Journal of Mathematical Chemistry 23 (1998) 229–238

A general procedure for building the transmembrane domains of G-protein coupled receptors

Juan J. Perez [a], Marta Filizola [a,b] and Maria Caritenì-Farina [b]

[a] *Dept. d'Enginyeria Quimica, UPC, ETS d'Enginyers Industrials, Av. Diagonal 647, 08028 Barcelona, Spain*

[b] *Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche (CRISCEB), Seconda Universitá degli Studi di Napoli, Via Costantinopoli 16, 80138 Napoli, Italy*

The only results available at present about the structural features of G-protein coupled receptors are the low resolution electron projection maps obtained from microscopy studies carried out on two-dimensional crystals of rhodopsin. These studies support previous suggestions that these integral proteins are constituted by seven transmembrane domains. The low resolution electron density map of rhodopsin can be used to extract information about helix relative positions and tilt. This information, together with a reliable procedure to assess the residues involved in each of the transmembrane regions, can be used to construct a model of rhodopsin at atomic resolution. We have developed an algorithm that can be used to generate such a model in a completely automated fashion. The steps involved are: (i) locate the centers of the helices according to the low resolution electron density map; (ii) compute the tilt of each helix based on the elliptical shape observed by each helix in the map; (iii) define a local coordinate system for each of the helices; (iv) bring them together in an antiparallel orientation; (v) rotate each helix through the helical axis in such a way that its hydrophobic moment points in the same direction as the bisector formed between three consecutive helices in the bundle; (vi) rotate each helix through an axis perpendicular to the helical one to assign a proper tilt; (vii) translate each of the helix to its center deduced from the projection map. A major advantage of the procedure presented is its generality and consequently can be used to obtain a model of any G-protein coupled receptor with the only assumption that the shape of the bundle is the same as found in rhodopsin. This avoids uncertainties found in other procedures that construct models of G-protein coupled receptors based on sequence homology using rhodopsin as template.

1. Introduction

G-protein coupled receptors (GPCRs) are a large group of integral membrane proteins of pivotal importance, directly involved in the transmission of signals to the interior of the cell [30,33]. These receptors exhibit seven hydrophobic domains, predicted to correspond to the transmembrane spanning regions of the protein. Moreover, systematic analysis of the hydrophobicity profile of these transmembrane domains, using both helical wheel and Fourier analysis, clearly suggests that these domains correspond to helical secondary structures.

Due to the difficulties associated to the isolation and purification in the quantities demanded by structural analysis, structural information about these integral proteins is very scarce. Moreover, it is notoriously difficult to obtain crystals suitable for structural analysis by X-ray diffraction. Microscopy studies of 2D crystals, have provided an electron density projection map of rhodopsin [28,29], the only GPCR structurally characterized to date. Unfortunately, the resolution of the two structures available is very low (9 and 7 Å, respectively) to provide a detailed structural information necessary to understand ligand–receptor interactions. However, in spite of the low resolution of these maps, they permit identification of the individual helices, and consequently get insight into their spatial arrangement in the form of a bundle and furthermore, obtain geometrical information about the relative position and orientation of each of the helices.

This structural information, as well as other results can be used to generate 3D models of these receptors at the atomic level. These structures are very useful since, they can be used to propose site directed mutagenesis experiments or suggest structural changes on different families of ligand and in general, contributing to increase our level of understanding of ligand–receptor interactions. Futhermore, as a feedback of these predictions, the outcome of the experiments proposed with these models would be very valuable to decide the validity of the approximations used in their generation.

