trp (Popot et al., 1989). Similar information is not available for rhodopsin. The full assignment and tilt of the helices will be derived in the calculations presented in this paper.

z-position restraints

These restraints control the position of the helices and a ligand relative to the boundaries of the membrane. For bR, neutron diffraction provides additional information about depth of a ligand binding site as discussed in Table 2 (see footnote 9). These are referred to as z-position (ligand) restraints. For both bR and rhodopsin the purified membranes are assumed to be approximately 45 Å thick (Henderson and Unwin, 1975). From this it is possible to deduce limitations on the position of the helix ends with respect to the membrane plane. These are referred to as z-position (membrane) restraints.

Loop length restraints

For some pairs of helices, restraints can also be applied on the basis of the number of amino acids forming the loop between helices. In bR, the loop lengths are as follows: loop AB (between helices A and B), 5 residues; loop BC (helices B and C), 17 residues; loop CD, 7 residues; loop DE, 4 residues; loop EF, 9 residues; and loop FG, 11 residues. For rhodopsin, the lengths of loops that can be considered are further constrained, making the assumption that all members of the GPCR family are expected to have the same basic 3-D arrangement of helices. Multiple sequence alignment of the members of the GPCR family of sequences by Baldwin (1993) showed that the smallest number of residues found for the six loops between the seven helices were as follows: loop AB, 5 residues; loop BC,

TABLE 2 Structural restraints used for bacteriorhodopsin

A: *xyz*-distance restraints used for bacteriorhodopsin

Atom 1	Atom 2	Upper limit[a] (Å)	Footnote
C_β-Asp-85	C_β-Lys-216	9.34	1
C_β-Asp-85	C15-RET	7.04	1
LR-RET	DE-F	10.00	2
C_α-Tyr-57	LS-RET	15.58/12.00	3, 4
C_α-Arg-82	LS-RET	9.95/12.00	3, 4
C_α-Arg-82	LC-RET	10.17	3
C_α-Tyr-83	LS-RET	9.95/12.00	3, 4
C_α-Tyr-83	LC-RET	10.17	3
C_α-Trp-86	LS-RET	9.95/12.00	3, 4
C_α-Trp-86	LC-RET	10.17	3
C_α-Thr-89	LS-RET	9.95/12.00	3, 4
C_α-Thr-89	LC-RET	10.17	3
C_α-Thr-90	LS-RET	9.95/12.00	3, 4
C_α-Thr-90	LC-RET	10.17	3
C_α-Asp-115	LR-RET	11.80/12.00	3, 4
C_α-Trp-137	LR-RET	11.43/12.00	3, 4
C_α-Trp-138	LR-RET	11.43/12.00	3, 4
C_α-Ser-141	LR-RET	11.43/12.00	3, 4
C_α-Trp-182	LR-RET	12.08/12.00	3, 4
C_α-Tyr-185	LR-RET	12.08/12.00	3, 4
C_α-Trp-189	LR-RET	12.08/12.00	3, 4
C_α-Asp-212	LS-RET	12.78/12.00	3, 4
DI-A	DI-B	15.70	5
DI-C	DI-D	18.2	5
DE-D	DE-E	14.70	5
DI-E	DI-F	20.70	5
C_β-Arg-82	C_β-Asp-212	8.95	6
C_β-Tyr-185	C_β-Asp-212	9.82	6
C_β-Thr-46	C_β-Asp-96	6.52	6
C_β-Thr-89	C_β-Asp-212	6.52	6
C_α-Lys-216	C15-RET	7.40	15
C_α-Lys-216	C14-RET	8.69	15
C_β-Lys-216	C15-RET	4.93	15
C_β-Lys-216	C14-RET	6.18	15

[a] The lower limit of all restraints is 4Å.

B: *xy*-distance restraints used for bacteriorhodopsin[a]

Restraint atom 1	Restraint atom 2	Lower limit (Å)	Upper limit (Å)
LS-RET	DC(1)	11.54	19.27
	DC(2)	9.97	13.12
	DC(3)	2.62	5.35
	DC(4)	10.75	13.11
	DC(5)	10.75	14.99
	DC(6)	6.56	8.97
	DC(7)	6.30	9.64
LC-RET	DC(1)	16.14	23.96
	DC(2)	11.55	14.72
	DC(3)	2.36	5.75
	DC(4)	9.18	11.38
	DC(5)	8.92	12.98
	DC(6)	5.90	8.43
	DC(7)	7.87	11.77
LR-RET	DC(1)	18.63	25.96
	DC(2)	16.00	19.00
	DC(3)	5.77	9.50
	DC(4)	6.30	8.30
	DC(5)	4.45	7.76
	DC(6)	5.64	8.69
	DC(7)	12.07	16.33

[a] See footnote 7 for how these restraints were derived.

C: *xy*-position restraints used for bacteriorhodopsin[a]

Helix atom	X (Å)	Y (Å)	Tolerance (Å)
DC(1)	−15.88	0.53	±1.06
DI(1)	−20.88	2.54	±1.76
DE(1)	−10.88	−1.23	±1.76
DC(2)	−11.06	7.55	±0.70
DI(2)	−12.46	5.35	±1.76
DE(2)	−9.74	9.74	±1.76
DC(3)	−2.19	6.41	±0.70
DI(3)	−2.28	1.93	±1.76
DE(3)	−2.11	10.79	±1.76
DC(4)	7.37	8.60	±0.70
DI(4)	7.37	8.60	±0.70
DE(4)	7.37	8.60	±0.70
DC(5)	10.88	0.88	±1.06
DI(5)	5.53	−4.21	±1.76
DE(5)	15.97	6.23	±1.76
DC(6)	3.16	−5.53	±1.06
DI(6)	−1.58	−9.13	±1.76
DE(6)	7.63	−2.19	±1.76
DC(7)	−6.84	−3.95	±1.06
DI(7)	−11.06	−3.77	±1.76
DE(7)	−2.72	−3.86	±1.76

[a] Lower and upper limits of these restraints are calculated as follows: X_l (Y_l) = X (Y) − tolerance, X_u (Y_u) = X (Y) + tolerance. See footnote 8 for how these restraints were derived.

D: *z*-position restraints used for bacteriorhodopsin

Atom	Lower limit (Å)	Upper limit (Å)	Footnote
LS-RET	−5.30	−2.30	9
LR-RET	−11.70	−9.30	9
DI-X[a]	11.70	22.50	10
DE-X[a]	−11.70	−22.50	10

[a] X indicates any helix A to G.

E: ORI outside restraints used for bacteriorhodopsin

Helix	ORI outside restraint to residue	Footnote
A	–	
B	–	
C	–	
D	–	
E	Ser-132	11
	Tyr-133	11
	Val-136	11
	Ala-139	11
F	Ser-183	12
G	–	

F: ORI inside restraints used for bacteriorhodopsin

Helix	ORI inside restraint to residue	Footnote
A	Gly-16	13
B	Thr-46	14
	Tyr-57	14
C	Arg-82	14
	Tyr-83	14
	Asp-85	1, 14
	Trp-86	14
	Thr-89	14
	Thr-90	14
	Asp-96	14
	Ala-98	11
D	Asp-115	14
E	Arg-134	11
	Trp-137	14
	Trp-138	11, 14
	Ser-141	11, 14
F	Trp-182	14
	Tyr-185	12
	Trp-189	14
G	Thr-205	14
	Asp-212	14
	Lys-216	15

Table 2 *Continued*

1. The effect of neutral mutations at position 85 on the *p*K of the Schiff base deprotonation (Otto et al., 1990) and the role of Asp-85 in the purple-to-blue transition (Subramaniam et al., 1991) suggest that this residue may be the primary counterion and oriented inside close to the protonated Schiff base. This is also confirmed by FTIR difference spectroscopy (Rothschild, 1992).
2. Cross-linking a photoactivatable retinal analogue to bR showed that the major sites of the retinal contact are Ser-193 and Glu-194 (Huang et al., 1982). These are in the loop following helix F, hence the restraint between retinal and the dummy atom representing the extracellular end of helix F. The maximal distance of 10 Å has been assumed.
3. These are binding pocket restraints derived assuming that mutations that affect the spectral characteristics of the retinal are close to the molecule. The data were taken from the following sources: Arg-82, Asp-85, Trp-86, Thr-89, Asp-115, Trp-137, Trp-182, Trp-189, Asp-212 (Khorana, 1988); Thr-90 (Marti et al., 1991); Ser-141 (Altenbach et al., 1991; Marti et al., 1991); Tyr-57, Tyr-83 (Mogi et al., 1987); Tyr-185 (Duñach et al., 1990); and Trp-138 (Altenbach et al., 1991). The restraints were applied between C_α atoms and the closest of the retinal LC, LS, and LR dummy atoms as judged from the neutron diffraction retinal projection map (Heyn et al., 1988) except for helix C for which two restraints between C_α and both LC and LS atoms were taken as those retinal atoms were roughly equidistant from helix C. The upper limit of the restraints have been evaluated as $(a^2 + b^2 + c^2)^{1/2}$ where a is the upper limit of an *xy*-distance restraint between a given retinal dummy atom and a projection of the helix centroid as derived from the projection map (see footnote 7), with b as 8.1 Å (1.5 times the helical pitch), and c as 2.28Å (the radius of a helical wheel of C_α positions).
4. These restraints represent a simplified binding pocket for the preliminary calculations to assign helices for bR. Uniform, weak *xyz*-distance restraints were applied between C_α atoms and the closest of LS and LR. Knowing the approximate 2-D orientation of the retinal defined by the Schiff base being between Asp-85 on helix C and Lys-216 on helix G, and the chromophore ring being positioned near helix F, implies that the LS atom is closest for residues belonging to helices C and G, and the LR atom is closest for helix F. The positions of atoms LR and LS with respect to helices B (Tyr-57), D (Asp-115), and E (Trp-137, Trp-138, and Ser-141) depend on a particular helix assignment and are used as restraints only for the best helix assignment after inspection of the retinal position (see text). The upper limits of the restraints defining the simplified binding pocket were assumed to be 12 Å.
5. Four loop-length restraints are applied for loops no longer than 10 residues. These restraints are used only in the helix assignment calculation. See Methods for how these interhelix distance restraints were derived on the basis of interhelical loop length.
6. FTIR difference spectroscopy of site-directed mutants identified four buried Asp residues important in different steps of the bR photocycle (Rothschild, 1992). Ionized Asp-212 and Asp-85 interact with positively charged Arg-82 and the protonated Schiff base adding the restraint for Asp-212 (for Asp-85 see footnote 1). Mutants of Asp-212 reduced or abolished bands assigned to alterations in Tyr-185. T46V perturbed the state of Asp-96 in the L and M intermediates and T89D caused proton transfer between Asp-89 and Asp-212 early in the photocycle. These restraints are used only in the helix assignment calculation.
7. The 2-D distances between the LR, LS, and LC deuteration centers on retinal (Fig. 1) and centers of helix images have been taken from Fig. 4 of the original neutron diffraction paper (Heyn et al., 1988). The centroid of a helix associated with an image k (Fig. 3) is designated as DC(k). The position of image 1 has been extrapolated from the 2-D electron density map (Henderson and Unwin; 1975) as it is not clearly visible on the picture derived from neutron diffraction. The lower and upper limits of the restraints have been evaluated from the uncertainty associated with the experimental results as judged from the original figure mentioned above. The precision of numbers presented is a result of a scaling procedure and should not imply the accuracy of the evaluation.
8. The 7-Å resolution map presented as Fig. 4 of the original paper (Henderson and Unwin, 1975) shows three bR molecules grouped around the threefold axis. As the angle of view causes the three molecule images to appear different, we have decided to use only one of them, that on the upper left. It is schematically represented in Fig. 3. For each of the 2-D helix images we took the coordinates of a centroid and ends that correspond to the projections of the helix centroid and ends, respectively, with some consideration of the other molecule images. DC(k), DI(k), and DE(k) label the appropriate helix dummy atoms belonging to the helix associated with the image k (Fig. 3). The size of tolerances reflect the appearance of the images. The precision of numbers presented is a result of a scaling procedure and should not imply the accuracy of the evaluation.
9. *z*-position (ligand) restraints. The deuteration center corresponding to Schiff base (LS) is located at a distance of 3.8 ± 1.5Å from the membrane plane whereas the β-ionone ring deuteration center (LR) is located at the same side of the membrane plane at a distance of 10.5 ± 1.2Å (Hauss et al., 1990). The cross-link between the ring of the photoactivatable retinal analogue and bR occurs close to the extracellular end of helix F (Huang et al., 1982); thus it can be assumed that the locations of these two deuteration centers are on the extracellular side of the membrane plane and an appropriate restraint can be applied.
10. *z*-position (membrane) restraints. We have assumed that no helix end can protrude from the 45-Å-thick membrane (Henderson and Unwin, 1975) and that it cannot be deeper in the membrane than more than two helix turns (10.8 Å). It is for this reason that the *z*-coordinates of dummy atoms representing intracellular helix ends have been restrained to the range 11.7–22.5 Å. The extracellular helix ends have been restrained to an analogous range of negative values.
11. The collision frequency of O_2 with a spin label attached to each of residues 125–142 showed highest accessibility for Ser-132, Tyr-133, Val-136, and Ala-139 and lowest for Arg-134 and Trp-138 (Altenbach et al., 1991). In the absence of a spin label, the mutation Ser-141 → Cys induced a sizeable shift in the visible absorption maximum whereas a small but measurable shift occurred for the Trp-138 → Cys mutant. After attachment of the spin label both mutants showed large shifts. These results suggest that residues Arg-134, Trp-138, and Ser-141 are oriented inside and residues Ser-132, Tyr-133, Val-136, and Ala-139 are oriented outside. Residue Ala-98 is given an inside restraint as it showed low accessibilities to both membrane-permeant and impermeant probes (Greenhalgh et al., 1991), suggesting it is buried.
12. Y185F mutation exhibits a pH-dependent absorbance spectrum (Duñach et al., 1990), suggesting the residue is oriented inside. Mutations of Trp-182, Tyr-185, Pro-186, and Trp-189 of helix F produce large alterations in the photocycle and are thus inside whereas the replacement of Ser-183 causes relatively small perturbations (Ahl et al., 1989), suggesting it is outside.
13. Neutron diffraction coupled with ^2H labeling of Leu and Trp residues in helices A and B produced a 2-D projection map that determined the orientation of helix A with respect to its axis with Gly-16 facing inside the bR molecule (Popot et al., 1989).
14. Mutagenesis studies showed residues with mutations that either altered the visible absorption spectrum (see footnote 3 for references) and/or affected proton pumping. Residues with mutations that affect proton pumping and do not affect absorption spectrum are Thr-46, Thr-205 (Marti et al., 1991), and Asp-96 (Khorana, 1988). It is assumed that these residues are oriented inside.
15. bR consists of an all-*trans* retinal covalently bound to the ε-amino group of Lys-216 in helix G through a protonated Schiff base. This residue is thus oriented inside and constrained to be close to the retinal.

