universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Odorant receptors“

verfasst von / submitted by

Girgis Noser

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Pharmazie (Mag.pharm.)

Wien, 2016 / Vienna, 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 996 449
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Gleichwertigkeit Pharmazie
Betreut von / Supervisor:	Univ. Prof. Mag. Dr. G. Buchbauer

Contents

Acknowledgement	III
Zusammenfassung	IV
Abstract	V
General Introduction in German	1
General Introduction in English	2
1.1 Anatomy of olfactory system	3
1.2 Olfactory Epithelium	4
1.3 The olfactory bulb (OB)	4
1.4 Olfactory sensory neuron cell structure	5
2. G-protein-coupled receptors (GPCRs)	6
2.1 G-Protein-coupled receptors, classification	7
2.2 GPCRs structure.	10
2.3 Bovine Rhodopsin model	11
2.4 GPCRs activation and signaling.	15
2.4.1 Mechanism of signaling formation.	15
2.4.2 G-protein-dependent and independent signaling pathways	18
2.4.3 Terminating the signal	18
2.4.4 Types of ligand according to biological response	18
3. Olfactory receptors (ORs)	19
3.1 ORs genes as the largest gene known superfamily in GPCRs.	19
3.2 ORs genes classification and nomenclature	21

3.3 ORs genes distribution 24

3.4 ORs genes expression. 24

3.5 ORs 3D structure. 25

3.6 ORs-odorant interaction. 28

3.7 Signal transduction of olfaction. 30

3.8 Combinatorial coding and olfactory pattern. 33

4. Homology modeling 35

4.1 Human olfactory receptor OR1G1 37

4.2Description of some programs, server and website for modeling and docking 41

4.3 Secondary structure prediction of OR1G1 46

4.4 Conserved sequence motifs in OR1G1 48

4.5 Template and templates selection 49

4.5 Multiple sequence alignment, MSA 52

4.6 Building of the model 58

4.7 Evaluation and refinement of the model 60

4.8 Comparisons between Rhodopsin and OR1G1 62

5 Predicting of binding site and docking 63

5.1 Some of previous studies of the ORs binding site 63

5.2 Docking of menthol as agonist for OR1G1 using MOE 66

6 Summary 69

References 71

Appendix 79

Curriculum Vitae 88

Acknowledgement

Ich danke Herrn Univ. Prof. Dr. phil. Mag. pharm. Gerhard Buchbauer dafür, dass er mich auf dieses Diplomarbeitsthema aufmerksam gemacht, mein Interesse für dieses Thema geweckt, dass er mir dabei geholfen hat ein Konzept für diese Diplomarbeit auszuarbeiten.

Ein herzliches Dankeschön an Ao. Univ. Prof. Dr. Peter Wolschann für seine geduldige Betreuung während der praktischen Arbeit, das zur Verfügung-Stellen der entsprechenden wissenschaftlichen Literatur und Nachkorrektur der schriftlichen Arbeit, sowie seine Hilfsbereitschaft bezüglich sprachlicher Verständnisprobleme.

Ebenfalls möchte ich Dr. Anna Stary-Weinzinger dafür danken, dass sie mir bei Fragen bezüglich der praktischen Arbeit immer großzügig und schnell Auskunft gegeben und Hilfe bereitgestellt hat.

Auch möchte ich meiner Frau dafür danken, dass sie mir dabei geholfen hat, Zeit zum Verfassen meiner Diplomarbeit zu finden.

Dankeschön an Juliana Wohlmuther für das Übersetzen von Englisch auf Deutsch.

Zusammenfassung

Bis heute wurden die Geruchs-Rezeptoren weit unterschätzt, obwohl der Geruchssinn eine bedeutende Rolle im Leben von Mensch und Tier spielt. Speziell Tiere sind auf ihren Geruchssinn angewiesen um geeignetes von untauglichem Futter zu unterscheiden, Gefahren zu meiden, sich vor Feinden zu schützen, miteinander zu kommunizieren und einen passenden Partner zu finden. Eine große Zahl dieser Funktionen des Geruchssinns haben sich im Laufe der menschlichen Evolution verändert. Der Mensch ist nicht mehr so sehr auf seinen Geruchssinn angewiesen, da die visuellen und auditiven Sinne mehr an Bedeutung gewonnen haben. Trotzdem spielt der Geruchssinn weiterhin eine bedeutende Rolle in der menschlichen Wahrnehmung. Der Großteil der menschlichen Geruchs-Rezeptoren wird von Chromosom 1 und 11 codiert, dagegen findet keine Codierung an Chromosom 20 und Y statt. In der Vergangenheit wurde lange Zeit angenommen, dass Geruchs-Rezeptoren nur der Wahrnehmung und Verarbeitung von Gerüchen dienen, neue Studien haben aber ergeben, dass Geruchs-Rezeptoren nicht nur in der Nase vorkommen, sondern auch in anderen menschlichen Organen, wie der Haut, der Leber, dem Herz, den Nieren, der Lunge, dem Dickdarm, im Gehirn und sogar im männlichen Samen. Sie spielen eine wichtige Rolle bei mehr physiologischen Abläufen, als bisher angenommen. Jede Geruchssinn-Zelle weist nur eine Art von Geruchs-Rezeptoren auf und ist hoch spezialisiert auf eine kleine Gruppe von Geruchsmolekülen. Die prozessierten Nerven-Signale jeder Geruchssinn-Zell-Gruppe werden erst zu jeweils dem gleichen Glomerulus und anschließend zu anderen Teilen des Gehirns weitergeleitet, wo sie Signal-Muster bilden. Somit können wir bewusst Gerüche wahrnehmen. Richard Axel und Linda Buck haben dies entdeckt und wurden dafür im Jahre 2004 mit dem Nobel-Preis für Physiologie und Medizin ausgezeichnet. Die Geruchs-Rezeptoren gehören zur größten Gruppe der G-Protein gekoppelten Rezeptoren (GPCR) und ihre Hauptaufgabe ist die Geruchs-Wahrnehmung. Sie sind heptahelicale, transmembranale Proteine, deren C-Terminus intrazellulär und N-Terminus extrazellulär vorliegt. Bis jetzt sind die Kristall-Strukturen der meisten Geruchs-Rezeptoren noch unbekannt, aber über Computer-gestützte Programme und Datenbanken, die homologe Strukturen miteinander vergleichen, lassen sich die 3D-Strukturen der Rezeptoren vorhersagen. Dadurch kann die Bindungsstelle berechnet werden und auch das Bindungs-Verhalten zu verschiedenen Geruchs-Molekülen. In dieser Arbeit wird die 3-dimensionale Struktur des Rezeptors OR1G1 durch das Homologie-Modellierungs-Programm MOE berechnet und die Bindung von Geruchsmolekülen untersucht.

