An Olfactory Receptor Pseudogene whose Function emerged in Humans—A case study in the Evolution of Structure-function in GPCRs

Peter C. Lai[1], Gautam Bahl[2], Maryse Gremigni[3], Valery Matarazzo[3], Olivier Clot-Faybesse[3], Catherine Ronin[3], and Chiquito J. Crasto[4,5,*,†]

[1]Division of Natural Science, Mathematics, and Computing, Bard College at Simon's Rock, Great Barrington, Massachusetts, USA

[2]Department of Radiology, Wayne State University/ Detroit Medical Center, Detroit, Michigan, USA.

[3]Laboratoire de Neuroglycobiologie, UMR 6149 CNRS, Université de Provence, Pole 3C, 3 Pl. V.Hugo, 13331-Marseille Cedex 3– France

[4,*,†]Department of Neurobiology and [5]Yale Center for Medical Informatics, Yale University
School of Medicine, New Haven, CT 06516. USA

*Corresponding author

†Current Address: Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294; email: ccrasto@genetics.uab.edu; Phone: 205-996-7083; Fax: 205-996-4056-5708

Abstract

Human olfactory receptor, hOR17-210, is identified as a pseudogene in the human genome. Experimental data has shown however, that the gene product of cloned hOR17-210 cDNA was able to bind an odorant-binding protein and is narrowly tuned for excitation by cyclic ketones. Supported by experimental results, we used the bioinformatics methods of sequence analysis, computational protein modeling and docking, to show that functionality in this receptor is retained due to sequence-structure features not previously observed in mammalian ORs. This receptor does not possess the first two transmembrane helical domains (of seven typically seen in GPCRs). It however,

possesses an additional TM that has not been observed in other human olfactory receptors. By incorporating these novel structural features, we created two putative models for this receptor. We also docked odor ligands that were experimentally shown to bind hOR17-210 model. We show how and why structural modifications of OR17-210 do not hinder this receptor's functionality. Our studies reveal that novel gene rearrangement that result in sequence and structural diversity in has a bearing on OR and GPCR function and evolution.

Keywords: Olfactory receptors, functional pseudogene, computational modeling, docking.

Introduction

GTP-binding **P**rotein **C**oupled **R**eceptors (GPCRs) are proteins that traduce the cell membrane and are responsible for catalyzing or initiating a cellular response in the form of a signal transduction process following an extracellular stimulus.[1,2] GPCR function is wide and varied. GPCRs are ubiquitously found in mammals, plants and fungi.[3] Olfactory receptors (OR) constitute the largest gene families in mammalian genomes[4]. Structurally, these entities are believed to be rhodopsin-like GPCRs, characterized by seven transmembrane helical regions that are connected by three extracellular and three intracellular loops, an extracellular N-terminus and an intracellular C-terminus. Earlier experimental observations have shown that there exists a many-many binding/activating relationship between ORs and odors. One odor may bind and activate more than one OR, while an OR might be activated by more than one odor.[4-15] Olfactory receptors' (OR)

interactions with odorous ligands are widely accepted as the first specific step in the early events leading to olfactory perception. Research has also suggested that interaction of odorant ligands with the binding region of an OR[16,17] may cause the receptor to evolve from a structurally inactive to active state. The olfactory system can differentiate odorous molecules based on structural and chemical diversity and concentration.

After the publication of the first draft of the human genome, several groups, working independently, identified the human olfactory repertoire[18-21]. As other mammalian genomes became available, the OR repertoires of these species were also identified[22-24]. These genomic OR genes were identified as either putatively functional or non-functional and pseudogenic. Initially, more than 60% of the human olfactory receptor genes were flagged as pseudogenes, while that number for the mouse OR repertoire stands at less than a third.[22,23] As additional analysis is being carried out, the number of mammalian functional receptors however, is being constantly revised.[25,26] There is evidence that primate evolution is marked by loss of olfactory functionality, as evidenced by a greater percentage of functional ORs in the evolutionary parent than the daughter.[27]

OR gene, hOR17-210 was genomically identified as pseudogenic (OR1E3P in the HUGO and HORDE (Human Olfactory Receptor Database Exploratorium house at http://bioportal.weizmann.ac.il/HORDE/) databases. This genomic pseudogene sequence was identified earlier as possessing a two-nucleotide frame shift[28]. A cDNA clone of this frame-shifted sequence was subsequently shown to successfully initiate a G-protein mediated signal transduction cascade in the presence of a mixture of odorants (especially ketone compounds[29]) commonly perceived by humans.

We used bioinformatics strategies to show that this receptor protein has sequence-structure features that are atypical of previously studied ORs or GPCRs. Despite these differences, this receptor remains functional. The novel sequence-structural features are: 1) hOR17-210 has only six transmembrane helical regions (TMs) instead of the typical seven. The first two TMs typically observed in models of ORs and GPCRs are missing. 2) While this presumably reduces the number of TMs to five, there exists an additional TM, which occurs after what is typically observed as the C-terminus in other ORs. The sequence for this TM is only found in two other ORs—a chimpanzee and a cow homolog, which themselves have additional unique structural features. 4) The amino acid sequence motifs for ORs that have been implicated in G-protein coupling and olfactory sensory neurons targeting[30], however, remain structurally and sequentially conserved. 5) Unlike mammalian ORs and GPCRs studied to date, the C-terminus is predicted to be extracellular.

We show how and why these structural modifications may not hinder the function of this OR. We created two putative computational models of this receptor. Our models incorporate the novel sequence-structural features for this OR. We also carried out computational docking studies using the preferred of the two models with selected odor ligands that are known to experimentally excite hOR17-210.[29]

Materials and Methods

Sequence Analysis and Transmembrane Domain Prediction

- Figure 1a shows the results of a comparative sequence analysis between the cDNA sequence functionally studied and the pseudogene (OR1E3P) identified

from the genome in the HORDE database[31]. The former is listed in GENBANK (http://www.ncbi.nih.gov/) under Accession Number AAC99555, the latter, Accession Number U53583. OR1E3P has a nucleotide sequence in a missing 5' region located upstream from the cloned cDNA sequence. The missing region is as follows:

ATGATGAAGA AGAACCAAAC CATGATCTCA GAGTTCCTGC TCCTGGGCCT TCCATCCAAC CTGAGCAGCA GAATCTGTTC TATGCCTTGT TCTTGGCCGT GTATCTTACC ACCCTCCTGG GGAACCTCCT CGTCATTGTC CTCATTCGAC TGGACTCCCA CCTCCAC.