In the last few years, several approaches to construct 3D models of GPCRs have been proposed [1,4–8,10,11,13–17,19–25,27,31,32,34–36]. Early models of these proteins were constructed by sequence homology modeling using bacteriorhodopsin as template. Atomic coordinates of this protein are available from the electron density projection map recorded from 2D crystals at 3.5 Å resolution [12]. However, although bacteriorhodopsin is an integral membrane protein exhibiting seven transmembrane domains, on the one hand, it is not a GPCR, and on the other, does not exhibit a significant sequence identity with any GPCR. These two issues together with the recent findings of substantial differences with rhodopsin in its helix arrangement, directed modeling studies in recent years to construct GPCRs without using bacteriorhodopsin as template. One alternative route was to use rhodopsin as template. However, since its coordinates are not available, it is necessary first to construct a reliable atomic model of it. Models of rhodopsin have been constructed by manual manipulation of the bacteriorhodopsin coordinates, imposing geometrical restraints inferred from experimental or theoretical studies. These resulting structures have been then used as templates for modeling other GPCRs. Unfortunately, these models have a limited usefulness since first, a low sequence identity is observed between GPCRs of different families and second, transmembrane regions have different lengths in diverse GPCRs. These two issues make questionable the construction of GPCRs by homology modeling. Alternatively, sequence divergence analysis [1] is a technique that avoids the difficulties associated with homology modeling of GPCRs, but it is hampered by the need to compare a large number of sequences sharing more than about a 60% of sequence identity.

In the present work, we proposed a general procedure to construct the transmembrane domains of any GPCR. The method does not require the use of a template protein

and avoids the requirements of the sequence divergence analysis procedure. The construction of the transmembrane domains of the protein is based on the assumption that, they are arranged in a seven helix bundle in the same spatial disposition as in rhodopsin. In the procedure, each helix is treated as an rigid object and global manipulations on the helix are induced on every atom through a local coordinate system defined on each helix. The steps involved in the procedure are: (i) location of the centers of the helices according to the low resolution electron density map; (ii) computation of the tilt of each helix, according to the elliptical shape exhibited by each helix in the projection map; (iii) definition of a local coordinate axis for each of the helices; (iv) orient them in an antiparallel fashion; (v) rotation of each helix through the helical axis in such a way that its hydrophobic moment is pointing in the same direction of the bisector formed between three consecutive helices in the bundle; (vi) rotation of each helix through an axis perpendicular to the helical one to assign a proper tilt; (vii) translation of each of the helix to its center deduced from the projection map.

2. Information deduced from the rhodopsin electron density projection map

The very first information required to model GPCRs is to decide the way the helices are arranged in the bundle. This information cannot be deduced from the bovine rhodopsin electron density projection map. Indeed, since the map exhibits two symmetry-related images both of which may represent a view from the same side of the membrane, there are 10,080 possible ways to assign them. In the present work, helices were assigned following the assumptions proposed by Baldwin as shown schematically in figure 1 [2]. Combining the low resolution map of rhodopsin with a detailed analysis of 204 GPCR sequences, together with the analysis of length ranges of the interhelical loops, she was able to fit each helix to the peaks in the projection map of rhodopsin and to propose tentative 3D arrangements of the helices by means of

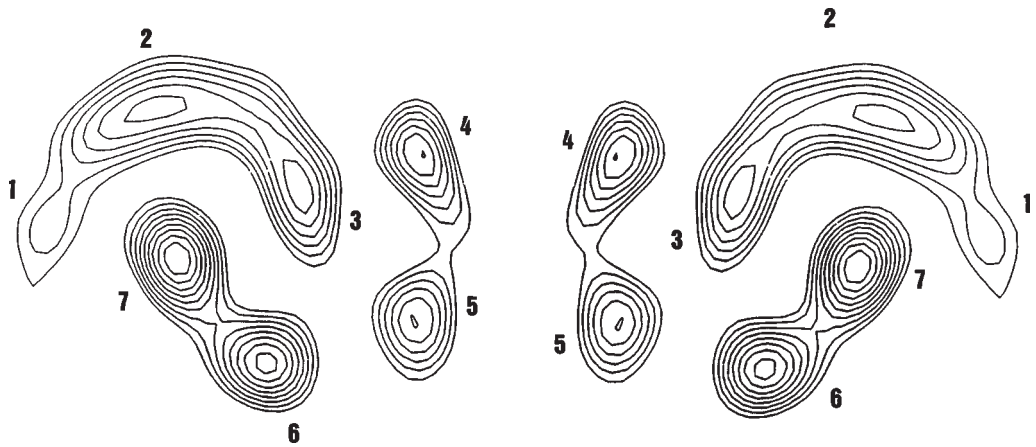

Figure 1. Representation of the asymmetric unit of the rhodopsin electron density projection map. Numbering of the helices was carried out following Baldwin's suggestions (see text).