TABLE 3 Structural restraints used for rhodopsin

A: *xyz*-distance restraints used for rhodopsin

Atom 1	Atom 2	Upper limit[a] (Å)	Footnote
C_β-Glu-113	C_β-Lys-296	9.73	1
C_β-Glu-113	C15-RET	7.43	1
C_β-Ala-117	C_β-Lys-296	10.47	2
C_β-Ala-117	C15-RET	8.14	2
C_α-Gly-90	C_β-Lys-296	10.05	3
C_α-Gly-90	C15-RET	7.69	3
C_α-Ala-292	C15-RET	7.41	3
C_α-Glu-122	C_β-His-211	8.50	4
C_α-Phe-208	C3-RET	7.25	6
C_α-His-211	C3-RET	7.25	6
C_α-Glu-122	LR-RET	12.00	7, 16
C_α-Trp-265	LR-RET	10.00	5, 7
C_α-Tyr-268	LR-RET	12.00	7, 16
C_α-Ala-164	LR-RET	12.00	10, 16
C_α-Phe-261	LR-RET	12.00	10, 16
C_α-Ala-269	LR-RET	12.00	10, 16
DI-A	DI-B	15.70	14
DI-C	DI-D	20.70	14
DE-F	DE-G	20.70	14
C_α-Lys-296	C15-RET	7.40	15
C_α-Lys-296	C14-RET	8.69	15
C_β-Lys-296	C15-RET	4.93	15
C_β-Lys-296	C14-RET	6.18	15

[a] Lower limit of all restraints is 4 Å.

B: *xy*-position restraints used for rhodopsin[a]

Helix atom	X (Å)	Y (Å)	Tolerance (Å)
DC(1)	−13.14	0.45	±2.27
DI(1)	−12.34	3.96	±2.27
DE(1)	−14.04	−4.19	±2.27
DC(2)	−7.02	8.72	±2.27
DI(2)	−0.45	11.44	±3.17
DE(2)	−13.25	5.78	±3.17
DC(3)	4.42	4.53	±2.27
DI(3)	6.46	0.00	±2.27
DE(3)	2.38	8.83	±2.27
DC(4)	13.70	7.81	±1.36
DI(4)	13.70	7.81	±2.27
DE(4)	13.70	7.81	±2.27
DC(5)	15.18	−4.53	±0.91
DI(5)	15.18	−4.53	±2.27
DE(5)	15.18	−4.53	±2.27
DC(6)	4.08	−9.40	±0.91
DI(6)	4.08	−9.40	±2.27
DE(6)	4.08	−9.40	±2.27
DC(7)	−3.74	−2.60	±0.91
DI(7)	−3.74	−2.60	±2.27
DE(7)	−3.74	−2.60	±2.27

[a] Lower and upper limits of these restraints are calculated as follows: X_l (Y_l) = X (Y) − tolerance, X_u (Y_u) = X (Y) + tolerance. See footnote 9 for how these restraints were derived.

C: *z*-position restraints used for rhodopsin[a]

Atom	Lower limit (Å)	Upper limit (Å)
DI-X	11.70	22.50
DE-X	−11.70	−22.50

[a] See Methods and footnote 10 to Table 2 for origin of these restraints.

D: ORI outside restraints used for rhodopsin[a]

Helix	Residue	Helix	Residue
A	Met-49	E	Ile-213
B	Ala-80		Leu-216
	Val-81		Ile-217
	Gly-89	F	Ala-260
C	Ile-123		Leu-266
D	Ala-158		Gly-270
	Ala-169	G	Ile-290
	Val-173		Val-300
			Met-308

[a] See footnote 13 for the source of these restraints.

E: ORI inside restraints used for rhodopsin

Helix	Residue	Footnote
A	Phe-45	8
	Gly-51	8
	Asn-55	11, 12
	Thr-58	8
B	Asn-73	11, 12
	Asn-78	11, 12
	Leu-79	11
	Asp-83	12
	Gly-90	3, 12
C	Glu-113	1
	Gly-114	11
	Phe-115	5
	Ala-117	2, 5
	Gly-121	11
	Glu-122	4, 5, 7, 12
	Leu-125	8
	Trp-126	5, 7
	Ser-127	5, 11
D	Ala-164	10
E	Phe-208	6
	His-211	4, 6, 12
	Ile-219	11
	Tyr-223	11
F	Phe-261	10
	Trp-265	5, 7, 11
	Tyr-268	7, 11
	Ala-269	10
G	Ala-292	3, 7
	Lys-296	11, 12, 15
	Asn-302	11, 12
	Ile-305	11
	Tyr-306	11

Table 3, *A-E* summarizes the restraints used in the generation of the model of rhodopsin. The following footnotes describe the source of these restraints.

1. Glu-113 is the only residue that can act as the counterion for the protonated Schiff base (Nathans, 1990a, b; Zhukovsky and Oprian, 1989), so the residue must be inside and in contact with retinal. These restraints were quantified assuming that the maximal distance between Lys-296 N and Glu-113 O is 3.2 Å as resonance Raman spectroscopy (Palings et al., 1987) and solid state NMR (Smith et al., 1990) suggests that, unlike bR, the hydrogen bond strength at the Schiff base of rhodopsin is similar to that of strongly H-bonded model compounds.

2. A double replacement mutant, E113A/A117E, shifts the position of the Schiff base counterion by one helix turn. This mutant protein is functional although with an altered photochemical cycle (Zvyaga et al., 1994, 1993), implying Ala-117 is oriented inside and near the Schiff base.

13 residues; loop CD, 10 residues; loop DE, 12 residues; loop EF, 12 residues; and loop FG, 8 residues.

Analysis of the projection maps for bR (Henderson and Unwin, 1975) and rhodopsin (Schertler et al., 1993) revealed that the distance between two helix ends on the same side of the membrane cannot be longer than 35 Å and that the angle between two helix axes cannot be greater than 55°. These values were then used as boundaries for a search of a set of nonhomologous proteins from the protein databank (Bernstein et al., 1977) for nonhelical loops between 4 and 19 residues in length between antiparallel helices of at least 10 residues each. The results are shown in Fig. 5, which shows that for short loops the average distance S grows very slowly with the loop length. As the number of occurrences decreases significantly with the loop length, reliable conclusions can be drawn only for loops no more than 10 residues in length. For such loops, an *xyz*-distance restraint has been applied between the ends of the two helices contributing to a particular loop. The upper limit of this restraint is set equal to the maximal distance S over the loops with lengths that are less than or equal to the loop under consideration. Thus for loops of length from 4 to 10 residues the upper limits are 14.7 Å, 15.7 Å, 18.2 Å, 18.2 Å, 20.7 Å, 20.7 Å, and 20.7 Å, respectively.

Orientation restraints

These are used to restrain a residue to be either pointing into a helix bundle (oriented inside) or pointing out (oriented outside).

Ligand orientation and conformation for bR

The all-*trans* retinal in bR was found to be approximately perpendicular to the membrane plane (Earnest et al., 1986; Heyn et al., 1988). It was also shown that the polyene chain makes an angle of 25–30° with the membrane

TABLE 3 *Continued*

3. Replacement of Lys-296 by 12 amino acids results in constitutive activation of the apoprotein except for K296R (Cohen et al., 1993). A similar result is obtained if Glu-113 is mutated (Robinson et al., 1992). Other constitutive mutations in rhodopsin are A292E (Dryja et al., 1993) and G90D (Rao et al., 1994). This suggests that constitutive activation of the apoprotein is caused by disruption of the hydrogen bond between Lys-296 and Glu-113 (Rao et al., 1994). This is consistent with the finding that the double mutant G90D/E113Q shows only a slight blue shift as compared with E113Q. This suggests that Gly-90 is close to Lys-296 and able to provide an alternative counterion for the protonated Schiff base upon a proper mutation (Rao et al., 1994). Thus residues Ala-292 and Gly-90 are oriented inside and are close to Lys-296 and Schiff base.
4. Site-directed mutagenesis experiments showed that Glu-122 is not a counterion to the protonated Schiff base. However, multiple sequence alignment of 59 opsins (Oliveira et al., 1993) showed that this Glu is always partnered by His-211, suggesting an interaction between the two residues.
5. Phe-115, Ala-117, Glu-122, Trp-126, and Ser-127 in helix C and Trp-265 in helix F were all identified as major sites of cross-linking to the photoactivatable analogue of 11-*cis* retinal in rhodopsin (Nakayama and Khorana, 1990). The residues in helix C span approximately 2.5 α-helical turns and all cannot be in close proximity to the β-ionone ring at the same time. Therefore there is only one restraint between the β-ionone ring and Trp-265 with all of the residues oriented inside.
6. Serines at either position 208 or 211 are implicated in hydrogen bond stabilization of different conformations of the 3-OH retinal in *Drosophila* rhodopsin variants (Oprian, 1992). Therefore Phe-208 and His-211 are assumed to be close to the C3 atom of the β-ionone ring of the retinal and oriented inside.
7. Mutations of Glu-122 and Trp-265 alter the spectral characteristics of the chromophore whereas replacements of residues Glu-122, Trp-126, Trp-265, Tyr-268, and Ala-292 affects transducin activation (Nakayama and Khorana, 1991). All of these residues are assumed to be oriented inside. In addition, mutations of Glu-122, Trp-265, and Tyr-268 cause hydrolysis of the Schiff base upon continuous illumination in the rhodopsin-like pigment with a cyclohexatrienylidine retinal analogue (Ridge et al., 1992), which leads to the assumption that they should be positioned next to the retinal.
8. Approximately 50, mainly point mutations have been identified in the rhodopsin gene of autosomal dominant retinitis pigmentosa (Kaushal and Khorana, 1994). As the mutations F45L, G51V, T58R, and L125R are transmembrane mutants that express at wild-type level and form the normal rhodopsin chromophore with 11-*cis* retinal, the residues are assumed to be oriented inside.
9. Restraints for each of 21 dummy atoms representing the ends and centroids of the helices were derived from the electron density projection map at 9-Å resolution presented in Fig. 2 of the original paper (Schertler et al., 1993). The helices associated with images 4, 5, 6, and 7 (Fig. 4) are essentially vertical whereas the others are significantly tilted resulting in some overlap. For the vertical helices the centroid and ends are restrained to the center of the corresponding peak of electron density, with smaller errors for helix centroids than for helix ends and slightly greater errors for the more diffuse image of helix 4. The dummy atoms for the tilted helices have been given positions with associated errors to reflect the appearance of the images. DC(k), DI(k), and DE(k) label the appropriate helix dummy atoms belonging to the helix associated with the image k (Fig. 4). The precision of numbers presented is a result of a scaling procedure and should not imply the accuracy of the evaluation.
10. Mutations at the equivalent positions of Ala-164, Phe-261, and Ala-269 have been identified as responsible for the approximately 30-nm difference in the spectral peaks of the pigments underlying human red-green color vision (Neitz et al., 1991). These residues should therefore be oriented inside and close to the retinal molecule.
11, 12, and 13. These orientation restraints were all derived from the multiple sequence alignment of 59 opsins available from the TM7 file server (Oliveira et al., 1993). Sites that have the same amino acid in at least 90% of the sequences are assumed to be oriented inward (noted as 11), as are positions that accommodate polar residues in at least 20% of the sequences (noted as 12). Here, the term polar refers to charged residues and those capable of forming more than one hydrogen bond, namely, Lys, Arg, Asp, Glu, Asn, Gln, and His (Baldwin, 1993). There are 13 sequences of dim light rhodopsin from different species that show very high homology, and positions with differences should be oriented outward (noted as 13) as suggested in Baldwin, 1993, with the difference defined as fulfilling two of the three following criteria: (1) more than 50% variability of the particular position in the entire multiple sequence alignment, (2) variation occurring in at least 3 of 13 closely related sequences, and (3) at least three different residues seen at the position in the 13 related sequences.
14. Three loop restraints are applied. These are used only in the helix assignment calculation. See Methods for description of how these interhelix distance restraints were derived on the basis of interhelical loop length.
15. There is a covalent link between Lys-296 and 11-*cis* retinal, which is inside the seven-helical bundle.
16. Uniform weak *xyz*-distance restraints were applied between C_α atoms and the closest of the retinal atoms representing the β-ionone ring and Schiff base, LR and C15, respectively, in a manner similar to the bR restraints described in footnote 4. Residues belonging to helix E (Phe-261, Tyr-268, and Ala-269) are assumed to be close to the LR atom on the basis of the cross-link experiment (Nakayama and Khorana, 1990) (see footnote 5). In the preliminary helix assignment calculations, Glu-122 on helix C is weakly attached to C15 as helix C contains a counterion to protonated Schiff base. In the case of the optimal helix assignment the restraints involving Glu-122 (helix C) and Ala-164 (helix D) are attached to the LR atom after the inspection of the retinal position (see text).

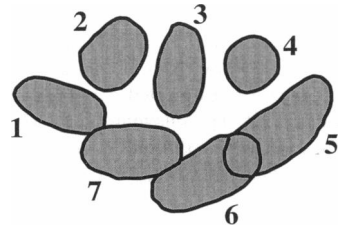

FIGURE 3 Schematic representation of electron density projection map of bR at 7 Å resolution (Henderson and Unwin, 1975) used as a source for *xy*-position restraints and a numbering scheme of helix images.

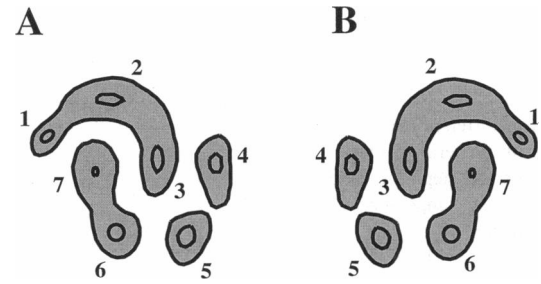

FIGURE 4 Schematic representation of electron density projection map of rhodopsin at 9 Å resolution (Schertler et al., 1993) used as a source for *xy*-position restraints and a numbering scheme for the helix images. Two symmetry-related molecule images are visible, A and B.

plane (Earnest et al., 1986). Furthermore, the orientation of the chromophore is such that the N→H bond of the Schiff base points toward the same membrane surface as the vector from the Schiff base to the β-ionone ring (Hauss et al., 1990). This surface proved to be the extracellular one (Huang et al., 1982). All of these data allow for the initial orientation of a retinal molecule in the plane perpendicular to the membrane plane with a polyene chain roughly parallel to the membrane plane with the C19 and C20 methyl groups pointing at the intracellular side of the membrane. Consequently the maximal step of rotation around the axis parallel to the polyene chain is reduced (see Table 4).