Schlagwort: Geruchs-Rezeptoren, Homologie-Modellierung, Bindungs-Verhalten, OR1G1, Menthol.

Abstract

Until now the odorant receptors are greatly underappreciated. Olfaction is very important for the live of humans and animals, especially animals depend more on olfaction than humans in many function such as to distinguish suitable food, to unfit foodstuff, to avoid dangers, to protect from enemies, to communicate that way, and to find a receptive mate. A lot of these functions have changed through the human evolution. Humans do not depend on olfaction to remain alive or to exist because they rely more on the visual and auditory senses now, but nevertheless it is still very important fact in humans live in comparison with other senses. The olfactory sense of animal is more advanced than that of humans. Most of human odorant receptors are distributed on chromosomes 1 and 11, on the other hand no odorant receptors are distributed on chromosomes 20 and Y. In the past it was believed that the odorant receptors have a function only for smell or olfaction process, but a lot of new researches have demonstrated that odorant receptors are not only found within the human nose, but also in many other human organs like the skin, liver, heart, kidney, lungs, colon, brain, and even sperm and play important role in several physiological functions. Each odorant receptor cell has only one type of odorant receptors and are highly specialized for a few odorant molecules; each odorant receptor cell sends their nerve processes to the same glomerulus and after that to other part of the brain where it is forming a pattern. Therefore, we can consciously experience the smell. Richard Axel, and Linda Buck discovered that and were awarded with the Nobel Prize in physiology or medicine in 2004. Odorant receptors belong to the largest family of GPCRs and underlie odorant perception. They are heptahelical transmembrane proteins with an intracellular carboxy-terminus and extracellular amino-terminus. Until now there is no crystal structure of most odorant receptors known, but several computer aided programs and database servers help to predict the 3 D structure of odorant receptors, through homology modeling with other template (or templates) to predict the 3D structure of ORs, and also through computer aided programs. One can predict the binding site, and allow docking with different odorants. In this work we try to predict the 3D structure of OR1G1 through homology modeling program MOE and docking with some odorant.

Keyword: Odorant receptors, Homology modeling, Binding site, OR1G1, Menthol.

General introduction in German

Der Mensch weist im Vergleich zu Tieren wie z.B. dem afrikanischen Elefant, dem Hund und der Ratte einen relativ schwach ausgeprägten Geruchssinn auf. Trotzdem können von ihm eine große Zahl verschiedene Gerüche wahrgenommen und unterschieden werden. Die Geruchsmoleküle, die an die Geruchs-Rezeptoren binden und sie dadurch stimulieren, weisen folgende spezifische Eigenschaften auf:

- eine gewisse Wasserlöslichkeit

- ausreichend Dampfdruck

- geringe Polarität

- Lipophilie

- Oberflächenaktivität

- ein Molekulargewicht kleiner als 300 Dalton.

Sie verfügen über eine oder mehrere funktionelle Gruppen (Osmophores), wie Carboxyl-, Aldehyd-, Hydroxyl-, Keton- und Ester-Gruppen und/oder Heteroatome, wie z.B. Stickstoff und Schwefel. Die Anatomie des Geruchs-Organs unterscheidet sich signifikant von Tierart zu Tierart. Das menschliche Geruchs-Epithel beinhaltet OSNs (olfactory neuron cells; Geruchs-Neuronen-Zellen), die jeweils über 20 bis 30 Zilien verfügen, auf deren Membranen sich die Geruchs-Rezeptoren finden. Jede OSN weist nur eine Art von Geruchs-Rezeptoren auf und kann deswegen nur eine kleine Anzahl an Geruchsmolekülen erkennen und unterscheiden. Das menschliche Genom weist 1000 unterschiedliche Gene für Geruchs-Rezeptoren auf, aber nur geschätzte 400 Gene codieren wirklich für funktionale Geruchs-Rezeptor-Proteine, die anderen 600 sind Pseudogene. Die Geruchs-Rezeptoren sind die größte Gruppe innerhalb der Familie der G-Protein-gekoppelten Rezeptoren. Sie sind heptahelicale, transmembranale Proteine, deren N-Terminus sich extrazellulär und C-Terminus intrazellulär befindet. Ein Geruchsstoff wird durch Bindung an den Geruchsrezeptor erkannt, wodurch ein elektrochemisches Signal initiiert wird, das die Information an spezifische Gehirn-Areale weitervermittelt und dort ein Signal-Muster verursacht. Die Geruchs-Rezeptoren werden durch die größte Gen-Familie codiert, deren Gene auf allen Chromosomen außer 20 und Y zu finden sind. Speziell viele dieser Gene sind auf den Chromosomen 11 und 1 lokalisiert.

General introduction in English

Humans have a poor olfactory system compared with the animals, such as African elephant, dog, and rat. A huge number of distinct odorants can be perceived and discriminated by the human odorant receptors. Odorant molecules which can stimulate the odorant receptors must have specific following properties:

1. Some water solubility
2. Sufficient high vapor pressure
3. Low polarity
4. Some ability to dissolve in fat (Lipophilicity)
5. Surface activity
6. Low molecular weight less than 300 daltons
7. Contain one or more of functional groups (Osmophores) such as carboxyl group (-COOH), aldehyde group (-CHO), hydroxyl group (-OH), ketone group (>C=O), ester group (-COO-) or hetero atoms such as nitrogen or sulfur.

The anatomy of olfactory system differs significantly by different animal groups. The human olfactory epithelium contains olfactory sensory neuron cells (OSNs), which have 20-30 cilia and on this cilia membrane odorant receptors are found. Each OSN have only one type of odorant receptors and can perceive and discriminate a few number of odorant molecules. In human there are about 1000 different ORs genes but have approximately less than 400 functional genes coding for odorant receptors and the others 600 genes are pseudogenes

The odorant receptors are the largest family of G-protein-coupled receptors superfamily, a structure of odorant receptors are heptahelical transmembrane protein with an extracellular amino - terminus and an interacellular carboxy – terminus.