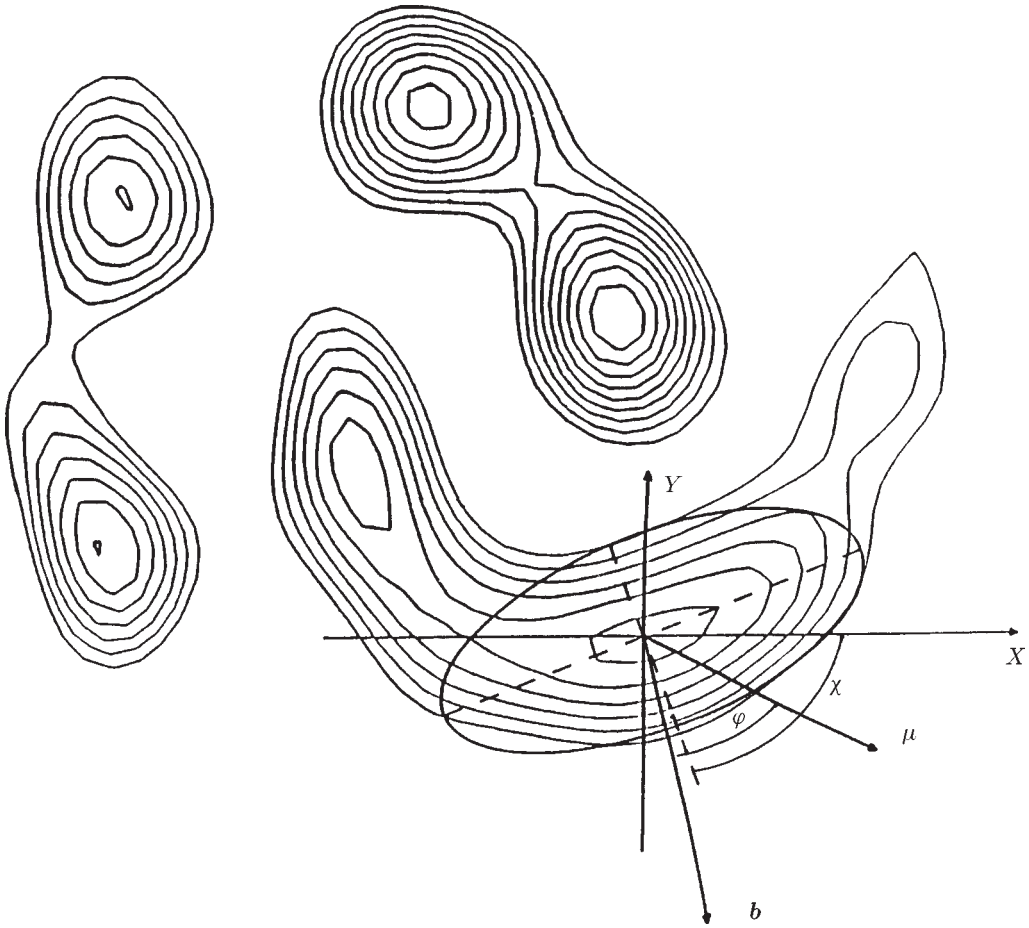

Figure 2. Illustration of ellipse inscription carried out on top of each of the seven helices. ***b*** indicates the direction of the bisector of the angle defined between the centers of the previous, the present and the following helix centers, μ is the hydrophobic moment, φ is the angle between μ and ***b***, and χ is the angle between the ellipse minor axis and the X-axis.

helical wheel projection models. Furthermore, whereas the seven helices of rhodopsin can be characterized from the electron density map, its sense of orientation (clockwise or counterclockwise) remains unclear. However, the counterclockwise orientation has recently become much more probable, as suggested by the results a set of mutation experiments [3]. In any case, the algorithm presented in the present work can handle any of the two models.

