Modelling G protein-coupled receptors

L. Oliveira[1], T. Hulsen[2], D. Lutje Hulsik[2], A.C.M. Paiva[1], G. Vriend[2]

[1]Escola Paulista de Medicina, Sao Paulo, Brazil
[2]CMBI, KUN, The Netherlands

Running title: Modelling GPCRs

Correspondence to:
Gerrit Vriend
CMBI KUN
Toernooiveld 1
6525 ED Nijmegen
The Netherlands
Tel +31-24-3653391
Fax +31-24-3652977
Email Vriend@CMBI.KUN.NL

INTRODUCTION

G-protein-coupled receptors (GPCRs) are a superfamily of integral membrane proteins. They have seven transmembrane helices (I-VII), three extracellular loops (II-III, IV-V and VI-VII), three cytosolic loops (I-II, III-IV and V-VI), a extracellular N-terminal domain (N), and a cytosolic C-terminal (C) domain (Fig. 1)[1]. The C terminal domain contains helix VIII, which starts directly after helix VII and runs parallel to the cytosolic membrane surface. GPCRs are usually divided into six major classes called A-F[2]. Class A is by far the largest and contains, for example, opsins and amine, olfactory, bradykinin, angiotensin II, opioid, somatostain, chemokine, and interleukine receptors[3]. Class B contains secretin-like, class C metabotropic glutamate, class D pheromone, and class E cAMP receptors. Class F contains other, often hard to categorize receptors[2], such as, for example, the frizzled receptor.

About 2000 GPCR sequences are currently available from the public databases. About a dozen new sequences become available every month. A large number of on-going genome sequencing projects guarantees that the monthly database influx of GPCR sequences will remain high in the next few years. The GPCR mutation database contains data on approximately

10000 mutations[4]. Ligand-binding information is available for more than 30000 ligand–receptor combinations[2].

This wealth of sequences, ligands, and mutation data is in marked contrast to the small amount of structural information. currently, only the structure of bovine opsin is known[1]. This lack of structural information leads to many modelling studies. In the past, bacteriorhodopsin[e.g.5-8] was often used as a modelling template, but recently the bovine opsin three-dimensional coordinates have become available. Bovine opsin is a much better template for GPCR homology modelling than bacteriorhodopsin. However, bovine opsin is not yet the perfect template for every GPCR modelling project, and some of the problems with this template will be discussed. Models produced Before the Crystal structure became available in August 2000 will be referred to as BC-models, and the newer models will be called AC-models (AC for After Crystal structure). Both BC-models and AC-models will be discussed. The structure of bovine opsin will be discussed from a modellers point of view to provide a set of hints for future modelling studies of family A GPCRs (AGPCRs).

GPCR modelling before the bovine opsin structure

Most BC-models were based either on the coordinates of bacteriorhodopsin[e.g.5-8], or on the Cα coordinates of bovine opsin extracted from a low-resolution electron diffraction map[8-10] by Baldwin[11]. The former had the problem that the helices in bacteriorhodopsin are packed differently from those in GPCRs[10,12] and the latter had the problem that the electron diffraction density map was not made public so one had to rely on Baldwin's interpretation[11]. Today we know that her interpretation was very good, but at the time, many modellers were worried about this fact. A few models were based on first principles, sometimes guided by a 6 Å resolution electron diffraction map of bovine and frog opsin[9,10,13].

Many different methods were employed to guess the ends of the helices, but the idea that a helix could continue beyond the membrane region was generally missed. Not many models had helix kinks incorporated, and π–helices and other large deviations from ideality that can occur in the middle of helices (as in bovine opsin) never were considered.

In the late 80's and the 90's, the GPCR modelling community developed a series of rules for building BC-models:

1. Determine which template to use, or design your own helix-packing model[14-16];

2. Use threading or moment calculations to determine the mapping of the GPCR sequence onto the selected template. Moment calculations can be based on hydrophobic moments[17], conservation moments[18], etc., or a combination of those[19]. Threading can be based on general rules, helix bundle rules[20,21], or even bacteriorhodopsin-specific rules[22];
3. Find experimental data that agree with the model and add it to convince the referees that this is the only correct model.

The release of the high-resolution coordinates of bovine opsin[1] is opening up a new world of possibilities. These coordinates provide a template for classical homology modelling. Consequently, all the old models and rules for modelling can be thrown away, and we can make a fresh start.

Global aspects of the bovine opsin structure

The high-resolution structure of rhodopsin reveals a seven-helix transmembrane (7TM) bundle surrounding an ellipsoidal central cavity lined by helices I-III and V-VII (see figure 1). Helix IV is not part of the cavity wall and only makes contacts with helix III. The wide side of the ellipsoidal cavity is flanked by the pairs of helices II-III and VI-VII. The VI-VII pair is perpendicular to the lipid membrane, while the II-III pair is inclined about 30° from the normal. The two narrow sides of the cavity are flanked by the

helices I and V, respectively. Helix I has contacts along its full length with the helices II and VII. At the cytosolic side, helix V touches helices III and VI, and on the extracellular side helix V has contacts mainly with helix VI.

The central cavity is accessible from the cytosol, but the hairpin that connects helices IV and V prevents access from the periplasm. This hairpin (IV-V) lies between the helices, roughly parallel to the membrane surface. It has contacts with sidechains of most of the helices. The most prominent contact is a disulphide bridge (Cys315-Cys480) between this hairpin and helix III.

METHODS

Data that are too numerous to be put in print, and hypotheses too speculative to be published can be found at **http://www.gpcr.org/articles/2002_4/**[12]. This website also contains a detailed recipe for building models along the lines described in this article.

A standardized numbering system

Any discussion about more than one molecule at the same time suffers from the most irritating problem facing bioinformatics: residue numbering. We have introduced a standardized residue-numbering scheme for AGPCR transmembrane helices[23]. This numbering scheme was adopted at a GPCR meeting at the EMBL in 1993. It has guided the GPCRDB development ever since, as well as the articles from a (small) number of laboratories. This numbering scheme was designed to indicate clearly residues that are conserved among all AGPCRs so originally only the helical residues needed to be numbered. Now that it is clear that features like the IV-V hairpin and helix VIII are conserved among many AGPCRs, the numbering scheme needs to be extended[2].

