The arrangement of the transmembrane helices in the secretin receptor family of G-protein-coupled receptors

Dan Donnelly *

Department of Pharmacology, University of Leeds, Leeds LS2 9JT, UK

Received 2 April 1997; revised version received 25 April 1997

Abstract The members of the secretin receptor family of G-protein-coupled receptors share no significant sequence similarity to the more familiar rhodopsin-like family. However, multiple sequence alignment analysis reveals seven hydrophobic regions with significant α-helical periodicity. Residues that are likely to be buried on the interior of the helical bundle and others that are likely to contact the lipid bilayer are identified. A predicted arrangement of the helical bundle is described in which, by comparison with the arrangement in the rhodopsin family, helices 2 and 7 are more buried within the bundle while helix 3 is more exposed to the lipid bilayer.

Key words: G-protein-coupled receptor; Secretin; Rhodopsin; Periodicity; Model; α-Helix

1. Introduction

The investigation of the structure and function of G-protein-coupled receptors (GPCR) has been the subject of numerous studies over the past decade [1–5]. However, the vast majority of this work has concentrated upon the rhodopsin-like GPCRs and, in particular, the β2-adrenergic receptor and rhodopsin itself. Significant advances in the elucidation of the 3-dimensional structure of rhodopsin have been made in recent years which have, in addition to multiple sequence alignment analysis and indirect structural studies, provided a preliminary picture of the arrangement of the helices in the transmembrane bundle [6–11].

The receptors of the rhodopsin family are characterised by a number of highly conserved regions, or fingerprints, that can be used to align sequences that would otherwise be too divergent for meaningful comparison [12]. However, the cloning of the secretin receptor and a variety of related GPCRs suggested a distinct GPCR family (the SecR family) that did not match the rhodopsin family fingerprint [13,14]. Since these receptors possess the seven putative transmembrane α-helices that characterise GPCRs, it is reasonable to assume that the general fold of the transmembrane bundle in the two families is likely to be conserved. However, the absence of significant sequence identity allows for the possibility of significant movements between the relative positions of equivalent α-helices, as has been observed in the globin family [15]. In addition, the absence of a reliable sequence alignment of the two families means that the deduced internal faces of the helices and the construction of molecular models of the SecR family must be deduced from first principles, rather than by comparison with those based upon rhodopsin [2]. Furthermore, the paucity of experimental data relating to the structure of SecR family GPCRs means that automated modelling procedures using experimentally derived distance constraints are inappropriate [16]. This paper describes the analysis of a multiple sequence alignment of members of the SecR family which has enabled a prediction of the global arrangement of the helix bundle and the positioning of individual residues within this domain. Since the number of structure/function studies of this family of receptors is increasing rapidly, the models presented here will be useful for the planning of experiments, especially those which require knowledge of the contact points between transmembrane helices.

*Fax: (44) 113-233-4331.
E-mail: d.donnelly@leeds.ac.uk

2. Methods

Members of the SecR family were identified using the PRINTS database [17] and complete sequences were extracted from the OWL sequence database v28.0 [18]. Sequence alignments were carried out using the program MALIGN [19] and trees were constructed using PHYLIP [20] with distances calculated as −100(log[percentage identity]/100). Fourier transform analysis was carried out using the computer programs within PERSCAN v7.0 as described previously [21].

The sequence alignment was scanned with a window of 18 residues using PERSCAN and the window with the highest α-helical periodicity (AP value [21]) within each hydrophobic region was defined as a transmembrane region. Note that if any sequences are identical within the window used in the PERSCAN analysis, the computer program only includes one such sequence in order to avoid bias. The sequence pattern at each position in each alignment was used to estimate the extent to which that position is buried or lipid-accessible, based upon substitution tables calculated from alignments of proteins of known structure as described previously [21]. The periodicity in the predicted accessibilities was then calculated using a Fourier transform procedure to yield both an estimation of the extent of the α-helical periodicity (AP_{max}) and a predicted direction of the internal face of the helix (θ_{max}) relative to the first residue in that helix.