Ideally, if the bundle was an arrangement of seven regular antiparallel helices, the projection map should be expected to show seven circles. Instead, as shown in figure 2, an ellipse can be more appropriately defined on top of each helix. In the present modeling approach, we assumed that ellipse eccentricity is only due to the tilt θ of the helix. Inspection of figure 3 suggests that, tilts can be easily computed.

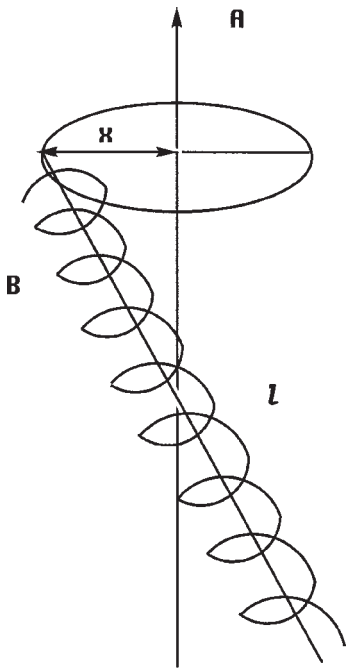

Figure 3. Illustration of the procedure followed to measure the tilt of each helix. x is the value of the ellipse major semi-axis and l is the length of the helix.

Table 1
Bundle parameters deduced from the 2D electron density map of rhodopsin assuming the coordinate origin placed on the center of helix VI.

Helix #	Helix center (x, y)/Å	Tilt angle θ/degrees	Tilt orientation χ/degrees	Bisector vectors (x, y)/Å
Helix I	9.894, −15.714	−9.49°	−59.74°	−9.435, −0.838
Helix II	0.582, −20.564	16.35°	75.96°	2.02, −11.366
Helix III	8.148, −12.998	22.67°	0.0°	5.906, −8.922
Helix IV	17.266, −9.312	15.72°	0.0°	9.246, −3.353
Helix V	11.252, 1.940	14.62°	−15.94°	5.602, 11.462
Helix VI	0.0, 0.0	13.86°	33.69°	−7.936, 8.209
Helix VII	2.134, −10.282	14.39°	5.71°	−8.77, 5.775

Let A be the direction of the Z-axis and B the direction of the helical axis reflecting the tilt of the helix. The ratio between the value of the ellipse principal axis x and a half of the helix length l provides the angle of tilt θ:

$$\theta = \arcsin\left(\frac{2x}{l}\right),$$

where the sign of θ is decided upon considerations of the lengths of the loops joining two consecutive helices. Moreover, the direction of the ellipse minor axis defines the

rotation axis required to provide the proper tilt to each one of the helices. This direction is characterized by the angle χ between this direction and the X-axis (figure 2). Finally, the points where the two ellipse axis cross, determine the coordinates of the helix centers, assumed to lie in the plane $z = 0$. All these parameters are listed in table 1.

3. Definition of a helix local coordinate system

The first step to be performed to build a GPCR is to assess accurately the sequence of each of the seven transmembrane domains. For this purpose we make use of the program PHDhtm [26], a neural network system capable of locating the transmembrane helices in integral membrane proteins with 95% accuracy. Its predictions are based on multiple sequence alignment of several GPCRs and have been shown to be very reliable.