Ligand conformation and orientation for rhodopsin

The idealized 11-*cis* retinal molecule was built with QUANTA (Molecular Simulations, Inc., Waltham, MA). The conformation on the C6–C7 bond was set to s-*cis* (Smith et al., 1987) with a ring twisted 55° relative to the retinal chain. The dihedral angle on the C12–C13 bond was set to −150° (Honig et al., 1979; Kakitani et al., 1985). The atom-numbering scheme is presented in Fig. 1. As there is no evidence about the orientation of the ligand in rhodopsin it is placed in the initial configurations in the *z* plane with C19 pointing either on the intra- or extracellular side. Both types of initial configurations are evenly distributed.

Penalty function

The penalty function P is given by:

$$P = repel + restr$$

where *repel* is used to prevent unduly close nonbonded contacts and *restr* is associated with geometrical restraints. The details of the penalty function are given in Appendix B.

In addition to the penalty function our algorithm rejects structures where the angle between one or more helix axes and *z* axis (the axis perpendicular to the membrane plane) is bigger than 40° on the basis of inspection of bR and rhodopsin projection maps (Henderson and Unwin, 1975; Schertler et al., 1993).

Optimization method

A combined simulated annealing (Kirkpatrick et al., 1983)/Monte Carlo optimization (Metropolis et al., 1953) technique was chosen to sample conformation space to optimize the penalty function. In simulated annealing a conformational searching is started at a high temperature. This gives the system enough energy to float above barriers of the effective potential energy function, thus melting the system being optimized. Conformational search is then continued while cooling (annealing) the system gradually and slowly. If the initial temperature is high enough and the cooling down process slow enough the system should reach its ground state (Kirkpatrick et al., 1983). The simulated annealing Monte Carlo (SAMC) technique has been successfully applied to structural biology problems such as distance-constrained molecular docking (Yue, 1990), packing arrangement of secondary structures (Chou and Carlacci, 1991), secondary structure prediction of short peptides (Wilson and Cui, 1990), and protein fragments (Nakazawa et al., 1992; Okamoto et al., 1991).

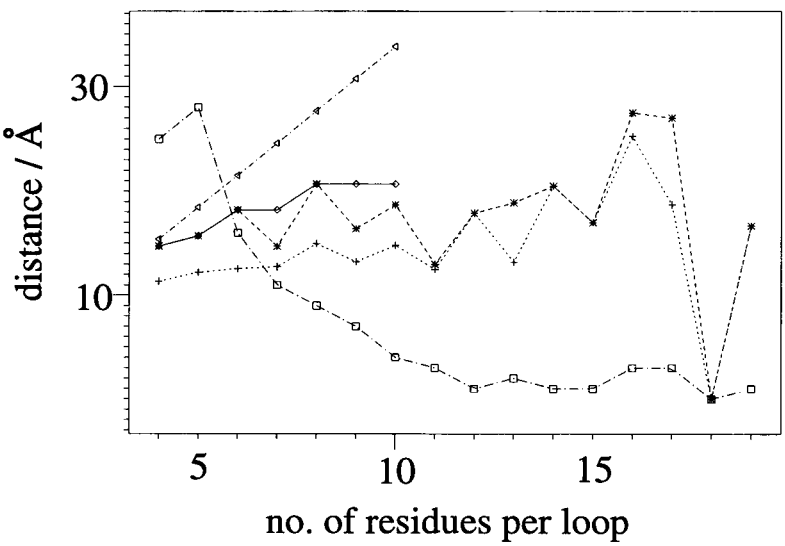

FIGURE 5 Distances between carboxyl end and amino end of loop-linked helices versus loop length: + and *, average distance (+) and maximal distance (*) obtained from the search of the Brookhaven Protein Databank; ◇, restraint upper limit applied in this work; △, extended loop distance; and □, number of hits in database search

TABLE 4 Parameterization of the simulated annealing Monte Carlo protocol

Parameter	Value
Number of different SAMC trajectories	50
Number of temperature runs	25
Number of MC steps per temperature run	1000
Initial effective temperature T_e	8
Cooling factor	0.90
Maximal step size for translation moves	±1 Å
Maximal step size for x' and y' rotations	±5°
Maximal step size for z' rotation	±90°
Maximal step size for the ligand's rotations	±10°
Maximal step size for x' ligand rotation[a]	±2.5°
Maximal step size reduction factor	0.95
Number of temperature runs with no *repel* potential	3
Number of temperature runs with simplified ORI potential	3
repel force constant k_{rep} (after the first three temperature runs)	0.025 Å^{-4}
van der Waals radius scale factor	0.8
Distance restraint force constant k_{dist}	50 Å^{-2}
xy-position restraint force constant k_{pos}	50 Å^{-2}
Simplified ORI restraint force constant k^S_{ori}	5 Å^{-2}
ORI restraint force constant k_{ori}	50 Å^{-2b}/5 Å^{-2c}

[a] Used only for bacteriorhodopsin; x' axis is roughly parallel to the polyene chain.
[b] Used only for bacteriorhodopsin.
[c] Used only for rhodopsin.

Configuration

We can now define a configuration of the system, which is any arrangement of seven rigid helices and a ligand in 3-D xyz Cartesian coordinate space where the xy plane is the membrane plane. Thus the overall system has 48 degrees of freedom corresponding to translations along global x, y, and z axes, and rotations about local x', y', and z' axes for each of the eight rigid body elements. Fig. 6 illustrates these degrees of freedom for one helix in the membrane plane.

Sampling

In principle a Monte-Carlo sampling technique generates a Markov chain of configurations that are weighted with their Boltzmann probability factors $\exp(-E/kT)$ where E is the energy of a given configuration, k is the Boltzmann constant, and T is the temperature. In our case a configuration probability factor is given by $\exp(-P/T_e)$ where P is the penalty function of a given configuration and T_e is an effective temperature that is simply a control parameter in the same units as the penalty function (Kirkpatrick et al., 1983). The role of this parameter is analogous to the role of temperature in statistical mechanics, which is to increase (high temperature) or decrease (low temperature) the probability of configurations.

In our calculations the Markov chain of configurations is generated in the following way. Given a last accepted configuration, the new one is generated by taking one of the following steps: (1) two consecutive translations along the x and y axes, (2) translation along the z axis, (3) two consecutive rotations around the x' and y' axes, or (4) rotation around the z' axis.

The step sizes of each translation and rotation are chosen randomly within the maximal step value associated with each degree of freedom. The steps 1–4 are taken randomly except that the probability of taking steps 1 and 3 is twice as big as that of steps 2 and 4 to maintain a uniform distribution of configurations. This is because steps 1 and 3 are equivalent to a single translation along a randomly chosen axis in the xy plane and a single rotation around such an axis, respectively. The random number generator used is based on the subtractive method of Knuth as presented in Press et al., 1992.

After each step has been taken a penalty function difference ΔP between the proposed configuration and the last accepted one is calculated. Then a Metropolis-type decision is taken in which the step is automatically accepted if ΔP is negative or accepted with a probability of $p = \exp(-\Delta P/T_e)$ if ΔP is positive. Technically the latter action is achieved by randomly generating a number q between 0 and 1 and accepting the configuration if $q < p$ (Metropolis et al., 1953). The chain of configurations is built for 1000 Monte Carlo steps.

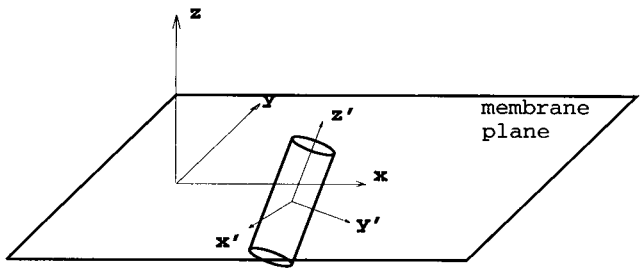

FIGURE 6 Coordinate system used for translation and rotation of rigid body elements.

Cooling

It can be seen from the Metropolis decision schema that the probability of acceptance of the uphill step ($\Delta p > 0$) is relatively high when the effective temperature is high and diminishes when the system is cooled down. Our cooling schedule involves 25 temperature runs starting with T_e set to 8. After each temperature run has been finished the T_e value is decreased by 10%. For each temperature run the Markov chain of configurations is generated with 1000 Monte Carlo steps as described above. The configuration with the lowest value of the penalty function is saved and used as the initial configuration in the next temperature run. At the end of each temperature run the configuration acceptance ratio is calculated for each coordinate and if it drops below the value 0.1/48 (where 48 is the number of degrees of freedom), the maximal step length associated with a given coordinate is reduced by 5%. The first three temperature runs are carried out with the repulsive potential *repel* switched off to allow the rigid body elements to pass through one another. Then a repel force constant of 0.025 Å^{-4} is applied. The parameters of the optimization protocol are presented in Table 4. All calculations are performed using a purpose-built program PANDA developed in our laboratory. One SAMC trajectory performed on bR takes approximately 60 seconds on Silicon Graphics Indy 4400.

Statistical analysis

The simulated annealing protocol may occasionally give rise to a final configuration being trapped in a local minima. To avoid such situations we generate a number of Monte-Carlo-simulated annealing trajectories for an ensemble of different initial configurations of the system. As a result, a family of final configurations is created, which is then subjected to further analysis.

The initial configurations are generated in the following way. Helices are arranged perpendicularly to the membrane xy plane with their axes overlapping the z axis of the coordinate system as well as one another. They are oriented so that their intracellular ends are positioned on the upper side of the membrane plane ($z > 0$) whereas the extracellular ends are positioned on the lower side ($z < 0$). The positions of the helix centroid projections on the xy plane are then randomized within 2 Å from the initial ones and the orientations of each helix with respect to its z axis are fully randomized. A ligand is placed centrally in the xz plane and its orientation around the z axis is fully randomized.

The final configurations are assessed by their penalty function value and those with abnormally high values are discarded. The mean structure is calculated by averaging the accepted final configurations. The mean configuration is created by superposition of rigid body elements on the mean structure.

The analysis of final configurations was performed with the programs QUANTA (Molecular Simulations, Inc.) and SQUID (Oldfield, 1992).

RESULTS

Bacteriorhodopsin test case

A series of calculations were performed on the bR system for two main reasons. The first was to develop the simulated annealing methods described above and to optimize the various parameters in the protocol. The second reason was to demonstrate that a configuration for bR can be generated that is close to that observed in the EM structure using only experimental and theoretical data available before the determination of the structure. In these calculations, the reference coordinates of the EM structure are taken from the structure deposited in the Brookhaven databank (Bernstein et al., 1977) with the code 1brd except for the z coordinates of helix D, which have been increased by 3 Å as suggested by R. Henderson (personal communication).

Assignment of helices to their images in the projection map

The major question addressed here is how to assign helices to their 2-D images in Fig. 3, so that we could successfully use the experimental information from a 2-D projection map (Henderson and Unwin, 1975). Although the assignment of helices into their images in the map was correctly predicted (Engelman et al., 1980; Popot et al., 1989) well in advance of the determination of the structure (Henderson et al., 1990), the first set of calculations demonstrate that our method can produce the correct set of assignments for the bR helices.

There are in all 10,080 possible helix assignments (including the uncertainty in knowing which is the intracellular and extracellular side of the membrane). This is reduced to just 120 combinations with knowledge of the orientation of the projection map with respect to the extra- and intracellular membrane surfaces (Henderson et al., 1978), as well as the assignment of helices A and B to images 1 and 2, respectively (Popot et al., 1989). Each of the assignments is referred to by a seven-digit number in which a particular digit describes a helix and its position describes an image on the projection map (see Fig. 3). For example, assignment 1265473 means assigning helices A, B, F, E, D, G, and C to images 1, 2, 3, 4, 5, 6, and 7, respectively. The proper assignment revealed by the high resolution bR structure is 1234567 (Henderson et al., 1990).

For each of the 120 different helix assignments we construct a set of xy-position restraints using Table 2 *C*. Then we ask which of the assignments best fits the other available experimental information. To answer that we incorporate newly constructed xy-position restraints into a penalty function and identify the assignment that gives a minimal value of this function. The details of this calculation are presented below.

For each assignment, five Monte Carlo trajectories were calculated using all ORI, xy-position, z-position, and xy-distance restraints and the majority of the xyz-distance restraints. The simplified binding pocket restraints were used except for those involving helices B, D, and E, namely, Tyr-57, Asp-115, Trp-137, Trp-138, and Ser-141, as at this stage it was impossible to assign which parts of the ligand molecule were the closest to those helices (see Table 2, footnote 4). Also, as the tilt direction for the helices associated with images 1, 2, 3, 5, 6, and 7 were not known (only image 4 presents a helix that is fully perpendicular to the membrane plane), the appropriate xy-position restraints were modified. The x and y coordinates for DI and DE atoms in a particular helix (Table 2 *C*) were replaced by those of atom DC and a tolerance of ± 10 Å was applied. This should enable the six helices of unknown tilt to find a favorable arrangement. The repel potential was switched on after the first three temperature runs, and a higher repel force constant of 0.25 $Å^{-4}$ was used to avoid steric overlap.

For each of the 120 assignments the configuration with the lowest penalty function was selected and the assignments ranked according to this value. The ten best assignments had a penalty function from 84 to 600 and were taken for further analysis. Six of them have helix C assigned to image 3 and helix G to image 7 whereas the other four have helices C and G exchanged. For each of the assignments the tilts of the helices associated with images 1, 2, 3, 5, 6, and 7 were assigned after inspection of the mean configurations and 25 Monte Carlo trajectories calculated with the full set of xy-position restraints.

Fig. 7 shows the penalty function sorted over the 25 final configurations for the best six assignments. It is clear that assignments 1234567 and 1235467 are better than any other as their final structures all have lower penalty function values than any other assignment. Eight final configurations

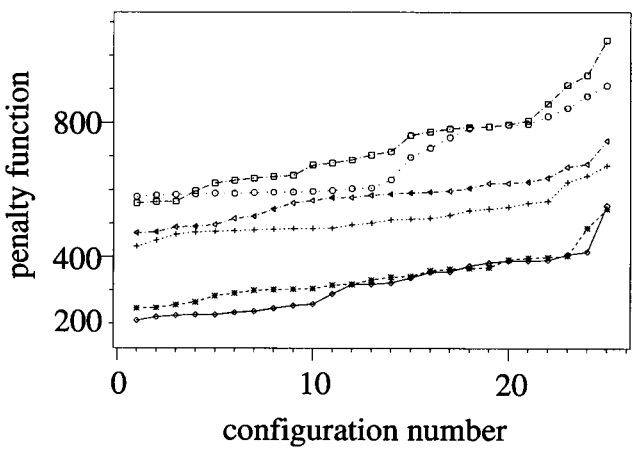

FIGURE 7 Penalty function (sorted) of final 25 configurations for 8 best helix assignments of bR. These assignments are as follows: ◇, 1234567; *, 1235467; +, 1236547; ◁, 1236457; □, 1234657; and ○, 1235647.

of assignment 1234567 have a lower penalty function than the best configuration of the 1235467 assignment. Furthermore, the 1235467 assignment requires that the second and third intracellular loops between helices need to cross each other, which would be sterically very difficult to achieve. A schematic representation of the best configuration obtained in these calculations is presented in Fig. 8. The results demonstrate that our methodology can determine the correct assignment for bR helices.