The SecR family sequences were grouped into either five major groups or 15 smaller subsets of closely related receptors based upon the clustering in the dendrogram (Fig. 1). A buried position is defined where either (i) a residue (other than glycine or proline) is completely conserved over three of the five major groups or (ii) at least one charged residue (Asp, Glu, His, Lys or Arg) exists within the central ten residues of the helix. A position is defined as being lipid accessible if it is variable within one of the 15 smaller subsets of closely related receptors. A variable position is defined as one where, for at least one of the smaller subsets, all the residues do not fall into one of the following nine residue groupings: (A, L, V, I, M), (L, V, I, M, F), (F, Y, W, H), (A, G, P, S, T), (S, T, C), (N, Q, H), (S, T, N, Q), (H, R, K), (D, E, N, Q).

3. Results and discussion

The Fourier transform analysis of the transmembrane sequence alignments are summarised in Table 1. A value of AP_{max} greater than 2 suggests that the α-helical periodicity is significant [21,22]. As can be seen from Table 1, all the

PII S0014-5793(97)00546-2

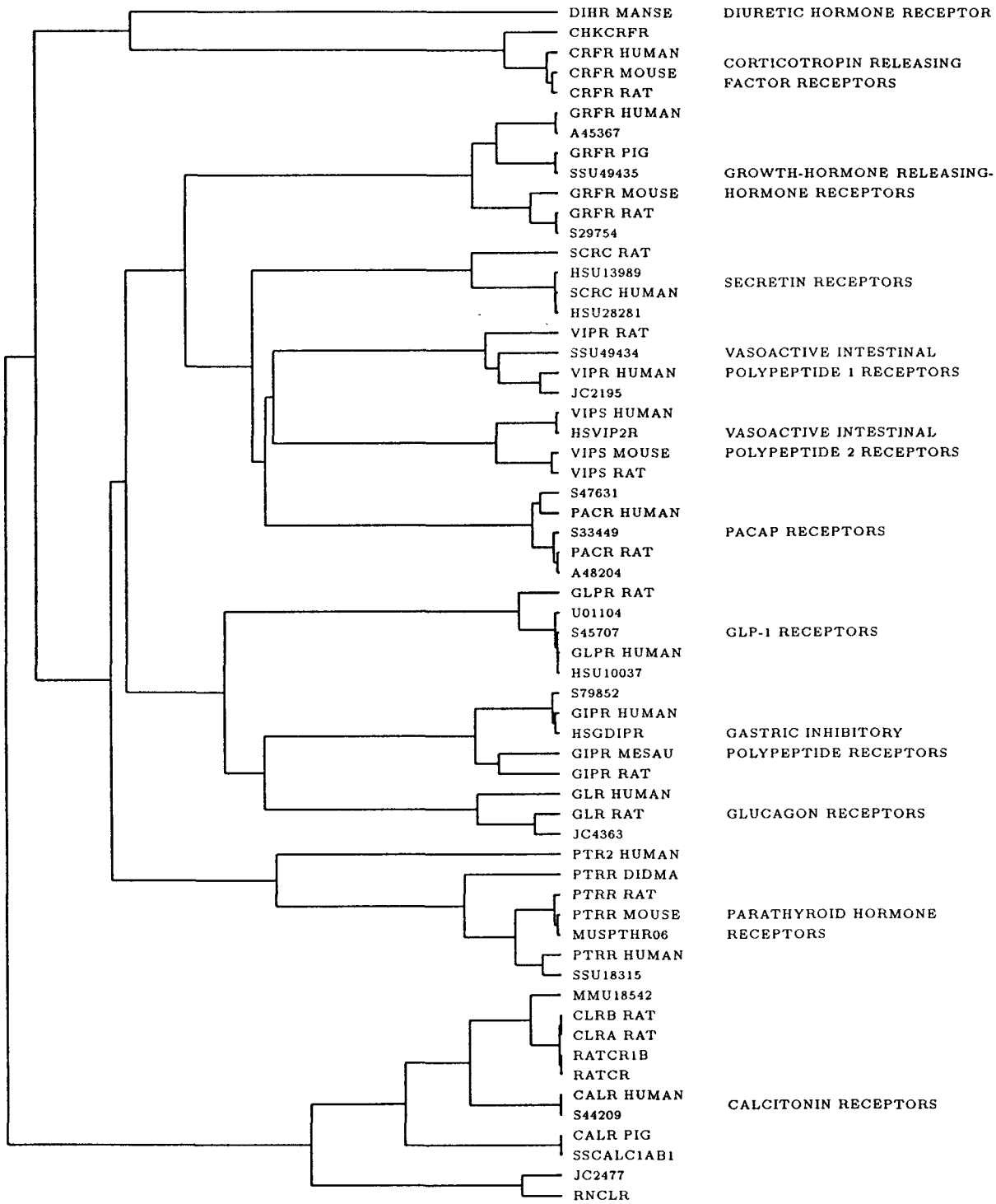

Fig. 1. Dendrogram depicting the clustering of the SecR family GPCRs based upon their relative levels of sequence identity.

AP_{max} values are greater or equal to 3.38 which suggests that all seven of the hydrophobic regions in the SecR family form transmembrane α-helices.

The Fourier transform analysis predicts the *direction* of the internal face of each helix (Table 1) but it does not estimate the *extent* of the buried face. For example, although a Fourier transform analysis of the rhodopsin-like GPCRs was able to show significant α-helical periodicity and predict the internal face of each helix [23], a more detailed analysis of the family was required, alongside the projection map of rhodopsin, to predict the extent to which each helix was buried within the helix bundle [7]. In the latter analysis, Baldwin used a large sequence alignment of rhodopsin-like GPCRs to predict whether individual residue positions on a helix were buried or lipid-accessible, which in turn allowed the extent of the exposure of the entire helix to be estimated. It was then possible to assign each helix to a peak in the projection map of rhodopsin.