The odorant can be discriminated by the ORs, throu binding of the odorant with these olfactory receptors which initiate the chemoelectrical transduction of odorant stimulus to specific part in the brain and forming the olfactory pattern.

The odorant receptor are encoded by the largest gene family, this genes are located on all chromosomes except chromosomes 20 and Y, in the other hand founded high distribution on chromosome 11 and 1.

1.1 Anatomy of olfactory system

The human olfactory region is found at the upper section of the human nose. The olfactory region of the two nasal passages is about 2.5 square centimeters (fig. 1). This area contains about 50 million olfactory sensory neurons cells (OSNs) (Leffingwell et al. 2002)

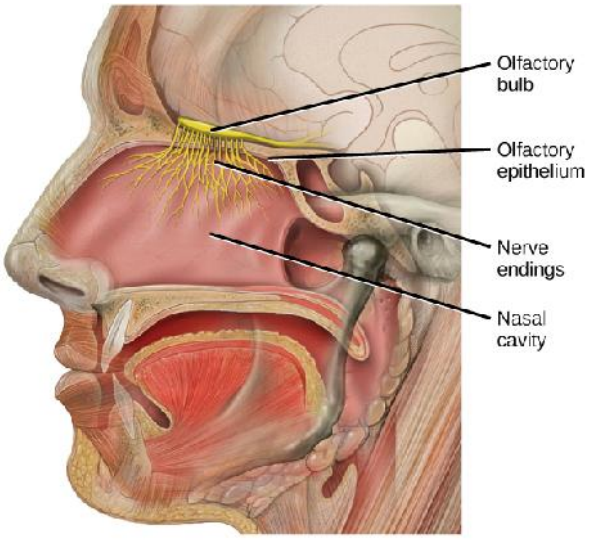


Fig. 1 Human olfactory region at the upper part or the nasal cavity (adapted from Neuroscience News December 9. 2013)

These regions are different in each species, for example in humans these are located at the upper part of the nasal cavity, in insects are 88% at the antenna, and 7% in the maxillary pulps, and fish have ORs in tubular chambers of the mouth. On the other hand the olfactory regions are also different in terms of size, for example these of dogs are about forty times larger than those of humans region, and mice have an olfactory region larger than humans, therefore they are not related to the body size (fig. 2).

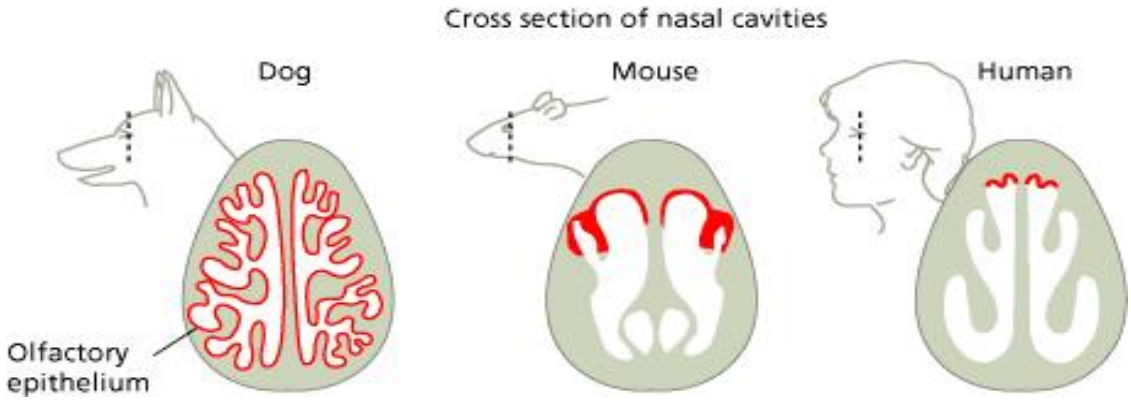


Fig. 2 the olfactory region size in Dog, Mouse and Human

Adapted from
www.nobelpriye.org/nobel_prizes/medicine/laureates/2004/illpres/7_species.htm

Different species show different numbers of odorant receptors, for example different ethnographic groups could have a different number of odorant receptors, mice have around 1500 ORs, and humans have near 1000 ORs. Some of human olfactory genes have been lost during evolution. Fish show a relatively small number of ORs of about 100 ORs. Even within the same species the numbers of odorant receptors are individually different.

1.2 Olfactory Epithelium

The olfactory epithelium is containing three cell types: basal cells, supporting cells, and olfactory sensory neurons cells (OSNs) (1) (fig 3)

1. Basal cells are stem cells, known as progenitor cells, which give rise to new olfactory sensory neurons cells (OSNs)

2. Supporting cells, which provide trophic, metabolic, and mechanical support for olfactory sensory neurons cells. These produce the mucus which cover the nasal epithelium.

3. Olfactory sensory neurons cells (OSNs) are actually bipolar neurons, these cells are the most important cells. Each of them possesses a thin dendritic rod that contains specialized cilia extending from the olfactory vesicle and a long central process which form the fila olfactoria. OSNs are replaced periodically and have a short lifetime about 30 - 60 days and also can be proliferate in response to toxic agents exposure, injury or nasal infection (2)

1.3 The olfactory bulb (OB)

The OB is located in ferior to the basal frontal lobe. It rlays between peripheral and central nervous system The layers (from outside toward the center of the bulb) are differentiated as following: Glomerular layer, External plexiform layer Mitral cell layer, Internal plexiform layer and Granule cell layer (Fig. 3)(1)

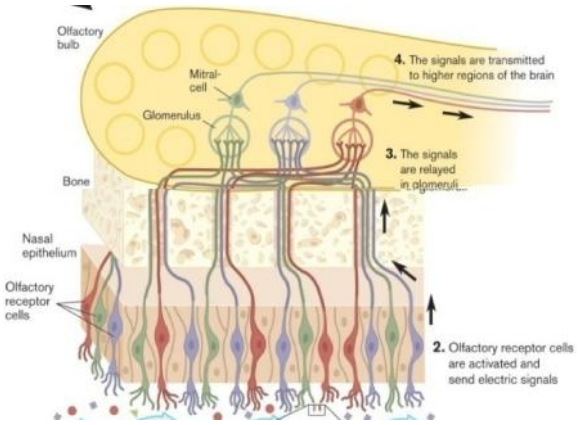


Fig. 3 Olfactory bulb and OSNs adapted from www.nobelpriye.org/nobel_prizes/medicine/laureates/2004/press.html

1.4 Olfactory sensory neuron cell structure

Humans have about 50million odorant receptor neurons. Odorant receptors can detect a huge number of different odorant molecules (3). Each single olfactory sensory neuron cell expresses only one gene of all the genes which are coded for ORs (Fig 4). Odorant receptor neurons are bipolar neurons with many dendrites at the side of the upper part of the nasal cavity and an axon passes through the cribiform plate going to the olfactory bulb. The olfactory receptor cell is a transduction cell for the odorant molecules within the olfactory system (4).

