

## RESEARCH ARTICLES

# Structure of the Integral Membrane Domain of the GLP1 Receptor

Thomas M. Frimurer and Robert P. Bywater\*

MedChem Research IV, Novo Nordisk Park, Novo Nordisk A/S, Måløv, Denmark

**ABSTRACT** A three-dimensional (3D) model of the integral membrane domain of the GLP1 receptor, a member of the secretin receptor family of the G-protein-coupled receptor superfamily is proposed. The probable arrangement of the seven helices in this receptor was deduced from a detailed analysis of all the sequences in the secretin receptor family. The analysis includes: 1) identifying the transmembrane helices, 2) charge distribution analysis to estimate to which extent the transmembrane helices are buried, 3) Fourier transform analysis of different property profiles within the transmembrane helices to determine the orientation of exposed and buried faces of the helices, 4) alignment of sequences with those of the rhodopsin-like family using the novel “cold spot” method reported herein, 5) determination of lengths of transmembrane helices and their connecting loops and the constraints these impose on packing, tilting and organization, 6) incorporation of mutagenesis and ligand specificity data. We find that there is a close similarity between the structural properties of receptors of the secretin family and those of the rhodopsin-like family as typified by the frog rhodopsin structure recently solved by electron cryomicroscopy. *Proteins* 1999;35:375–386. © 1999 Wiley-Liss, Inc.

**Key words:** G-protein-coupled receptor; secretin family; GLP1 receptor model; structure prediction; transmembrane helices

### INTRODUCTION

The superfamily of the G-protein-coupled receptors (GPCRs) can be divided into several pharmacologically distinct categories of which the three that feature most prominently in biomedical research are the rhodopsin-like family (RLF), secretin-like family (SRF) and metabotropic glutamate-like receptors. All GPCRs possess an integral membrane heptahelical domain (7TM) where the transmembrane helices (TMs) are linked by loops that extend outwards on both sides of the membrane. The latter two families have in addition a large extracellular N-terminal domain (Nter). Members of SRF have significant sequence similarity and are, with the exception of EMR1 and CD97,

very uniform in length. Nter is typically 120 residues long and contains six highly conserved cysteine residues and multiple potential glycosylation sites. The endogenous ligands for these receptors are polypeptide hormones which can be grouped into a small number of subsets whose members are of similar length and sequence. In this work we focus on the receptor (GLP1R) for the peptide hormone GLP1 as a representative for the SRF. Our aim is to construct a three-dimensional (3D) model of GLP1R and propose experiments to subsequently test and iteratively improve the model.

Methods for the construction of 3D models of GPCRs have been discussed by Ballesteros et al.<sup>1</sup> and we follow these general guidelines with a combination of a homology-based approach with one of more ab initio character.

It has been possible to align the GPCRs within each of the various families using advanced multiple sequence alignment tools such as MaxHom<sup>2</sup> and the iterative profile alignment procedure in the WHAT IF package<sup>3</sup> which was used to produce the alignments stored in the GPCR database, GPCRDB<sup>4</sup>. What has not been so easy however is to find an alignment between any of these families.<sup>5</sup> The sequence identity is too low, <20 %, i.e., below the threshold at which it is possible to use sequence information alone to draw conclusions about structure, function, or phylogeny.<sup>2</sup> At this level one obtains roughly the same similarity score regardless of how the sequences to be aligned are positioned alongside each other. The low sequence identity has prompted some researchers to eschew the use of the bacteriorhodopsin<sup>6</sup> or bovine or even frog rhodopsin structures<sup>7</sup> as templates for building models of secretin receptors. Donnelly<sup>5</sup> adopted a rule-based approach without the bias of having any template. Tams et al.<sup>8</sup> constructed a two-dimensional (2D) consensus model of a secretin-like receptor using a method involving considerations of free energy transfers of side chains between aqueous and lipid environments and side chain volumes.

*Abbreviations:* For amino-acid residue names the standard single letter code has been used throughout. GLP1 stands for the hormone glucagon-like peptide-1 (7:36) also known as insulinotropin.

\*Correspondence to: Robert P. Bywater, MedChem Research IV, Novo Nordisk Park, Novo Nordisk A/S, DK-2760 Måløv, Denmark.  
Received 21 September 1998; Accepted 26 February 1999

The method was first tested on bacteriorhodopsin in a "postdictory" manner and then applied to the secretin receptors.

It is becoming widely accepted that what is conserved between proteins of the same family or superfamily is *structure and function* rather than *sequence*. There is no reason why this should not apply in the case of the GPCRs. There are in fact many similarities between the RLF and SRF and we shall investigate whether the folds are essentially the same.

Recently, the coordinates of a model<sup>9</sup> based on the medium-resolution electron crystallography map<sup>7</sup> of the transmembrane domain of frog rhodopsin have been made available. The overall arrangement (topology) of the TMs in this structure is analogous to that of bacteriorhodopsin, although the relative positions and tilting of the individual TMs are different. In our work structural features of SRF such as the most exposed helices, helix length, and their relative tilting are predicted and compared to the corresponding predicted structural features of RLF. Our analysis includes: 1) identifying the TMs, 2) charge distribution analysis to estimate to which extent the TMs are buried, 3) Fourier transform analysis of different property profiles within the TMs to determine the orientation of exposed and buried faces of helices, 4) alignment of sequences with those of the rhodopsin-like family using the novel "cold spot" method which aligns protein sequences on the basis of conservation rather than similarity, 5) lengths of TMs and loops and the constraints this imposes on packing, tilting, and organization, 6) incorporation of mutagenesis and ligand specificity data, 7) joining the TMs by loops selected from a loop database.<sup>10</sup>

Certain principles of membrane protein structure are beginning to emerge<sup>11-16</sup> and these considerations have led to the following assumptions.

### Positions of Charged and Polar Residues

This residue class comprises all charged residues and all those capable of forming more than one hydrogen bond.<sup>11</sup> However, S, T, and Y residues are excluded from this group. S and T can satisfy their hydrogen bonding potential by bonding to the main-chain of the TMs and therefore could be on the lipid-facing surface. Similarly, Y has been observed to be exposed to lipid in porins.<sup>17,18</sup> Thus residues assigned to the above class are D, N, E, Q, H, R, and K. Positions close to the ends of helices could face the lipid and still accommodate polar residues as long as they interact with the polar head group in the lipid bilayer or else form a "cap" by binding to a backbone peptide group.

### Hydrophobic Residue Positions

The aliphatic residue types A, V, I, L, and M and the aromatics F, Y, and W tend to cluster in lipid facing or core-forming positions.<sup>19</sup>

### Conserved Residue Positions

The conserved residue positions are expected to play an important role in maintaining the function of the receptor, both for correct folding and for common functional proper-

ties such as the need to recognize and bind G proteins. In GPCRs residues that are buried are more conserved than those that are exposed.<sup>20,21</sup> Conserved positions that are important at intrahelical *loci*, e.g., a possible kink<sup>1</sup> introduced by P, do not convey any information concerning the preference for the lipid phase or TM contact.

### Variable Residue Positions

Sequences from different species that bind the same ligand have very high sequence identity (90% or more). In contrast, positions with significant sequence differences most likely occur at functionally unimportant sites. In the transmembrane segment these positions are considered mainly to be located at the lipid facing side of the TMs.