Final bacteriorhodopsin calculations

A final set of calculations was performed using the helix assignment 1234567 with all available ORI, *z*-position, *xy*-position, and *xy*-distance restraints and several *xyz*-distance restraints. The configuration schematically presented in Fig. 8 shows well defined tilts for all helices. These were used in assigning *xy*-position restraints. This configuration allowed several additions to the *xyz*-distance restraints. It is now possible to assign binding pocket restraints between Tyr-57 to ligand atom LS and Asp-115, Trp-137, Trp-138, and Ser-141 to ligand atom LR. The loop length restraints were not used as with a proper helix assignment they are no longer needed. In addition, the four restraints from FTIR difference spectroscopy (see footnote 6 of Table 2) were left out to exclude this more detailed information, which is not available for the rhodopsin calculations.

Fifty SAMC trajectories from different initial configurations were generated. The penalty function values of the fifty final configurations are distributed evenly in the range from 63.0 to 98.7, thus there is no reason to discard any of them from further analysis. Fig. 9 *A* shows the twenty-five best configurations overlapped on the C_α and ligand atoms. The average values of different restraint violations are presented in Table 5. The mean configuration of the final fifty structures has been calculated and is compared in Fig. 9 *B* with the configuration determined by EM (Henderson et al., 1990). Hereafter, the term EM configuration refers to a configuration created by superposition of ideal helices on the EM structure. The rms deviation on C_α and ligand atoms is 1.67 Å from the EM configuration and 1.87 Å from the EM structure, which shows that the error caused by using idealized helices is very small indeed. To quantify the structural convergence of final structures, the rms deviation

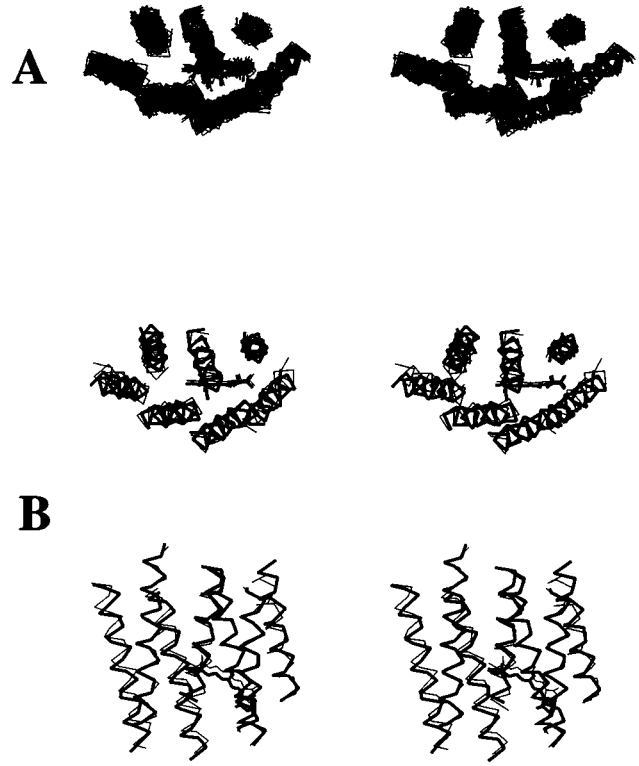

FIGURE 9 (*A*) Stereo plot of 25 best final structures of bR (C_α atoms only) overlapped on C_α and ligand atoms. The view is from the intracellular side. (*B*) Orthogonal stereo views of the C_α atoms for the mean configuration calculated for bR (*bold line*), overlapped with the EM configuration (*thin line*) by least squares superposition of all C_α and ligand atoms. The EM configuration is created by superposition of the ideal helices on the experimental EM structure. The top view is from the intracellular side.

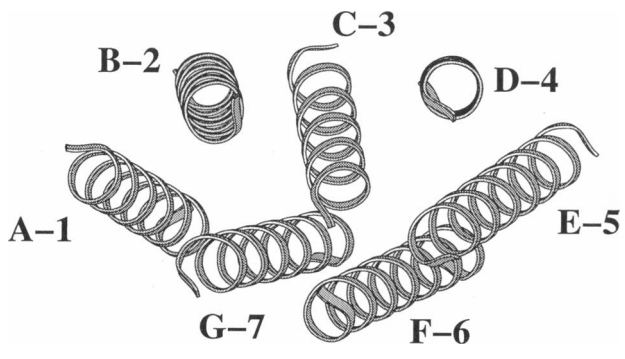

FIGURE 8 The best configuration (C_α atoms only) of 1234567 assignment, obtained with bR helix assignment calculations. The favored tilt of helices A, B, C, E, F, and G associated with helix images 1, 2, 3, 5, 6, and 7, defined in Fig. 3, is clearly visible. The view is from the intracellular side. This figure was generated with Molviewer (M. J. Hartshorn, personal communication).

TABLE 5 Characterization of the family of the final 50 bR configurations

RMSV[a] of *xyz*-distance restraints	0.030 ± 0.013Å
RMSV of *xy*-distance restraints	0.041 ± 0.014Å
RMSV of *xy*-position restraints	0.051 ± 0.013Å
RMSV of *z*-position restraints (membrane)	0.028 ± 0.017Å
RMSV of *z*-position restraints (ligand)	0.002 ± 0.005Å
RMSV of ORI restraints[b]	2.54° ± 0.480Å
RMSD[c] (mean)	1.79 ± 0.42 Å
RMSD[d] (EM)	1.67 Å

[a] RMSV, root mean square violation of the upper or lower limits of particular restraints as defined in Appendix B.
[b] The violation of ORI restraints, expressed in an angular rather than a distance form, as defined in Appendix B.
[c] Convergence of the family of final configurations, expressed as rms deviation from the mean configuration fitted on C_α and ligand atoms.
[d] The rms difference between the mean configuration and the EM configuration fitted on C_α and ligand atoms.

from the mean configuration was calculated for the family of final configurations. This value is 1.79 ± 0.42 Å.

The comparison with the EM configuration gives slightly different results for different helices. The rms deviation averaged over C_α atoms of each helix are 2.60 Å for helix A, 1.99 Å for helix B, 1.22 Å for helix C, 0.64 Å for helix D, 1.96 Å for helix E, 0.51 Å for helix F, 1.17 Å for helix G, and 1.50 Å for the retinal.

Reproducibility of the results

To assess the reproducibility of results we have performed the entire bR final calculation, described in a previous section, five times using different initial random number seeds. All of them gave essentially the same results. The accuracy described in terms of rms deviation of the mean configuration from the EM configuration is 1.68 Å ± 0.03 Å over all five runs. The convergence varies from 1.79 Å to 1.90 Å and the standard deviations do not exceed 0.5 Å over 50 final configurations. The rms deviation (on C_α and ligand atoms) between different pairs of mean configurations is 0.33 Å ± 0.07 Å over all 10 pairs, which is very small considering the low resolution of modeling.

Stability of results: restraint tolerances

It is also interesting to consider the effect of the tolerances of different groups of restraints on the quality of the final results. This seems especially important as it would help to answer the problem of the precision in the evaluation of the restraint bounds. Here, we have performed the tests in which the limits of a particular group of restraints were increased by 0.1 Å and 0.5 Å. Groups of restraints subjected to tolerance changes are as follows: (1) *xyz*-distance restraints, (2) *xy*-distance restraints, (3) *xy*-position restraints, (4) *z*-position restraints, and (5) all four groups together. The ORI restraints are excluded from this analysis, as their construction does not require the evaluation of tolerances. The calculations were performed in the same way as described in a previous section, except that the *xyz*-distance restraints contained the subset of simplified binding pocket restraints (Table 2 *A*, footnote 4) rather than the binding pocket restraints (Table 2 *A*, footnote 3). Results presented in Table 6 show that extending the restraint limits by 0.1 Å is insignificant. Relaxing of the restraints from each group separately results in mean configurations that are very similar to the one obtained with the original restraints. The appropriate rms deviations are less than 0.4 Å, which is the margin of the reproducibility of the method (previous section). Only in the case of simultaneous change did this number grow to 0.53 Å, which is still acceptable in terms of low resolution modeling. Relaxing of *xy*-distance, *xy*-position, or *z*-position restraints by 0.5 Å brings about slight deterioration of the quality of the final configuration. However, even the simultaneous change of limits by 0.5 Å results in a mean configuration that is only 0.67 Å distant from the one generated with no limit change.

TABLE 6 Comparison of the results for bacteriorhodopsin using restraints with modified limits

	Restraint limits extended by 0.1 Å		Restraint limits extended by 0.5 Å	
Restraints with extended limits	RMSD EM bR[a] (Å)	RMSD all[b] (Å)	RMSD EM bR[a] (Å)	RMSD all[b] (Å)
None[c]	1.69		1.69	
xyz-distance	1.68	0.25	1.65	0.40
xy-distance	1.71	0.30	1.81	0.53
xy-position	1.67	0.34	1.91	0.46
z-position	1.80	0.38	1.80	0.46
All[d]	1.80	0.53	1.96	0.67

[a] The rms difference (on C_α and ligand atoms) between the mean configuration and the EM bR configuration fitted on C_α and ligand atoms.
[b] The rms difference (on C_α and ligand atoms) between the mean configuration and the mean configuration generated with no extension of restraint limits, fitted on C_α and ligand atoms.
[c] Run with the same restraint set but no limit extension.
[d] Run with all four groups of restraints extended.

Stability of results: force constants

The force constants in our penalty function have been chosen somewhat heuristically. As shown in Table 4 we use a uniform force constant of 50 $Å^{-2}$ for all distance, position, and ORI restraints. In the case of simplified ORI restraints the force constant is reduced 10-fold as these restraints are very strict and are hardly satisfied (see Appendix B). To determine the effect of using different force constant values on the quality of final results we performed the final bR calculations using the restraint force constant of (1) 5 $Å^{-2}$ and (2) 500 $Å^{-2}$. We have also performed calculations with the repel force constant of (3) 0.003 $Å^{-4}$ and (4) 0.250 $Å^{-4}$. The mean configurations of runs 1, 2, 3, and 4 have rms deviation of 1.82 Å, 1.85 Å, 1.71 Å, and 1.81 Å from the EM configuration, respectively. Their rms deviation from the mean configuration of the final bR run discussed above are all less than 1 Å. These results show that our choice of force constants is reasonably optimal and that the method is not very sensitive to the force constant selection.

Stability of results: presence of different groups of restraints

The importance of different groups of restraints on the quality of final configurations has been assessed by performing several calculations in which different subsets of restraints are left out. For each calculation, 50 SAMC runs were performed and a mean configuration was generated. Table 7 shows the rms deviation between these mean configurations and the EM configuration (Henderson et al., 1990). Throughout these calculations *xyz*-distance restraints for a simplified binding pocket are used as they are more appropriate for a general case of GPCR.

Treatment of the case where *xy*-position restraints are left out is slightly different and requires extra description. First, leaving out *xy*-position restraints implies leaving out *xy*-distance restraints derived from neutron diffraction experi-

TABLE 7 Comparison of the results for bacteriorhodopsin using different subsets of restraints

Restraints used in calculating structures	No. of selected configurations	RMSD[a] (Å)	convergence[b] (Å)
ALL[c]	50	1.67	1.79 ± 0.42
all[d]	46	1.69	1.85 ± 0.27
all without *xy*-distance	48	1.94	1.72 ± 0.36
all without neutron diffraction restraints[e]	46	2.28	1.77 ± 0.42
all without *xyz*-distance	50	2.38	2.36 ± 0.61
all without *xy*-position[f]	27	3.20	2.65 ± 0.58
all without *z*-position (membrane)	47	3.34	2.92 ± 0.91
all without ORI	50	3.32	2.20 ± 0.41
all with 7 ORIs only[g]	46	2.29	1.90 ± 0.37
Model A[h]	29	3.21	2.61 ± 0.59
Model B[i]	47	2.34	1.77 ± 0.37

[a] The rms difference (on C_α and ligand atoms) between the mean configuration and the EM bR configuration fitted on C_α and ligand atoms.

[b] Convergence expressed as rms deviation (on C_α atoms) of a family of selected configurations from the mean configuration.

[c] ALL refers to all restraints including *xyz*-distance restraints for binding pocket (Table 2, footnote 3), as used in the final bR calculation.

[d] *all* refers to ALL restraints except that *xyz*-distance restraints for the binding pocket (Table 2 footnote 3) have been replaced by the restraints for the simplified binding pocket (Table 2, footnote 4).

[e] *all* restraints without *xy*-distance restraints and *z*-position (ligand) restraints, both derived from neutron diffraction experiments.

[f] *all* restraints without *xy*-position restraints; here, *xy*-distance restraints were also excluded as they depend on the 2-D projection map. It is assumed here that each helix in a bundle is positioned next to its neighbor in the sequence. This information is translated into an extra set of weak *xy*-distance restraints; also a different construction of initial configurations is used (see text).

[g] *all* restraints, but a set of 27 original ORI restraints is replaced by 7 ORI restraints (1 ORI restraint per helix).

[h] This is the same set of restraints as described in footnote *f* except that *z*-position (ligand) restraints are left out as well.

[i] *all* restraints without *xy*-distance and *z*-position (ligand) restraints derived from neutron diffraction experiment (as under footnote *e*), except that the tolerances for *xy*-position restraints are taken from the rhodopsin projection map (Table 3 *B*).

ments as well, because construction of the latter ones requires knowledge of the 2-D helix arrangement (Heyn et al., 1988). Second, the results and the treatment depend on the knowledge of the order of the helices in the bundle. If this order is unknown and consequently no adequate information is incorporated into simulation, then the final configurations are widely scattered in configurational space with no clear conclusions about helix arrangement (results not shown). However, if the order of helices is known, which is the case as demonstrated in this paper for both bR and rhodopsin, and we believe it is the same throughout the GPCR family, then it gives rise to extra information being held in the simulation. In our test this information is supplied at two stages. First, for each SAMC run the initial configuration is built differently to the procedure presented in Methods. Here, helices are positioned perpendicular to the membrane plane at the apices of a regular heptagon 10 Å from its centroid, so that each helix is lined up next to its neighbor in the sequence, clockwise looking from the intracellular side. Then the initial randomization is performed in the same way as previously described (see Methods). Second, to keep a bundle-like arrangement throughout the calculations, *xy*-distance restraints are introduced between neighboring helices in the bundle. The distances between helix ends on the same side of the membrane plane and between helix centroids are restrained to be no longer than 12 Å.