The numbering scheme is based on the idea that the hundreds' of the residue numbers correspond to the helix in which the residues reside, and the most conserved residue in each helix is rounded to a near multiple of ten. For example, the Asn of the highly conserved GN motif in helix I is numbered 130; the conserved L from the LXXXD motif in helix II is 220; the ultra conserved R from the DRY motif in helix III is 340; etc (see fig 3, and http://www.gpcr.org/).

Bovine opsin and bacteriorhodopsin structure comparison

The structures of bovine opsin (PDBid=1F88)[1], and bacteriorhodopsin[e.g.5-8] are very different. Both contain a sevenhelix bundle, but these bundles are sufficiently differently organized to make any detailed structure comparison meaningless[10-12]. However, in order to evaluate the quality of models based on the bacteriorhodopsin template, this superposition must be made. We therefore did this structure superposition manually. The results are shown in figure 2. Alternative solutions in which bacteriorhodopsin was shifted by one helical turn relative to bovine opsin were not much worse according to our visual inspection (see figure 2b).

The recipe for determining the quality of bacteriorhodopsin based BC-models is as follows:

1. Extract from the GPCRDB the alignment of the sequence of the GPCR model with the sequence of bovine opsin;
2. Use the structure superposition to align the bovine opsin sequence onto the bacteriorhodopsin sequence.
3. Extract from the modelling article how the authors aligned their GPCR with bacteriorhodopsin. (If this alignment is not given, it can be extracted from a superposition of the bacteriorhodopsin based

GPCR model on the real bacteriorhodopsin structure). This produces the alignment used for the modelling.

4. A comparison of the 'optimal' alignment with the alignment that has been used gives a good indication of the model quality.

This recipe for the determination model quality differs somewhat from what is normally used. The two reasons for this are:

1. Only the structure of bovine opsin is known, and not the structures of all the BC-models. Therefore we cannot calculate RMS deviations etc., as is done, for example, in the CASP modelling competition[24].
2. RMS deviations do not make sense if the alignment on average is out by about four residues.

MODELLING RESULTS

The quality of BC-models

The bacteriorhodopsin and bovine opsin structures are so different that any quantitative comparison is meaningless. Alternative structure superposition

solutions shifted by one helical turn imply a shift by 3 or 4 residue positions. The superposition solution that we selected was the one that had a reasonable overlap between the two retinal molecules. A shift in the structure superposition would lead to a the shift in the sequence alignment that is alternating positive and negative in the seven helices. As can be seen from the alignments in figure 3, such a shift would not make any of the alignments better. Therefore, the subjective nature of the superposition is not influencing our conclusions. We believe that all GPCR models (including our own models) that were based on the bacteriorhodopsin template were very bad, and none of these models can possibly have made a positive contribution to rational drug design projects. A more extensive reality check of BC-models can be found in the article section of the GPCRDB (http://www.gpcr.org/articles/).

The models based on Cα coordinates[11] extracted from the low-resolution electron diffraction map[8-10] were reasonably good, though none of them were really good enough for rational drug design purposes. Figure 4 shows the superposition of the bovine opsin structure and one BC-model built using the Cα coordinates extracted from the low-resolution electron diffraction map. It can be seen that the overall features are modelled reasonably well. The Cα and all-atom modelling errors (i.e. displacements

between the model and the X-ray structure) are 2.5 and 3.2 Å, respectively. The models that were based on the bacteriorhodopsin template, on the other hand, never bore any relation with reality. Figure 3 shows a series of alignments produced by the recipe described in the methods section. We list only models that we produced (and deposited) ourselves, or that we helped to build. This is partly because these models were among the better amongst the bacteriorhodopsin-based BC-models, but also, it is inappropriate to criticise results from a decade ago. However, the fact that models that had residues misaligned by as many as 21 positions could be published, holds a warning for the future.

No BC-model had the IV-V hairpin modelled correctly between the helices. All modellers 'knew' that the loop IV-V was external at the extracellular side, but also were aware of the Cys315-Cys480 disulphide bridge. Bizzare reasoning was often used to reconsile these two contradicting 'facts' and to justify the position of helix III. This led to errors as large as a 21-residue shift in the sequence alignment. The experimental data enabling the correct prediction that the IV-V hairpin is located between the helices was available to the BC-modellers, because it was known that His474 and Lys477 in this hairpin in opsins form a chloride binding site that regulates which wavelength can be absorbed optimally by the retinal[25]. It

could have been reasoned that if this site modifies the wavelength, it should be located near the retinal. Unfortunately, the 'knowledge' that the loops should stick out into the extracellular solvent 'beat' the knowledge about the chloride site, and no IV-V hairpin has ever been modelled correctly.

The quality of AC-models

We were surprised to find several modelling studies performed after the release of the bovine opsin three-dimensional coordinates in which very little knowledge about this template was incorporated. Ballesteros *et al.*[26] recently wrote that amine receptors can be modelled from the bovine opsin template. They neglect the IV-V hairpin, crystal contacts, and the fact that many residues cannot be detected in the X-ray structure. Orry *et al.*[27] docked endothelin in the endothelin receptor. The article was published in the AC period. Their endothelin receptor model was based on a rhodopsin model by Pogozheva[21]. They write in a note added after submission that since the paper was first submitted the bovine opsin structure became available, and they claim that their model and the bovine opsin structure are similar. Their model is not published, but from the figures in the article, it can be seen that the endothelin molecule is docked roughly around the space where one would expect the IV-V hairpin, and this hairpin is modelled as a hyper-

exposed loop. These are just two examples of neglect of details of the bovine opsin structure. It is to be hoped that future GPCR modelling experiments can use some of the information provided in this article to advantage.

GPCR MODELLING USING THE BOVINE OPSIN STRUCTURE

The availability of the bovine opsin structure opens new alleys for modelling GPCRs. However, some warnings are needed. First, the structure of the extra-membrane loops in the bovine opsin structure seem to large degree to be determined by crystal contacts. Second, the bovine opsin structure is the inactive form of the protein, while the active form is a much more appropriate modelling goal for pharmaceutical purposes. Third, it is far from certain if the bovine opsin structure can be used as a template for all AGPCRs.

Modelling studies start with a sequence alignment between the bovine opsin template and the AGPCR model sequence. The percentage sequence identity between bovine opsin and many other AGPCRs can be as low as 20%. Normally, when the sequence identity between the model and the template falls below 30%, the sequence alignment is the main bottleneck in

the homology modelling procedure. AGPCRs form an exception to this rule. As was already hinted at in the methods section, each helix contains one or two highly conserved residues that allow for an unambiguous alignment of the helices of the model and the bovine opsin template in the majority of cases.