### Environment-Dependent Substitution Tables

GPCRs reside partly in an aqueous environment, and are partly embedded in a lipid membrane. Therefore, different physicochemical constraints apply to the residues, depending on their spatial location. In general, the same mutations at different positions in a protein are not equally likely to occur, and therefore one single scoring matrix for all positions in a sequence alignment is often not adequate. The first attempt to characterize and quantify these structural constraints were made by Overington et al.<sup>20</sup> who extended Dayhoff's idea<sup>22</sup> and generated multiple substitution matrices from families of homologous proteins of known structure as function of local environment. A major further development was the use of one exchange matrix for each position in the sequence alignment.<sup>14,21,23-25</sup> This so-called structure-based profile can of course only be made if a three-dimensional structure for at least one of the family members is available. We used environment-dependent substitution tables<sup>14,21,25</sup> derived from accessible and inaccessible residue positions in aligned proteins to predict whether the substitution pattern in each position of a sequence alignment are typical of a buried or exposed residue.  $\alpha$ -helices that have one face buried and the other exposed show a periodicity of buried and exposed residues corresponding to the periodicity of the helix.

## METHODS

### Secondary Structure Predictions

A complete alignment of all the members of the SRF were obtained from the GPCRDB at the EMBL server <http://www.gpcr.org/7tm>. The PHD program<sup>26</sup> used for secondary structure prediction and TM assignment is accessible at <http://www.embl-heidelberg.de/predictprotein/predictprotein.html>.

### Helix-Facing Properties

One of the earliest published methods to display the amphipathy of a helix was the helical wheel.<sup>27</sup> A similar but more quantitative plot employs the concept of hydrophobic moment.<sup>28</sup> Other plots have also been proposed<sup>29,30</sup> in which a Fourier transformation of the hydrophobicity is

TM1		TM2		TM3		TM4		TM5		TM6		TM7	
L +/*		L -/*		R/K ++/-/*		E +/-/*		Y +/-		R +/-/*		F ++/--/*	
Y ++/**				L		Y +		W ++/--		A +/-/*		I -	
1 I +		26 A +/-/*		1 L *		26 L +		1 L		26 H ++/--/*		1 K ++/**	
2 I		25 A +/-/*		2 M ++		25 Y +		2 I		25 E --/**		2 L +/-/*	
3 Y -/**		24 D/Q --/**		3 Q +/-/*		24 K/R ++/*		3 I +		24 D --/+		3 F ++/*	
4 T +/-*		23 K ++/*		4 Y		23 V		4 R ++/-/*		23 M		4 T	
5 V		22 I		5 C		22 I		5 L		22 V ++/*		5 E +/-/*	
6 G		21 F		6 V ++		21 G		6 P		21 F *		6 L ++	
7 Y *		20 V +		7 A		20 W ++		7 I +		20 A		7 S	
8 A		19 S +		8 A		19 P		8 L		19 F		8 F	
9 L		18 L		9 N **		18 V		9 F		18 I		9 T -/**	
10 S		17 A +/-/*		10 Y/F		17 V		10 A -		17 V		10 S ++	
11 F		16 R ++/**		11 Y		16 F +		11 I		16 E --/*		11 F/L	
12 S		15 L		12 W		15 L		12 G		15 H ++/-/*		12 Q **	
13 A		14 I		13 L		14 L		13 V		14 T		13 G	
14 L		13 F/L		14 L/F		13 P		14 N **		13 G *		14 L/F	
15 V		12 S		15 V		12 V		15 F/L		12 L/F		15 M	
16 I		11 A		16 E --		11 G		16 L		11 L		16 V	
17 A		10 F/I		17 G		10 W		17 I		10 P		17 A/S	
18 S		9 L/M +		18 V		9 G		18 F/L		9 I +		18 I/V	
19 A		8 N ++		19 Y -		8 I		19 V		8 L *		19 L	
20 I		7 L ++		20 L		7 S		20 R ++/**		7 T		20 Y/F +	
21 L *		6 H ++		21 Y ++/*		6 V +		21 V		6 L		21 C	
22 L		5 I/L		22 T +/-*		5 Y		22 I		5 T		22 F	
23 G ++/*		4 Y		23 L +/-*		4 L +		23 C ++/*		4 S +		23 V	
24 F/l		3 N **		24 L/A		3 R ++		24 I/V +		3 K/R ++		24 N	
25 R/K ++		2 R ++/*		25 A +/-*		2 F ++		25 V/L		2 A -		25 N ++/**	
26 H ++/-/*		1 T +/-*		26 F +/-		1 I ++		26 V +		1 L/A +		26 E +/-/*	
L/R		C *		S		W ++/-/*		S ++/**		R/K ++		V	
H ++				V +/-*		Q ++/-/*		K ++		C ++		Q/R ++/**	
				F +/-*									
				S +/-*									

Fig. 1. Positions of each helix are numbered on the left, downwards for the predicted transmembrane helices I, III, V, and VII and upwards for II, IV, and VI. In this way the extracellular receptor or the membrane is at the top of the figure. The symbols (+) and (++) indicate that the position is occupied by positively charged residues in <10% (+) or in >10% of the sequences (++) . The symbols (-) and (--) indicate negatively charged

positions corresponding to the definition for positively charged positions. The absence of a symbol indicates that there is never a polar residue observed at that position. Positively charged residues are R and K and negatively charged residues are D and E. The highly conserved residue sites are labelled to the right.

**TABLE I. Predicted Location of the Central 18 Residues of Each of the Seven Transmembrane Helices and Their Calculated AP Values for the Variability, Conservation, Hydrophobic, and Substitution Profiles Respectively<sup>†</sup>**

TMs	Location GPCRDB	Variability AP value	Conservation AP value	Hydrophobic AP value	Substitution AP value
I	153–170	6.50	6.50	2.41	6.95
II	183–200	4.98	4.98	2.48	4.75
III	233–240	6.08	6.08	3.04	4.84
IV	269–288	5.30	5.30	2.04	3.48
V	311–328	5.90	5.90	3.35	3.70
VI	355–372	3.73	3.73	2.22	2.79
VII	389–406	1.28	1.28	4.08	3.43

<sup>†</sup>The annotation of the location numbers are consistent with the alignment of the SRF in the GPCRDB.<sup>4</sup>

carried out in order to identify any periodic function such as the amphipathy of helices.

In this study the PERSCAN v7.0 program developed by Donnelly<sup>31,32</sup> was applied to analyze the moment or the

amphipathic character of the SRF helices. The program has earlier been applied to predict TMs with success in a number of studies<sup>14,25,31,32</sup> and described in detail,<sup>31</sup> hence only a brief description will be given here. The program