The results presented in Table 7 show that none of the analyzed groups of restraints are necessary to reach a converged set of final configurations; however, different restraints have different impacts on the quality of final configurations. Three sets of restraints, namely, ORI, *xy*-position, and *z*-position (membrane) restraints seem to be extremely important as their absence leads to the most significant distortion of the final structures. The mean final configurations produced in these three tests are ~3.2 Å distant from the EM configuration in terms of rms deviation. The omission of *xyz*-distance or neutron diffraction restraints (*xy*-distance and *z*-position (ligand)) have lesser effect on the quality of the mean configurations, which are ~2.3 Å distant from the EM configuration. Finally, *xy*-distance restraints from neutron diffraction are the least significant. It is interesting to note that the improvement in the definition of the binding pocket restraints, which is possible with the additional *xy*-distance restraints from neutron diffraction (binding pocket restraints versus simplified binding pocket restrains, see Table 2 *A*, footnotes 3 and 4), has no significant effect on the quality of the final configurations.

The other tests (results not shown) demonstrated that none of the single group of restraints is capable of producing a reasonably good structure, even if accompanied by *z*-position (membrane) restraints. The final mean configurations do not come closer than 4.2 Å to the EM confguration. Here, the ORI restraints, *xyz*-distance restraints, and *xy*-position restraints seem to be equally important.

We have also performed calculations for the case in which the amount of bR constraint data was reduced to simulate the amount of data likely to exist for GPCRs. To do so we have left out *xy*-distance and *z*-position restraints derived from neutron diffraction experiments. Two models are presented: (A) without and (B) with *xy*-position restraints from a 2-D projection map. The latter uses *xy*-position restraint tolerances from rhodopsin data (Table 3 *B*) that are, by and large, higher than those for bR. The results are presented in Table 7. Comments on these results can be found in the Discussion. Here, we simply point out that a comparison of the results for model B with those of the test, in which *xy*-distant and *z*-position restraints are left out (Table 7, row 4), shows that using significantly higher tolerances for *xy*-position restraints does not result in significant deterioration of the quality of final structures.

The problem of stability of results with respect to numbers of restraints in each group, in particular answering the question of the minimal amount of restraint data, is more

complicated. Not only does it require analysis of the effect of different numbers of good restraints but also the side effect of incorporation of a small number of spurious restraints. Thorough investigation on this subject should be addressed elsewhere. Partial analysis is presented in Table 7 that shows the results of the test in which the total number of 27 ORI restraints has been reduced to 7, one ORI inside restraint per helix. One can see that although the quality of the mean configuration has deteriorated it is still only 2.3 Å from the EM configuration.

Rhodopsin

Assignment of helices to their images in the projection map

In contrast to bR there are no experimental data on helix assignments for rhodopsin. The projection map (Schertler et al., 1993) contains two symmetry-related images of the rhodopsin molecule (see Fig. 4), either of which may present the view from the same side of the membrane. This gives 10,080 possible helix assignments. Assignments to molecule image A are coded by seven digits in a way analogous to the bR case, whereas codes of assignments to molecule image B contain an additional character m (for mirror), e.g., 1234567m. The assignment 1234567 (analogous to the bR assignment) has been suggested by the analysis of the lipid exposure of seven helices in the 2-D model (Alkorta and Du, 1994; Baldwin, 1993). It was, however, obtained under the critical assumption that each helix must be positioned next to its neighbors in the sequence. This assumption was formulated by Baldwin (Baldwin, 1993) on the basis of multiple sequence alignment of GPCRs and the analysis of range of lengths of interhelical loops. Our database search for interhelical loops presented above shows, however, that the loops are not sufficiently short to assume packing of all sequential helix pairs in a hairpin conformation.

We have performed calculations to assign helices in rhodopsin by considering all possible assignments after elimination of those unlikely for a number of reasons. Detailed investigation of the projection map reveals that the ends of helices associated with pairs of images 1,5; 1,4; and 2,5 (see Fig. 4) cannot be closer than 24.7 Å, 21.5 Å, and 16.8 Å, respectively. Bearing in mind the loop length restraints imposed on loops AB, CD, and FG (see Table 3 and Methods) we reject a priori those assignments for which helices A and B are associated with images 1,5; 1,4; or 2,5; helices C and D with images 1,5 or 1,4; and helices F and G with images 1,5 or 1,4. Helices C and G are linked via a salt bridge between Glu-113 and Lys-296, whereas Gly-90 must be positioned close to Lys-296 (Rao et al., 1994). Model studies of these interactions have revealed that the maximal distances between helix axes of the helix pairs C-G, B-G, and B-C are all shorter than 20 Å. It is therefore reasonable to exclude all assignments for which these three pairs of helices are associated with images 1,5 or 1,4. In addition, we eliminate all assignments for which loop connections within one face cross one another as they are very unlikely (Engelman et al., 1980; Taylor et al., 1994). All of the conditions mentioned above reduces the number of possible helix assignments to 1742. These have been analyzed with calculations similar to those performed for the bR test case.

Five Monte Carlo trajectories have been calculated for each assignment. All ORI, *xy*-position, *z*-position, and *xyz*-distance restraints were used except for a binding pocket restraint for Ala-164 as it was not clear which part of the ligand is the closest to Ala-164 (see Table 3, footnote 16). The *xy*-position restraints were modified because at this stage the tilt directions for the helices associated with images 1, 2, and 3 were not known. This was addressed in a similar manner to the bR test case. The appropriate *xy*-position restraints were modified so that *x* and *y* coordinates for DI and DE atoms in a particular helix (Table 3 *B*) were replaced by those of the DC atom and a tolerance of ± 10 Å was applied. The *repel* potential was switched on after the first three temperature runs, and a higher repel force constant of 0.25 $Å^{-4}$ was used to avoid steric overlap. Test calculations revealed that the penalty function term associated with ORI restraints is ~ 10 times higher than in the case of bR, which makes the penalty function significantly dominated by violation of ORI restraints. This is caused by the high number of ORI restraints, which makes them difficult to satisfy. We have therefore reduced the ORI force constant to the value of 5 $Å^{-2}$.

For each of the 1742 assignments a configuration of the lowest penalty function was then selected and the assignments were ranked according to it. The 16 best assignments (which had penalty function values from 33 to 96) were selected for further analysis. For each of the assignments the tilts of helices associated with images 1, 2, and 3 were assigned after the inspection of the best configuration. This enabled the addition of an *xyz*-distance restraint between Ala-164 and retinal of the form of LR - C_α-Ala-164 or C15 - C_α-Ala-164. Similarly, an *xyz*-distance restraint C15 - C_α-Glu-122 was converted into LR - C_α-Glu-122 wherever applicable. Twenty-five SAMC trajectories were then calculated for each of the 16 assignments with full *xy*-position restraints. Penalty functions of the best 9 assignments are presented in Fig. 10. The best assignment has proved to be the one analogous to the bR one, namely, 1234567. The penalty function value of the best configuration is 36.8 as compared with 51.4 of the second assignment 1236547. Fifteen of the twenty-five final configurations of the first assignment have lower penalty functions than the best configuration of the second one. A schematic representation of the best configuration obtained in these calculations is presented in Fig. 11.

Final rhodopsin calculations

The final calculations have been performed using helix assignment 1234567, with helices A, B, and C, tilted as shown schematically in Fig. 11. With an exception of loop-

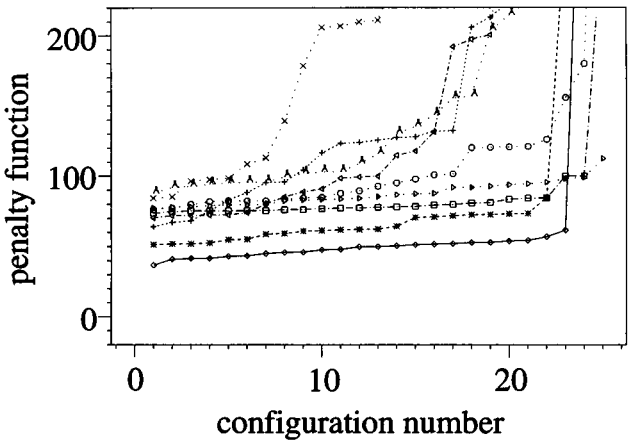

FIGURE 10 Penalty function (sorted) of the final 25 configurations for the 9 best helix assignments for rhodopsin. These assignments are as follows: ◇, 1234567; *, 1236547; +, 6521473; ◁, 2134567; □, 4231567; ○, 4271653m; ▷, 1435672; ×, 5672143; and Å, 6534127m.

length restraints, which we intended to use only in helix assignment calculation, the full set of restraints was applied as presented in Table 3. We have performed 50 SAMC trajectories for 50 different initial configurations. Penalty function values of the 50 final configurations are distributed smoothly in the range from 33.3 to 50.3. The 25 best configurations overlapped on the C_α and ligand atoms, and the mean configurations are presented in Fig. 12.

The average values of different restraint violations are presented in Table 8. It is striking to notice that the ORI restraints are satisfied more poorly than the others. The ORI penalty function (Appendix B) restrains residues to inside or outside of the seven-helix bundle. In fact, certain residues declared inside can actually be positioned between the adjacent helices. We have not, however, created a special interhelical class of ORI-restrained residues as it is often hard to say if, e.g., a conserved residue should be positioned inside the bundle or between the adjacent helices. Moreover, an interhelical restraint penalty function would have to be bimodal, which means an increased number of local

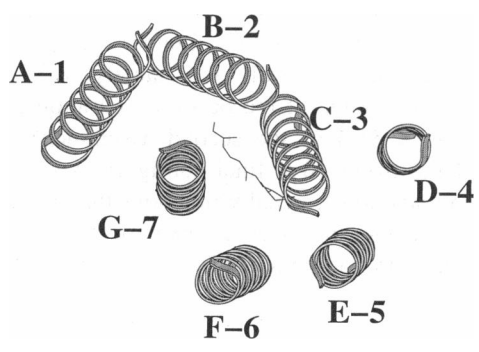

FIGURE 11 The best configuration (C_α atoms only) of 1234567 assignment, obtained from the rhodopsin helix assignment calculations. The favored tilts of helices A, B, and C associated with helix images 1, 2, and 3, defined previously in Fig. 4, are clearly visible. The view is from the intracellular side.

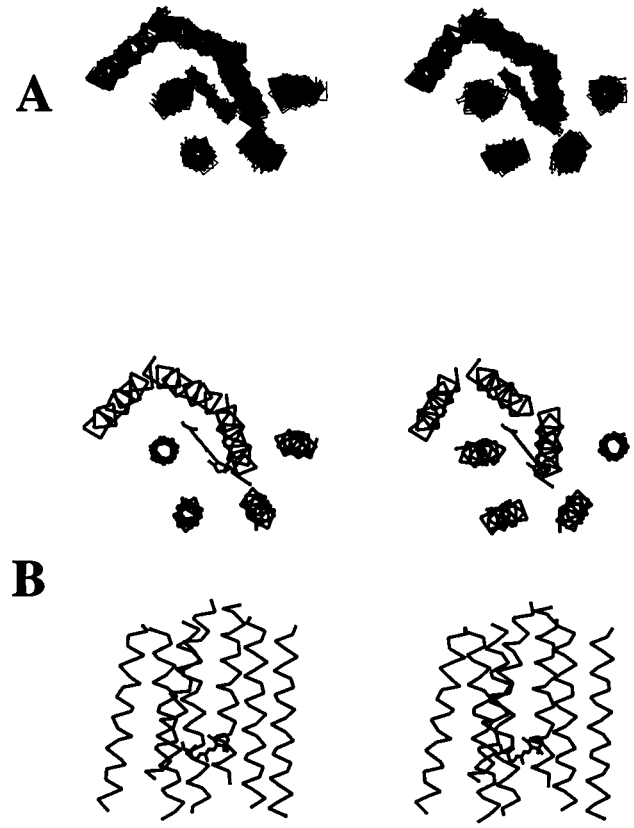

FIGURE 12 (*A*) Stereo plot of 25 best final structures of rhodopsin (C_α atoms only) overlapped on C_α and ligand atoms. The view is from the intracellular side. (*B*) Orthogonal stereo views of the C_α atoms for the mean configuration averaged over the whole family of 50 final rhodopsin configurations. The top view is from the intracellular side.

minima are available. In the absence of an interhelical class of ORI restraints, any residues positioned between adjacent helices may produce violations of the ORI restraints. Indeed, the final mean configuration shows 10 residues that violate ORI inside restraints and all of them are positioned between adjacent helices.

Although the ligand was initially given an even distribution of orientations between the two possible, i.e., (1) *xz* plane and C19 pointing at the extracellular side and (2) *xz* plane and C19 pointing at the intracellular side were evenly distributed, the final configurations show a 36:14

TABLE 8 Characterization of the family of the 50 final rhodopsin configurations

RMSV[a] of *xyz*-distance restraints	0.011 ± 0.010Å
RMSV of *xy*-position restraints	0.024 ± 0.011Å
RMSV of *z*-position restraints	0.005 ± 0.008Å
RMSV of ORI restraints[b]	9.55° ± 0.68°
RMSD[c] (mean)	1.59 ± 0.22 Å

[a] The rms violation of the upper or lower limits of particular restraints as defined in Appendix B.

[b] The violation of ORI restraints, expressed in an angular rather than distance form, as defined in Appendix B.

[c] Convergence of the family of final configurations expressed as rms deviation from the mean configuration fitted on C_α atoms.

preference for orientation 1. Furthermore, of the best 10 configurations the ligand with orientation 2 is present only once, ranked at number 8. These results suggest that the 11-*cis* retinal in rhodopsin lies roughly in the *xz* plane with C19 pointing at the extracellular side as presented in Fig. 12 *B*.

Reproducibility of the results

We have assessed the reproducibility of the rhodopsin results by repeating the entire final calculation four times using different initial random number seeds. The final configurations proved to be very similar as the rms difference between different pairs of mean configurations is 0.29 Å ± 0.10 Å over all 10 pairs. The convergence varies from 1.47 Å to 1.59 Å and the standard deviations do not exceed 0.5 Å over the number of selected final configurations.

DISCUSSION

The aim of this work is to propose a rule-based automated technique for aggregating helices in seven-helix membrane proteins on the basis of existing experimental and theoretical data to build a structural template that can then be used for further modeling. As far as we are aware, this is the first unbiased, objective modeling procedure for this class of protein systems. The results presented in this paper demonstrate that the technique described is capable of determining the 3-D low resolution structure of seven-helix membrane proteins from various theoretical and experimental data.

The results presented in this paper showed high reproducibility of the method for both bR and rhodopsin. The comparison of the results for bR with the EM structure showed that the accuracy of the method, which is a function of the methodology used as well as the distribution and the quality of restraints, proved high. The restraints used in both cases ensured reasonably good convergence of the final configurations. The additional tests performed on bR showed that the final results are not too sensitive to the restraint limits and to the choice of force constants.