With a few exceptions (discussed below), none of the extramembrane regions can be modelled by homology using the bovine opsin structure as a template. This is for two reasons. First, most cytosolic extramembrane parts could not be seen in the electron density map, and most extracellular parts most likely have a structure induced by crystal packing forces (see figure 5). Second, the sequence identity between most AGPCRs and bovine opsin is too low to arrive at any reliable alignment. At three extramembrane locations, features can be observed that give hope that some modelling will be possible. These are the highly conserved:

1. Trp280 and Gly295 in the extracellular loop SII-III;
2. The loop IV-V and the Cys315-Cys480 disulphide bridge;
3. Tyr734 at the bend between the helices VII and VIII and the adjacent sequence motif Phe800, Arg(Lys)801 in helix VIII;

Trp280 and Gly295 in the extracellular loop II-III

Loop II-III contains the residues Trp280 and Gly295, which are conserved in the majority of the known AGPCRs (amine, peptide, prostanoid, etc., receptors). In rhodopsin, this segment forms a surface-accessible type I β-turn. The conservation and location of this turn makes it a good candidate for a recognition site for some external factor. In some AGPCR classes, very long insertions can be observed in this region, indicating that this motif is not required for one of the main functions of AGPCRs. Modelling these insertions is obviously not possible, but their external location ensures that this is not likely to influence the rest of the model.

The IV-V hairpin

The extracellular IV-V hairpin in bovine opsin is a crossed hairpin which makes sidechain interactions around the inner surface of the 7TM bundle. The most prominent contact is the disulfide bridge Cys315-Cys480. It looks as if this hairpin forms a shield separating the ligand binding site from the extracellular medium. Its presence leaves us with the question of how the ligand gets from the external medium into the binding pocket. We think that this can happen in only one of two ways:

1) Through the space between the helices VII, I and II, and the beta turn 474-478 in the IV-V hairpin, through a space flanked by the 472-475 strand, the part of the IV-V hairpin directly after Cys480, and the helices VI-VII;
2) Motion of the IV-V hairpin leaves the entry to the binding cavity open long enough for a ligand to enter.

The majority of the AGPCR sequences have a loop between the helices IV and V that has a reasonable length for forming a IV-V hairpin (see table in the web material), and nearly all of these IV-V loops have a cysteine at position 480. It therefore seems safe to assume that the vast majority of all AGPCRs have this hairpin structure between the ligand binding pocket and the external surface. Several other AGPCRs (e.g. melanocortin, cannabinoid, edg, mas) lack the Cys315 and Cys480 residues and are therefore unlikely to have a similar IV-V hairpin to bovine opsin. Some of these receptors have very short IV-V loops and might even have the ligand binding site fully exposed to the periplasm. Such a wide-open ligand binding site would allow large ligands to enter. Examples are the ACTHR, MSHR, and mc3-5r receptors. Mas, cannabinoid and edg receptors have intermediate length IV-V loops that are not long enough to form the equivalent of the bovine opsin IV-V hairpin, but are long enough to obstruct part of the ligand entry path.

Modelling the IV-V hairpin should start by aligning the sequences, using as reference the Cys480 while minding the following points:

- The conformation of the hairpin between positions 472 and 483 and the disulphide bridge Cys315-Cys480 should be as in bovine opsin;
- Insertions or deletions should be made adjacent to helices IV and V because around there is 'plenty of room' in the bovine opsin structure.

Modelling the IV-V hairpin will only be difficult for a few receptors. Adenosine 1 and dopamine 1a receptors have insertions just before Cys480, fmlp receptors have insertions just after Cys480, and odorant receptors have insertions at both those locations. Some histamine, alpha-adrenergic, adenosine 2,3, and melatonin receptors lack residues after the Cys480 (see the table in the web material). In a few receptors, deletions in the IV-V area might even require some minor modifications in the last turn of helix IV or in the first turn of helix V.

The IV-V hairpin sequence is highly variable among the AGPCR sub-families. It is therefore unlikely that any function can be found in this area that is common to many receptors. Many subclasses of receptors, however, have developed some function or another in this area. A good example is the chloride binding site in opsins, which supposedly regulates which wavelength can optimally be absorbed by the retinal[28]. In red and green but

not in blue and violet opsins, this site is formed by positions 474 and 477, occupied by residues of His and Lys, respectively[25].

A PROSITE[29] search reveals many potential glycosylation sites in AGPCRs. Remarkably, several receptors (e.g. angiotensin II and bradykinin B1 and B2) have a glycosylation site at Asn476 in the IV-V loop. Mutation of the Asn in each glycosylation site reveals a slowly increasing effect on the ability of the receptor to insert into the membrane. Mutation of Asn476 contributed more strongly to this effect than any other mutation[30]. When these experiments were done, it went unremarked, because the loop IV-V was expected to protrude into the periplasm. Now that the bovine opsin structure is known, these experiments take on a new significance because a glycosyl group would appear to cause steric hindrance. We do not understand fully the structural aspects of this glycosylation, but it seems clear that any modeller will have to find out whether or not his/her model GPCR is glycosylated before anything can be done with the IV-V loop.

C-terminus

The loops I-II, III-IV, and V-VI are not well resolved in the bovine opsin structure. So, these loops cannot be used as modelling templates. The cytosolic helix VIII seems to be common to many AGPCRs. The precise

position of this helix in the bovine opsin structure is still unclear because of crystal packing effects, but we believe that this helix and single residue between the helices VII and VIII are a valid template for modelling. It is worth noting that we predicted the presence and location of helix VIII prior to the release of the bovine opsin three-dimensional coordinates[31].

In 97% of all AGPCRs, a tyrosine is observed at position 734[32], which is very close to the sharp turn between the helices VII and VIII. This Tyr734 has a crucial role in the receptor G-protein interaction process[33]. It is part of the conserved NPXXY motif on the cytosolic side of helix VII. A few receptor subfamilies (e.g. endothelin) do not have the NPXXY motif but NPXXXY. This is why the tyrosine that normally sits 3 positions after the highly conserved Pro730 is numbered Tyr734 and not Tyr733. If the phase of helix VII stays the same in these NPXXXY receptors, we have to assume either that the tyrosine points in a different direction, or the cytosolic end of helix VII is very irregular in these receptors.