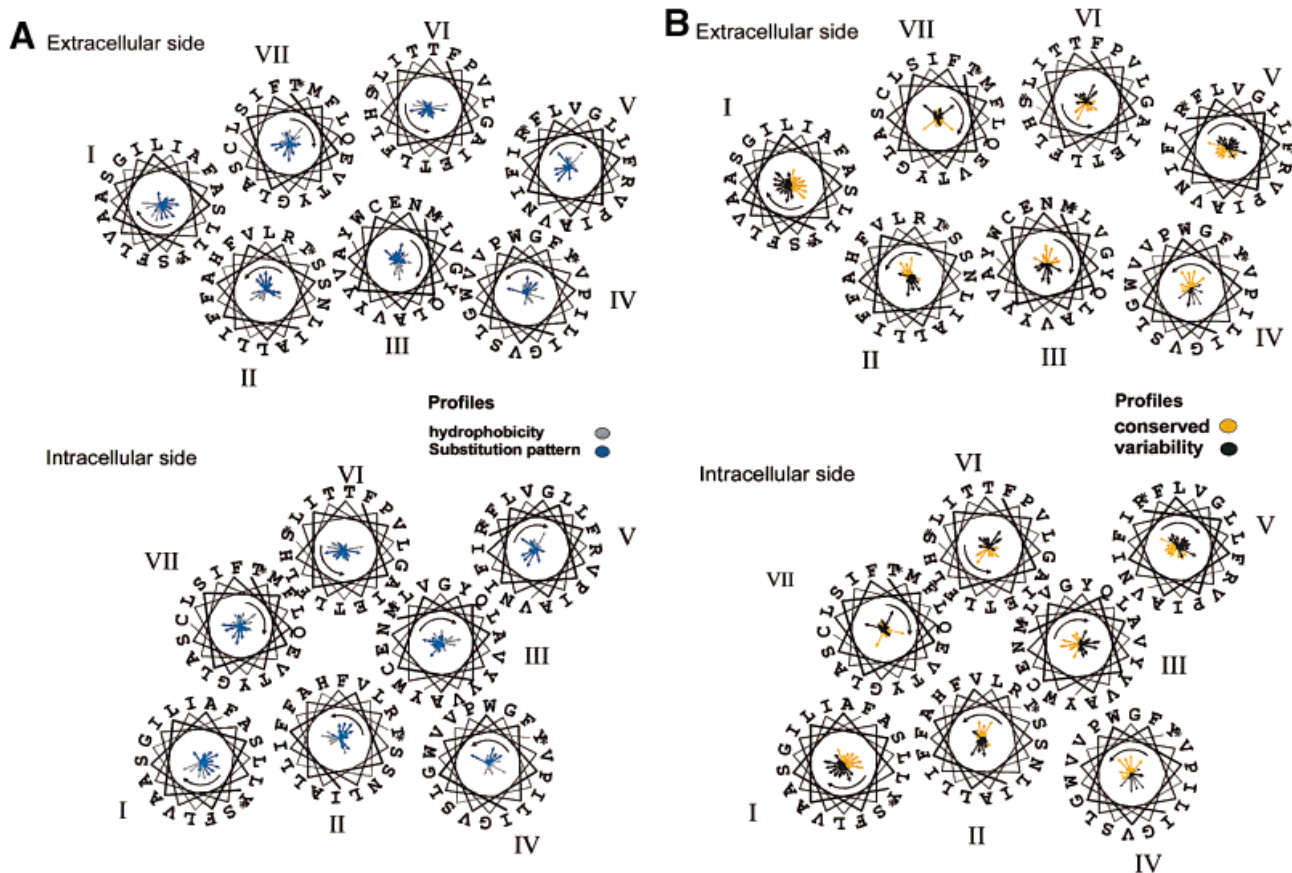


Fig. 2. (A) Sequence comparison information summarized around helical wheels for each of the seven helices in the secretin receptor family. The query sequence in the figure are taken from the GLP1\_HUMAN sequence and represent the central 18 residues in each helix viewed from the extracellular side. The organization of the helices has been taken from the recent high-resolution electron microscopic structure of frog rhodopsin,<sup>7</sup> whereas the orientation of the individual helices has been chosen so the predicted internal face points towards the center of the helical bundle.

**TABLE II. The Predicted Minimum and Maximum Length of the Three Intracellular (I1, I2, and I3) and Three Extracellular (X1, X2, and X3) Loops for the SRF and RLF Respectively<sup>†</sup>**

Loops	I1	X1	I2	X2	I3	X3
SRF min loop length	3	14	9	14	16	2
SRF max loop length	3	25	10	15	21	9
RLF min loop length	5	13	10	12	12	12
RLF max loop length	11	22	18	43	420	162

<sup>†</sup>The minimum and maximum length of the loops in the RLF are obtained from the study of Baldwin.<sup>11</sup>

searches for helical periodicity in sequence alignments and predicts the internal face of any helix found. Environment-dependent substitution tables (Sj), hydrophobicity scales (Hj), variability (Vj), conservation (Cj), or accessibility profiles (Ij) can be used as property profiles to predict helical periodicity. The periodicity of the property profile is calculated by a standard Fourier transform procedure. A property,  $U_j$  is assigned at each position in a sequence

The vector arrows (in the center of each helical wheel) represent the orientation of the predicted buried and lipid-accessible face of the individual helices derived from the substitution and hydrophobicity property profile respectively. (B) the comparison of the calculated conserved and variable face/positions of the seven helices is shown. These faces are opposite to each other as might be expected for helical regions in an apolar environment.<sup>25</sup>

alignment over a window size  $N$  and the moment  $M$  can then be calculated as:

$$M = \left[ \left( \sum_{j=1}^N U_j \sin(j\omega) \right)^2 + \left( \sum_{j=1}^N U_j \cos(j\omega) \right)^2 \right]^{1/2}, \quad (1)$$

where  $\omega$  is the angle between adjacent side-chains when the sequence is considered as a regular structure and viewed down an axis defined by the  $C_\alpha$  atoms. When calculating the periodicity in the values of  $U_j$ , the Fourier transform power spectrum is calculated by

$$P(\omega) = \left[ \left( \sum_{j=1}^N U_j^n \sin(j\omega) \right)^2 + \left( \sum_{j=1}^N U_j^n \cos(j\omega) \right)^2 \right]^{1/2}, \quad (2)$$

where

$$U_j^n = U_j - \bar{U} \quad (j = 1, 2, \dots, N), \quad (3)$$



## GLP1-Receptor:

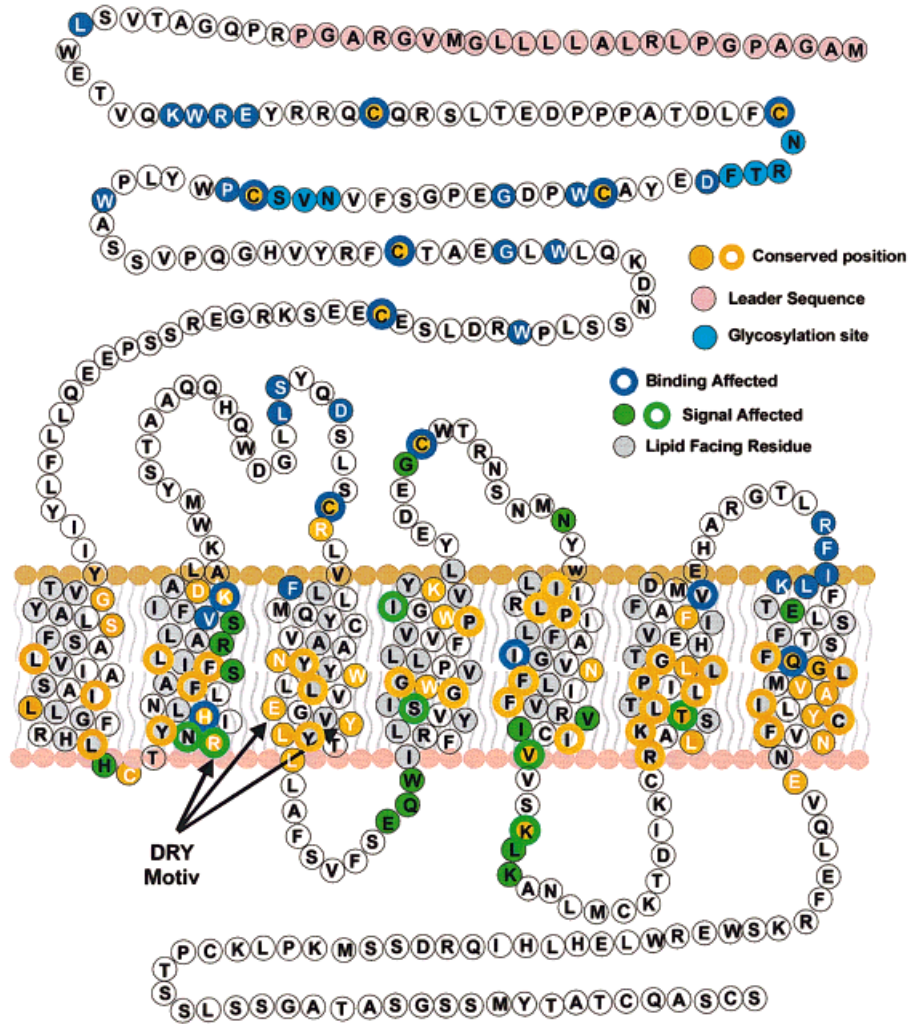


Fig. 3. A schematic representation of the GLP1 ligand and receptor, showing the N-terminal, the predicted seven transmembrane helices connected by the intra and extracellular loops, and the C-terminal domain. The N-terminal contains six cysteines that form putative disulfide bonds. Residues surrounded by a blue or green ring are experimentally determined to be binding- and signal-affected respectively. Residues with a dark grey background are predicted to face the lipid membrane. Conserved positions are in orange.