On the other hand, the sequence used in our informatics-based work and which was shown to be functional possesses the following additional base pairs at the 3' terminus.

TAGTAGGTGTAGTAAAGTTGATAATGAAATATCACTCTAAA TCAGTGG CTTAA

- When the genomic gene sequences is translated using the TRANSLATE tool available through Swissprot's Expasy web site (http://ca.expasy.org/tools/dna.html), OR1E3P (genomic OR17-210) contained several stop codons (denoted by /) after the first 132 residues. (Figure 1b) The TRANSLATE program also translates a given nucleotide sequence in three frames. A two-nucleotide frame shift yielded the same peptide sequence as AAC99555. The sequence in entry AAC99555, when directly translated yields the functional OR. The bolded region in Figure 1c shows the cDNA cloned protein sequence that was used in our analysis and in the experimental functional studies [29].

- As a prelude to creating our computational model, we used Hidden Markov Models to predict transmembrane helices in hOR17-210. Figure 2 highlights regions that are predicted as TMs by two transmembrane prediction algorithms: TMHMM [32] and HMMTOP [33] TM. Both were identified as the best α-helical TM prediction programs in an analysis of over ten such programs [34]. The figure shows that both programs agree in their identification of only six TMs and an extracellular C-terminus (red circles).
- We carried out a sequence comparison of hOR17-210 with rat OR I7, Olr266 in Genbank (Accession Number P23270, Figure 3). OR I7, having been among the first cloned and identified ORs is also exceptionally well characterized, both experimentally [35] and computationally, structurally [36,37]. We use rat I7 here to represent ORs with structural features that are typical of GPCRs. The I7 TM regions are highlighted as predicted by both TMHMM and HMMTOP. The figure shows that I7 has TM1 and 2. Both of these TMs are missing in hOR17-210; the latter has an additional predicted TM after the C-terminus. We denote this additional TM as TM7'. Interestingly, the region of TM2 correctly predicted as a helical TM region for rat I7 is not predicted as a TM in hOR17-210 despite the apparent sequence similarities. Experiments showed that this region was indeed extracellular and not a TM [29].
- We carried out a comprehensive BLAST search for hOR17-210 against GENBANK. Sequence identity was found between hOR17-210 and its predicted chimpanzee and cow (Olr466) OR homologs. Figure 4 shows the results of the alignment. The TM regions are highlighted. All three sequences possess the

TM7' region. The notable difference between the three sequences is that the predicted chimpanzee and cow homologs lack OR17-210's frame shift mutation and so have seven intact TMs. And just as in hOR17-210, the cow and chimpanzee homologs are missing the typically observed TM2.

Constructing a structural model of hOR17-210

We created two computational, structural models of hOR17-210 (Figure 5a and 5b). Our modeling strategies incorporated the new structural features discussed above. In addition to using homology modeling strategies [38] to create our preliminary model, we tested a new paradigm for rationalizing the hydrophobic nature of the inside of the receptor (Equation 1). Since the first two TMs were missing, during homology modeling, the first five TMs of hOR17-210 were positioned in the helical assembly in orientations occupied by TMs 3, 4, 5, 6 and 7, respectively, in typical ORs. The key challenge however, was the location of orphaned region, TM 7' which could be packed in either of the spaces formerly having been occupied by TM 1 or 2.

- TM helical regions were predicted by using Hidden Markov Models through the programs: TMHMM2.0 [39] and HMMTOP2.0 [33].
- Due to the missing TMs 1 and 2 and the addition of an orphaned TM7', the secondary structures of helices in positions 3, 4, 5, 6 and 7 were refined by alignment of the first five regions of hOR17-210 against the previously predicted secondary structure of rat OR-I7. Each predicted OR helix was aligned to the center of the homologous helix in the rhodopsin structure, with no gaps. Two variants of helical packing were explored: in the first variant, the final orphaned

helix of hOR17-210 was aligned against helix 2 in rhodopsin; in the second variant, the last (orphaned) helix of hOR17-210 was aligned against helix 1 in rhodopsin.

- 3D models were generated using the homology modeling program, Modeler 8v2 [38]. We used the highest resolution structure (2.2 Å) of dark-adapted bovine rhodopsin [40] (Protein Data Bank ID: 1U19) as a template. For the helical construction, each predicted OR helix was sequentially aligned to the center of the homologous helix in the rhodopsin structure, with no gaps. Two variants of helical packing were explored: in the first variant, the final orphaned helix of hOR17-210 was aligned against helix 2 in rhodopsin; in the second variant, the last (orphaned) helix of hOR17-210 was aligned against helix 1 in rhodopsin.
- Each helix of each variant of the structural model was minimized with typical α-helix H-bond distance constraints using the consistent valence force field (CVFF) and conjugate gradient algorithm in Accelrys Discover suite of programs (http://www.accelrys.com/products/insight/). The helices were individually submerged in water during the energy minimization step to relax helical features specific to rhodopsin.
- The hydrophobic moments at each residue around a helix were calculated using the following expression:

$$\Theta_\theta = \sum_{i=o}^{360-\theta} \mu_\theta . \cos i \qquad \text{Equation [1]}$$

where, μ_θ is the effective aggregate hydrophobicity at each point around a helical wheel computed by summing the arc contributions to the hydrophobicity moment on that residue from all other points along the helical wheel for a given TM. In

order to establish the correct frame of reference with respect to the entire helical assembly for hydrophobic moments derived from this algorithm, these were initially calculated for the TMs in rhodopsin and subsequently mapped to the actual rotational orientations within the helical bundle. We observed that in rhodopsin, that the largest Θ valued residues in TM 1, 3, 4, and 7 pointed toward the binding pocket; for TM2, the largest Θ pointed away from the binding region; and for TMs 5 and 6, the largest aggregate hydrophobic moments pointed toward each other. The hydrophobicities for hOR17-210 were computed and the TM helices were rotated and oriented using this rationale.

- After helix construction and rotation, the helices were used as the input template into the Modeller software for *ab initio* assignment of the intra- and extracellular loop residues. The resulting structure was then rigorously minimized using the Accelrys Discover program by constraining only the motion of the alpha-carbon atoms of the protein in order to maintain the integrity of the transmembrane helices.