Test case calculations for bR permitted optimization of the method with respect to several parameters (see Table 4) by generating a final bR C_α atom template close to the high resolution bR structure (Henderson et al., 1990). The results have shown that the rms difference between the final configuration and the target EM bR configuration is only 1.7 Å on C_α atoms, and the helix arrangement is close as may be seen from Fig. 9 *B*. The rms difference is less then 2 Å for all helices except for helix A and significantly less then 1 Å for helices D and F. The relatively poor performance for helix A may be explained by the lack of *xyz*-distance restraints for this helix as there is no evidence that it contributes to the retinal binding site. There is only one ORI restraint for helix A and the *xy*-distance restraints are relatively loose as seen in Table 2 *B*, which may account for the underdetermination of the position of helix A. However, in general there is no correlation between the number of restraints and the quality of the final helix position. For example, helix D, the rms deviation of which is 0.64 Å, has only one ORI restraint and one *xyz*-distance restraint whereas helix E, the rms deviation of which is 1.96 Å, has eight ORI restraints and three *xyz*-distance restraints. It is worth mentioning that the position of helix D is actually better described than the original position of helix D in the 1brd structure (Bernstein et al., 1977), which required a 3-Å translation out of the membrane plane as suggested by R. Henderson (personal communication). The quality of the final model structure is particularly striking considering the simplicity of the system used with a crude protein representation and rigid and idealized helices with no allowance for kinks and local distortions. The idealized helix approach seems to work rather well for bR as the rms difference between the EM bR structure and the EM bR configuration is only 0.86 Å on C_α atoms.

The test case results were produced by using experimental data that were obtained with no knowledge of the high resolution EM bR structure. For example, we have refrained from using data on replacements of residues Leu-93 (Subramaniam et al., 1991) and Met-143 (Ihara et al., 1994) because these mutations were prompted by the knowledge of the binding site geometry to analyze its changes during retinal isomerization. Although more data of higher quality had been published, we have refrained from using it as we did not want to use data of significantly higher quality than those that can currently be obtained for rhodopsin or other GPCRs. For example, we have used an electron density projection map at 7 Å published in 1975 (Henderson and Unwin, 1975), although a 2.8-Å map was available (Baldwin et al., 1988). This retains parity with the quality of results available for rhodopsin.

The methodology is successful in the detection of the proper bR helix assignment to their images in the electron density projection map. Although the assignment had been correctly predicted before the high resolution bR structure was solved, it is important to show that the method is capable of solving the helix assignment problem as it makes the prediction for rhodopsin more credible.

The comparison of restraint selections for bR and bovine rhodopsin presented in Tables 2 and 3 show that in both cases there are restraints having major effects in limiting configurational space. These restraints are the *xy*-position restraints from electron density projection maps, *z*-position restraints imposed by the presence of membrane, and ORI restraints. Although in the case of bR we have extra *xy*-distance and *z*-position restraints from neutron diffraction, the former are not crucial for the quality of the final configurations (see Table 7). Rhodopsin on the other hand shows more ORI restraints because of the multiple sequence alignment analysis. They are more equally distributed between helices than in bR, which together with their higher number may give rise to a better determination of the helix orientation. Indeed, perhaps for that reason, the convergence of rhodopsin final configurations is higher than for bR.

The method has been used to assign helices to their images in the electron density projection map of bovine rhodopsin (Schertler et al., 1993). This was computationally more demanding because no experimental data on the assignment were available. A large number of possible assignments were assessed without the assumption that each helix must be positioned next to its neighbors in the sequence (Baldwin, 1993). The final assignment, however, is consistent with the 2-D helix arrangement obtained by the analysis of the lipid exposure of seven helices in the 2-D model (Alkorta and Du, 1994; Baldwin, 1993).

Comparison of our final bR template with that for rhodopsin (see Figs. 9 and 12) reveals important differences in the spatial organization of the helices. It also demonstrates the difference between our rhodopsin model and GPCR models based upon the bR structure. On the intracellular side of our model, helix C is deeply buried between helices B, D, E, F, and G. Helix A is tightly packed with helices B and G. On the extracellular side, however, helix C is pushed out of its buried position as it is now partially occupied by the ligand. The extracellular end of helix C occupies a more bR-like position although its distance to the helix G end is much shorter than in bR. The general features of the model are similar to those of Donnelly's model (Donnelly et al., 1994), although there are some subtle differences. For example, helix D seems to be positioned slightly closer to the ligand binding site. Helix arrangement is also consistent with the one proposed by Baldwin (1993).

Recently a new sharpened projection density map of bovine rhodopsin at a 9-Å resolution has been published (Unger and Schertler, 1995). This map is consistent with the earlier one used by us (Schertler et al., 1993) except that a helix associated with image 5 (see Fig. 4) appears to be more tilted. The arc-shaped feature of images 1, 2, and 3 is extended by image 5 although the tilt of the helix corresponding to the latter image is considerably less than those of images 1, 2, and 3 (see Figs. 7 and 8 in Unger and Schertler, 1995). The rhodopsin template generated in this paper seems consistent with this finding. As one can see from Fig. 12, the direction of the tilt of helix E matches that seen in the new projection map. We believe that the rhodopsin low resolution model presented in this paper gives a solid template for further modeling of rhodopsin as more experimental data become available as well as for modeling other GPCRs.

The penalty function used throughout the calculations can easily be extended to work with other kinds of data. For example, if an approximate 3-D template is known, then the *xy*-position restraints can easily be extended to 3-D. Similarly, extra terms can be introduced for the restraining crossing angle between adjacent helices.

In the penalty function presented in this paper we have not used the concepts of periodicity in hydropathy, sequence variability, or substitution pattern (Donnelly et al., 1993; Eisenberg et al., 1982; Taylor et al., 1994) to explicitly guide the orientations of helices along their axes. Instead we employ ORI restraints, the advantage of which is that they deal with all restrained residues separately rather than restraining moments, which are the averaged quantities. We feel that the concepts mentioned above are useful mainly in 2-D helix arrangement and using them in 3-D modeling for purposes other than initial helix orientation would be an oversimplification. Indeed we have made tests with the additional penalty term for keeping the helix buried face vectors oriented inside using the α-helical periodicity concept of Donnelly (Donnelly et al., 1993). This produced no improvement for the bR results, although it may be appropriate to use it in the case for which other data are insufficient (e.g., low number of ORI restraints).

The optimization problem addressed in this work is not a typical simulated annealing problem. For these proteins it is known that the approximate organization of the helices will be roughly parallel to one another and positioned essentially perpendicular to the membrane plane. Furthermore, we know that the region of configurational space containing the global minimum is roughly described by *xy*- and *z*-position restraints. We also know which helix ends should be on the extra- and intracellular sides. All this knowledge allows us to make an initial configuration that is not entirely random but already biased toward the final structure. This means that the initial configuration does not have to be completely randomized, reducing the time spent searching uninteresting regions of configurational space. Instead, the initial configuration is relatively close to the final configuration with the helices arranged close and parallel to one another and the *z*-axis with their ends positioned on the correct sides of the membrane plane. Such initial configurations ensure rapid convergence on the interesting regions of configurational space. Consequently, there is no need to melt the system, and the effective temperature $T_e = 8$ is high enough to ensure an acceptance ratio (0.44) adequate for reaching the global minimum. Although the choice of the initial configuration narrows significantly the configurational space to be searched, this configuration is far enough from the global minimum so as not to bias the final result.

In a typical optimization problem the configuration with the lowest penalty function value is regarded as a final result. In our calculations, however, the final result is represented by the configuration closest to the average over the family of accepted configurations. There are two reasons for this. First, our penalty function is an approximation containing many parameters such as force constants and restraint lower and upper limits, which have been chosen in a rather heuristic way. Also, the protein representation used is far from being an accurate one. Consequently, it is an approximation to say that the configuration corresponding to the global minima of the penalty function best satisfies the structural restraints coming from experimental and theoretical knowledge. Second, the data we use for creation of the penalty function for bR may not necessarily favor the EM bR configuration. Consequently, the global minima configuration does not have to be the best approximation of the EM bR configuration even if obtained with the most accurate penalty function. Therefore we think that the final results seem to be better approximated by a family of

configurations close to the configuration of the lowest penalty function rather than by a best configuration itself. Indeed the rms deviation of the lowest penalty configuration from the EM bR configuration is 2.46 Å (as compared with the 1.67 Å of the mean configuration). Furthermore, the configuration closest to the EM bR one (rms deviation = 1.90 Å) is far down the list of final configurations. It is partly for that reason that 1000 Monte Carlo steps per temperature run seem to be enough. Tests showed that, although using 5000 MC steps leads to better convergence of final configurations, it results in losing the accuracy of the mean configuration.

It is worthwhile to discuss the precision of the restraint limits needed to generate accurate results. The numbers presented in Tables 2 and 3 are often quoted with two decimal digits. This is a result of either the preliminary modeling or scaling procedure. The tests performed on bR (Table 6) showed that relaxation of the restraints by 0.1 Å is insignificant, whereas the relaxation by 0.5 Å still generates acceptable configurations less than 2 Å distant from the EM configuration. This suggests that precision higher than that of one decimal place is not really necessary for achieving good quality results.

It seems interesting to consider how appropriate the method is for application to other GPCRs. The current implementation of the method uses rigid helices. The high resolution structure of bR shows that the helices are close to ideal; thus this approach is justified at least as a first approximation. If a particular helix geometry is known, it may be easily used instead of idealized helices but will be treated as a rigid structure throughout the calculations. Multiple sequence alignment of GPCRs shows the existence of highly conserved prolines in helices D, E, F, and G, which suggests that helix kinks may occur at those positions. On the other hand, helices D, F, and G are observed to be the least tilted in the rhodopsin projection maps (Schertler et al., 1993; Unger and Schertler, 1995), which suggests that those kinks cannot have that much affect. In future versions of the program we intend to allow kinks wherever a conserved proline occurs.

Both bR and rhodopsin are rather well characterized in terms of experimental knowledge compared with other GPCRs. It seems that, at present, the application of the method to other GPCRs will rely mainly on four classes of data: (1) ORI restraints from site-directed mutagenesis, analysis of multiple sequence alignment, or analysis of α-helical periodicity in sequence; (2) *xyz*-distance restraints between part of a ligand and transmembrane residues, derived from possible cross-linking experiments or site-directed mutagenesis; (3) *z*-position restraints imposed on helix ends by the membrane; and (4) knowledge of the order of helices in a bundle, assumed to be the same as in rhodopsin. A test performed on bR shows that using only these four classes of data results in a mean configuration that is 3.2 Å distant from the EM configuration (Table 7). The deterioration of the quality of the final configuration is mainly due to the lack of *xy*-position restraints from the 2-D projection map. It is true that the only GPCR for which such a map exists is bovine rhodopsin. However, a detailed sequence analysis leads to the conclusion that the basic 3-D helix arrangement is common (to a first approximation) among GPCRs (Baldwin, 1993; Donnelly et al., 1994). It is therefore possible to improve the results by introducing extra information from the 2-D rhodopsin projection map. This case has been assessed by performing another test on bR in which *xy*-position restraints, with tolerances taken from the rhodopsin data, are added to the reduced input data described above. The improvement of the quality of results is significant as the final mean configuration is only 2.3 Å distant from the EM configuration. Alternatively, one could use a 3-D template built for rhodopsin as a source of 3-D *xyz*-position restraints. These restraints could be used in addition to the experimental data available for a particular GPCR.

Another important issue is quality and efficacy of restraints available for GPCRs. Because at present the main source of data seems to be site-directed mutagenesis and multiple sequence alignment, a question arises on applicability of these types of data for constructing structural restraints. In general, restraints can be easily defined if the effect of a particular mutation in a mutagenesis experiment is direct, and both ORI and weak *xyz*-distance restraints can be applied. In the case of bR and rhodopsin it was possible to identify groups of mutations causing direct effect, e.g., those affecting spectral characteristics of the retinal. It seems significantly more difficult to perform this identification for other GPCRs, which makes it important to know how to handle the more common case of mutations having an indirect effect. In such cases we assume throughout this work that, if a mutation concerning a transmembrane residue affects receptor function, then a particular residue should not be oriented outside a seven-helix bundle. Taking into account a bimodal character of our ORI restraints (Appendix B) it means technically orientation inside. A similar interpretation can be applied to multiple sequence alignment data. Conserved are those residues with spontaneous mutations that have been rejected in the evolution process; hence, they must have affected the receptor function. Thus, conserved residues should be oriented inside (or not outside). Although in general both assumptions concerning site-directed mutagenesis and multiple sequence alignment data seem to be true, one cannot rule out the situation in which they are not. This makes ORI restraints particularly likely to contain a number of spurious restraints coming from misinterpretation of data. However, if, by and large, the number of correct restraints prevails, then the effect of the spurious restraints will be reduced as they are in a state of competition with the correct ones. It is for that reason that we advocate using a large number of ORI restraints.

We thank Mike Hartshorn for programs to calculate the mean structure as well as MOLVIEWER; Liz Potterton for help in interhelical loop analysis using QUANTA; Glenn Proctor for help in automation of the calculations for different helix assignments; Jan Zelinka for use of his ZMOLVIEWER

TABLE A Parameterization of C_α virtual atoms and C_β virtual atoms for each amino acid and ligand atoms

Virtual atom[a]	van der Waals radius (Å)	$b(C_\alpha$–$C_\beta)$ (Å)	$\theta(C_\alpha$–C_α–$C_\beta)$[b]	$\theta(C_\beta$–C_α–$C_\alpha)$[c]
C_α	2.702			
Ala	2.165	1.53	121.0	110.1
Val	2.746	1.97	121.9	115.9
Cys	2.522	2.07	115.2	118.9
Pro	2.700	1.88	82.1	127.0
Ser	2.367	1.89	118.0	108.7
Thr	2.660	1.95	117.6	114.6
Asp	2.725	2.47	119.9	120.2
Asn	2.762	2.47	115.7	126.3
Leu	2.984	2.62	120.6	124.5
Ile	3.185	2.35	117.8	120.3
Met	3.020	2.95	118.3	121.6
His	2.998	3.14	119.9	120.5
Phe	3.253	3.40	115.2	122.7
Tyr	3.351	3.77	113.5	123.2
Glu	2.793	3.14	118.0	118.4
Gln	2.827	3.12	118.1	121.3
Trp	3.364	3.86	114.3	116.9
Arg	3.284	4.13	116.9	120.2
Lys	3.166	3.51	119.3	121.2
CR[d]	2.100			
CH1E[e]	2.365			
CH2E[f]	2.235			
CH3E[g]	2.165			

[a] C_α atom and C_β atom for specified residues.
[b] For C_β belonging to the residue i, this is an angle between $C_\alpha(i-1)$–$C_\alpha(i)$–$C_\beta(i)$.
[c] For C_β belonging to the residue i, this is an angle between $C_\beta(i)$–$C_\alpha(i)$–$C_\alpha(i+1)$.
[d] For C1, C5, C6, C9, and C13 retinal atoms.
[e] For C7, C8, C10, C11, C12, C14, and C15 retinal atoms.
[f] For C2, C3, and C4 retinal atoms.
[g] For C16, C17, C18, C19, and C20 retinal atoms.

program; and Leo Caves, Mike Hann, Peter Murray-Rust, Manuel C. Peitsch, and Pam Thomas for stimulating discussions.
This work was supported by Glaxo-Wellcome Research and the BBSRC and DTI through the LINK protein engineering program.