$\bar{U}$  is the average value of  $U_j$  over the window. The alpha periodicity index AP is then calculated as

$$AP = 1/30 \int_{95^\circ}^{106^\circ} P(\omega) d\omega / 1/180 \int_{0^\circ}^{180^\circ} P(\omega) d\omega, \quad (4)$$

AP is a ratio of the extent of the periodicity in the helical region of the spectrum compared with that over the whole spectrum. Analogous values of AP are used in other published studies;<sup>33,34</sup> *inter alia* an AP value greater than 2 indicates<sup>33</sup> a helical region.

### Homology Modelling

We recall our philosophy that “structure is better conserved than sequence” and we stretch that idea to include

the assumption that the SRF might be aligned to the RLF. On this basis, a homology model of the TMs was constructed using the coordinates for the frog rhodopsin model.<sup>9</sup> From these coordinates the ends of the helices can be identified. Now we have two independent means of comparing the positions of the ends of the helices, the electron crystallography data and the data obtained from the use of the PHD program. The assignment of the ends of the helices by themselves is by no means sufficient for an alignment to be made, and since there is no, or at least very low, sequence identity we resort to a new alignment method. Wherever there is a *pair of conserved positions* within one family whose members are *the same distance apart* in sequence space as the members of a pair of

conserved positions in another family then this identifies critical sites at which structure is most likely to be preserved and at which alignment can therefore be based. This is done *quantitatively* using the statistics for residue variability at each position for the entire RLF and SRF given in the HSSP tables in GPCRDB. When such "cold spots" are found to be located at the same distance apart in TMs from different families it makes sense to "pivot" the alignment on these positions.

In the method used here, the focus is on finding sites that are conserved within each family and which are located relative to each other in such a way as to preserve structural integrity of, in this case, the transmembrane helices. Only after the sites of conservation that come into register have been found do we consider questions like sequence similarity. Thus apparent sequence similarity is not allowed to bias our alignment.

### Construction of 3D Models

The WHAT IF protein modelling program<sup>3</sup> was used for model building. Homology modelling of the TMs was carried out using the alignments obtained as above. Loops were fitted using the DGLLOOP procedure in WHAT IF. A database of loops is searched in which the following constraints are in force: 1) loop length, 2) end-to-end distance defined by (in this case) the ends of the helices to be joined, 3) suitable geometry, 4) sequence is not paramount but loops containing P or G residues in the same sequence positions are selected if the other criteria are met. The standard side-chain rotamer library was used since it has been shown<sup>16</sup> that there are no significant differences between rotamer preferences in the current set of known membrane protein structures and the large set of known water-soluble globular proteins.

## RESULTS AND DISCUSSION

### Vertical Positions of the Transmembrane Helices

We define vertical as being normal to the plane of the membrane. The vertical position of the TMs was predicted using the PHD program. The full-length helix data is given in Figure 4, and a truncated version with the 18 innermost residues is shown in Table I.

### Predicted TM Lengths

For a helix to span the (circa 30 Å thick) membrane, with a standard 1.5-Å rise per turn of helix, 20 residues would be required. The predicted length of TMs I–VII, embedded within the membrane as defined by the PHD program are 19, 18, 25, 18, 23, 18, and 18 residues respectively. This refers to the strictly membrane embedded region, but the program also predicts that the helices extend at both ends. The helices that are expected to be most tilted are, in descending order, TMs III, V, and I, while the least tilted are TMs II, IV, VI, and VII. These structural characteristics agree closely with observations in the recent rhodopsin structure,<sup>7</sup> where the most-tilted helices are assigned to I, II, III, and V, while the least-tilted helices are IV, VI, and VII. Helices III and V appear to be significantly the longest in both SRF and RLF. Comparison of the shortest helices is

more questionable since the predicted length of these helices vary by only a few residues. Helix VI appears to be the shortest in SRF. The individual TMs are described in the legend to Figure 4.

### Exposed Helices

The extent to which the helices are exposed or buried has been predicted by analyzing the distribution of the polar residues in the individual TMs of 56 SRF sequences from the GPCRDB. This was done in a similar way as the analysis of the rhodopsin receptor.<sup>11</sup> The results are shown in Figure 1, where the sequences of each of the predicted seven helical segments are represented as vertical lines that are numbered downwards for the helices I, III, V, and VII and upwards for II, IV, and VI. The numbering in the following refers to this figure, e.g., 2:6 means TM II position 6.

A region of at least 28 residues has been selected about each helix and residue positions in each are classified as being always occupied by a hydrophobic amino acid (blank); occupied by a positively (+) or negatively (–) charged amino acid in less than 10% of the sequences or occupied by a positively (++) or negatively (--) charged amino acid in more than 10%. The distribution of positively charged residues clearly conforms to the well documented<sup>35</sup> "positive inside rule." There is also a preponderance of charged residues at sites which indicate where the helices must pass close to or through the head groups of the lipid bilayer.

Clusters of entirely hydrophobic residues are identified by a block of grey color on each TM. Helix I, IV, and V each seem to have relatively large surfaces containing hydrophobic residue positions while helix II, III, VI, and VII have a higher content of polar residues in the central part. This indicates that TMs I, IV, and V are the most lipid-exposed helices while the TMs II, III, VI, and VII are expected to be more buried in the SRF. A comparison of the decreasing number of polar residue positions in the TMs, the order is: III > VII = II > VI > V = IV = I for both SRF and RLF.<sup>11</sup>

### Orientations of the Transmembrane Helices

Table I shows the position of the 18 innermost residues in each TM and the AP values calculated from the different property profiles; variability  $V_j$ , conservation  $C_j$ , hydrophobicity  $H_j$ , and substitution patterns  $S_j$ , for all seven predicted TMs. All of the predicted transmembrane domains are strongly predicted to be helical, having AP values >2 with the exception of TM VII. This TM only shows significant helix propensity for  $H_j$  and  $S_j$  but not for  $V_j$  and  $C_j$  which are calculated to 1.28. The central 18 residues for the predicted seven TMs are represented as helical wheels in Figure 2. Figure 2A shows the central region and the calculated  $S_j$  and  $H_j$  vectors for GLP1R while Figure 2B shows  $V_j$  and  $C_j$ . The vectors calculated from profiles  $H_j$  and  $V_j$  are both predicted to face the lipid, with only small variations in the individual vector sums for all the helices. The vectors calculated for profiles  $C_j$  and  $S_j$ , i.e., the conservation and substitution pattern are predicted to face the opposite side (the internal face) of the

helices. The vector sums of these individual profiles only have minor differences in their orientations. The above results show that in these TMs the hydrophilic/conserved side faces the interior and the variable/hydrophobic side faces the lipid. No polar residues are found in the lipid-facing part of the central region but, as stated earlier, the polar residues in positions 5, 6, 7 and 20, 21, 22 can be accommodated because of proximity to the head groups.