THE ACTIVE FORM OF AGPCRs

Modelling the active form of AGPCRs depends critically on the hypothesized mechanism of that activation process. We therefore start with a summary of possible activation mechanisms. All activation models have to consist of essentially the same three steps:

1. Admission of the ligand into the ligand binding pocket;
2. The receptor moving from the inactive state into the active state, or the active state being frozen by the ligand;
3. The G protein being activated.

Admission of the agonist into the receptor central cavity

Studies on various aspects of an AGPCR multiple sequence alignment show that the space occupied by retinal in the bovine opsin structure is the same as where all or part of the endogenous agonists (and probably also non-endogenous agonists and antagonists) bind in nearly all AGPCRs[34]. From this, we conclude that the activation mechanism must also be essentially the same for nearly all AGPCRs. However, if the retinal binding site corresponds to the agonist binding site in AGPCRs, and the loop IV-V is structurally conserved among most AGPCRs, how can the ligand enter the agonist-binding pocket? Two simple mechanisms can be imagined. First, it is conceivable that loop IV-V is involved in a flip-flop mechanism that

occasionally leaves enough space for ligands to enter into the binding pocket. Second, it is also possible that loop IV-V does not move and that ligands have to squeeze into the pocket through small spaces left either between loop IV-V and helices IV and V, or between the same loop and helices I, II and VI.

Expansion of the central cavity as a mechanism for activation

All AGPCRs share a set of conserved residues at the G-protein coupling interface. These residues are so highly conserved that they must form the major active site[35], and the role of agonists must be to generate a signal that reaches this site[32]. As it is very likely that all AGPCRs have a common G-protein activation mechanism, it is also likely that the activation by the agonists uses a similar mechanism, and it is even likely that this mechanism is similar to the opsin activation by light-induced retinal isomerization. For rhodopsin, it has been found that the light-induced cis-11 to all-trans isomerization pushes the retinal towards helix III leading to a rearrangement in the 7TM bundle that activates the GPCR. In opsins, activation is associated with the formation of meta-rhodopsin MII, a form of rhodopsin in which Glu224 and Glu329 have elevated pKa values[36]. The corollary in AGPCRs would be that an agonist molecule is squeezed inside the 7TM

central cavity, leading to a displacement of helices mimicking the retinal isomerization effects. Depending on which agonist entry model one believes in, two hypotheses can be considered for the activation process:

1. If the cavity is open because of retraction of loop IV-V, the agonist can enter the pocket freely and will be pushed into the cavity when the loop moves back to the closed position;
2. If the central cavity of the AGPCRs is closed by loop IV-V, then the equilibrium between the two receptor states should be considered. The open state (when the 7TM bundle is expanded and the entrance is enlarged), and the state in which the helix bundle is compressed and the agonist entrance is closed.

Both of these hypotheses lead to a similar reasoning about helix motions:

- Residue positions prevalently containing aromatic sidechains (Phe, Tyr and Trp) are found at the extracellular side of the 7TM bundle of all AGPCRs, pointing inwards as well as towards neighbouring helices[32];
- Aromatic sidechains in the agonist structures are crucial to trigger activation of angiotensin II-, bradykinin- and other receptors[37-42];
- Helices with aromatic sidechains form less well packed bundles[43-44].

NEW RULES FOR MODELLING AGPCRs

For most studies, it will be enough to model the seven transmembrane helices and the IV-V hairpin. If extracellular loops are needed in the model, there is some hope for a few receptors that these loop models can be based on the bovine opsin structure, but in most cases it will not be possible to model these loops. The works by Yeagle[45-48] made clear that determination of the structure of the loops independently from the rest of the structure is not likely to be a successful option either.

The alignment of the helices should be based on the highly conserved motifs that are present in all helices. Extrapolating from the performance of GPCR modellers over the years, we can only advise to sticking to the bovine opsin helix backbone coordinates. Any attempt to 'improve' this template for other GPCRs will undoubtedly make things worse rather than better.

The IV-V hairpin should be modelled using the bovine opsin as template. If this loop is not present in the model, it seems doubtful that a reliable model can be built at all.

The bovine opsin three-dimensional coordinates represent the inactive form of this receptor. To model the (pharmaceutically much more interesting) active form of AGPCRs one should not rely on molecular

dynamics, but rather on the outcome of experiments that can be interpreted unambiguously. The following recipe for modelling the active form of a receptor is just one possibility. It will not prove correct, but is the best we can do with current data. A much more detailed description can be found in the article section of the GPCRDB[2].

To arrive at a potential model for the activated form of the receptor, the helices VI and VII were rotated together in such a way that helix VII approaches helix III at the cytosolic side, and helix VI approaches helix V. This motion displaces helices VI and VII so that they are slightly further away from the other five helices than in the bovine opsin structure. As a consequence, a salt bridge found in the inactive bovine opsin structure between Arg340 and Glu600 is broken and a hydrogen bond could be formed between Arg340 and Tyr734. This model is in agreement with the sparse experimental data on the activation of opsins by cis-trans isomerisation of retinal, and it has frequently been referred to as a possible mechanism of activation for AGPCRs[49-51].

Modelling GPCRs in the AC era is easier than it was in the BC era. Nevertheless, it is not yet straightforward, and a thorough analysis of the bovine opsin structure and an alignment of all relevant AGPCR sequences will still be needed for each model. We have proposed a recipe for

modelling the active form AGPCRs from the inactive bovine opsin template structure. This proposal will suffer from many problems that the BC-models also suffered from: too few experimental data. Anyone interested in modelling the active form of an AGPCR will need to follow the literature on the activation process, and incorporate all available data in the model, and not just the data that agree with the model that is at hand at that moment!

ACKNOWLEDGEMENTS

The GPCRDB is an EC sponsored project (PL 950224). The authors thank Margot Beukers, Bob Bywater, Oyvind Edvardsen, Florence Horn, Rob Hooft, Wilma Kuipers, Chris Sander, Mike Singer, Nora vd Wenden, and Ad IJzerman for stimulating discussions. The authors thank David Thomas for critical reading of the manuscript.

REFERENCES

1. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 2000 289:739-745.
2. Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G. GPCRDB: an information system for G protein-coupled receptors. NAR 1998 26:275-279.