From the odorant receptor neuron cell many short cilia are extended and these cilia indulge and dendrite into the mucus layer which is covering the olfactory epithelium. At the surface of these cilia the Odorant receptors ORs are found. Odorant receptors belong to the GPCR superfamily (Buck and Alex 1991). The odorant receptors which are found in the same ORN are only one type of OR. The axons of ORNs which are of the same OR type lead to the same glomeruli in the olfactory bulb.

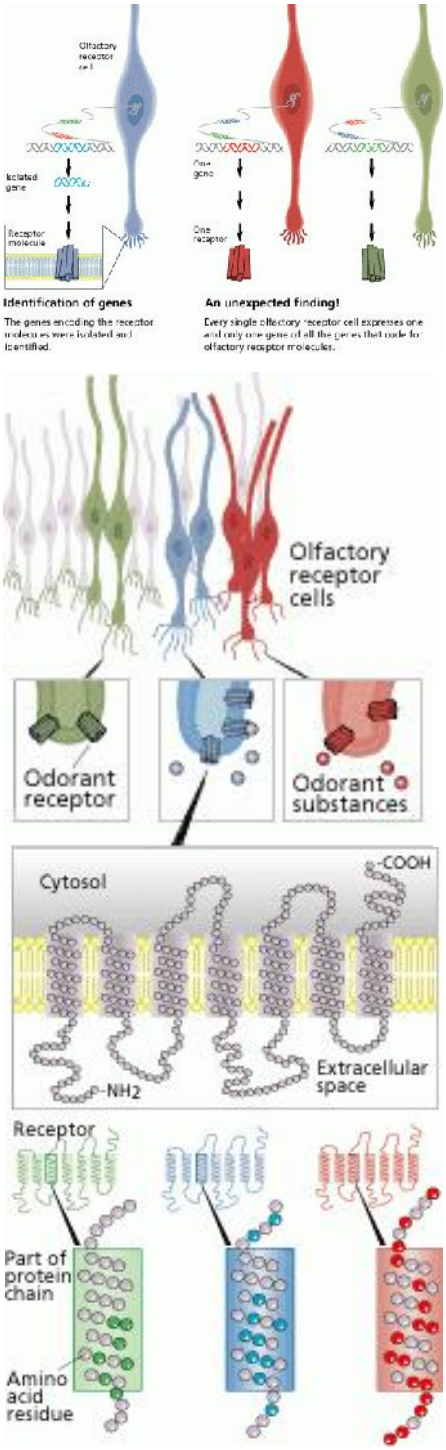


Fig. 4 Olfactory sensory neurons cell (OSNs) in the surface of its cilia found the (ORs) www.nobelpriye.org/nobel_prizes/medicine/laureates/2004/illpres/3_receptors.html

2. G-protein-coupled receptors (GPCRs)

G protein-coupled receptors (GPCRs) are accountable for a plurality of cellular restraints to outer stimuli. They act as mediates for our sense of vision, smell, taste, and pain (5), and they also contribute to cell recognition and communication processes so they are a distinct superfamily for drug targets. A huge number of diseases exist due to GPCRs malfunction and more than 30% of our medicines act on these receptors. It is important to study the structure of these receptors to develop smart drugs with high selectivity, high efficacy and low side effects. GPCRs are symbol for the largest superfamily diverse of membrane receptors. More than 1000 humane genes encode G protein-coupled receptors (6) around 350 genes of GPCRs regulate growth factor, hormones, and other endogenous compounds. About 150 genes of human GPCRs are of unknown function (orphans). Heterogeneous ligands can bind with GPCRs and activate them, for example hormones, pheromones, ions, photons, taste and odorant molecules, etc.

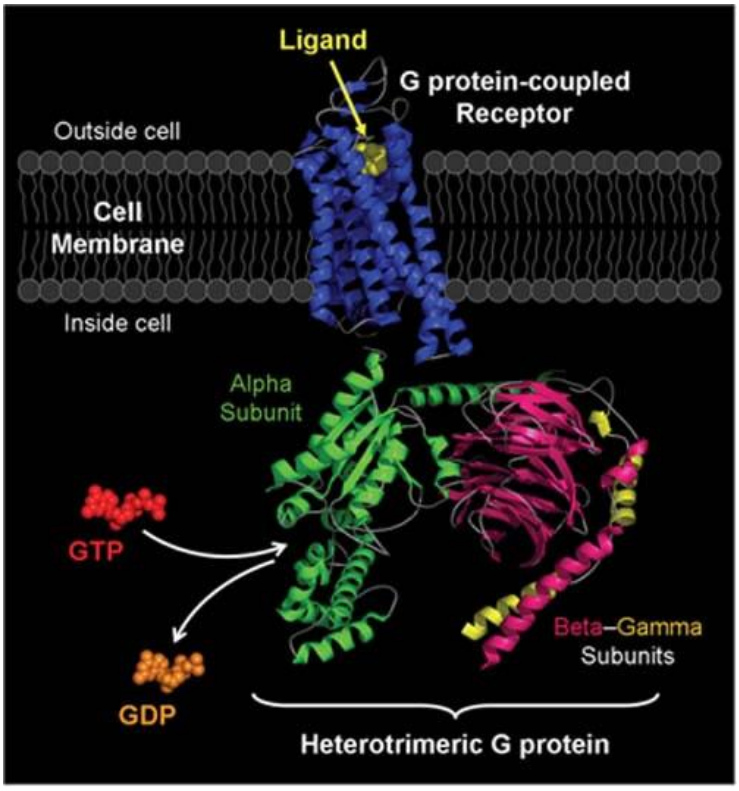


Fig. 5 Three dimensional crystal structure of a G-Protein coupled receptor (GPCR)

Illustration by William Oldham

http://www.mc.vanderbilt.edu/lens/article/?id=112&pg=999

2.1 G-Protein-coupled receptors, classification

Until now the exact number of GPCRs is obscure, from sequence analysis of genome it is predicted that about 4% of the coding genome is responsible for GPCRs. This means that around thousand different human genes code for GPCRs (6)

Many classification designs have been suggested,

2.1.1 At first GPCRs are classified into 3 fundamental classes (A, B, and C) without noticeable shared sequence homology between classes.