Dihedral angles of a regular right-handed alpha helix ($\varphi = -57^\circ$, $\psi = -47^\circ$) are assigned to each of the putative transmembrane polypeptide segments, with exception of prolines where $\varphi = -65^\circ$. For convenience, side chains were kept in the extended conformation. Subsequently, a local coordinate system is defined. First, the mean position of the helix is calculated from the cartesian coordinates of all the helix atoms:

$$\bar{x} = \frac{1}{N}\sum_i x_i, \qquad \bar{y} = \frac{1}{N}\sum_i y_i, \qquad \bar{z} = \frac{1}{N}\sum_i z_i,$$

where N is the number of atoms in the helix and x_i, y_i and z_i are the coordinates of atom i. Mean position coordinates are subsequently subtrated from those of every atom, resulting in a translation of the helix to a new coordinate system where the mean position is the new origin of coordinates.

In order to define a helical coordinate system a variance matrix is defined as follows:

$$\Delta = \begin{bmatrix} \Delta_{xx} & \Delta_{xy} & \Delta_{xz} \\ \Delta_{yx} & \Delta_{yy} & \Delta_{yz} \\ \Delta_{zx} & \Delta_{zy} & \Delta_{zz} \end{bmatrix}, \quad \text{where } \Delta_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}).$$

This matrix provides an estimation of the dispersion of the atomic coordinates with respect to the mean point computed above. Diagonalization of this matrix yields three orthogonal eigenvectors that once normalized, are used to define the new local coordinate system. These vectors also define the rotation matrix necessary to transform the atomic coordinates to the new local coordinate system.

4. Helix manipulations

GPCRs consist of a single polypeptide chain containing seven hydrophobic segments looping back and forth across the lipid bilayer. This determines that the transmembrane domains must be organized in a bundle of antiparallel helices with the

exception of the first and the seventh, arranged to form a hydrophobic outer surface, facing the lipid membrane and a hydrophilic inner interior that originates a binding pocket. Consequently, the first transformation to be performed on the individual helices consist of a 180° turn on helices #2, 4 and 6 to get all the helices in an antiparallel orientation.

Once helices are properly oriented, next the hydrophobic moment μ of each helix is determined. The hydrophobic moment of a helix is computed as the summation of unitary vectors, defined on each residue as the difference of the side chain mean point coordinates and those of the Cα, and multiplied by the hydrophobic value of the residue as a weighting factor [9]:

$$\mu_x = \sum_i \xi_i(x_{is} - x_{i\alpha}), \qquad \mu_y = \sum_i \xi_i(y_{is} - y_{i\alpha}), \qquad \mu_z = \sum_i \xi_i(z_{is} - z_{i\alpha}),$$

where ξ_i are the hydrophobicity value of residue i in the Kyte and Doolittle's scale [18] and x_{is}, y_{is}, z_{is} are the coordinates of the mean point of the side chain of residue i.

The hydrophobic moment defines the most hydrophobic side of the helix and consequently the one facing the lipid bilayer. Inspection of the projection map suggests that not all the helices have the same exposure to the lipid environment, being helix #3 the least exposed. This different exposure can be quantitatively assessed by the value of the angle formed between the centers of three consecutive helices. The larger the angle, the more exposed is the helix to the lipid phase. In order to use the hydrophobic moment to indicate the orientation of the helix, the simplest assumption is to force it to point in the direction of the bisector of the angle formed between the centers of the previous helix, the one under consideration and the following one deduced from the electron density map. Bisector coordinates are listed in table 1. Let φ be the angle between the hydrophobic moment μ and the direction of the bisector $\boldsymbol{b}$ (figure 2). In order to make both vectors to point in the same direction, a rotation R of the angle φ around the helical axis should to be performed to each helix. However, the need to apply in a subsequent operation a tilt to each helix, requires to rotate the helix an angle $\varphi + \chi$, being χ is the angle between the ellipse minor axis and the X-coordinate axis, that will be later subtracted:

$$R = \begin{bmatrix} \cos(\varphi+\chi) & -\sin(\varphi+\chi) & 0 \\ \sin(\varphi+\chi) & \cos(\varphi+\chi) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Once helices are transformed, an additional rotation through the minor axis of the ellipse defined on top of each helix, needs to be performed to provide the proper tilt. This operation requires the simultaneous rotation of the tilt angle θ around an axis located on the XY-plane and undo the previous rotation of $-\chi$ around the Z-axis:

$$R = \begin{bmatrix} \cos(-\chi) & -\sin(-\chi) & 0 \\ -\cos\theta\sin(-\chi) & \cos\theta\cos(-\chi) & -\sin\theta \\ \sin\theta\sin(-\chi) & \sin\theta\cos(-\chi) & \cos\theta \end{bmatrix}.$$

After this is performed, the final step consist of translating the centers of the helices according to the coordinates listed in table 1.