### Overall TM Arrangement

An analysis of the minimum intracellular and extracellular loop lengths in GPCRs suggests ways in which the helices can be mutually positioned in 3D. The maximum and minimum length of the connecting loops between the TMs in SRF vary within the ranges given in Table II together with the corresponding data for RLF.<sup>11</sup> In SRF, the minimum length of the first, second, and third intracellular (I1, I2, and I3) loops are 3, 9, and 16 residues respectively. For the first, second, and third extracellular loops (X1, X2, and X3) the minimum lengths are 14, 14, and 2 residues. I1, X1, I2, X2, and I3 have very similar minimum length (I3 differs by at most 4 residues, the others by no more than 2) in both families while X3 is 10 residues longer in RLF. This implies that the ends of the helices very likely are in a similar juxtaposition in both families.

The importance of the disulfide bridge formed between the first and second extracellular loops has been addressed for rhodopsin,<sup>36</sup> muscarinic receptors,<sup>37</sup> and for adrenergic receptors,<sup>38</sup> and is believed to be present in the majority of GPCRs. The minimum length of X2 of SRF is 14 residues (see Table II) and the minimum number of amino acid residues between the cysteine in the loop and top of TM V can be as low as 8. Therefore the extracellular end of TM III has to be close to the extracellular end of TM V.

All of these structural features support the contention that the seven transmembrane helical domain of SRF is organized in a similar way to that of RLF.

### Incorporation of Mutagenesis Data

The reliability of a model is improved considerably if experimental data on the structure are taken into account in devising it. Combination of our structure prediction for SRF together with the many structural similarities with RLF, leads to the 2D serpentine model shown in Figure 3, in which the predicted lipid-facing regions are indicated by grey shading. Sites that are important with respect to structure, function, or in agonist and/or antagonist binding (as determined by site-directed mutagenesis, chemical labelling, or other experimental studies<sup>39–44</sup>) are expected to face inwards and form a potential ligand-binding site. In Figure 3 blue and green signify that binding and activity respectively are affected. Point mutations that lead to altered ligand binding, receptor expression or function are almost all located at the more conserved/hydrophilic side of the helices which face the interior or which contact the other helices in the seven transmembrane helical bundle. The three exceptions to this are I4:22 and S4:7 on TM IV and an D2:3 on TM II which all are close to the head group

region of the lipid and there may be departures from true  $\alpha$ -helix character. An alternative explanation, which we shall investigate in a future extension of this work, is that these sites could be signals for dimerization. It has been shown<sup>45</sup> that the glucagon receptor acts as a dimer, and there are other indications<sup>46,47</sup> that dimerization may be important in many GPCRs.

In general the majority of conserved residues, indicated in Figure 3 by red, are located in the cytosolic part of the transmembrane domain where they form clusters made up largely of aromatic side chains. This applies both to SRF (Fig. 3) and RLF, once again indicating that the folding characteristics for these two categories of GPCRs are similar. In the middle of the bilayer the polar residues all point inwards. Furthermore, the residues known to be important for binding and/or activity belong to this category. This supports a plausible orientation of the helices since structurally and functionally important residues are expected to face the other helices.

### Motifs

Residues that are characteristic for each helix in RLF have been compared to characteristic residues in SRF. Mutational studies of the E/DRY motifs (located in the intracellular end of TM III in RLF) have shown that this R340 is essential for the activation and the mutation of D to A in  $\alpha_{1B}$ -adrenergic receptor confers activity.<sup>48</sup> This motif in our alignment corresponds to a conserved sequence YLY in SRF (Fig. 4). This sequence contains none of the important functional residues associated with the E/DRY motif.

The R function in SRF could be furnished by the fully conserved R2:2 in the intracellular end of TM II predicted to be located at the same level in the membrane as the conserved R in the E/DRY triplet of RLF. As illustrated in Figure 3, the R2:2 in TM II of the secretin receptors has been shown to be important for receptor activation. The other missing function is the E/D function which in SRF could be taken over by the ExxY (3:16–19) motif in the intracellular end of TM III. E is placed above Y on the internal side of the helix. All of the residues, i.e., R, E, and Y face towards TMs VI and VII.

TM II and VII contain many polar sites (see Fig. 1), including the R2:16 in TM II and Q7:12 in TM VII, which have been predicted to be in contact in the secretin-like human PTH receptor.<sup>42</sup> If this is the true for GLP1R, then K2:23, and E7:5 in helix II and VII also face each other since each of these residues are positioned seven residues above R2:16 and Q7:27 respectively. Experiments on gonadotropin-releasing hormone receptor, a member of RLF, showed that the residue pairs N224 TM II and D729 TM VII could be inverted without loss of function.<sup>42</sup> These results indicate that TM II and TM VII are in close contact in both SRF and RLF.

### The Arginine Switch

As R340 in the E/DRY motif in RLF, is located both near the so-called polar pocket and the cytosol, it may have a "switching role," which is expected through alternative

side-chain conformations.<sup>49</sup> It is suggested that in RLF the switch is: 1) *off* when the R340 side-chain is located in a polar pocket surrounding this residue; and 2) *on* when the R340 side-chain is shifted toward the cytosol where it is

proposed to bind to a fully conserved D residue in the G protein.<sup>49</sup>

This switching mechanism could explain why in SRF the PTH receptor is constitutively active when the strictly

```

TM1 CONSERV RLF      *   * * * * *
PROFILE RLF VLLAALYSLLFLVGVLGNNLLVILVILR
          :::: |||X::X::X |XX:|.X X
PROFILE SRF IIYTVGYSLSLVALLVAIILLLFRRL
CONSERV SRF      *   * * * * *
INSIDE RLF      # # # ## # # #
INSIDE SRF      ## # ## # # #
PHD SRF HTTTTTTTTTTTTTTTTTTHH

TM2 CONSERV RLF *       * * * *
PROFILE RLF RTVTNYFLLNLAVADLLVALLVPPLAL
          X : .X:.XX:XX: |.: ..|..
PROFILE SRF NYIHMNLFVSFILRAISVFIKDAVLYP
CONSERV SRF * * * * *
INSIDE RLF      # # # # # # #
INSIDE SRF      ## # ## # ## #
PHD SRF HTTTTTTTTTTTTTTTTTTH

TM3 CONSERV RLF      * * *   * * * * * *
PROFILE RLF VLCKLVGFLDVLNGTASIFSLTAISVDRLAICH
          |X|:::.. . XX:|X.: XXX:::..
PROFILE SRF VGCKVAVVFFQYCVMTNYFWLLVEGLYLHTLLVVT
CONSERV SRF      * *   * * * * *
INSIDE RLF      # ## ### ### # #
INSIDE SRF      # # ## ## ## #
PHD SRF HHHHHHTTTTTTTTTTTTTTTTTTTHHHHHH

TM4 CONSERV RLF      *       * * * *
PROFILE RLF RRAKLVIALLVWVLSLLSLPPLFGW
          : X :::: X |X:: :X ..
PROFILE SRF KYFWWYLLIGWGLPAVFVVAWAIVR
CONSERV SRF      *       * * * *
INSIDE RLF      # # ## # #
INSIDE SRF      # # ## # #
PHD SRF HHHTTTTTTTTTTTTTTTTTTHH

TM5 CONSERV RLF      *   * * * * *
PROFILE RLF VIYSFIVSFYLPILLIILFCYGRIRYTLQKA
          :.: . X.X X|: ...X X.|.
PROFILE SRF LWWIIQGPILLSILVNFIILFINIIRILVTK
CONSERV SRF      * *   * * * * *
INSIDE RLF      # # # # # ## #
INSIDE SRF      # # # # # ##
PHD SRF HHHTTTTTTTTTTTTTTTTTTHH

TM6 CONSERV RLF      * *   * * * *
PROFILE RLF KAERKAAKTLAIVVGVFALCWLPPFIVALL
          : |X:|X:XX:X: :X X :: X..|.
PROFILE SRF DQYRRLAKSTLVLIPLFGIHYVVFVAFVPDE
CONSERV SRF      * *   * * * *
INSIDE RLF      # # ## # # ## #
INSIDE SRF      # # ## # # ##
PHD SRF TTTTTTTTTTTTTTTTTT

TM7 CONSERV RLF      * * * * * *
PROFILE RLF FLITLWLAYLNSCLNPPIYAFNLK
          ... |:| X.XX..X.XXX.|X
PROFILE SRF LYFELFLGSFGQFFVAVLYCFLNG
CONSERV SRF      * * * * * *
INSIDE RLF      # ## ## # ## #
INSIDE SRF      ## ## ## ## #
PHD SRF HHHTTTTTTTTTTTTTTTTTTH

```

Figure 4.