Of the two models created, our preferred model is one where TM7' is positioned in place of TM2. As shown in figure 4, the protein sequences in the chimpanzee and cow homologs both have strongly predicted TM1s and therefore, in order to maintain the helical bundle, TM7' would have to be placed in the position typically occupied by TM2.

Ligand docking

We docked eight ligands: beta -ionone, d- and l-camphor, 2- and 6-undecanone, heptanal, decanal, nonanol and nonanone (ligand positional parameters and those for the two model variants are available from corresponding author) in the binding pocket of our preferred

model of hOR17-210—with the TM7' homology-modeled in the place of missing TM2 Figures 6a (ringed ligands) and 6b (straight chain ligands) show the results of computational docking. Of these ligands, experimentally, the cyclic ketones show strong responses, the straight chain ketones show weak responses and the alcohols show no response at any concentration. (Personal Communications—CR). These docked ligands vary in length of carbon chain and functional groups (aldehyde, ketone, alcohol and ring structures). Ligand models were constructed using the InsightII suite of software (http://www.accelrys.com/insight/).

Ligand conformational energies were minimized using the Discover module in InsightII. We added hydrogen atoms to our I7 OR model [36] to create a system of pH 7.0. Atomic charges were assigned using Consistent Valence Force Field (CVFF). We used DOCK [41,42] to identify the ideal binding configurations of the ligands in the binding pocket [36] of the 17-210 human olfactory receptor model. Using every atom in the OR model as input for the DMS (Dot Molecular Surface) program [43], we calculated a solvent accessible molecular surface-area for the I7 model; and, DOCK's SPHGEN (**SPH**ere **GEN**erator) module identified cavity site-points in the receptor. We discarded spheres that represented cavities on the intracellular side of the receptor; these spheres were structurally "below" the TM3 and TM4 crossover plane in the model. The GRID module in DOCK was used to generate force fields and interaction parameters to compute intermolecular binding. DOCK used spheres that were retained to compute spatial restraints based on van der Waals interactions. Flexible_Ligand, a module in DOCK, allowed the modification of torsion angles in the ligand. Figures 6a and 6b show the docked ligands in the receptor model.

Results

The helical regions and the internal and external locations of the intra and extracellular loops and the N- and C-termini are highlighted in the figure 3, which compares the sequences of hOR17-210 and rat OR I7, one of the first ORs cloned [44], functionally analyzed [35], and modeled [36,37]. The highlighted regions in the I7 sequence in figure 3 are representative of what is typically known about ORs. Figure 2 shows that both HMMTOP and TMHMM transmembrane prediction programs agree that for hOR17-210: 1) the region beginning with **PMY**-- is not a TM, although it is predicted to be the second TM in most ORs; 2) the region we identify as the orphaned TM, TM 7' exists; and, 3) the C-terminus of the receptor must be atypically extracellular. TM 7' in OR17-210 extends past the C-terminus of ORI7. The amino acids of the C-terminus in I7 and the final intracellular loop (between TM 7 and 7') of OR17-210 are conserved.

The hOR17-210 sequence begins with a **MPMY** polypeptide region. Typically, in ORs, the **PMY** sequence motif marks the beginning of TM 2. HMMs for this functional hOR17-210 sequence predict however, that the region beginning with PMY (i.e., **MPMYLCLSNLSFSDLCFSSVTM** is not a helical TM and is extracellular). Experimentally, heterologous expression of a FLAG (DYKDDDDK polypeptide sequence) tagged receptor in insect cells confirmed an extracellular N-terminus [29]. This leads us to conclude that hOR17-210 is missing both the first and second TMs when compared to other mammalian ORs. The hOR17-210 transmembrane regions start from what would typically be TM3 in other ORs. The "MAYDRY" motif region, a highly conserved sequence within Class A GPCRs and known to be located at the intracellular end of this TM, has been shown to be essential for G-protein coupling and the initiation

of signal transduction following ligand-binding [30,45,46]. In the case of hOR17-210, the "MAYDRY" sequence has undergone mutation to MAYHCY; since this TM must be oriented extracellular to intracellular in order for correct positioning of this conserved motif, it allows the polarity of the remaining TMs predicted by the HMM to be determined with certainty.

A sequence similarity search of the TM 7' peptide region "**FVFKIVIVGILPLLNLVGVVKLI,**" returns only two matches. The first is a chimpanzee OR (GENBANK Accession Number XP_523775), which is homologous to hOR17-210; the second is a cow OR (GENBANK Accession Number XP_872923). HMM of these OR sequences predicts that, in addition to the presence of TM 7', the polypeptide regions beginning with PMY are also not TM helices. One major difference is that these two orthologs do, however, possess intact TM 1 helices. (Figure 4)

The constrained polarity of the final TM (TM 7') causes the C-terminus of OR17-210 to become extracellular. Extracellular C-termini have been predicted in Drosophila odorant receptors.[47] The sequence region **RNRDMRGNPGQSLQHKENFF** is the third intracellular loop in hOR17-210 (between TM 7 and TM 7'). We carried out a BLAST search using the above sequence of this loop, focusing the search to return only sequences for olfactory receptors. From over 2500 results, the "RNRDMRG" region is strongly conserved (greater than 70% identify and 100% positive matches, where R is often replaced by K) in the C-termini of most ORs (and is possibly involved in GPCR-G-protein interactions). hOR17-210 functionality is therefore, not affected by an extra-cellular C-terminus or the lack of a TM (the absence of TMs 1 and 2).

Figures 6a and 6b indicate that all the docked odorous ligands are clustered in the same spatial region bound by the first four TMs (3, 4, 5 and 6) of hOR17-210. The numbers in parenthesis indicate the TM numbers for typical ORs. An inspection of the interior of the receptor, which is modeled using the hydrophobicities determined using equation [1] indicate that there are no strongly polar residues pointing into the binding pocket, except His48 on TM1. This residue however, is greater than 10Å away from the nearest ligand atom. The nearest distances from every side-chain atom within the receptor's binding pocket to each atom of the docked ligands were calculated. The closest distance (between 0.8 and 1.5 Å) was for Ala108, specifically between ligand atoms and the methyl hydrogens in Ala108. Some of the interactions can be considered to be electrostatic in nature because they are between the ligand carbonyl oxygen and the Ala108 hydrogen atoms. Possible interactive distances were also observed between ligands and Phe122 and Cys123. These residues however, belong to the first extracellular loop, which in our model dips into the binding pocket. The contributions of these residues however, cannot be ascertained because of the dynamic nature of loop conformations.