About 85% of GPCRs genes (662 genes) belong to Class A (Rhodopsin-like), over half of these are odorant receptors.

Class B (Secretin-like) includes 15 genes

Class C (Glutamate receptor-like) includes 22 genes

Besides there are 92 other genes, 33 of them are Adhesion, 11 are Frizzled, 25 are Taste type-2, and 23 are unclassified.

Class A is divided into the olfactory receptors (391 genes), and known endogenous compound or orphan receptors (271 genes) fig 6. (7)

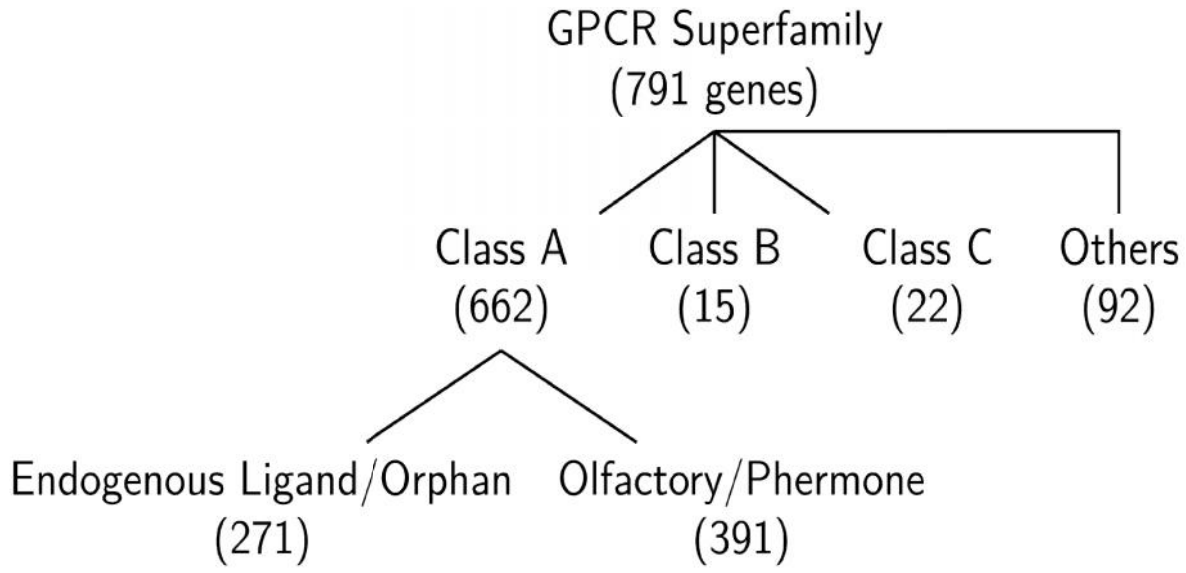


Fig. 6 Classification of GPCRs superfamily (7)

Regardless of the poorness of sequence homology between the classes, all GPCRs have a conjoint signal transduction mechanism and conjoint structure.

Class A (Rhodopsin-like) has been subdivided into nineteen subclasses (A1 – A19) (8) and an unclassified subclasses.

2.1.2 In 1994 Attwood and Findlay divided GPCRs into 6 classes (A-F) in accordance with their similarity of function and sequence homology (9)

Class A Rhodopsin-like.

Class B Secretin receptor family.

Class C Metabotropic glutamate/pheromone.

Class D Fungal mating pheromone receptors.

Class E Cyclic AMP receptors.

Class F Frizzled/Smoothened.

2.1.3 Finally supplied with GRAFS classification system through comprehensive phylogenetic analysis of the human repertoire. In GRAFS classification GPCRs are classified into five fundamental families Rhodopsin, Adhesion, Glutamate, Friyyled, and Secretin (10) (fig 7). Rhodopsin family is the greatest family containing 683 individual receptors in humans (11) and divided into 4 sub-classes alpha, beta, gamma, and delta (Fredriksson et al. 2003) ORs belong to delta sub-class and form a monophyletic group.

Rhodopsin family is characterized by short N-termini and binding with a broad assortment of ligands, like odorants, photons of light, flavors, biogenic amines, and small peptides etc (12).

Glutamate family is characterized by long N-termini which act as regions for ligand binding.

Adhesion family is distinguished by long N-termini which containing a plethora of multiple domains.

Frizzled family is characterized by long cysteine-rich N-termini. (13)