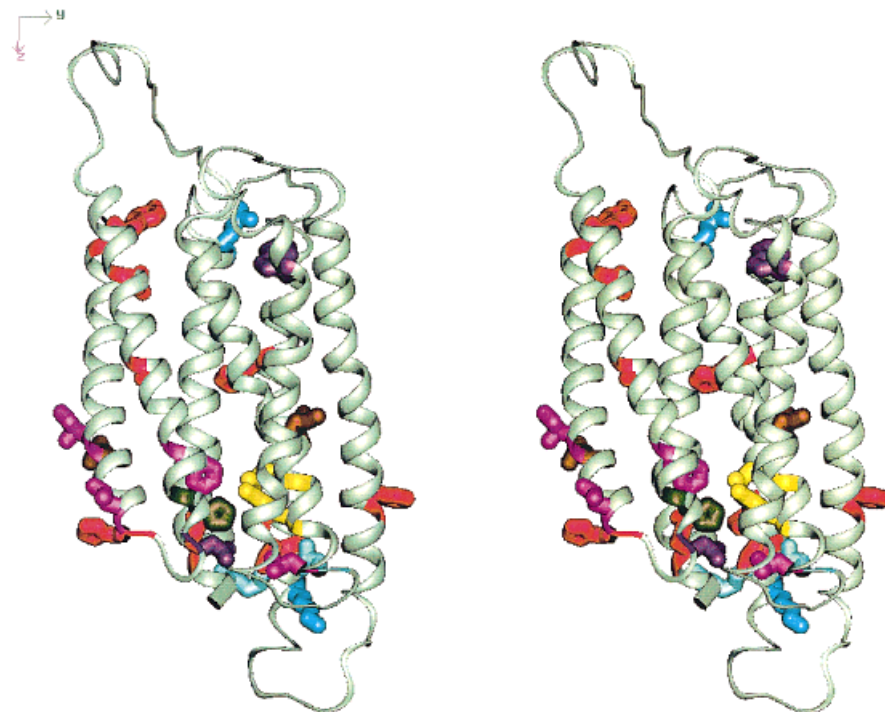


Fig. 5. Stereoview of the 7TM domain of the GLP1 receptor produced with the graphics program Quanta. The extracellular side is uppermost and the molecule is oriented orthogonally to the putative membrane (Z-axis downward). The backbone is shown as helix/ribbon and only the most significant side chains (see Results and Discussion section) are

displayed to show how they form clusters. Coloring scheme is: correlated mutations (red, pink, brown, blue, light blue for the different groups<sup>4</sup> of correlated residues), the H2:6R mutation site<sup>40,43</sup> (dark green) and the ExxY site (yellow).

conserved H2:6 is mutated<sup>40</sup> to R. We place this H2:6 on the same side of TM II as the switching R2:2, precisely one turn above it. The substituted R, now in the H2:6 position occupies the polar pocket and the lower R2:2 is forced to permanently face the cytosol. This R could be a common step for signal transduction in G protein-coupled receptors, corresponding to the R of the E/DRY motif in RLF.

This critical H residue has been mutated<sup>43</sup> to R and this conferred constitutive activity on the glucagon receptor.

#### Conserved Residues. Prolines, Glycine

In the study of rhodopsin<sup>11</sup> it was suggested that the somewhat different structure of bacteriorhodopsin and rhodopsin could be associated with the different positions

Fig. 4. Alignment for each TM of SRF and RLF. The conservation statistics for the two families, obtained from variability data ("VAR") in the GPCRDB alignment tables, were used to identify conserved regions within each family, which are labeled PROFILE in the figure.

The following symbols are used to relate the two consensus sequences:

- . Weak homology across the two families
- : Rather strong homology
- | Residue type identity
- X Highly conserved sites ("cold spots").

The most conserved positions in each family (70% or more in SRF, 60% or more in the much larger RLF) are marked with an asterisk on the line labelled CONSERV, those for RLF are colored blue and those for SRF are in orange. Note how the alignment in each TM set can be "pivoted" on these conserved "cold spots." Identities, symbol or X at conserved "cold spots," are shown in magenta. Next, buried residue positions (inside) of frog rhodopsin and the predicted buried positions of SRF are indicated by a # on the lines labelled INSIDE. Finally, the PHD prediction for each TM of SRF transmembrane helix (symbol T in green) and helix not in the membrane (symbol H in black) are shown.

Each of these TMs is described in summary below.

**Helix 1:** The predicted feature assigned to helix I is a 19-residue-long

helix. The helix is predicted to continue 5 to 6 residues at the cytoplasmic side and make a short 3–5 residue cytoplasmic loop connection to helix 2.

**Helix 2:** The predicted feature assigned to helix II is a helix containing 18 residues. The helix very likely continues a couple of residues at the extracellular side.

**Helix 3:** The predicted feature assigned to helix III is a helix containing 25 residues embedded in the transmembrane segment. The helix is predicted to continue a couple of residues on the external side and 7–8 residues on the internal side of the membrane. This makes it the longest predicted helix in the secretin receptor family.

**Helix 4:** The predicted features assigned to helix IV is a helix containing 18 residues embedded in the transmembrane segment. It is possible that the helix continues a couple of residues on each side of the membrane.

**Helix 5:** The predicted feature assigned to helix V is a helix containing 23 residues. It is possible that the helix continue a couple of residues on each side of the membrane. This makes the helix V the second longest predicted helix in the secretin receptor family.

**Helix 6:** Helix VI is predicted to be made up of 12 to 18 residues which makes this helix to the shortest in this family.

**Helix 7:** The length of helix VII is predicted to have 18 residues in the transmembrane part probably continuing a couple of residues on each side of the membrane.

of the fully conserved prolines in the TM helices. In RLF there conserved prolines in TM IV, V, VI, and VII while in SRF there are conserved prolines only in helices IV, V, and VI. In TM IV and V the prolines are at the same site in our alignment while in TM VI the prolines are about two turns of helix apart (see Fig. 4.) The conserved P in TM VII of RLF could be replaced by a fully conserved G in TM VII in SRF.

### Alignment of SRF With RLF and 3D Model

We aligned SRF with RLF using the “cold spot” method described in Methods (see Fig. 4). The distance between these conserved “cold spots” is very similar in both families. We also plot the predicted “inside” of each TM for SRF and RLF and these come into register with each other when the families are aligned in this way. Finally, the PHD prediction for SRF is printed alongside the alignment.

The “cold spot” method as introduced here is novel, but there are certain precedents to the idea. In cases where sequence similarity between protein families is too low for standard alignment techniques based on similarity, other features of protein sequences must be enlisted in the attempt to align these families. Matching of conservation and variability at individual sites within families are catered for in the MaxHom<sup>2</sup> and MULTAL<sup>50,51</sup> protein sequence alignment programs as well as sequence similarity between families.