Discussion

We created computational structural models for two possible variants to account for the atypical nature of the hOR17-210 (Figures 5a and 5b). Such a model, based on structural template matching (the sequence homology between I7 and rhodopsin is less than 40%) [10,14] may however, introduce rhodopsin structure-specific biases into the model. Biases include differences in lengths of loops [48] and kinks [49,50] in TM helical

domains. We attempted to limit the intra-helical biases by allowing each helix to structure-energetically relax in an aqueous medium individually before assembling the TM domains.

The first step in any GPCR modeling study is the identification of the TM regions. TM helices presumably protect the interior of the binding pocket from the surrounding lipid bilayer, while at the same time, ensuring that the signal-transducing structural features of the receptor are properly positioned inside and outside the cell. We aligned the first five TMs of hOR17-210 cloned sequence in the positions occupied by TMs 3, 4, 5, 6 and 7 of typical ORs, respectively. In each variant, TM 7' occupied the positions typically occupied by TM 1 and TM2, respectively. Our modeling strategy ensured that sequentially conserved (and possibly functionally implicated) regions were positioned as found in typical ORs. TM 7' in the two variants was positioned to maintain the structural integrity of the TM scaffold while protecting the interior of the OR and the odor ligand. We have indicated earlier that orthologs of hOR17-210 exist in cow and chimpanzee. Evidence of sequence predicted as TM7' and absence of TM2 in observed in only these three mammalian ORs. During hOR17-210 modeling, when presented with a choice of positioning TM7' in the position of TM1 or TM2, we posit that since the cow and chimpanzee orthologs retained TM1 and were missing only the typical TM2, the orphaned TM7' would favorably occupy in the position of TM2. (Figure 5b). For olfactory function to persist, the main GPCR scaffold needs to be maintained.

Katada et al. [51] have shown that the C-termini of ORs are involved in G-protein interactions. Our BLAST results have shown that the third intra-cellular loop shows strong sequence homology with several hundred vertebrate ORs, especially in the

"RNRDMRG" sequence motif, which is invariably in the C-terminus, specifically, in the region where the seventh TM ends and the C-terminus begins. A few of the more than 900 results also show homology in the rest of this intra-cellular loop and the homolgous ORs are always in the C-terminus. This indicates to us that if a certain motif of amino acids interacts with the G-protein, then this motif is present in OR17-210 intracellularly. We conclude therefore that OR function is not hindered because of the presence of an intracellular third loop.

An extracellular C-terminus for mammalian ORs has also not been experimentally shown. The presence of an orphaned TM puts the C-terminus extracellularly. This is confirmed by two TM prediction programs. Experiments involving the attachment of a poly-Histidine tag to the end of the OR would confirm the extracellular C-terminus for hOR17-210.

Identifying the active and inactive states for an OR and elucidating its role in olfaction at a molecular level necessitates an experimental determination of its protein structure, in addition to knowledge of its odorant repertoire. There is currently, no experimentally derived structure of an olfactory receptor. This lack of a structure engenders the assumption that GPCRs are structurally similar to rhodopsin. Every computational study of olfactory receptors and other GPCRs uses a rhodopsin structure (the x-ray derived structure with the highest resolution [40] as used in our modeling here or a lower resolution electron diffraction structure [52]). Modifications to remove rhodopsin-specific biases as detailed in the Materials and Methods section begin from this point.

Computational modeling, docking and simulation studies [15,36,37,53] have shown that the OR binding region is on the extracellular side—a pocket that is created by side

chains belonging to TM regions 3, 4 5 and 6—which is confirmed by our docking results. Our docking results indicate that the ALA108 is the only residue in the binding pocket that is within Van der Waals distances with the ligands. These docking results are preliminary at best. Since all the odors docked in the same general area of the binding region, we believe that the binding region is apt for odor-OR interactions. We have previously shown that dynamic simulation of odorous molecules in the olfactory receptor binding pocket provide instances of interactions with key amino acid residues in the binding pocket. These interactions are however, not always observed as a result of static docking. The time and computational effort required to complete the dynamic simulation of all odors identified that excite hOR17-210 strongly make it the subject of another paper. Certainly, site-directed mutagenesis results would provide us with better starting points in our docking and simulation studies. These results unfortunately, do not exist for this receptor.

Our standard model building protocols differ slightly from those previously established [36,53,54], by independently predicting the TM regions followed by removing of rhodopsin structure specific biases. In the case of hOR17-210, we are breaking new ground because we have identified and attempt to model a novel TM (TM 7') that has not been sequence-structurally observed. Once TM 7' was placed in position of TM2 (our preferred structure), it was subject to the same TM-modeling protocol as other TMs in our model. Further validation of the novel method we introduced with equation [1] of calculating hydrophobic moments for determining TM rotations would also be aided by future experimental work followed by fine tuning of our modeling strategies.

Future work with experimental functional analyses following key point mutations would aid in identifying the role of binding pocket residues in ligand interactions. Simulating the dynamic motion of ligand in the OR binding pocket where its interactions with key residues in the binding pocket can vary over the time period of the simulation would be useful to identify if other residues are involved in the ligand-OR interaction. Also, computational docking shows that ligands tested (Figures 6a-b) are clustered within a single region, we can only surmise from the docking results that this region is the preferred binding region for this OR..

Conclusion

This informatics-based study, supported by experimental results, identifies an OR possessing atypical sequence-structure features while still maintaining olfactory functionality. The human olfactory repertoire reveals that the ratio of functional ORs to pseudogenes is 1:2 (http://senselab.med.yale.edu/senselab/ORDB/humanOR.html). This number has been revised more recently and will be likely revised as more information becomes available and more genome level experiments are carried out.

Evolutionarily, hOR17-210 could occupy a position of transition between functional and pseudogenic ORs. This receptor is a possible illustration of how loss in OR function may occur, namely, through mutations that create unfavorable sequences for transmembrane helical assembly. Our study, we hope, will cause researchers to reassess the sequence-structure-function correlates in olfactory receptors, and also the necessity to incorporate structural features in the classification of ORs and GPCRs.