APPENDIX A

In our calculations each residue is represented by one virtual main-chain C_α atom and one side-chain C_β atom. The C_α virtual atom is located at the C_α atom position whereas the C_β atom is located at the centroid of a side chain. The positions of C_β virtual atoms were calculated using a set of 83 protein structures from the Brookhaven databank (Bernstein et al., 1977), the resolution of which is less than 2 Å, and which were refined after 1982 (Oldfield and Hubbard, 1994). The van der Waals radii of the virtual C_α and C_β atoms were calculated from the HH reduced representation using the same rules (Herzyk and Hubbard, 1993). The van der Waals radii of ligand atoms are taken from the PARAM19 parameter set as presented in Brünger, 1990. The details of the representation used are presented in Table A1.

APPENDIX B

The penalty function P is given by:

$$P = repel + restr$$

where *repel* is used to prevent unduly close nonbonded contacts and *restr* is associated with geometrical restraints.

The simplified repulsive potential *repel* is given by (Brünger, 1990):

$$repel = \sum_{i<j} repel_{ij}$$

where

$$repel_{ij} = \begin{cases} 0 & \text{if} \quad r_{ij} \geq sR_{ij} \\ k_{rep}(s^2R_{ij}^2 - r_{ij}^2)^2 & \text{if} \quad r_{ij} < sR_{ij} \end{cases}$$

r_{ij} is the distance between two atoms i and j, whereas R_{ij} is a sum of their van der Waals radii, k_{rep} is a repulsive force constant, and s is a van der Waals radius scale factor. k_{rep} takes different values depending on the stage of the optimization protocol. As the protein representation is rather crude, very low force constants have been chosen for van der Waals repulsive interactions to avoid errors caused by oversimplification of the system.

Violations of geometrical restraints are penalized by the term *restr*:

$$restr = \sum dist_{xyz} + \sum dist_{xy} + \sum pos_{xy} + \sum pos_z + \sum ori_{in} + \sum ori_{out}$$

where each term is associated with different groups of restraints: *xyz*-distance, *xy*-distance, *xy*-position, *z*-position, ORI inside, and ORI outside restraints, respectively. Each of them is a sum of the contributions caused by singular restraints; hence the summations are performed over the number of restraints belonging to a particular group. A single contribution into the penalty function takes the form of a square-well function of the restraint violation. In the case of satisfied restraint, meaning its value falls between its lower and upper limits, the violation is zero and it does not contribute to the penalty function. Below we describe singular violations of different restraints and their contributions into the penalty function.

xyz-distance and *xy*-distance restraints

A particular distance r is allowed to fall between the lower r_l and upper r_u limits. If it is outside this range then its violation is represented by

$$\text{viol} = \begin{cases} r - r_u & \text{if} \quad r > r_u \\ r_l - r & \text{if} \quad r < r_l \end{cases}$$

and its contribution to the penalty function is

$$dist_{xyz(xy)} = k_{dist}\,\text{viol}^2$$

where k_{dist} is a force constant for distance restraints.

xy-position restraints

The x and y coordinates are allowed to be in the range $x_l \leq x \leq x_u$ and $y_l \leq y \leq y_u$. If one or both of them are outside these ranges then the violation for each coordinate is

$$\text{viol}_x = \begin{cases} x - x_u & \text{if} \quad x > x_u \\ x_l - x & \text{if} \quad x < x_l \end{cases}$$

$$\text{viol}_y = \begin{cases} y - y_u & \text{if} \quad y > y_u \\ y_l - y & \text{if} \quad y < y_l \end{cases}$$

The violation of *xy*-position restraints is defined as

$$\text{viol}_{xy} = (\text{viol}_x^2 + \text{viol}_y^2)^{1/2}$$

and its contribution to the penalty function is

$$pos_{xy} = k_{pos}\,\text{viol}_{xy}^2$$

where k_{pos} is a force constant for position restraints.

z-position restraints

A particular z coordinate is allowed to be in the range $z_l \leq z \leq z_u$. If it is outside this range then the violation is given by

$$\text{viol}_z = \begin{cases} z - z_u & \text{if } z > z_u \\ z_l - z & \text{if } z < z_l \end{cases}$$

and their contribution to the penalty function is

$$pos_z = k_{pos}\,\text{viol}_z^2$$

where k_{pos} is a force constant for distance restraints.

ORI restraints

We have constructed two types of ORI restraints, which we use at different stages of the calculation protocol. The first one referred to as simplified ORI restraints is based on the assumption that the terms inside and outside of the bundle, as seen by a particular residue, can be roughly detected by the 2-D direction toward or away from the local centroid of the bundle (where local refers to a plane that contains C_α atom of the restrained residue and is parallel to the *xy* plane). An inside-oriented residue satisfies the restraint only if its C_α atom aims at the local centroid whereas the outside-oriented residue has to aim in the opposite direction. This simplistic approach is used only at the initial stages of our calculations when helices are arranging themselves with respect to one another and the bundle is not yet well defined. The construction of the simplified ORI restraints makes them very strong, which assures roughly correct helix orientation at early stages of the calculations. The second type of ORI restraints, referred to as ORI restraints, describes the inside as an angle between two vectors anchored in a particular helix and pointing at two neighboring helices. An inside oriented residue satisfies the restraint if its C_α atom is positioned inside such an angle calculated in the plane parallel to the *xy*-plane. The outside-oriented residue has to have its C_α atom positioned outside that angle. This approach is used at later stages after the helix bundle has been formed. The ORI restraints are weaker than the simplified ORI restraints, which puts them more on a par with other groups of restraints described above. Below we present more detailed descriptions leading to definitions of violations.

Simplified ORI restraints (Fig. A1)

The centroid of points of intersection of helix axes with plane P_k (where P_k is defined by $z = z_k$ where z_k is a *z*-coordinate of the C_α atom of a residue restrained with the *k*th ORI restraint) is calculated (point A_k).

A vector of a length of a helix radius, from a point of intersection of the particular helix axis with the P_k plane (point S_k), directed toward point A_k for ORI inside restraints defines a target position for the C_α atom of a residue in question (point C_k^i). In the case of ORI outside restraints the target position is defined by orientation of the vector defined above away from point A_k (target point is C_k^o).

xy-distance restraint with the upper limit equal to zero is applied between the C_α atom and C_k^i or C_k^o positions, for inside and outside restraints, respectively. Violation of this restraint is defined by

$$\text{viol} = r_{\alpha k}$$

where $r_{\alpha k}$ is a distance between the C_α atom and C_k^i or C_k^o points. Thus this restraint is satisfied only if point C_α overlaps with points C_k^i or C_k^o for inside and outside restraints, respectively.

The contribution to the penalty function is

$$ori_{\text{in/out}} = k_{\text{ori}}^s\,\text{viol}^2$$

where k_{ori}^s is a force constant for simplified ORI restraints.

ORI restraints (Fig. A2)

Two vectors of a length of a helix radius, from point S_k and directed toward points of intersection of helix axes of two neighboring helices with plane P_k, define an angle β and points B_k and D_k. These points provide graphical representation of the lower and upper limits of the ORI restraint.

If the C_α atom of a residue in question is positioned inside the angle β for ORI inside restraints, and outside this angle for ORI outside restraints, the restraint is satisfied and does not contribute to the penalty function.

If the conditions mentioned above are not satisfied, then an *xy*-distance restraint of the upper limit equal to zero is applied between the C_α atom and one of the points B_k or D_k whichever is nearer. Violation of this restraint is defined by

$$\text{viol} = r_{\alpha k}$$

where $r_{\alpha k}$ is the distance between the C_α atom and the nearer of B_k and D_k points.

The contribution to the penalty function is

$$ori_{\text{in/out}} = k_{\text{ori}}\,\text{viol}^2$$

where k_{ori} is a force constant for ORI restraints. To present violation of ORI restraint for final configurations in a more understandable way we also express it as the angle C_α-S_k-B_k or C_α-S_k-D_k, depending on which of the points B_k and D_k is nearer to C_α (Fig. A2). These violations are presented in Tables 5 and 8.

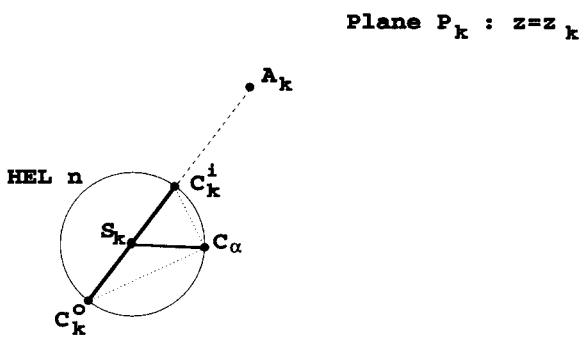

FIGURE A1 Graphical explanation of simplified ORI restraints. Depending on whether the *k*th ORI restraint imposed on C_α is inside or outside, the *xy*-distance restraint of upper limit equal to zero is applied between C_α and C_k^i or C_k^o points, respectively.

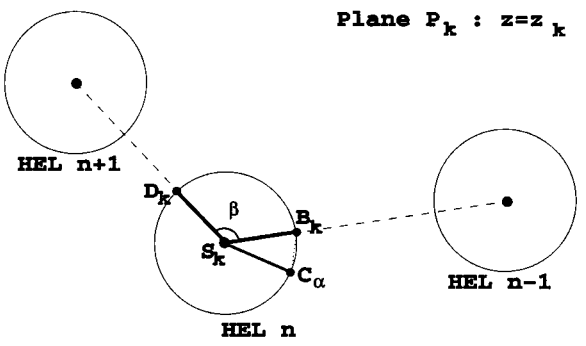

FIGURE A2 Graphical explanation of ORI restraints. If C_α is bound with ORI outside restraint, no penalty function is applied; if C_α is bound with ORI inside restraint, then *xy*-distance restraint of upper limit equal to zero is applied between C_α and B_k.

It is important to emphasize that ORI restraints are significantly weaker than simplified ORI restraints as they allow more room for the restraint satisfaction. Although it may seem that the ORI restraint violation is better expressed in terms of angle rather than distance, the latter was selected for computational reasons.

For the bR test case we have used $k_{dist} = k_{pos} = k_{ori} = 10k^s_{ori} = 50$ Å^{-2}. For the rhodopsin case, where the number of ORI restraints was significantly higher, we have used $k_{dist} = k_{pos} = 10k_{ori} = 10k^s_{ori} = 50$ Å^{-2}. Although uniform force constants were used for each group of restraints, there is a possibility of rescaling individual force constants for a particular restraint.

REFERENCES

Ahl, P. L., Stern, L. J., Mogi, T., Khorana, H. G., and Rotschild, K. J. 1989. Substitution of amino acids in helix F of bacteriorhodopsin: effects on the photochemical cycle. *Biochemistry.* 28:10028–10034.

Alkorta, I., and Du, P. 1994. Sequence divergence analysis for the prediction of seven-helix membrane protein structures: II. A 3-D model of human rhodopsin. *Protein Eng.* 7:1231–1238.

Altenbach, C., Marti, T., Khorana, H. G., and Hubbell, W. L. 1991. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science.* 248:1088–1092.

Attwood, T. K., and Findlay, J. B. C. 1993. Design of a discriminating fingerprint for G-protein-coupled receptors. *Protein Eng.* 6:167–176.

Baldwin, J., Henderson, R., Beckman, E., and Zemlin, F. 1988. Images of purple membrane at 2.8 Å resolution obtained by cryo-electron microscopy. *J. Mol. Biol.* 202:585–591.

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

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Brünger, A. T. 1990. Xplor Version 2.1 User Manual. Yale University Press, New Haven, CT

Chabre, M. 1985. Trigger and amplification mechanisms in visual phototransduction. *Annu. Rev. Biophys. Chem.* 14:331–360.

Chou, K.-C., and Carlacci, L. 1991. Simulated annealing approach to the study of protein structures. *Protein Eng.* 4:661–667.

Cohen, G. B., Oprian, D. D., and Robinson, P. R. 1993. Constitutive activation of opsin: influence of charge at position-134 and size at position-296. *Biochemistry.* 32:6111–6115.

Cronet, P., Sander, C., and Vriend, G. 1993. Modeling of transmembrane seven helix bundles. *Protein Eng.* 6:59–64.

Curtis, C. A. M., Wheatley, M., Bansal, S., Birdsall, N. J. M., Eveleigh, P., Poyner, D., and Hulme, E. C. 1989. Propylbenzilylcholine mustard labels an acidic residue in transmembrane helix-3 of the muscarinic receptor. *J. Biol. Chem.* 264:489–495.

Dahl, S. G., Edvardsen, O., and Sylte, I. 1991. Molecular dynamics of dopamine at the D2 receptor. *Proc. Natl. Acad. Sci. USA.* 88: 8111–8115.

de Groot, H. J. M., Harbison, G. S., Herzfeld, J., and Griffin, R. G. 1989. Nuclear magnetic resonance study of the Schiff base in bacteriorhodopsin: counterion effects on the ^{15}N shift anisotropy. *Biochemistry.* 28:3346–3353.

Donnelly, D., Findlay, J. B. C., and Blundell, T. L. 1994. The evolution and structure of aminergic G protein-coupled receptors. *Receptors Channels.* 2:61–78.

Donnelly, D., Johnson, M. S., Blundell, T. L., and Saunders, J. 1989. An analysis of the periodicity of conserved residues in sequence alignment of G-protein coupled receptors: implication for the three-dimensional structure. *FEBS Lett.* 251:109–116.

Donnelly, D., Overington, J. P., Ruffle, S. V., Nugent, J. H. A., and Blundell, T. L. 1993. Modeling α-helical transmembrane domains: the calculation and use of substitution tables for lipid facing residues. *Protein Sci.* 2:55–70.

Dryja, T. P., Berson, E. L., Rao, V. R., and Oprian, D. D. 1993. Heterozygous missense mutation in the rhodopsin gene as a cause of congenital stationary night blindness. *Nat. Genet.* 4:280–283.

Duñach, M., Berkowitz, S., Marti, T., He, Y.-W., Subramaniam, S., Khorana, H. G., and Rotschild, K. J. 1990. Ultraviolet-visible transient spectroscopy of bacteriorhodopsin mutants: evidence for two forms of tyrosine-185→phenylalanine. *J. Biol. Chem.* 265:16978–16984.