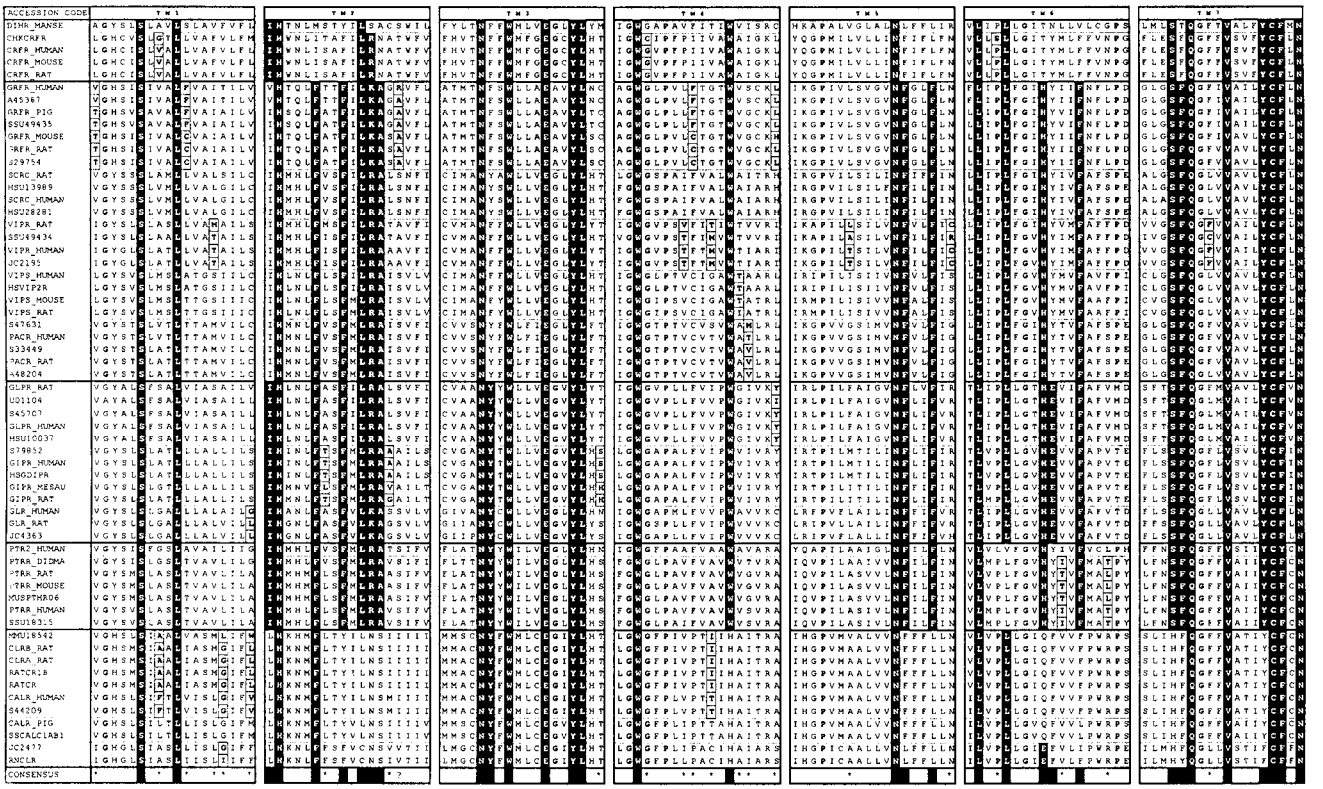

Fig. 2. Sequence alignment of the 7 putative transmembrane regions of the SecR family of GPCRs. Bold horizontal lines separate the 5 major clusters, while dotted horizontal lines separate the 15 smaller groups of closely related sequences. White letters upon black backgrounds represent residue positions predicted to be buried. Boxed letters represent residue positions predicted to be lipid-accessible. The consensus prediction is shown on the final row with buried positions depicted with black squares and lipid-accessible residues depicted with asterisks (*). Conserved Gly and Pro residues are shown in bold.

In order to predict the arrangement of the helices in the SecR family of GPCRs, a similar procedure was used to complement the Fourier transform analysis. The rules for defining whether a residue was either buried or in contact with the bilayer were based upon general principles of protein structure as determined from a variety of studies [e.g. [7,21,22,24]]. Charged residues are unlikely to contact the lipid region of the bilayer although those on the extremities of the transmembrane helices may interact with the polar phospholipid headgroup region. Hence, the presence of at least one charged residue within the central 10 positions was considered as defining that position as buried.

A second property used to identify the location of residues relies upon the observation that residues buried in the core of a protein, which are under greater conformational restraint, tend to be more conserved than those that are exposed to the lipid [21–24]. However, the special conformational properties of glycine and proline [25] could allow these residues to influence the properties of an α-helix even when they are on the external lipid-accessible face; hence they were given special consideration. In