We have used bioinformatics methods to show how and why a receptor appearing pseudogenic in several portions of the population can be functional in others. Functionality was confirmed by measuring the experimental, varying excitatory responses to odor ligands with different of functional groups. While hOR17-210 appears to lack the first two TMs typically observed in ORs, it possesses an additional TM present in only two other non-human olfactory receptors in all of GenBank. This TM, named TM 7', may preserve olfactory function within this OR (when functional) by maintaining the TM structure, thus protecting the binding odor ligand. The intracellular positions of regions identified as possibly responsible for olfactory function, due to their highly conserved nature, are preserved. hOR17-210 possibly straddles the point in mammalian OR evolution where loss of function occurs.

Bibliography

1. Ji, T.H., Grossmann, M. & Ji, I. G protein-coupled receptors. I. Diversity of receptor-ligand interactions. *J Biol Chem* **273**, 17299-302 (1998).
2. Muller, G. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. *Curr Med Chem* **7**, 861-88 (2000).
3. King, N., Hittinger, C.T. & Carroll, S.B. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361-3 (2003).
4. Mombaerts, P. Molecular biology of odorant receptors in vertebrates. *Annu Rev Neurosci* **22**, 487-509 (1999).
5. Malnic, B., Hirono, J., Sato, T. & Buck, L.B. Combinatorial receptor codes for odors. *Cell* **96**, 713-23 (1999).
6. Touhara, K. et al. Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc Natl Acad Sci U S A* **96**, 4040-5 (1999).
7. Abaffy, T., Malhotra, A. & Luetje, C.W. The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J Biol Chem* **282**, 1216-24 (2007).
8. Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J Neurosci* **22**, 3033-43 (2002).

9. Bruch, R.C. & Rulli, R.D. Ligand binding specificity of a neutral L-amino acid olfactory receptor. *Comp Biochem Physiol B* **91**, 535-40 (1988).
10. Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J Neurosci* **25**, 1806-15 (2005).
11. Mombaerts, P. Targeting olfaction. *Curr Opin Neurobiol* **6**, 481-6 (1996).
12. Mombaerts, P. Odorant receptor gene choice in olfactory sensory neurons: the one receptor-one neuron hypothesis revisited. *Curr Opin Neurobiol* **14**, 31-6 (2004).
13. Oka, Y., Omura, M., Kataoka, H. & Touhara, K. Olfactory receptor antagonism between odorants. *Embo J* **23**, 120-6 (2004).
14. Rothman, A., Feinstein, P., Hirota, J. & Mombaerts, P. The promoter of the mouse odorant receptor gene M71. *Mol Cell Neurosci* **28**, 535-46 (2005).
15. Touhara, K. Odor discrimination by G protein-coupled olfactory receptors. *Microsc Res Tech* **58**, 135-41 (2002).
16. Shepherd, G.M. The cognitive neurosciences. in *The Cognitive Neurosciences: A Handbook for the Field* (eds. Gazzaniga, M.S. & Bizzi, E.) 105-102 (MIT Press, Cambridge, Mass., 1995).
17. Buck, L.B. Information coding in the vertebrate olfactory system. *Annu Rev Neurosci* **19**, 517-44 (1996).
18. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Res* **11**, 685-702 (2001).
19. Niimura, Y. & Nei, M. Evolution of olfactory receptor genes in the human genome. *Proc Natl Acad Sci U S A* **100**, 12235-40 (2003).
20. Malnic, B., Godfrey, P.A. & Buck, L.B. The human olfactory receptor gene family. *Proc Natl Acad Sci U S A* **101**, 2584-9 (2004).
21. Zozulya, S., Echeverri, F. & Nguyen, T. The human olfactory receptor repertoire. *Genome Biol* **2**, RESEARCH0018 (2001).
22. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5**, 124-33 (2002).
23. Young, J.M. et al. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biol* **4**, R71 (2003).
24. Quignon, P. et al. The dog and rat olfactory receptor repertoires. *Genome Biol* **6**, R83 (2005).
25. Niimura, Y. & Nei, M. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* **346**, 13-21 (2005).
26. Niimura, Y. & Nei, M. Extensive gains and losses of olfactory receptor genes in Mammalian evolution. *PLoS ONE* **2**, e708 (2007).
27. Sharon, D. et al. Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* **61**, 24-36 (1999).
28. Menashe, I., Man, O., Lancet, D. & Gilad, Y. Population differences in haplotype structure within a human olfactory receptor gene cluster. *Hum Mol Genet* **11**, 1381-90 (2002).

29. Matarazzo, V. et al. Functional characterization of two human olfactory receptors expressed in the baculovirus Sf9 insect cell system. *Chem Senses* **30**, 195-207 (2005).
30. Imai, T., Suzuki, M. & Sakano, H. Odorant receptor-derived cAMP signals direct axonal targeting. *Science* **314**, 657-61 (2006).
31. Fuchs, T., Glusman, G., Horn-Saban, S., Lancet, D. & Pilpel, Y. The human olfactory subgenome: from sequence to structure and evolution. *Hum Genet* **108**, 1-13 (2001).
32. Visiers, I., Ballesteros, J.A. & Weinstein, H. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol* **343**, 329-71 (2002).
33. Tusnady, G.E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849-50 (2001).
34. Moller, S., Vilo, J. & Croning, M.D. Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics* **17 Suppl 1**, S174-81 (2001).
35. Araneda, R.C., Kini, A.D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nat Neurosci* **3**, 1248-55 (2000).
36. Singer, M.S. Analysis of the molecular basis for octanal interactions in the expressed rat 17 olfactory receptor. *Chem Senses* **25**, 155-65 (2000).
37. Lai, P.C., Singer, M.S. & Crasto, C.J. Structural activation pathways from dynamic olfactory receptor-odorant interactions. *Chem Senses* **30**, 781-92 (2005).
38. John, B. & Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res* **31**, 3982-92 (2003).
39. Melen, K., Krogh, A. & von Heijne, G. Reliability measures for membrane protein topology prediction algorithms. *J Mol Biol* **327**, 735-44 (2003).
40. Okada, T. X-ray crystallographic studies for ligand-protein interaction changes in rhodopsin. *Biochem Soc Trans* **32**, 738-41 (2004).
41. Gschwend, D.A., Good, A.C. & Kuntz, I.D. Molecular docking towards drug discovery. *J Mol Recognit* **9**, 175-86 (1996).
42. Gschwend, D.A. & Kuntz, I.D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* **10**, 123-32 (1996).
43. Richards, F.M. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **6**, 151-76 (1977).
44. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-87 (1991).
45. Burger, M. et al. Point mutation causing constitutive signaling of CXCR2 leads to transforming activity similar to Kaposi's sarcoma herpesvirus-G protein-coupled receptor. *J Immunol* **163**, 2017-22 (1999).
46. Scheer, A., Fanelli, F., Costa, T., De Benedetti, P.G. & Cotecchia, S. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *Embo J* **15**, 3566-78 (1996).