The use of conserved sites *within* families where the consensus residue type is not maintained *between* them has parallels in 3D protein modelling. Examples of this are correlated mutations observed as being important for preserving protein structure/function<sup>44,52</sup> and the “evolutionary trace” method.<sup>53</sup> Residues required for function are fully conserved while residues critical for preserving the structure necessary for that function can drift in tandem as long as a change at one site is compensated by a change at another site. Further, considerations of variability and conservation at different sites and within their immediate surroundings in 3D have been shown to be important for protein fold identification and structure prediction<sup>20,54</sup> and the preservation of function.<sup>55</sup> These findings lend credence to our use of the “cold spot” method for identification of key sites in sequences that, although not related by strong sequence similarity, nevertheless belong to the same fold.

In our alignment (see Fig. 4) which formed the basis of the homology modelling, sequence similarities *across* the two families become revealed that by themselves were not significant enough to align the families by conventional methods but which now support the alignment obtained by the cold spot method.

We constructed an explicit atomic model of the transmembrane region of the GLP1R, based on the structural framework of rhodopsin, the structural properties extracted in this study and the alignment obtained by the pivoting method described above. The 3D structure is shown in Figure 5. Groups of residues which are highly correlated<sup>44</sup> (see also correlation data and snake plots in <http://swift.embl-heidelberg.de/7tm/seq/002/002.html>)

in the family are observed to cluster in 3D. Also highlighted are significant sites referred to above, the ExxY site and the H2:6 residue<sup>40,43</sup> which faces it in 3D.

### Comparison of Models

A comparison of our model with the 2D models of Donnelly<sup>5</sup> and Tams et al.<sup>8</sup> show that the central positions of all the helices were within three residues displacement laterally except for TM III where the relative position of this helix is shifted by 6 residues. Despite very different approaches used in the three independent studies, the center of TM II is predicted to be at relatively the same vertical position. The orientation of all the TMs is in all three cases very similar.<sup>5,8</sup>

### Considerations Regarding 3D Models of Membrane Proteins

There is no a priori reason to assume that the folding behavior of membrane proteins conforms exactly to that of water-soluble globular proteins. It has been shown<sup>15,16</sup> that helix crossing angles occupy a much smaller range in membrane helical structures than in water-soluble globular proteins and there are some differences in residue type preferences<sup>12,13,16</sup> while side-chain rotamer preferences are not significantly different in GPCRs as compared with water-soluble globular proteins.<sup>16</sup> Our model conforms to these general findings.

### Relevance of the Proposed Structure for Mechanism of Secretin Hormone Action

Our model represents a static structure for the 7TM domain of the receptor, but nothing in our model precludes possible structural changes concomitant with ligand binding. The original template structure<sup>9</sup> was for a dark-state rhodopsin model, i.e., inactive. In going from an inactive structure to an active one several alternatives are possible including changes in helix crossing angles or in sloping, kinking, rotation, or vertical translation of helices, or possibly the formation of dimers<sup>45,46</sup> or domain-swapped dimers.<sup>47</sup> Any or all of these structural features can be incorporated into the model as experimental data accrues.

### Complete Structure of Secretin Receptors

The structure we propose is for the 7TM domain only. As pointed out in the Introduction, one of the distinguishing features of secretin receptors is the large N-terminal domain. We have not addressed the issue of determining this part of the receptor structure in this work as it is a different problem, rendered difficult by the lack of homology to any protein of known structure. The work is in progress<sup>58</sup> and will be reported elsewhere, here we only note that the next problem is to decide how Nter docks onto the 7TM domain.

### CONCLUSION

Members of the SRF display several common properties with those of the much larger RLF, despite the very little sequence similarity between these two families:

1. The pattern of helix length in SRF and RLF are similar. In SRF TMs III, V, and I are most tilted, while the least-tilted helices are TMs II, IV, VI, and VII. These structural characteristics agree with observations in RLF, where the most-tilted helices are assigned to I, II, III, and V, while the least-tilted helices are IV, VI, and VII. Helices III and V appear to be significantly the longest and most-tilted helices in both families.

2. The extent to which the helices are exposed or buried has been estimated by analyzing the distribution of the polar residues in each individual TM of the SRF. The helices I, IV, and V are the most lipid-exposed helices while the helices II, III, VI, and VII are more buried in SRF. The number of polar residue positions in the TMs diminishes in the order: III > VII = II > VI > V = IV = I for both RLF and SRF.

3. The proposed orientation of the helices, in conjunction with experimental data available from site-directed mutagenesis and other studies suggest a plausible orientation in the sense that all of the structural data make up a consistent picture of the structure of the membrane domain of SRF. The residues known to be important for binding and/or activity form a coherent cluster in a central location.

4. The alignment between SRF and RLF obtained using the "cold spot" technique gives the same relative orientation of the helices as predicted by PERSCAN and the helix lengths of the two methods match.

5. The minimum loop lengths are comparable between the two families, suggesting that the overall arrangement of the helices are very similar and therefore that the rhodopsin structure is a good template.

The results of this analysis suggest that the structure of the transmembrane domain of the SRF is very similar to that of rhodopsin.<sup>7</sup> The proposed arrangement is based on predictions and is therefore still speculative and three-dimensional crystallographic data is required to determine the structure in detail.

The value of a model is that it simplifies the description of a system: it focuses attention onto potentially critical features, and the intention is that it will be superseded by better models. Based on our analysis we are currently producing chimeras and mutants suitable for use with the published Zn<sup>2+</sup>-binding<sup>56</sup> and spin-label<sup>57</sup> methods. These experimental results will allow us to refine our model and shed light on details of function such as ligand binding, activation, and coupling to G-proteins.

Finally, our model is a consensus model for the 7TM domains of the entire SRF and therefore our predictions can be transferred directly to other members and tested experimentally in that particular case.

The coordinates of the model can be made available upon request to the authors and will be deposited in the GPCRDB.

## ACKNOWLEDGMENTS

We wish to thank many colleagues for help with this work: Dr. Joyce Baldwin for kindly supplying coordinates of her bovine rhodopsin model<sup>9</sup> and Dr. Dan Donnelly for

the use of his PERSCAN software.<sup>31,32</sup> Prof. Thue Schwartz, Dr. Gerrit Vriend, Dr. Donnelly, Dr. Lotte Bjerre Knudsen, and Dr. Henning Thøgersen kindly read this manuscript and provided valuable criticism. Novo Nordisk has participated as an end-user in the EC-funded GPCRDB project (project number PC96-0224). We thank Dr. Florence Horn for collecting the mutant data and for expert curation of the GPCRDB.