measuring the extent of conservation, it is more significant if a residue is conserved across sequences that are otherwise divergent, whereas a variable position is more significant in sequences that are closely related [7]. Therefore, the clustering of the sequences in the dendrogram was used to group them either into subsets of closely related sequences *within* which variable positions are more significant, or else into larger more divergent groups *between* which conserved positions are more significant.

The dendrogram calculated from the full sequence alignment of the SecR family is shown in Fig. 1. The clustering pattern allows for the division of the SecR sequences into five major categories (at approximately the 45% identity level) or else into 15 smaller subsets (at approximately the 75% identity level). Fig. 2 shows the alignment of the seven putative transmembrane regions with the five larger groups separated by bold horizontal lines and the 15 smaller subsets by dotted horizontal lines. The residues used to define buried positions are shown with a white letter upon black background while those used to define a lipid accessible position are shown with boxed letters. An accepted numbering convention for the future analysis of mutations in this family of receptors would be extremely useful. A convention in this paper, similar to that used in other receptor families, is based upon the alignment in Fig. 2 in which the helix number is followed by the position of the residue in that helix — for example, position 3.11 represents the eleventh position in transmembrane helix 3 (TM3).

The final row of Fig. 2 displays the overall consensus pre-

Table 1
Analysis of α-helical periodicity in the transmembrane regions

	AP_{max}	θ_{max}
TM1	5.52	148°
TM2	4.24	320°
TM3	3.78	38°
TM4	3.89	238°
TM5	4.20	70°
TM6	3.47	234°
TM7	3.38	253°