3. Watson S, Arkinstall S. The G-protein linked receptor FactsBook 1994 Eds Steve Watson and Steve Arkinstall Academic Press Harcourt Brace & Company, London, San Diego, New York, Boston, Sydney, Tokyo, Toronto.
4. Beukers MW, Kristiansen I, IJzerman AP, Edvardsen I. TinyGRAP database: a bioinformatics tool to mine G protein-coupled receptor mutant data. TiPS 1999 20:475-477.
5. Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipid cubic phases. Science 1997 277:1676-1681.
6. Luecke H, Richter HT, Lanyi JK. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. Science 1998 280:1934-1937.
7. Takeda K, Sato H, Hino T, Kono M, Fukuda K, Sakurai I, Okada T, Kouyama T. A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. J.Mol.Biol. 1998 283:463-474.
8. Henderson R, Schertler GFX. The structure of bacteriorhodopsin and its relevance to the visual opsins and other seven-helix G protein-coupled receptors. Philos.Trans.R.Soc.Lond.B.Biol.Sci. 1990 326:379-389.
9. Unger VM, Schertler GFX. Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. Biophys.J. 1995 68:1776-1786.
10. Schertler GF, Villa C, Henderson R. Projection structure of rhodopsin. Nature 1993 362:770-772. And Unger VM, Hargrave PA, Baldwin JM, Schertler GFX. Arrangement of rhodopsin transmembrane alpha-helices. Nature 1997 389:203-206.
11. Baldwin JM. The probable arrangement of the helices in G protein-coupled receptors. EMBO J. 1993 12:1693-1703.
12. http://www.gpcr.org/articles/2002_4/index.html
13. Schertler GFX, Hargrave PA. Projection structure of frog rhodopsin in two crystal forms. PNAS 1995 92:11578-11582.
14. Filizola M, Perez JJ, Carteni-Farina M. BUNDLE: A program for building the transmembrane domains of G protein-coupled receptors. J.Comp.-Aid.Mol.Des. 1998 12:111-118.
15. Prusis P, Schiöth HB, Muceniece R, Herzyk P, Afshar M, Hubbard RE, Wikberg JES. Modelling of the three-dimensional structure of the human melanocortin 1 receptor, using an automated method and docking of a rigid cyclic melanocyte-stimulating hormone core peptide. J.Mol.Graph.Mod. 1997 15:307-315.
16. Bramblett RD, Panu AM, Ballesteros JA, Reggio PH. Construction of a 3D model of the cannabinoid CB1 receptor: determination of helix ends and helix orientation. Life Sci. 1995 56:1971-1982.
17. Donnelly D, Overington JP, Ruffle SV, Nugent JH, Blundell TL. Modelling alpha-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. Prot.Sci. 1993 2:55-70.
18. Pardo L, Ballesteros JA, Osman R, Weinstein H. On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors. PNAS 1992 89:4009-4012.

19. Rick RP, Handschumacher MD, Sung SS, Tan M, Glynias MJ, Schluchter MD, Novotny J, Graham RM. Evolutionary conservation of both the hydrophilic and hydrophobic nature of transmembrane residues. J.Theor.Biol. 1995 172:245-258.
20. Herzyk P, Hubbard RE. Combined biophysical and biochemical information confirms arrangement of transmembrane helices visible from the three-dimensional map of frog rhodopsin. J.Mol.Biol. 1998 281:741-754.
21. Pogozheva ID, Lomize AL, Mosberg HI. The transmembrane 7-alpha-bundle of rhodopsin: distance geometry calculations with hydrogen bonding constraints. Biophys.J. 1997 72:1963-1985.
22. Cronet P, Sander C, Vriend G. Modelling of transmembrane seven helix bundles. Prot.Enging. 1993 6:59-64.
23. Oliveira L, Paiva ACM, Vriend G. A common motif in G protein-coupled seven transmembrane helix receptors. J.Comp.-Aid.Mol.Des. 1993 7:649-658.
24. Venclovas C, Zemla A, Fidelis K, Moult J. Comparison of performance in successive CASP Experiments. 2001 PROTEINS Suppl. 5:163-170.
25. Wang Z, Asenjo AB, Oprian DD. Identification of the Cl^--binding site in the human red and green colour vision pigments. Biochemistry 1993 32:2125-2130.
26. Ballesteros JA, Shi L, Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. Mol.Pharmacol. 2001 60:1-19.
27. Orry AJW, Wallace BA. Modelling and docking the endothelin G protein-coupled receptor. Biophys.J. 2000 79:3083:3094.
28. Sun H, Macke JP, Nathans J. Mechanisms of spectral tuning in the mouse green cone pigment. PNAS 1997 94:8860-8865.
29. Bucher P, Bairoch A. A generalized profile syntax for biomolecular sequence motifs and its function in automatic sequence interpretation. Proc.Int. Conf.Intell.Syst.Mol.Biol. 1994 2:53-61.
30. Lanctot PM, Leclerc PC, Escher E, Leduc R, Guillemette G. Role of N-glycosylation in the expression and functional properties of human AT1 receptor. Biochemistry. 1999 38:8621-8627.
31. Oliveira L, Paiva ACM, Vriend G. A low resolution model for the interaction of G protein with G protein-coupled receptors. Prot.Enging. 1999 12:1087-1095.
32. Oliveira L, Paiva PB, Paiva ACM, Vriend G. Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. PROTEINS *submitted.* See also http://www.gpcr.org/articles/2002_2/index.html
33. Rosendorff A, Ebersole BJ, Sealfon SC. Conserved helix 7 tyrosine functions as an activation relay in the serotonin 5HT(2C) receptor. Brain Res.Mol.Brain Res. 2000 84:90-96.
34. Oliveira L, Paiva ACM, Vriend G. Correlated mutation Analyses on Very Large Sequence Families. ChemBioChem *submitted.* See also: http://www.gpcr.org/articles/2002_3/index.html
35. Oliveira L, Paiva PB, Paiva ACM, Vriend G. From sequence to structure and function with entropy-variability plots. PROTEINS *submitted.* See also: http://www.gpcr.org/articles/2002_1/index.html
36. Arnis S, Fahmy K, Hofmann KP, Sakmar TP. A conserved carboxylic acid group mediates light-dependent proton uptake and signalling by rhodopsin. J.Biol.Chem. 1994 269:23879-23881.