47. Benton, R., Sachse, S., Michnick, S.W. & Vosshall, L.B. Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS Biol* **4**, e20 (2006).
48. Otaki, J.M. & Firestein, S. Length analyses of mammalian G-protein-coupled receptors. *J Theor Biol* **211**, 77-100 (2001).
49. Olender, T., Feldmesser, E., Atarot, T., Eisenstein, M. & Lancet, D. The olfactory receptor universe - from whole genome analysis to structure and evolution. *Genet Mol Res* **3**, 545-53 (2004).
50. Yohannan, S., Faham, S., Yang, D., Whitelegge, J.P. & Bowie, J.U. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **101**, 959-63 (2004).
51. Katada, S. & Touhara, K. [A molecular basis for odorant recognition: olfactory receptor pharmacology]. *Nippon Yakurigaku Zasshi* **124**, 201-9 (2004).
52. Krebs, A., Villa, C., Edwards, P.C. & Schertler, G.F. Characterisation of an improved two-dimensional p22121 crystal from bovine rhodopsin. *J Mol Biol* **282**, 991-1003 (1998).
53. Vaidehi, N. et al. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **99**, 12622-7 (2002).
54. Singer, M.S. & Shepherd, G.M. Molecular modeling of ligand-receptor interactions in the OR5 olfactory receptor. *Neuroreport* **5**, 1297-300 (1994).

List of Figures

1a) Results of the sequence alignment between OR17-210 cDNA found in GENBANK Accession number (AAC99555) and OR1e3P genomic DNA found in the HORDE database. The functional region in OR17-210 begins from nucleotide 170. This is caused by a two-residue frame shift in the genomic DNA. The sequence of the functional protein also contains an added region beginning from nucleotide number 977. This orphan TM 7' and the extracellular C-termini are contained in this region. The frame shift results in a stop codon beyond nucleotide 1030.

1b) Results of the translation of the genomic DNA for Or1e3P as found in the HORDE database. The presence of stop codons indicates that the receptor is pseudogenic as listed.

1c) Results of the translation of the genomic DNA of OR1e3P as found in the HORDE database, but following a 2-nucleotide frame shift. This sequence is however, is missing the 3' region that resulted in the orphan TM observed in ACC99555 (the sequence used in the functional studies).

2. Figure shows the results of Hidden Markov Model predictions of termini, intra and extra-cellular loops in hOR17-210. Two different profiling programs were used, TMHMM and HMMTOP. The yellow highlighted regions show the predicted TM helices. The blue colored regions indicate the extra-cellular N-termini which contains MPMY polypeptide motif, which marks the beginning of TM2 in most ORs. The region last highlighted region is the orphan TM 7' region. The red colored region shows that the C-terminus of this protein is extracellular.

3. Figure shows the results of a sequence alignment with rat I7, here represented as a typical olfactory receptor, and hOR17-210. The highlighted regions indicating the TM domains assigned to the model show that 17-210 lacks TMs 1 and 2, that the MPMY region is extracellular. It shows alignment between TM1 of hOR17-210 and I7, with the HCY of OR17-210 and the DRY of I7 aligned in the intracellular loop. The figure also shows strong alignment between the third intracellular loop of OR17-210 and the C-terminus of I7, before the origination of TM 7' in OR17-210. Yellow segments show final TM helical assignments used in the models.

4. Figure shows the results of sequence alignment of hOR17-210, its predicted chimpanzee homolog and predicted OR in cow (Orl466-homolog). Each OR shows atypical features of the polypeptide region beginning with PMY not being a TM, and the existence of a TM 7'. Additionally, the chimpanzee and the cow OR gene products show the existence of a TM 1. Residues highlighted in yellow are regions predicted by TMHMM to be transmembrane domains.