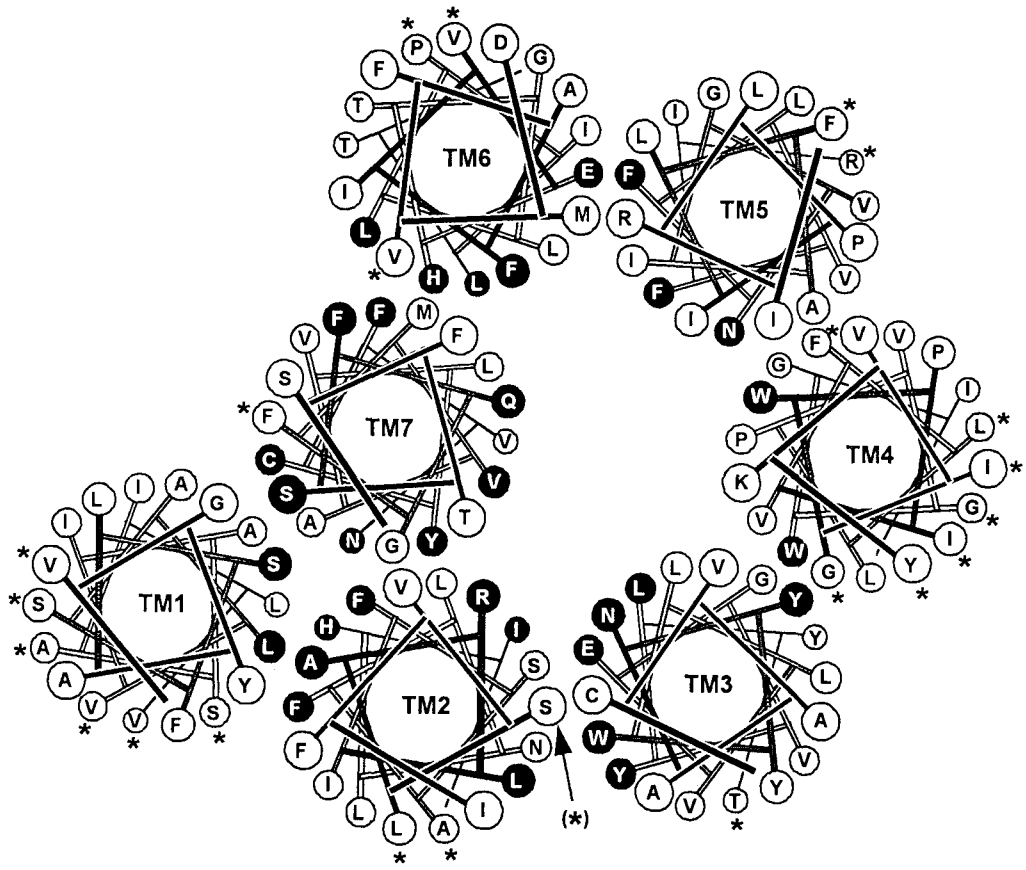

Fig. 3. Predicted arrangement of the SecR family transmembrane domain using the sequence of the rat GLP-1 receptor (**GLPR_RAT**; see legend to Fig. 2 for consensus prediction conventions).

diction with the buried positions shown with black squares and the lipid accessible positions shown with asterisks. If each sequence is represented as a helical wheel, it can be seen that, in general, the predicted buried and lipid-accessible faces of the helices are mutually exclusive (Fig. 3) with only two exceptions. In the first of these, there is an overlap between the buried and lipid-accessible faces on TM2. However, in this case the accuracy of the lipid-accessible prediction is questionable since it is based upon the presence of an arginine residue at position 2.15 of the human growth hormone-releasing hormone-receptor (**GRFR_HUMAN**). This is likely to be a sequence error since a second sequence for the same human receptor (**A45367**) suggests that this residue is an alanine as it is in the rat, mouse and pig homologues. Further evidence that this position is buried comes from mutagenesis work on the opossum PTH receptor in which the Ser^{233}–Ala substitution results in a greater than 50% reduction of the maximal stimulation of adenylyl cyclase activity as well as a rightward shift in the dose response curve [26]. The second overlap of the predicted buried and lipid-accessible faces is in TM6, although the overlap is of only one residue. This may indicate the direction of the tilt of TM6 since it suggests that the extracellular C-terminus of the helix tilts away from the bundle while the cytoplasmic N-terminus tilts towards it.