5. Results and discussion

The procedure has successfully been used to generate atomic models of rhodopsin and other GPCRs like the opiate or bradykinin receptors. In all of the cases tested, the procedure provides the residues known to influence the binding affinity of selective ligands facing to the interior of the bundle. Furthermore, docking of these ligands following directives of their proposed pharmacophores permit identification of residues involved in ligand–receptor interactions.

Once the models are constructed, they are subsequently refined by energy minimization followed by a 300 ps molecular dynamics simulation in order to relax side chain interactions between helices and with the ligand. Results of applying this procedure to different systems will be published elsewhere.

The procedure described above has the great advantage of being totally general and can be used to construct a model of any GPCR, avoiding manual manipulation and does not make use of sequence identities between different receptors, necessary for both homology modeling and sequence divergence analysis. These models may allow to make predictions about possible mutations that may importantly influence the binding of different ligands to their receptors as well as modifications on the ligands to make the receptor–ligand interaction stronger. Accuracy of the models constructed can be easily contrasted by comparing their predictions with the results of a large number of pharmacophores described in the literature, deduced from comparison of the molecular properties of series of ligands. Modeling the receptors represents an alternative view of the ligand–receptor interactions and may provide new insights into the design of new ligands based on the consistency of the results achieved.

6. Conclusions

The paper describes a general automated procedure to build the transmembrane regions of GPCRs that does not require any protein as template. The method is not hampered by the need of finding several receptors with high sequence homology. The procedure is based on the assumption that all GPCRs adopt the same helix arrangement as observed in the rhodopsin projection electron density map obtained from electron microscopy of 2D crystals. From this map, location of the seven helix centers as well as their tilts from the elliptical shape exhibited by each helix in the projection map, can be measured. Next step consist of assessing the sequence of the transmembrane regions by means of the profile fed neural network system from the EMBL and construct seven ideal alpha helices of these sequences. Next, helices are manipulated according to the following steps: (i) definition of a local coordinate axis for each of the helices; (ii) orient them in an antiparallel fashion; (iii) rotation of each helix through the helical

axis in such a way that its hydrophobic moment is pointing in the same direction of the bisector formed between three consecutive helices in the bundle; (iv) rotation of each helix through an axis perpendicular to the helical one to assign a proper tilt; (v) translation of each of the helix to its center deduced from the projection map.

Acknowledgements

The authors wish to express their gratitude to both the Italian and Spanish governments for a collaborative research grant HI-1995-0091B between the two laboratories.