Earnest, T. N., Roepe, P., Braiman, M. S., Gillespie, J., and Rothschild, K. J. 1986. Orientation of bacteriorhodopsin chromophore probed by polarized Fourier-transform infrared difference spectroscopy. *Biochemistry.* 25:7793–7798.

Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. 1982. The helical hydrophobic moment: a measure of the amphiphilicity of α-helix. *Nature.* 299:371–374.

Engelman, D. M., Henderson, R., McLachlan, A. D., and Wallace, B. A. 1980. Path of the polypeptide in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 77:2023–2027.

Findlay, J. B. C., and Eliopoulos, E. E. 1990. 3-Dimensional modeling of G-protein-linked receptors. *Trends Pharmacol. Sci.* 11:492–499.

Greenhalgh, D. A., Altenbach, C., Hubbel, W. L., and Khorana, H. G. 1991. Locations of Arg-82, Asp-85, and Asp-96 in helix C of bacteriorhodopsin relative to the aqueous boundaries. *Proc. Natl. Acad. Sci. USA.* 88:8626–8630.

Grötzinger, J., Engels, M., Jacoby, E., Wollmer, A., and Strassburger, W. 1991. A model for the C5a receptor and for its interaction with the ligand. *Protein Eng.* 4:767–771.

Harbison, G. S., Smith, S. O., Pardoen, J. A., Courtin, J. M. L., Lugtenburg, J., Herzfeld, J., Mathias, R. A., and Griffin, R. G. 1985. Solid-state C-13 NMR detection of a perturbed 6-s-*trans* chromophore in bacteriorhodopsin. *Biochemistry.* 24:6955–6962.

Hauss, T., Grzesiek, S., Otto, H., Westerhausen, J., and Heyn, M. P. 1990. Transmembrane location of retinal in bacteriorhodopsin by neutron diffraction. *Biochemistry.* 29:4904–4913.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. 1990. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* 213: 899–929.

Henderson, R., Jubb, J. S., and Whystock, S. 1978. Specific labelling of the protein and lipid on the extracellular surface of purple membrane. *J. Mol. Biol.* 123:259–274.

Henderson, R., and Unwin, P. N. T. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature.* 257:28–32.

Herzyk, P., and Hubbard, R. E. 1993. A reduced representation of proteins for use in restraint satisfaction calculations. *Proteins Struct. Funct. Genet.* 17:310–324.

Heyn, M. P., Westerhausen, J., Wallat, I., and Seiff, F. 1988. High-sensitivity neutron diffraction of membranes: location of the Schiff base end of the chromophore of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 85:2146–2150.

Hibert, M. F., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J. 1991. Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol. Pharmacol.* 40:8–15.

Honig, B., Dinur, U., Nakanishi, K., Bologh-Nair, V., Gawinowicz, M. A., Arnaboldi, M., and Motto, M. G. 1979. An external point-charge model for wavelength regulation in visual pigments. *J. Am. Chem. Soc.* 101: 7084–7086.

Huang, K.-S., Ramachandran, R., Bayley, H., and Khorana, H. G. 1982. Orientation of retinal in bR as studied by cross-linking using photosensitive analog of retinal. *J. Biol. Chem.* 257:13616–13623.

Ihara, K., Amemiya, T., Miyashita, Y., and Mukohata, Y. 1994. Met-145 is a key residue in the dark adaptation of bacteriorhodopsin homologs. *Biophys. J.* 67:1187–1191.

Ijzerman, A. P., Galen, P. J. M. v., and Jacobson, K. A. 1992. Molecular modeling of adenosine receptors. I. The ligand binding site on the A1 receptor. *Drug Design Disc.* 9:49–67.

Kakitani, H., Kakitani, T., Rodman, H., and Honig, B. 1985. On the mechanism of wavelength regulation in visual pigments. *Photochem. Photobiol.* 41:471–479.

Kaushal, S., and Khorana, H. G. 1994. Structure and function in rhodopsin. VII. Point mutations associated with autosomal dominant retinis pigmentosa. *Biochemistry.* 33:6121–6128.

Khorana, H. G. 1988. Bacteriorhodopsin, membrane protein that uses light to translocate protons. *J. Biol. Chem.* 263:7439–7442.

Kirkpatrick, S., C. D. Gellat, J., and Vecchi, M. P. 1983. Optimization by simulated annealing. *Science.* 220:671–680.

Kontoyianni, M., and Lybrand, T. P. 1993. Three-dimensional models for integral membrane proteins: Possibilities and pitfalls. *Drug Design Disc.* 1:291–300.

Lewell, X. Q. A. 1992. A model of the adrenergic beta-2 receptor and binding sites for agonist and antagonist. *Drug Design Disc.* 9:29–48.

MaloneyHuss, K., and Lybrand, T. P. 1992. Three-dimensional structure for the β_2 adrenergic receptor protein based on computer modeling studies. *J. Mol. Biol.* 225:859–871.

Marti, T., Otto, H., Mogi, T., Rosselet, S. J., Heyn, M. P., and Khorana, H. G. 1991. Bacteriorhodopsin mutants containing single substitutions of serine or threonine residues are all active in proton translocation. *J. Biol. Chem.* 266:6919–6927.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equation of state calculation by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Mogi, T., Stern, L. J., Hackett, N. R., and Khorana, H. G. 1987. Bacteriorhodopsin mutants containing single tyrosine to phenylalanine substitutions are all active in proton translocation. *Proc. Natl. Acad. Sci. USA.* 84:5595–5599.

Nakayama, T. A., and Khorana, H. G. 1990. Orientation of retinal in bovine rhodopsin determined by crosslinking using a photoactivatable analog of 11-*cis*-retinal. *J. Biol. Chem.* 265:15762–15769.

Nakayama, T. A., and Khorana, H. G. 1991. Mapping of the amino acids in membrane-embedded helices that interact with the retinal chromophore in bovine rhodopsin. *J. Biol. Chem.* 266:4269–4275.

Nakazawa, T., Kawai, H., Okamoto, Y., and Fukugita, M. 1992. β-Sheet folding of fragment (16–36) of bovine pancreatic trypsin inhibitor as predicted by Monte Carlo simulated annealing. *Protein Eng.* 5:495–503.

Nathans, J. 1990a. Determinants of visual pigments absorbance: identification of the retinylidine Schiff base counterion in bovine rhodopsin. *Biochemistry.* 29:9746–9752.

Nathans, J. 1990b. Determinants of visual pigments absorbance: role of charged amino acids in the putative transmembrane segments. *Biochemistry.* 29:937–942.

Neitz, M., Neitz, J., and Jacobs, G. H. 1991. Spectral tuning of pigments underlying red-green color vision. *Science.* 252:971–974.

Nordvall, G., and Hacksell, U. 1993. Binding-site modeling of the muscarinic m1 receptor: a combination of homology based and indirect approaches. *J. Med. Chem.* 36:967–976.

Okamoto, Y., Fukugita, M., Nakazawa, T., and Kawai, H. 1991. α-Helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A. *Protein Eng.* 4:639–647.

Oldfield, T. J. 1992. SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. *J. Mol. Graph.* 10: 247–252.

Oldfield, T. J., and Hubbard, R. E. 1994. Analysis of Cα geometry in protein structures. *Proteins Struct. Funct. Genet.* 18:324–337.

Oliveira, L., Paiva, A. C. M., and Vriend, G. 1993. A common motif in G-protein-coupled seven transmembrane helix receptors. *J. Comput.-Aided Mol. Design* 7:649–658.

Oprian, D. D. 1992. The ligand-binding domain of rhodopsin and other G protein-linked receptors. *J. Bioenerg. Biomembr.* 24:211–217.

Otto, H., Marti, T., Holz, M., Mogi, T., Stern, L. J., Engel, F., Khorana, H. G., and Heyn, M. P. 1990. Substitution of amino acids Asp-85, Asp-212 and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK of the Schiff base. *Proc. Natl. Acad. Sci. USA.* 87:1018–1022.

Palings, I., Pardeon, J. A., Vandenberg, E., Winkel, C., Lugtenburg, J., and Mathies, R. A. 1987. Assignment of fingerprint vibrations in the resonance Raman-spectra of rhodopsin, isorhodopsin, and bacteriorhodopsin: implications for chromophore structure and environment. *Biochemistry.* 26:2544–2556.

Pardo, L., Ballesteros, J. A., Osman, R., and Weinstein, H. 1992. On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors. *Proc. Natl. Acad. Sci. USA.* 89: 4009–4012.

Popot, J. L., Engelman, D. M., Gurel, O., and Zaccai, G. 1989. Tertiary structure of bacteriorhodopsin: positions and orientations of helices A and B in the structural map determined by neutron diffraction. *J. Mol. Biol.* 210:829–847.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992. Numerical recipes in FORTRAN. The art of scientific computing, 2nd ed. Cambridge University Press, Cambridge, UK.

Rao, V. R., Cohen, G. B., and Oprian, D. D. 1994. Rhodopsin mutation G90D and a molecular mechanism for congenital night blindness. *Nature.* 367:639–642.

Rath, P., Decaluwe, L. L. J., Boveegeurts, P. H. M., de Grip, W. J., and Rotschild, K. J. 1993. Fourier-transform infrared difference spectroscopy of rhodopsin mutants: light activation of rhodopsin causes hydrogen-bonding change in residue aspartic acid-83 during meta-II formation. *Biochemistry.* 32:10277–10282.

Ridge, K. D., Bhattacharya, S., Nakayama, T. A., and Khorana, H. G. 1992. Light-stable rhodopsin. II. An opsin mutant (Trp-265 → Phe) and a retinal analog with a nonisomerizable 11-*cis* configuration form a photostable chromophore. *J. Biol. Chem.* 267:6770–6775.

Rippmann, F., and Bottcher, H. Molecular modelling of serotonin receptors. *7TM* 1.

Robinson, P. R., Cohen, G. B., Zhukovsky, E. A., and Oprian, D. D. 1992. Constitutively active mutants of rhodopsin. *Neuron.* 9:719–725.

Rothschild, K. J. 1992. FTIR difference spectroscopy of bacteriorhodopsin: toward a molecular model. *J. Bioenerg. Biomembr.* 24:147–167.

Savarese, T. M., and Fraser, C. M. 1992. In vitro mutagenesis and the search for structure-function relationship among G protein-coupled receptors. *Biochem. J.* 283:1–19.

Schertler, G. F. X., Villa, C., and Henderson, R. 1993. Projection structure of rhodopsin. *Nature.* 362:770–772.

Seiff, F., Wallat, I., Ermann, P., and Heyn, M. P. 1985. A neutron diffraction study on the location of the polyene chain of retinal in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 82:3227–3231.

Seiff, F., Westerhausen, J., Wallat, I., and Heyn, M. P. 1986. Location of the cyclohexane ring of the chromophore of bacteriorhodopsin by neutron diffraction with selectively deutered retinal. *Proc. Natl. Acad. Sci. USA.* 83:7746–7750.

Smith, S. O., Palings, I., Copie, V., Raleigh, D. P., Courtin, J., Pardeon, J. A., Lugtenburg, J., Mathies, R. A., and Griffin, R. G. 1987. Low-temperature solid state ^{13}C NMR-studies of the retinal chromophore in rhodopsin. *Biochemistry.* 26:1606–1611.

Smith, S. O., Palings, I., Miley, M. E., Courtin, J., de Groot, H., Lugtenburg, J., Mathies, R. A., and Griffin, R. G. 1990. Solid-state NMR studies of the mechanism of the opsin shift in the visual pigment rhodopsin. *Biochemistry.* 29:8158–8164.

Soppa, J. 1994. 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.

Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., and Dixon, R. A. F. 1994. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* 63:101–132.

Subramaniam, S., Greenhalgh, D. A., Rath, P., Rothschild, K. J., and Khorana, H. G. 1991. Replacement of leucine-93 by alanine or threonine slows down the decay of the N and O intermediates in the photocycle of bacteriorhodopsin: implications for proton uptake and 13-*cis*-retinal → all-*trans*-retinal reisomerization. *Proc. Natl. Acad. Sci. USA.* 88: 6873–6877.

Sylte, I., Edvardsen, Ø., and Dahl, S. G. 1993. Molecular dynamics of the 5-HT_{1a} receptor and ligands. *Protein Eng.* 6:691–700.

Taylor, E. W., and Agarwal, A. 1993. Sequence homology between bacteriorhodopsin and G-protein coupled receptors: exon shuffling or evolution by duplication. *FEBS Lett.* 325:161–166.

Taylor, W. R., Jones, D. T., and Green, N. M. 1994. A method for α-helical integral membrane protein fold prediction. *Proteins Struct. Funct. Genet.* 18:281–294.

Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A., and Hibert, M. 1992. Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mamalian opsins. *J. Med. Chem.* 35:3448–3462.

Underwood, D. J., Strader, C. S., Rivero, R., Patchett, A. A., Greenlee, W., and Prendergast, K. 1994. Structural model of antagonist and agonist binding to the angiotensin II, AT1 subtype, G protein coupled receptor. *Chem. Biol.* 1:211–221.

Unger, V. M., and Schertler, G. F. X. 1995. Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. *Biophys. J.* 68:1776–1786.

Vriend, G. Molecular modeling of GPCRs. *7TM* 3.

Wald, G. 1968. The molecular basis of visual excitation. *Nature.* 219: 800–807.

Watson, S., and Arkinstall, S. 1994. The G-protein linked receptor. *In* FactsBook. Academic Press Ltd., London.

Wilson, S. R., and Cui, W. 1990. Applications of simulated annealing to peptides. *Biopolymers.* 29:225–235.

Yamamoto, Y., Kamiya, K., and Terao, S. 1993. Modeling of human thromboxane A2 receptor and analysis of the receptor-ligand interaction. *J. Med. Chem.* 820–825.

Yue, S.-Y. 1990. Distance-constrained molecular docking by simulated annealing. *Protein Eng.* 4:177–184.

Zhang, D., and Weinstein, H. 1993. Signal transduction by a 5-HT2 receptor: a mechanistic hypothesis from molecular dynamics simulations of the three-dimensional model of the receptor complexed to ligands. *J. Med. Chem.* 36:934–938.

Zhukovsky, E. A., and Oprian, D. D. 1989. Effect of carboxylic acid side chains on the absorption maximum of visual pigments. *Science.* 246: 928–930.

Zvyaga, T. A., Fahmy, K., and Sakmar, T. P. 1994. Characterisation of rhodopsin-transducin interaction: a mutant rhodopsin photoproduct with a protonated Schiff base activates transducin. *Biochemistry.* 33: 9753–9761.

Zvyaga, T. A., Min, K. C., Beck, M., and Sakmar, T. P. 1993. Movement of the retinylidine Schiff base counterion in rhodopsin by one helix turn reverses the pH dependence of the metarhodopsin I to metarhodopsin II transition. *J. Biol. Chem.* 268:4661–4667.