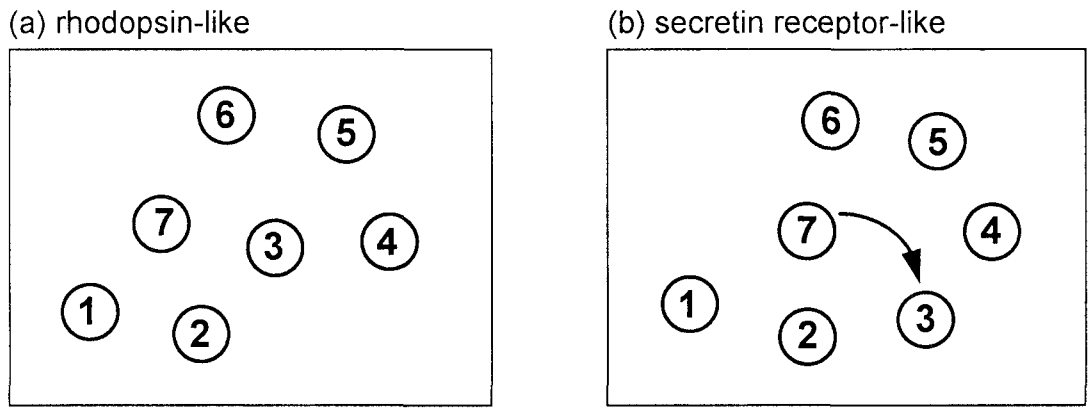

Fig. 4. Comparison of the arrangement of the transmembrane helices in the rhodopsin-like and SecR-like GPCRs.

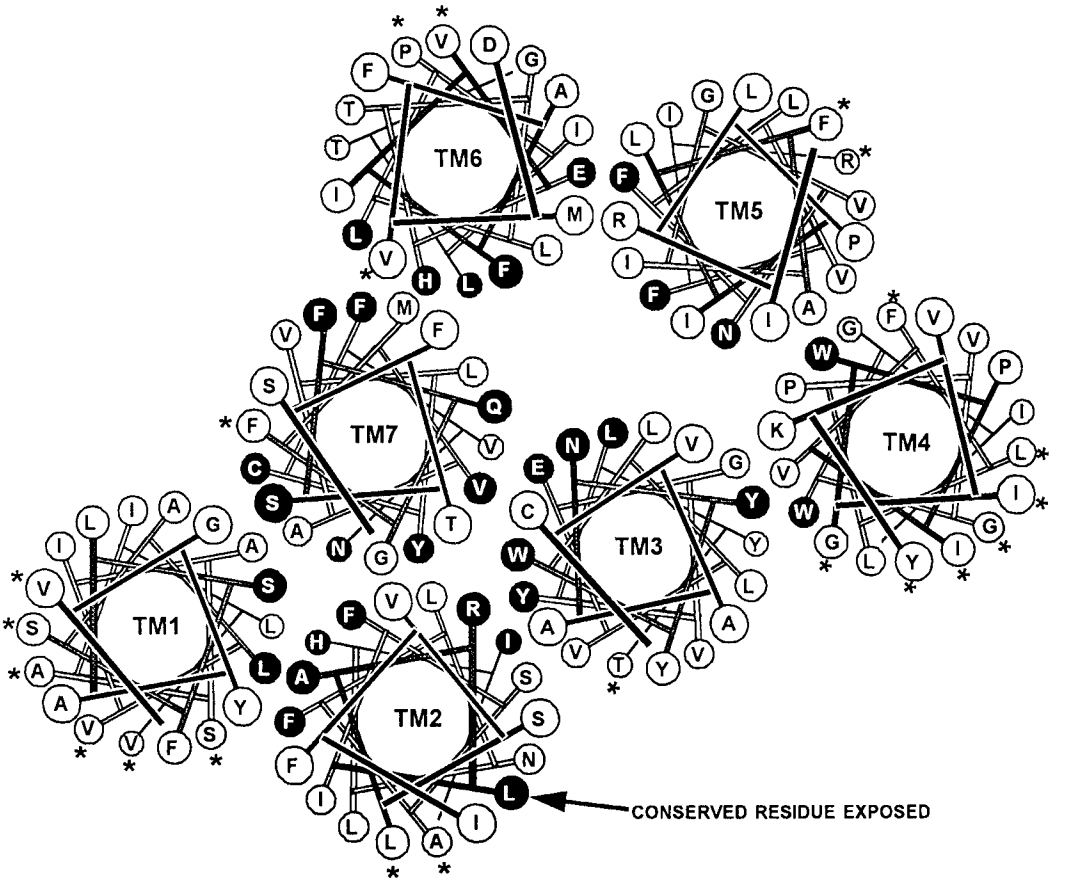

Fig. 5. Model of the transmembrane domain of the SecR-like GPCRs based upon the arrangement of rhodopsin [6] (see legend to Fig. 2 for consensus prediction conventions).