References

[1] I. Alkorta and P. Du, Protein Eng. 7 (1994) 1231–1238.
[2] J.M. Baldwin, EMBO J. 12 (1993) 1693–1703.
[3] A.G. Beck-Sickinger, Drug Discov. Today 1 (1996) 502–513.
[4] P.G. de Benedetti, M.C. Menziani, F. Fanelli and M. Cocchi, J. Mol. Struct. (Theochem) 285 (1993) 147–153.
[5] M.R. Brann, V.J. Klimkowski and J. Ellis, Life Sci. 52 (1993) 405–412.
[6] P. Cronet, C. Sander and G. Vriend, Protein Eng. 6 (1993) 59–64.
[7] S.G. Dahl, O. Edvardsen and I. Sylte, Proc. Natl. Acad. Sci. USA 88 (1991) 8111–8115.
[8] D. Donnelly and J.B.C. Findley, Curr. Opin. Struct. Biol. 4 (1994) 582–589.
[9] D. Eisenberg, R.M. Weiss and T.C. Terwilliger, Nature 299 (1982) 371–374.
[10] F. Fanelli, M.C. Menziani, M. Cocchi and P.G. de Benedetti, J. Mol. Struct. (Theochem) 333 (1995) 49–69.
[11] F. Fanelli, M.C. Menziani, A. Cocchi, A. Leonardi and P.G. de Benedetti, J. Mol. Struct. (Theochem) 314 (1994) 265–276.
[12] N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin and R. Henderson, J. Mol. Biol. 259 (1996) 393–421.
[13] J. Grotzinger, M. Engels, E. Jacoby, A. Wollmer and W. Strabburger, Protein Eng. 4 (1991) 767–771.
[14] P. Herzyk and R.E. Hubbard, Biophys. J. 69 (1995) 2419–2442.
[15] M.F. Hibert, S. Trump-Kallmeyer, A. Bruinvels and J. Hoflack, Recept. Mol. Pharmacol. 40 (1991) 8–15.
[16] C. Hutchins, Endocrine J. 2 (1994) 7–23.
[17] D.J. Kyle, S. Chakravarty, J.A. Sinsko and T.M. Stormann, J. Med. Chem. 37 (1994) 1347–1354.
[18] J. Kyte and R.F. Doolittle, J. Mol. Biol. 157 (1982) 105–132.
[19] L.J. Laakkonen, F. Guarnieri, J.H. Perlman, M.C. Gershengorn and R. Osman, Biochemistry 35 (1996) 7651–7663.
[20] C.D. Livingstone, P.G. Strange and L. H. Naylor, Biochem. J. 28 (1992) 7277–7282.
[21] M.C. Menziani, M. Cocchi, F. Fanelli and P.G. de Benedetti, J. Mol. Struct. (Theochem) 333 (1995) 243–248.
[22] M. Neumuller and F. Jahnig, Proteins Struct. Funct. Genet. 26 (1996) 146–156.
[23] G. Nordvall and U. Hacksell, J. Med. Chem. 36 (1993) 967–976.
[24] L. Pardo, J.A. Ballesteros, R. Osman and H. Weinstein, Proc. Natl. Acad. Sci. USA 89 (1992) 4009–4012.
[25] J.H. Perlman, L.J. Laakkonen, F. Guarnieri, R. Osman and M.C. Gershengorn, Biochemistry 35 (1996) 7643–7650.
[26] B. Rost, R. Casadio, P. Fariselli and C. Sander, Protein Sci. 4 (1995) 521–533.

[27] M.S.P. Sansom, H.S. Son, R. Sankararamakrishnan, I.D. Kerr and J. Breed, Biophys. J. 68 (1995) 1295–1310.
[28] G.F.X. Schertler and P.A. Hargrave, Proc. Natl. Acad. Sci. USA 92 (1995) 11 578–11 582.
[29] G.F.X. Schertler, C. Villa and R. Henderson, Nature 362 (1993) 770–772.
[30] C.D. Strader, T.M. Fong, M.R. Tota, D. Underwood and R.A.F. Dixon, Ann. Rev. Biochem. 63 (1994) 101–132.
[31] I. Sylte, O. Edvardsen and S.G. Dahl, Protein Eng. 6 (1993) 691–700.
[32] S. Trump-Kallmeyer, J. Hoflack, A. Bruinvels and M.F. Hibert, J. Med. Chem. 35 (1992) 3448–3462.
[33] S. Watson and S. Arkinstall, in: *Facts Book* (Academic Press, London, 1994).
[34] Y. Yamamoto, K. Kamiya and S. Terao, J. Med. Chem. 36 (1993) 820–825.
[35] Y. Yamano, K. Ohyama, M. Kikyo, T. Sano, Y. Nakagomi, Y. Inone, N. Nakamura, I. Morishima, D.F. Guo, T. Hamakubo and T. Inagami, J. Biol. Chem. 270 (1995) 14 024–14 030.
[36] D.Q. Zhang and H. Weinstein, J. Med. Chem. 36 (1993) 934–938.