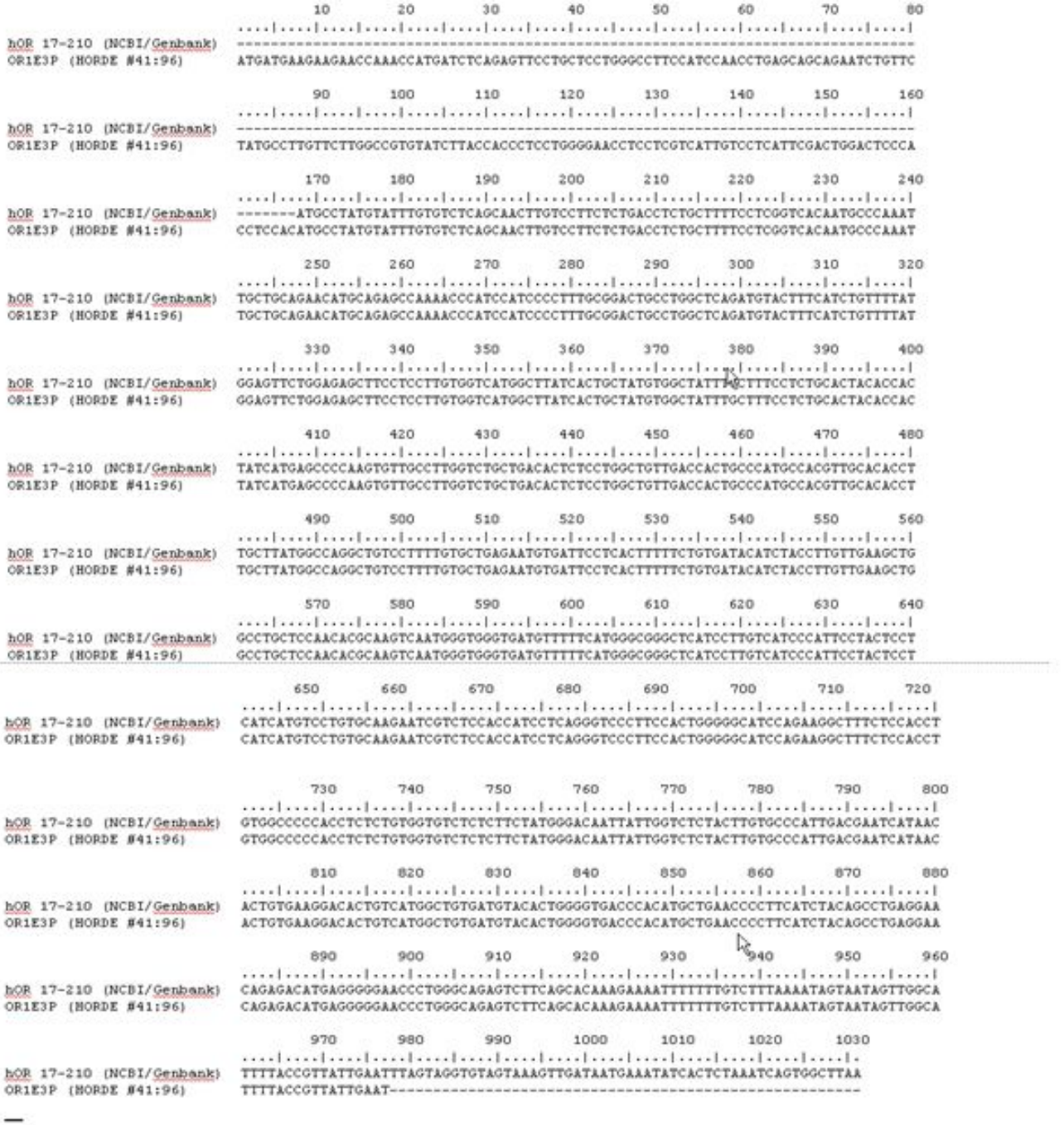
5a: Figure shows a structural model for OR17-210 with the TM 7' occupying the position typically occupied by TM 1 in rhodopsin-like GPCRs.

5b: Figure shows a structural model for OR17-210 with the TM 7' occupying the position typically occupied by TM 2 in rhodopsin-like GPCRs.

6a: Figure shows the docking of three ligands with ring structures: beta ionone (green), D-camphor (yellow) and L-camphor (pink). The figure shows the proximity of the docked ligands to ALA108 in white. The binding is expectedly in the region bound by TMs 3 (1), 4 (2), 5(3) and 6(4). The TM identifiers are numbers typical of ORs and GPCRs. The numbers in parentheses are TM numbers for hOR17-210.

6b: Figure shows the docking of five ligands straight chains: decanol (yellow), nonanone(green), nonanol (pink), 2-undecanone (orange) and 6-undecanone (white). The figure shows the proximity of the docked ligands to ALA108 in white. The binding is

expectedly in the region bound by TMs 3 (1), 4 (2), 5(3) and 6(4). The TM identifiers are numbers typical of ORs and GPCRs. The numbers in parentheses are TM numbers for hOR17-210.