Fig. 3 shows the seven helical wheels placed in a sequential anti-clockwise arrangement with all the predicted buried residues either orientated towards the interior of the helix bundle or in contact with neighbouring helices. The arrangement of the helices appears to differ from that observed for rhodopsin [6] in that helix 3 is more exposed while helices 2 and 7 are more buried (Fig. 4). Relative movements between helices in distantly related proteins have been observed, notably in the globin family [15], although their overall fold is conserved. It should therefore not be too surprising if differences are observed between the transmembrane helices of the rhodopsin-like and secretin receptor-like GPCRs since they share no significant sequence identity.

However, if the helices of the SecR family are arranged a priori in a similar fashion to those of rhodopsin, then it is possible to place the vast majority of the predicted buried residues in contact with neighbouring helices (Fig. 5). The only significant exception is position 2.11 which would now be on the lipid accessible face of TM2, despite containing a leucine residue in all but two of the sequences. It would be difficult to explain why such a conserved residue should exist on the lipid accessible face but the possibility that it plays an important role in some aspect of the receptor structure/function (e.g. the protein folding mechanism, dimer formation, etc.) cannot be completely ignored.

The SecR family is characterised by a relatively large N-terminal domain containing several highly conserved residues which include six cysteines that may form disulphide bonds. Although the transmembrane domain of some GPCRs plays a dominant role in forming the ligand binding pocket [1], there is growing evidence that the SecR family extracellular domain is the major factor in the peptide binding site [e.g. [27,28]]. Hence the models described here are unlikely to be useful in predicting the majority of the ligand interaction points. However, the transmembrane domain is critical for mediating the conformational changes required for the transduction of the extracellular signal to the intracellular G protein. Hence the SecR family models described here will be useful in predicting experiments for locating 'hot spots' for constitutively active receptors, as has recently been carried out for the angiotensin AT_1 receptor [29]. Moreover, the models will be useful in designing experiments which aim to identify contact points between transmembrane helices, as has been carried out for rhodopsin [11].

A potential helix–helix contact point has been identified in the PTH receptor between Arg^{233} (position 2.12) and Gln^{451} (position 7.06). Substitution of Arg^{233} caused upto a 200-fold reduction in agonist binding affinity without effecting antagonist binding. However, the agonist binding affinity was restored by a second mutation in which Gln^{451} was substituted with Lys [30]. The suggested functional coupling of these residues is compatible with the model in Fig. 3 (Cα–Cα distance of approximately 8 Å [1] but it is more difficult to envisage in

[1] The co-ordinates of 3-dimensional model based upon Fig. 3, as well as a tentative model in which the helices have been tilted and loops added, can be obtained from http://swift.embl-heidelberg.de/7tm/models/.

the model depicted in Fig. 5 where TM3 blocks the path between TM2 and TM7. In addition, the predicted buried residues at positions 2.2, 2.8, 2.12 and 2.15 have been shown to be functionally important in the PTH receptor [26].

Therefore, in summary, a model for the transmembrane domain of the SecR family of GPCRs is presented. The overall structure of the domain is based upon the fold of the rhodopsin family — that is, a 7 helical bundle in which the helices are arranged in an anti-clockwise fashion when viewed from the extracellular side of the bilayer [10]. However, the precise positioning of individual residues within this helical bundle was derived from first principles through the analysis of a multiple sequence alignment. If all the conserved residues are positioned in the interior of the helix bundle, then the shape of the bundle differs from that of rhodopsin in that TM3 is more exposed while TM2 and TM7 are more buried. Alternatively, if the helices are arranged a priori in a similar fashion to those in the rhodopsin model, then only one highly conserved position is significantly exposed to the bilayer. It is not clear which arrangement is most likely; the model described in Fig. 3 is more objective and rule-based and appears to agree more fully with the limited mutagenesis data. On the other hand, the model shown in Fig. 5 is based upon the arrangement of a structural homologue. Of course, both of the 2-dimensional models presented here do not take into account any relative tilting between the transmembrane helices which will affect, to a limited degree, the contact points between neighbouring helices [1]. However, such models are open to empirical testing and should be extremely useful for the planning of experiments that require knowledge of contact points between transmembrane helices [e.g. [9–11,29]].