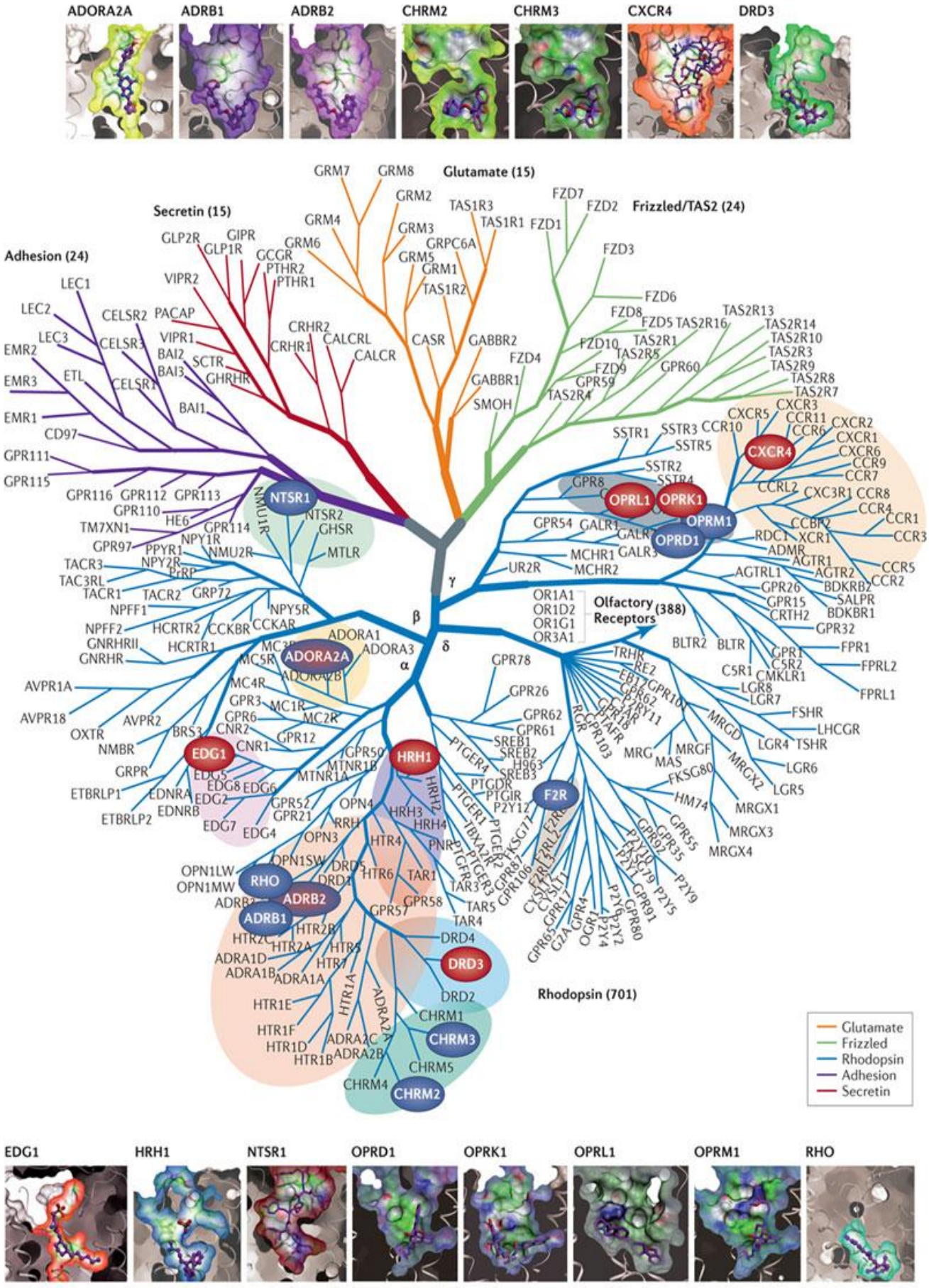


Fig. 7 GRAFS classification system adapted from (14)

Nature Reviews | Drug Discovery

2.2 GPCRs structure.

GPCRs are formed by a bundle of seven transmembran helices (7TM) spanning the cell membrane seven times. The 7TM are linked by three intracellular loops (3CL) at the side of the cell cytoplasm and are linked by three extracellular loops (3EL) at the side of extracellular membrane (fig 8).

The transmembrane segments are highly conserved sequence motifs, as well as the intracellular carboxyl terminus (COOH) and (CL3) between helix five and six have highly conserved sequence motifs. But the extracellular N-termini show high diversity in sequence motifs and length between different classes and conserved sequence motifs within members of the same class fig 8.

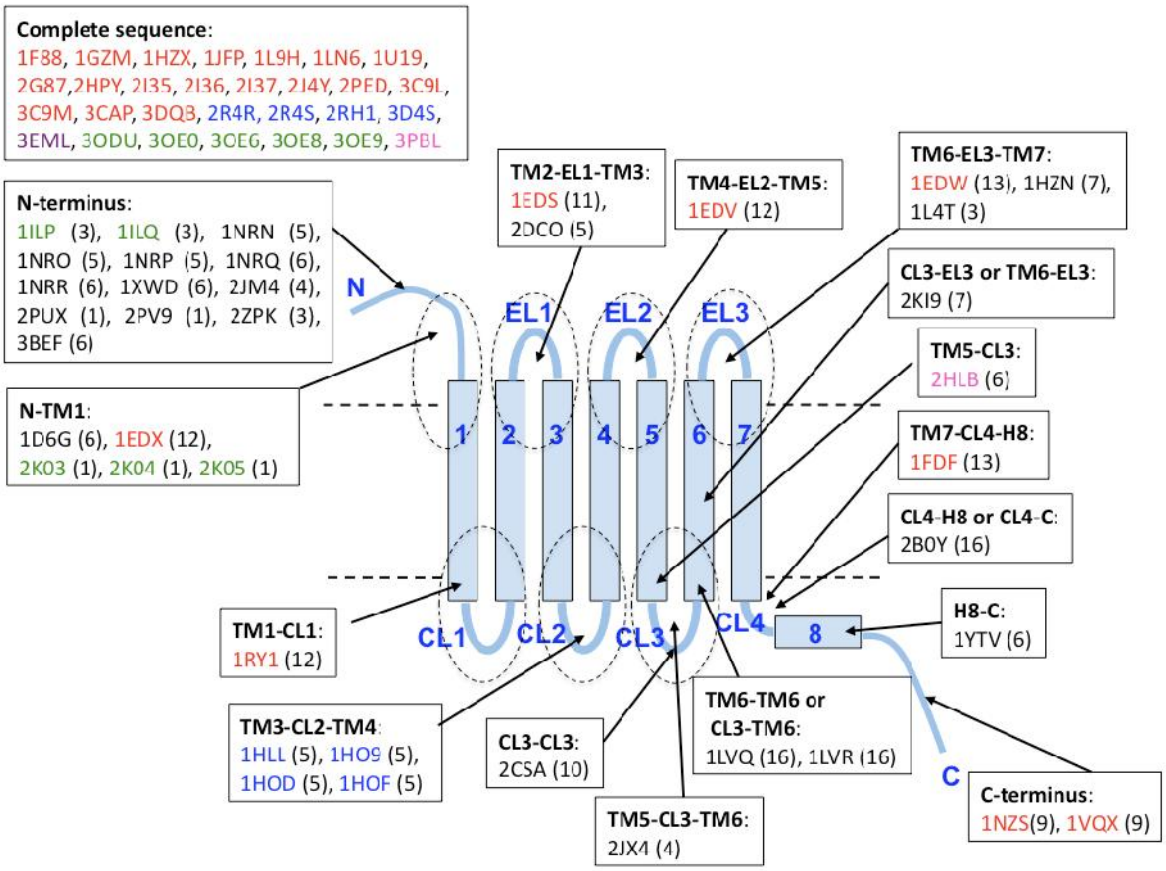


Fig 8. 29 structures of five subfamilies of class A GPCRs Bovine Rhodopsin colored in red, adrenergic receptor in blue, adenosine receptor in magenta, chemokine receptor in green, and dopamine receptor in pink. The numbers in parenthesis represent the hit count of PDB regions to the sequences identified in SEVENS database. For details see (15)

The understanding of GPCRs structures was considerably facilitated through the analysis of the crystal structure of Bovine Rhodopsin by using X-ray crystallography.