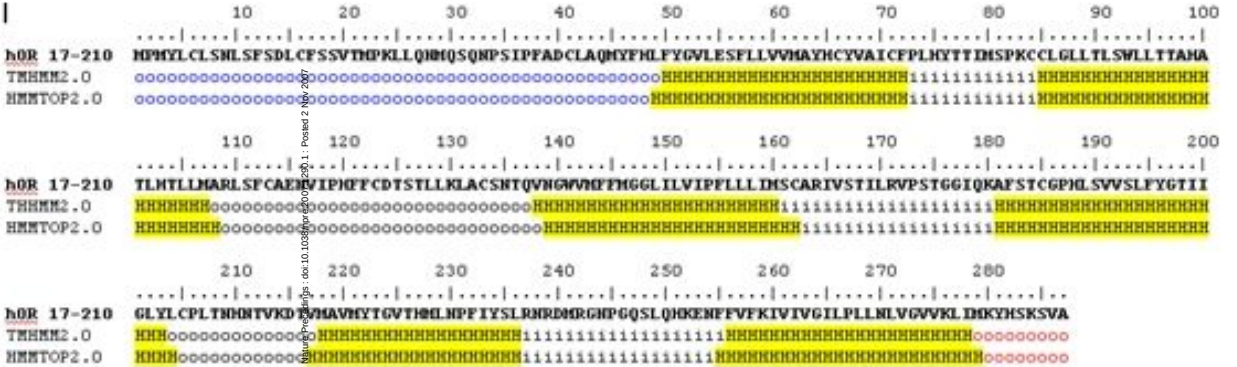


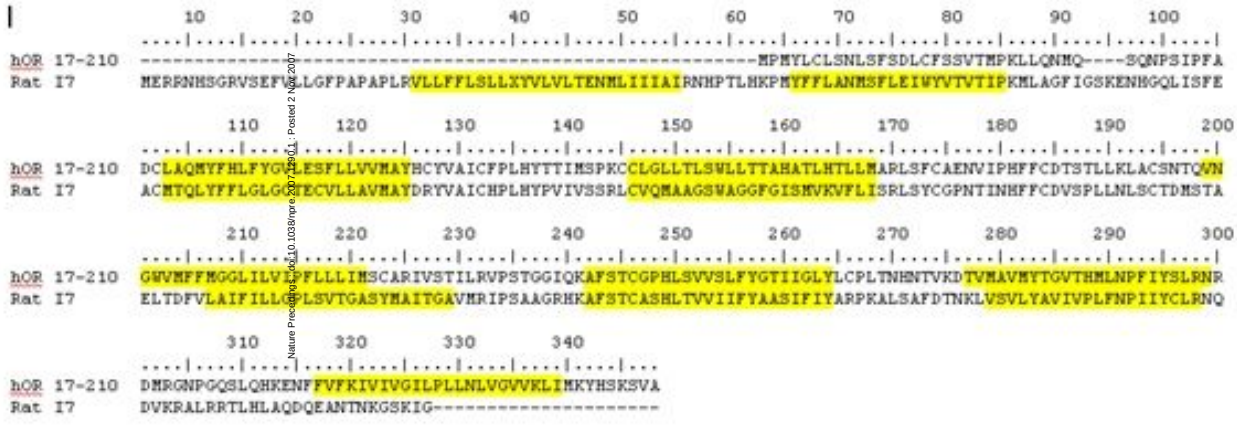
b) 5'3' Frame 1

MMKKNQTMISEFLLLGLPSNLSSRICSMPCSWPCILPPSWGTSSSLSSFDWTPTST
CLCICVSATCPSLTSAFPRSQCPNCCRTCRAKTPSPLRTAWLRCTFICFMEFWRAS
SLWSWLITAMWLFAFLCTTPLS/APSVALVC/HSPGC/PLPMPRCTPCLWPGCPFV
LRM/FLTFSVIHLPC/SWPAPTRKSMGG/CFSWAGSSLSSHSYSSSCPVQESSPPSSG
SLPLGASRRLSPPVAPTSLWCLSSMGQLLVSTCAH/RIITL/RTLSWL/CTLG/PTC/T
PSSTA/GTET/GGTLGRVFSTKKIFLSLK///LAFYRY/

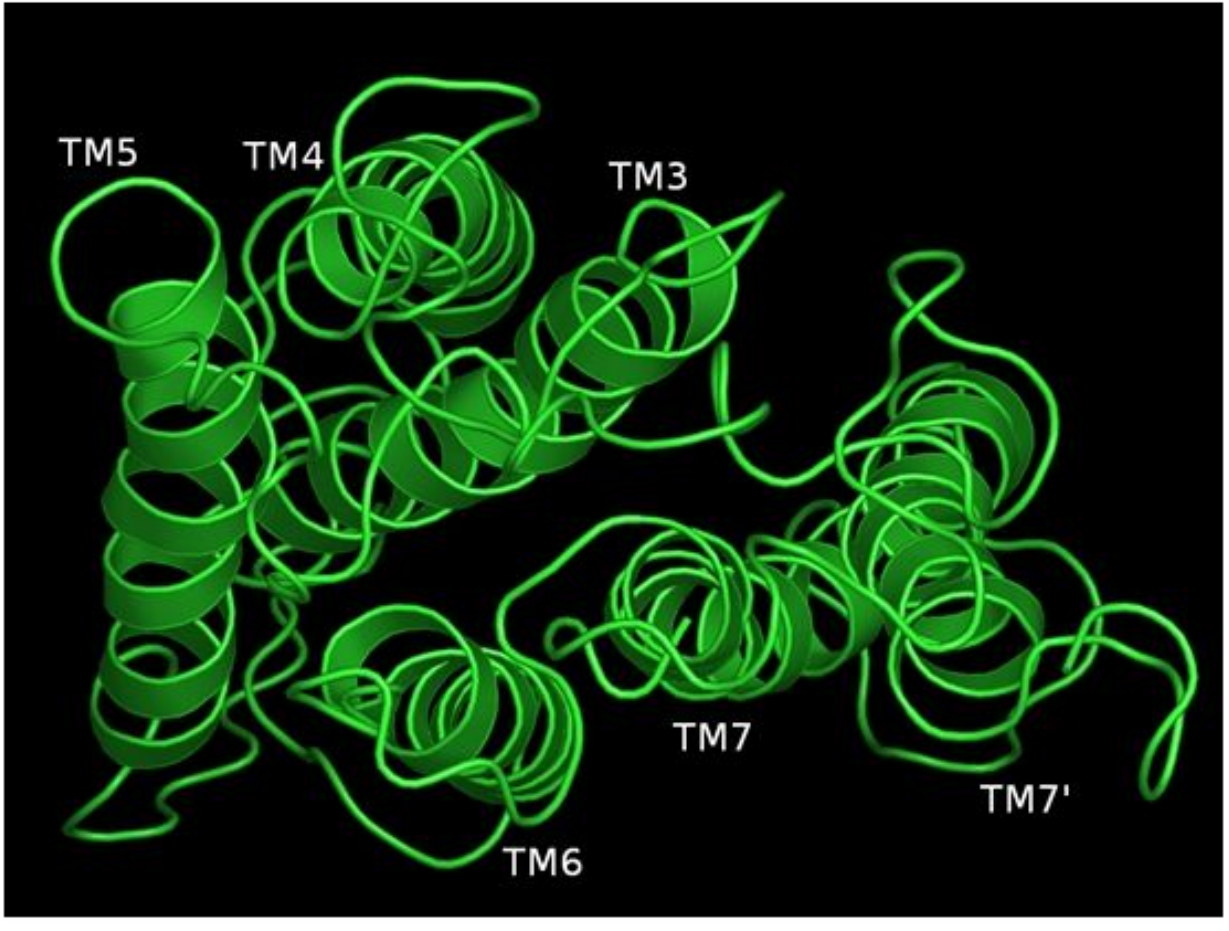
c)5'3' Frame 3

DEEEPNHDLRVPAPGPSIQPEQQNLFYALFLAVYLTTLLGNLLVIVLIRLDSHLHM
PMYLCLSNLSFSDLCFSSVTMPKLLQNMQSQNPSIPFADCLAQMYFHLFYGV
LESFLLVVMAYHCYVAICFPLHYTTIMSPKCCLGLLTLSWLLTTAHATLHTL
LMARLSFCAENVIPHFFCDTSTLLKLACSNTQVNGWVMFFMGGLILIPFLLL
IMSCARIVSTILRVPSTGGIQKAFSTCGPHLSVVSLFYGTIIGLYLCPLTNHNT
VKDTVMAVMYTGVTHMLNPFIYSLRNRDMRGNPGQSLQHKENFFVFKIVIV
GILPLLNLVGVVKLIMKYHSKSVA






10 20 30 40 50 60 70 80 90 100
bOR 17-210
Pred. Chimp.
Pred. Cow 466
110 120 130 140 150 160 170 180 190 200
bOR 17-210
Pred. Chimp.
Pred. Cow 466
210 220 230 240 250 260 270 280 290 300
bOR 17-210
Pred. Chimp.
Pred. Cow 466
310 320 330 340
bOR 17-210
Pred. Chimp.
Pred. Cow 466
Nature Precedings : doi:10.1038/npre.2007.1202.1 : Posted 2 Nov 2007


TM5
TM4
TM3
TM7
TM7'
TM6