37. Miura S, Feng YH, Husain A, Karnik SS. Role of aromaticity of agonist switches of angiotensin II in the activation of the AT1 receptor. J.Biol.Chem. 1999 274:7103-7110.
38. Hovelmann S, Hoffmann SH, Kuhne R, ter Laak T, Reilander H, Beckers T. Impact of aromatic residues within transmembrane helix 6 of the human gonadotropin-releasing hormone receptor upon agonist and antagonist binding. Biochemistry 2002 41:1129-1136.
39. Guerrini R, Calo G, Rizzi A, Bigoni R, Rizzi D, Regoli D, Salvadori S. Structure-activity relationships of nociceptin and related peptides: comparison with dynorphin A. Peptides 2000 21:923-933.
40. Shenderovich MD, Liao S, Qian X, Hruby VJ. A three-dimensional model of the delta-opioid pharmacophore: comparative molecular modelling of peptide and nonpeptide ligands. Biopolymers 2000 53:565-580.
41. Drouin JN, Gaudreau P, St-Pierre SA, Regoli D. Structure-activity studies of [des-Arg9]-bradykinin on the B1 receptor of the rabbit aorta. Can.J.Physiol.Pharmacol. 1979 57:562-566.
42. Nyronen T, Pihlavisto M, Peltonen JM, Hoffren AM, Varis M, Salminen T, Wurster S, Marjamaki A, Kanerva L, Katainen E, Laaksonen L, Savola JM, Scheinin M, Johnson MS. Molecular mechanism for agonist-promoted alpha(2A)-adrenoceptor activation by norepinephrine and epinephrine. Mol.Pharmacol. 2001 59:1343-1354.
43. Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ. Internal packing of helical membrane proteins. PNAS 2000 97:5796-5801.
44. Adamian L, Liang J. Helix-helix packing and interfacial pair-wise interactions of residues in membrane proteins. J.Mol.Biol. 2001 311:891-907.
45. Yeagle PL, Alderfer JL, Albert AD. Three-dimensional structure of the cytoplasmic face of the G protein receptor rhodopsin. Biochemistry 1997 36:9649-9654.
46. Yeagle PL, Alderfer JL, Salloum AC, Ali L, Albert AD. The first and second cytoplasmic loops of the G protein-receptor, rhodopsin, independently form beta-turns. Biochemistry 1997 36:3864-3869.
47. Yeagle PL, Alderfer JL, Albert AD. Structure of the third cytoplasmic loop of bovine rhodopsin. Biochemistry 1995 34:14621-14625.
48. Yeagle PL, Salloum A, Chopra A, Bhawsar N, Ali L, Kuzmanovski G, Alderfer JL, Albert AD. Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin. J.Pept.Res. 2000 55:455-465.
49. Shapiro DA, Kristiansen K, Weiner DM, Kroeze WK, Roth BL. Evidence for a model of agonist-induced activation of 5-hydroxytryptamine 2A serotonin receptors that involves the disruption of a strong ionic interaction between helices 3 and 6. J.Biol.Chem. 2002 277:11441-11449.
50. Huang P, Li J, Chen C, Visiers I, Weinstein H, Liu-Chen LY. Functional role of a conserved motif in TM6 of the rat mu opioid receptor: constitutively active and inactive receptors result from substitutions of Thr6.34(279) with Lys and Asp. Biochemistry 2001 40:13501-13509.

51. Cotecchia S, Rossie O, Fanelli F, Leonardi A, De Benedetti PG. The alpha1a and alpha1b-adrenergic receptor subtypes: molecular mechanisms of receptor activation and of drug action. Pharm.Acta.Helv. 2000 74:173-179.
52. Kuipers W, Van Wijngaarden I, IJzerman AP. A model of the serotonin 5-HT1A receptor: agonist and antagonist binding sites. Drug.Des.Discov. 1994 11:231-249.
53. Kuipers W, Oliveira L, Vriend G, IJzerman AP. Identification of class-determining residues in G protein-coupled receptors by sequence analysis. Receptors Channels 1997 5:159-174.
54. Rippmann F, Bottcher E. Molecular modelling of serotonin receptors. 7TM 1993 1.
55. Schwartz TW, Gether U, Schambye HT, Hjort SA. Molecular mechanisms of action of non-peptide ligands for peptide receptors. Curr.Pharm.Des. 1995 1:325-342.
56. Baldwin JM, Schertler GFX, Unger VM. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G protein-coupled receptors. J.Mol.Biol. 1997 272:144-164.
57. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Meth.Neurosci. 1995 25:366-428.
58. Vriend G. WHAT IF: A molecular modeling and drug design program. J.Mol.Graph. 1990 8:52-56.