Bovine Rhodopsin was the first GPCR whose crystal structure was detected with a good resolution. Regrettably, Bovine Rhodopsin remains one of the few solved crystal structure yet, in spite of many years of research, attempt, and trial.

GPCRs are flexible polypeptides and span in the fatty environment of the cell membrane, so it is not easy to obtain the crystal structure of GPCRs.

2.3 Bovine Rhodopsin model

Rhodopsin is a member of class A of GPCRs responsible for vision by the transfer of photons into a signal (16). Its highly resolution crystal structure of the inactive conformation was determined by Palczewsk et. al. in 2000 (17). The geometric structure of Rhodopsin like these of other GPCRs, contains 7TMs with 3EL and 3CL (Fig 9 A, B).

The second extracellular loop (EL2) is the tallest loop, which is significant for the ligand remaining in the binding bouquet, in order to act as lid to prevent the ligand from dissociating too fastly from the binding bouquet. EL2 has 2 beta sheets (Fig 9), and their stability in their positions due to:

1- Hydrophobic interaction between residues tyr178, pro180, met183, cys185 and cys187 of EL2, and the surrounding residues of helices 1,2,3, and 7
2- Disulfide bond between cys187 and cyc110 (Fig 9 B, C).
3- Hydrogen bonding among glu181 and tyr192 of EL2, tyr269 of helix six and water.
4- Salt bridge by asp190 and arg177 (18)

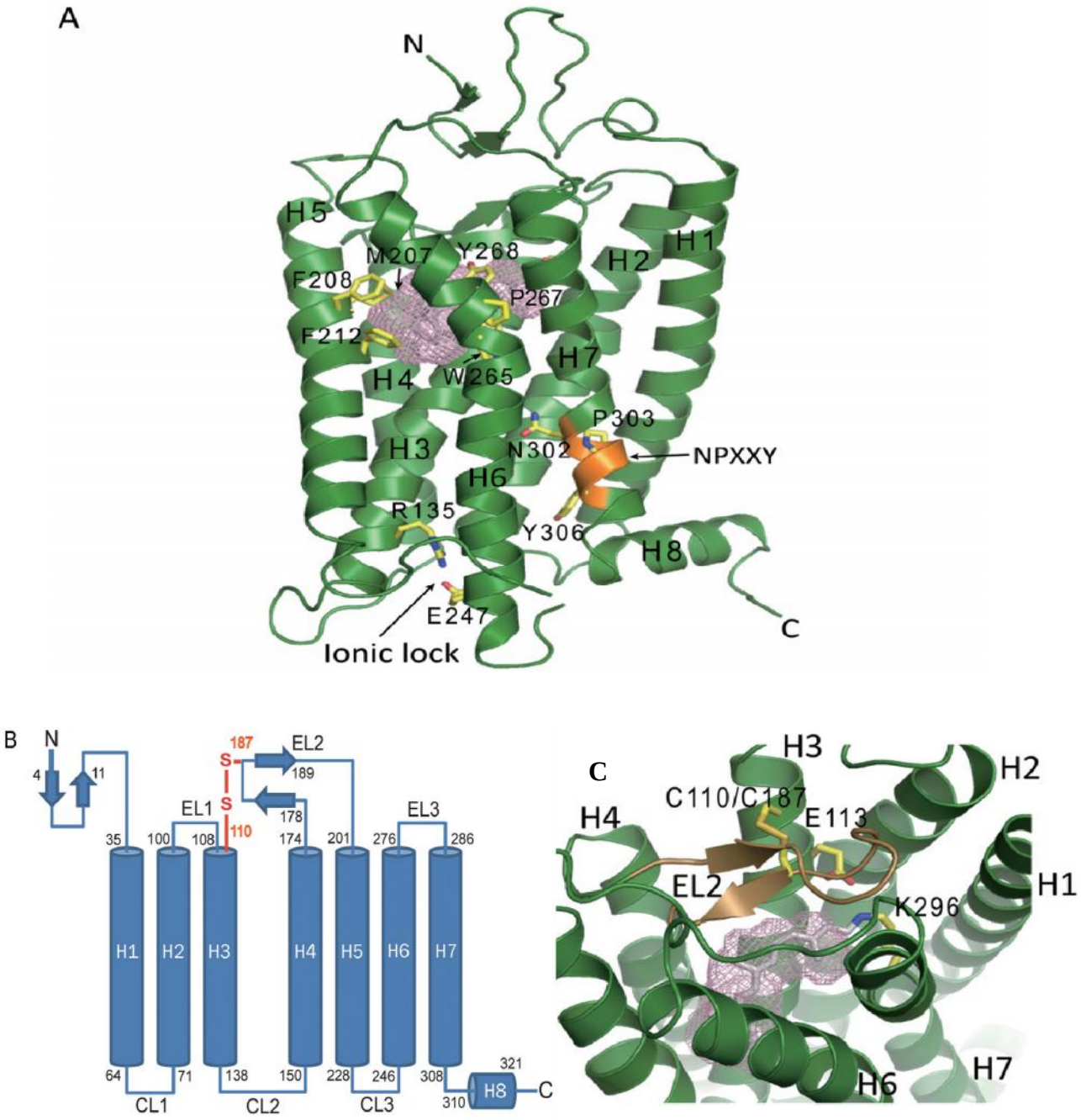


Fig. 9. The structure of Bovine Rhodopsin (PDB: 1F88).

(A) This is vertical picture showing the seven-transmembrane helices in 3D structure, the retinal in gray stick which is bound to the binding pocket shown as pink mesh. The side chain of the residues which are surrounding the binding pocket are shown in yellow color. Other features include the ionic bridge (yellow sticks) and the NPXXY motif (orange). (B) The secondary structure of Bovine rhodopsin contains 7TMs, N- terminus, carboxyl terminus, 3CLs and 3ELs. A disulfide bond connecting EL2 to helix 3 is shown in orange between C110 and C187. EL2 includes 2 beta-strands. (C) The lid of the pocket which is formed by EL2 is shown in dark brown. The disulfide bond can be seen between C110 and C187.