## REFERENCES

- Ballesteros J, Weinstein H. Integrated methods for modeling G-protein coupled receptors. *Meth Neurosci* 1995;25:366-428.
- Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 1991;9:56-68.
- Vriend G. WHAT IF: a molecular modelling and drug design program. *J Mol Graph* 1990;8:52-56.
- Horn F, Weare J, Beukers MW et al. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acid. Res* 1998;26:275-279.
- Donnelly D. The arrangement of the transmembrane helices in the secretin receptor family of G protein-coupled receptors. *FEBS Lett* 1997;409:431-436.
- Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckman E, Downing KH. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* 1990;213:899-929.
- Unger MV, Hargrave AP, Baldwin MJ, Schertler GFX. Arrangement of the rhodopsin transmembrane  $\alpha$ -helices. *Nature* 1997;389:203-206.
- Tams JW, Knudsen SM, Fahrenkrug J. Proposed arrangement of the seven transmembrane helices in the secretin receptor family. *Receptors and Channels* 1997;5:79-90.
- Baldwin JM, Schertler GFX, Unger VZ. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein coupled receptors. *J Mol Biol* 1997;272:144-164.
- Jones TA, Thirup S. Using known structures in protein model building and crystallography. *EMBO J* 1986;5:819-822.
- Baldwin JM. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J* 1993;12:1693-1703.
- Li SC, Deber CM. A measure of helical propensity for amino acids in membrane environments *Nat Struct Biol* 1994;1:368-558.
- Deber CM, Li SC. Peptides in membranes: helicity and hydrophobicity. *Biopolymers* 1995;37:295-318.
- Donnelly D, Overington JP, Stuart VR, Nugent HAJ, Blundell LT. Modeling  $\alpha$ -helix transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. *Protein Sci* 1993;2:55-70.
- Bowie UJ. Helix packing in membrane proteins. *J Mol Biol* 1997;272:780-789.
- Bywater RP, Thomas D, Vriend G. Residue preferences, side chain rotamer angles and helix-helix packing in membrane proteins. 1999. In press.
- Weiss MS, Abele U, Weckesser J, Welte W, Schiltz E, Schulz GE. Molecular architecture and electrostatic properties of a bacterial porin. *Science* 1991;254:1626-1630.
- Cowan SW, Schirmer T, Rummel G, et al. Crystal structures explain functional properties of two *E. coli* porins. *Nature* 1992;358:727-733.
- Rippmann F. Molecular modelling of G protein-coupled receptors: the ligand gives the clue. *7TM* 1994;4:1-17.
- Overington JP, Johnson MS, Sali A, Blundell TL. Tertiary structural constraints on protein evolutionary diversity. *Proc R Soc Lond B Biol Sci* 1990;241:132-145.
- Overington JP, Donnelly D, Johnson MS, Sali A, Blundell TL. Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci* 1992;1:216-226.
- Dayhoff MO, editor. *Atlas of Protein Sequence and Structure Vol. 5 Suppl. 3*. Washington DC: National Biomedical Research Foundation; 1978.
- Scharf M. PhD Thesis, Heidelberg: University of Heidelberg; 1989.



24. Bowie JB, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1991;253:164–170.
25. Donnelly D, Cogdell JR. Predicting the point at which transmembrane helices protrude from the bilayer: a model of the antenna complexes from photosynthetic bacteria. *Protein Eng* 1993;6:629–635.
26. Rost B, Sander C. Transmembrane helices predicted at 95% accuracy. *Protein Sci* 1995;4:521–533.
27. Shiffer M, Edmundson AB. Use of a helical wheel to represent the structures of protein and to identify segments with helical potential. *Biophys J* 1967;7:121–135.
28. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* 1982;299:371–374.
29. Finer-Moore J, Stroud RM. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. *Proc Natl Acad Sci USA* 1984;81:155–159.
30. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA* 1984;81:140–144.
31. Donnelly D, Overington JP, Blundell TL. The prediction and orientation of  $\alpha$ -helices from the sequence alignments: the combined use of environment-dependent substitution tables, Fourier transform methods and helix capping rules. *Protein Eng* 1994;7:645–653.
32. Donnelly D, Findlay JBC, Blundell TL. The evolution and structure of Aminoergic G protein-coupled receptor. *Receptors Channels* 1994;2:61–78.
33. Komiya H, Yeates TO, Rees DC, Allen JP, Feher G. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: symmetry relations and sequence comparison between different species. *Proc Natl Acad Sci USA* 1988;85:9012–9016.
34. Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J Mol Biol* 1987;195:659–685.
35. Andersson H, Von Heijne G. Membrane protein topology: effects of delta mu H<sup>+</sup> on the translocation of charged residues explain the "positive inside" rule. *EMBO J* 1994;13:2267–2272.
36. Karnik S, Khorana HG. Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *J Biol Chem* 1990;265:17520–17524.
37. Kurtenbach E, Curtis CAM, Pedder EK, Aitken A, Harris ACM, Hulme EC. Muscarinic acetylcholine receptors. Peptide sequencing identifies residues involved in antagonist binding and disulfide bond formation. *J Biol Chem* 1990;265:13702–13708.
38. Dohlman HG, Caron MG, DeBlasi A, Frielle T, Lefkowitz RJ. Role of extracellular disulfide-bonded cysteines in the ligand binding function of the beta 2-adrenergic receptor. *Biochemistry* 1990;29:2335–2342.
39. Gardella TJ, Juppner H, Wilson AK et al. Determinants of [Arg2]PTH-(1–34) binding and signaling in the transmembrane region of the parathyroid hormone receptor. *Endocrinology* 1994;135:1186–1194.
40. Schipani E, Kruse K, Juppner H. A constitutively active mutant PTH/PTHrP receptor in Jansen-type metaphyseal chondrodysplasia. *Science* 1995;268:98–100.
41. Vilardaga JP, di Paolo E, de Neef P, Waelbroeck M, Bollen A, Robberecht P. Lysine 173 residue within the first exoloop of rat secretin receptor is involved in carboxylate moiety recognition of Asp 3 in secretin. *Biochem Biophys Res Commun* 1996;218:842–846.
42. Turner PR, Bambino T, Nissenson RA. Mutations of neighboring polar residues on the second transmembrane helix disrupt signaling by the parathyroid hormone receptor. *Mol Endocrinol* 1996;10:132–139.
43. Hjorth SA, Ørskov C, Schwartz TW. Constitutive activity of glucagon receptor mutants. *Mol Endocrinol* 1998;12:78–86.
44. Horn F, Bywater RP, Krause G et al. The interaction of class B G-protein-coupled receptors with their hormones. *Receptors Channels* 1998;5:305–314.
45. Herberg JT, Codina J, Rich KA, Rojas FJ, Iyengar R. The hepatic glucagon receptor. Solubilization, characterization, and development of an affinity adsorption assay for the soluble receptor. *J Biol Chem* 1984;259:9285–9294.
46. Hebert TE, Moffett S, Morello JP, Loisel TP, Bichet DG, Barret C, Bouvier M. A peptide derived from a  $\beta_2$ -adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J Biol Chem* 1996;271:16384–16392.
47. Gouldson PR, Snell CR, Bywater RP, Higgs C, Reynolds CA. Domain swapping: a mechanism for functional rescue in G-protein coupled receptors. *Protein Eng* 1998;11:1181–1193.
48. Scheer A, Fanelli F, Costa T, De Benedetti PG, Cotecchia S. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *EMBO J* 1996;15:3566–3578.
49. Oliveira L, Paiva AC, Sander C, Vriend G. A common step for signal transduction in G protein-coupled receptors. *Trends Pharmacol Sci* 1994;15:170–172.
50. Taylor WR, Jones DT. Deriving an amino acid distance matrix. *J Theor Biol* 1993;164:65–83.
51. Taylor WR. Motif based protein sequence alignment. *J Comp Biol* 1994;1:297–311.
52. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interactions. *J Mol Biol* 1997;271:511–523.
53. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 1996;257:342–358.
54. Ison JC, Parish JH, Daniel SC, Blades MJ, Findlay JBC. A key residues approach to protein fold detection. 1999. In press.
55. Cardle L, Dufton MJ. Identification of important functional environs in protein tertiary structures from the analysis of residue variation in 3D: application to cytochromes c and carboxypeptidases A and B. *Protein Eng* 1994;7:1423–1431.
56. Elling CE, Møller-Nielsen S, Schwartz TW. Conversion of antagonist-binding site to metal-ion site in a tachykinin NK-1 receptor. *Nature* 1995;374:74–77.
57. Farrens DL, Altenbach CA, Yang K, Hubbell WL, Khorana HG. Requirement of rigid-body motions of transmembrane helices for light activation of rhodopsin. *Science* 1996;274:768–770.
58. Munro REJ, Taylor WR, Bywater RP. Ab initio folding of the N-terminal domain of the secretin receptors. 1999. In press.