Acknowledgements: I thank Dr. Robert Bywater, Novo Nordisk and the Royal Society for their support. I thank Dr. S.A. Baldwin for his comments on this manuscript.

References

[1] Strader, C.D., Fong, T.M., Tota, M.R., Underwood, D., Ann. Rev. Biochem. 63 (1994) 101–132.
[2] Donnelly, D., Findlay, J.B.C., Curr. Opin. Struct. Biol. 4 (1994) 582–589.
[3] Baldwin, J.M., Curr. Opin. Cell Biol. 6 (1994) 180–190.
[4] Schwartz, T.W., Curr. Opin. Biotechnol. 5 (1994) 434–444.
[5] Freedman, N.J., Lefkowitz, R.L., Recent Progr. Hormone Res. 51 (1996) 319–353.
[6] Unger, V.M., Schertler, G.F.X., Biophys. J. 68 (1995) 1776–1786.
[7] Baldwin, J.M., EMBO J. 12 (1993) 1693–1703.
[8] Donnelly, D., Findlay, J.B.C., Blundell, T.L., Receptors Channels 2 (1994) 61–78.
[9] Elling, C.E., Nielsen, S.M., Schwartz, T.W., Nature 374 (1995) 74–77.
[10] Elling, C.E., Schwartz, T.W., EMBO J. 15 (1996) 6213–6219.
[11] Hongbo, Y., Kono, M., McKee, T.D., Oprian, D.D., Biochemistry 34 (1995) 14963–14969.
[12] Attwood, T.K., Findlay, J.B.C., Prot. Eng. 6 (1993) 167–176.
[13] Attwood, T.K., Findlay, J.B.C., Prot. Eng. 7 (1994) 195–203.
[14] Segre, G.V., Goldring, S.R., TEM 4 (1993) 309–314.
[15] Lesk, A.M., Chothia, C., J. Mol. Biol. 136 (1980) 225–270.
[16] Herzyk, P., Hubbard, R.E., Biophys. J. 69 (1995) 2419–2442.
[17] Attwood, T.K., Beck, M.E., Prot. Eng. 7 (1994) 841–848.
[18] Bleasby, A.J., Wooton, J.C., Prot. Eng. 3 (1990) 153–159.
[19] Johnson, M.S., Overington, J.P., J. Mol. Biol. 233 (1993) 716–738.
[20] Felsenstein, J., Evolution 35 (1985) 783–791.
[21] Donnelly, D., Overington, J.P., Ruffle, S.V., Nugent, J.H.A., Blundell, T.L., Prot. Sci. 2 (1993) 55–70.
[22] Komiya, H., Yeates, T.O., Rees, D.C., Allen, J.P., Feher, G., Proc. Natl. Acad. Sci. USA 85 (1988) 9012–9016.
[23] Donnelly, D., Johnson, M.S., Blundell, T.L., Saunders, J., FEBS Lett. 251 (1989) 109–116.
[24] Rees, D.C., Komiya, H., Yeates, T.O., Allen, J.P., Feher, G., Ann. Rev. Biochem. 58 (1989) 607–633.
[25] G.E. Schulz, R.H. Schirmer, Principles of protein structure, Springer, New York, 1979.
[26] Turner, P.R., Bambino, T., Nissenson, R.A., Mol. Endocrinol. 10 (1996) 132–139.
[27] Cao, Y.-J., Gimpl, G., Farenholz, F., Biochem. Biophys. Res. Commun. 212 (1995) 673–680.
[28] Wilmen, A., Göke, B., Göke, R., FEBS Lett. 398 (1996) 43–47.
[29] Balmforth, A.J., Lee, A.J., Warburton, P., Donnelly, D., Ball, S.G., J. Biol. Chem. 272 (1997) 4242–4251.
[30] Gardella, T.J., Luck, M.D., Fan, M.-H., Lee, C., J. Biol. Chem. 271 (1996) 12820–12825.