FIGURES

A) **B)**

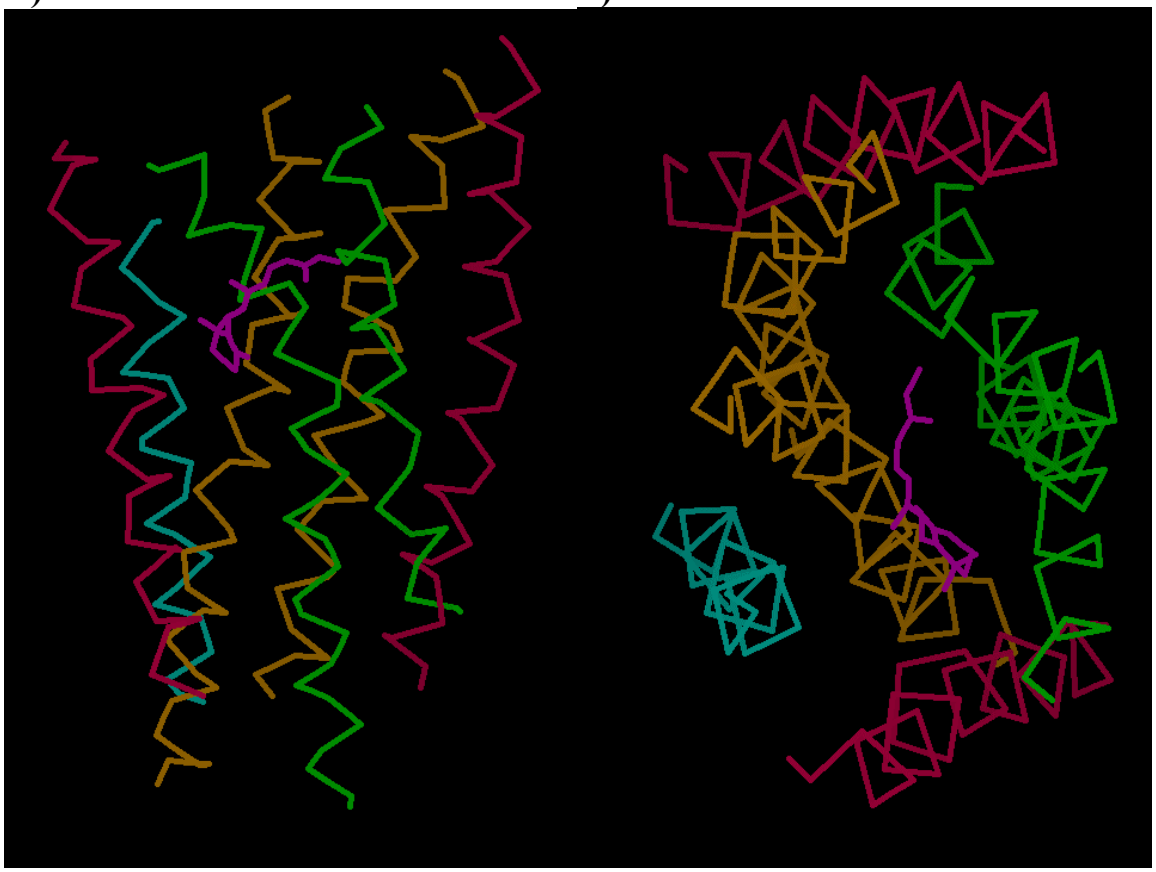
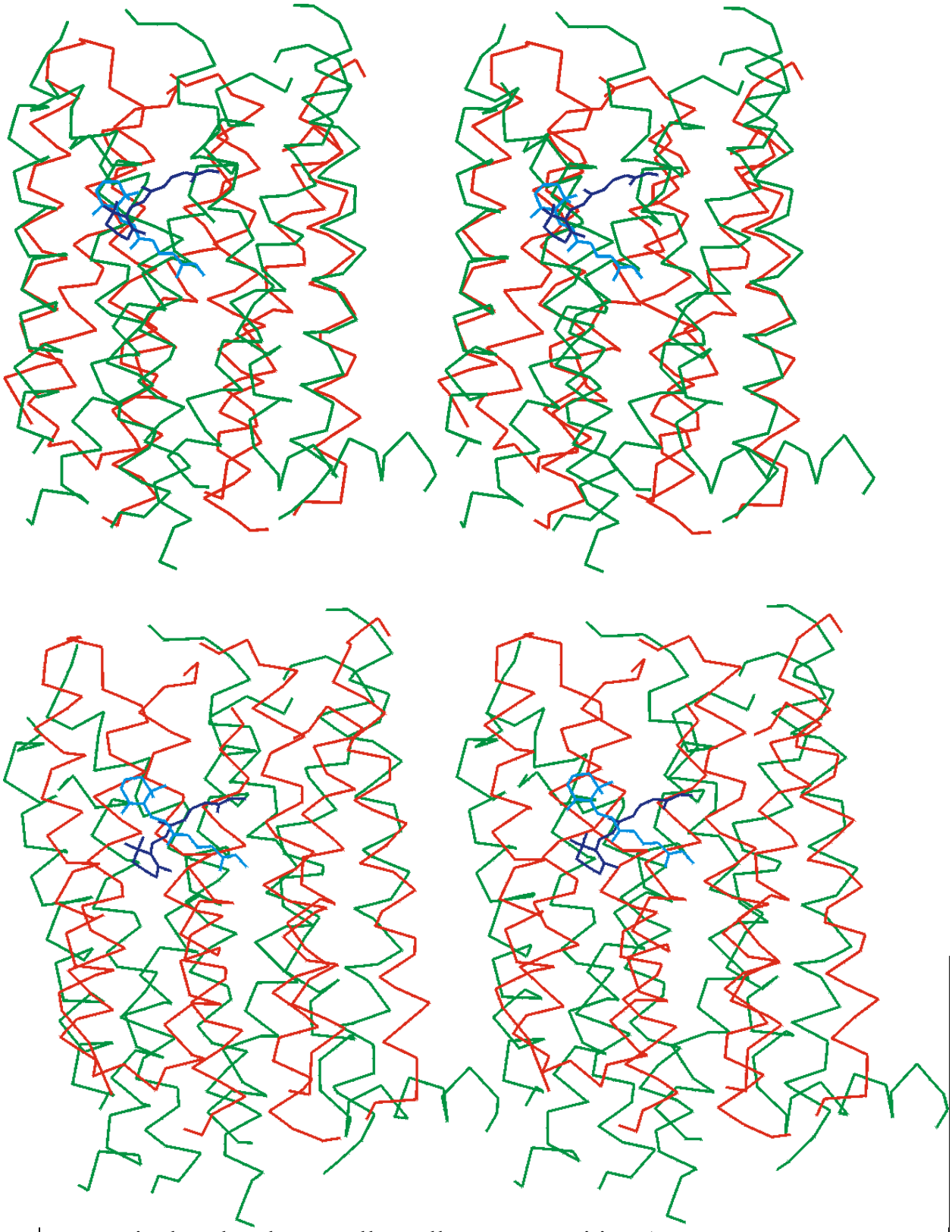


Fig. 1. The helix bundle in bovine opsin.
α-carbon trace of bovine opsin[1]. Retinal is shown in purple, helices II-III orange, VI-VII green, IV light blue, I and V red. A) side view. B) top view.



two retinal molecules equally well as superposition A.

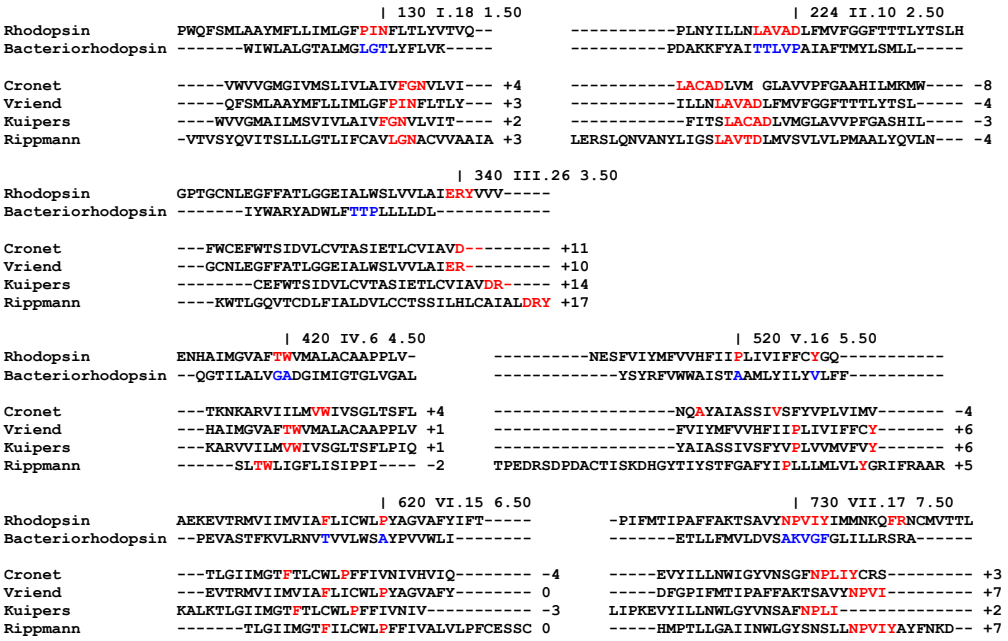


Fig. 3. Sequence alignment extracted from deposited GPCR models. The models shown are produced by Cronet[22], Vriend[23], Kuipers[52,53], and Rippmann[54]. The top two lines show the alignment of bovine opsin with bacteriorhodopsin as extracted from the superposition as shown in figure 2. The motifs containing the most conserved residues in the seven transmembrane GPCR helices are in red, and the corresponding bacteriorhodopsin sequences in blue. The corresponding residues in the four models are coloured red too. The vertical bars indicate the most conserved residue in each helix. The numbers behind these bars indicate the residue numbers in the GPCRDB[2], the Schwartz[55]/Baldwin[56], and the Ballesteros-Weinstein[57], and the numbering schemes, respectively. The numbers behind the sequences indicate the shift away from the perfect alignment.

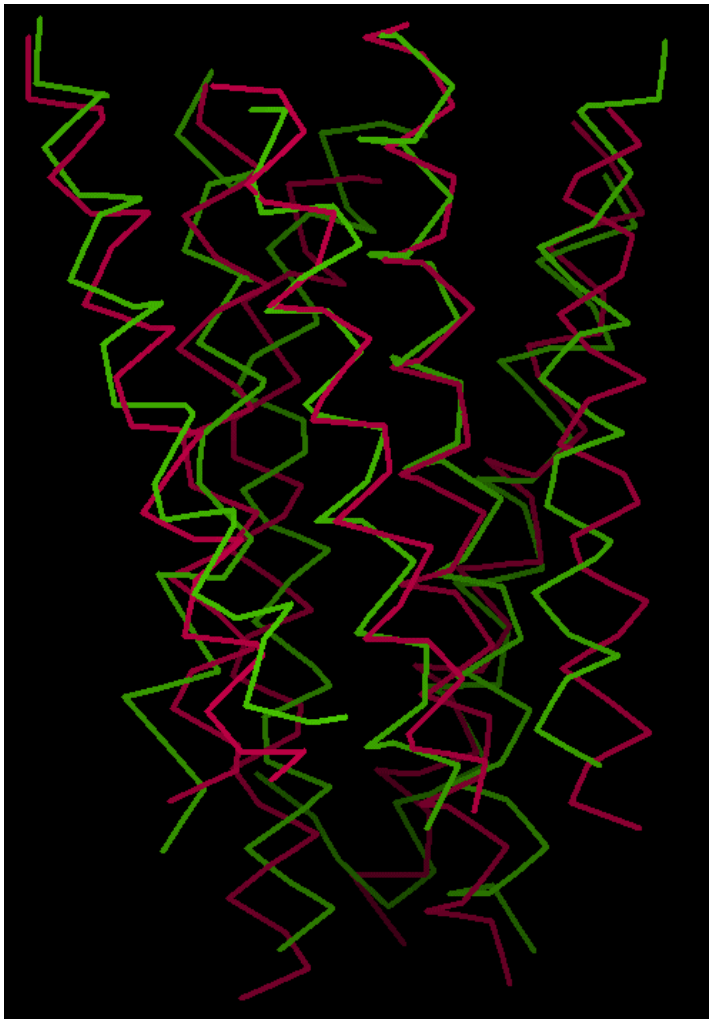


Fig 4. Superposed bovine opsin structure and model.
The bovine opsin structure[1] in red superposed in the GPCRDB[2] BC-model built with WHAT IF[58], based on the Cα coordinates provided by Baldwin[11]. The average Cα and all-atom modelling errors are 2.5 and 3.2 Å, respectively.

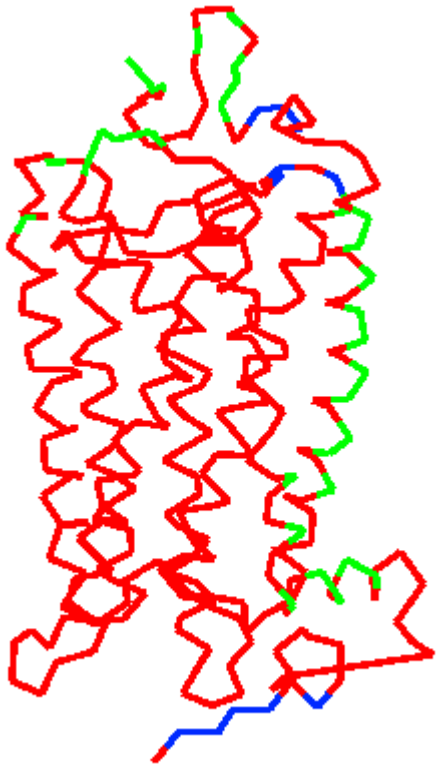


Fig 5. Non-natural contacts in the bovine opsin structure.
The Cα-trace of the A-subunit of the PDB file 1F88[1] is shown in red. Residues making crystal contacts with the B-subunit or any of the crystallisation inducing additives are in green. Residues making crystal contacts with other unit cells are in blue. Note that the extracellular loops, helix I, and helix VIII make many crystal contacts.