Adapted from Acta Pharmacologica Sinica 2012, 33, 291-299

When the conformation structure is inactive a salt bridge between Arg135 of helix three and Clu247 of helix six is preventing the C-termins of G alpha protein to arrive at its binding site. Through the activation of Rhodopsin by light, 11-cis retinal is transferred to All-trans retinal (Fig 10 A,B,C), thereby helix six is displaced and forms a hydrophobic cleft at the side of cytoplasmic end through breaking the ionic lock. The cleft which is formed allow the C-terminus of G alpha protein to bind with Rhodopsin (Fig 10 D,E).
Conserved NPXXY sequences on helix seven have a significant function, this special sequences move toward helix six and the cleft residue Tyr306 flips in the direction of helix six and breaks the ionic lock thereby helix six is displaced away and forms a cleft.

C-terminus of Rhodopsin is able to interact hydrophobically through its amino acids Phe313, Cys316, and Met317 with Leu57, Val61 of helix one and His65 of CL1 (Fig. 10). And the serine residues which are found in C-terminus of Rhodopsin (Ser334, Ser338, and Ser343) terminate the activation of rhodopsin when they have been phosphorylated and binding with errestin (19).

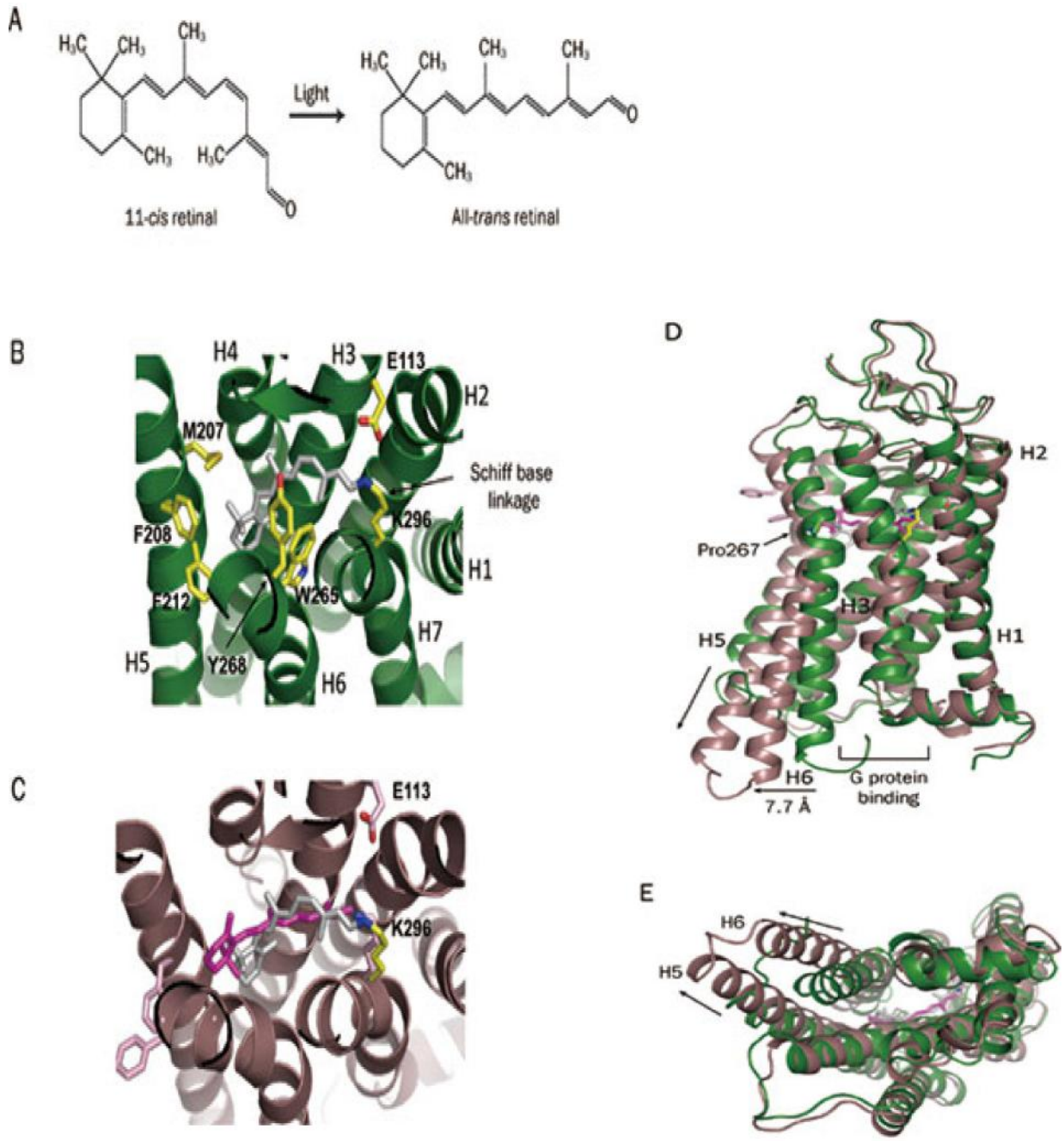


Fig. 10 Different structures of retinal and conformational changes in rhodopsin. (A)The ligand 11-cis-retinal is activated by light and turns to all-trans-retinal. (B) shows the binding site of the inactive form of retinal 11-cis-retinal in gray color, and the surrounding side chain in yellow color (PDB: 1F88). (C) The bound retinal after activation through the light (All-trans-retinal) and conformational change of rhodopsin (PDB: 3PQR)(D) The activation of the bound retinal causes a movement or displacement of cytoplasmic end of helix 6 and elongates of cytoplasmic end of helix 5. That forms a cleft for binding of C-terminus of G-protein. (E) The horizontal view of (D)

Adapted from Acta Pharmacologica Sinica 2012, 33, 291-299

2.4 GPCRs activation and signaling.

2.4.1 Mechanism of signaling formation.

The binding of the ligands at the side of extracellular region of GPCRs leads to several changes in the GPCRs structure which allows the C-termini of G alpha protein to bind at the side of intracellular part to form a ternary complex. Ternary complex acts as an activated signaling unit (Fig. 11) (20).

Grip of the ligand to GPCRs increases the affinity of GPCRs to the G-protein, and contact of the G-protein to the GPCRs increases the receptors affinity to the ligand (20)

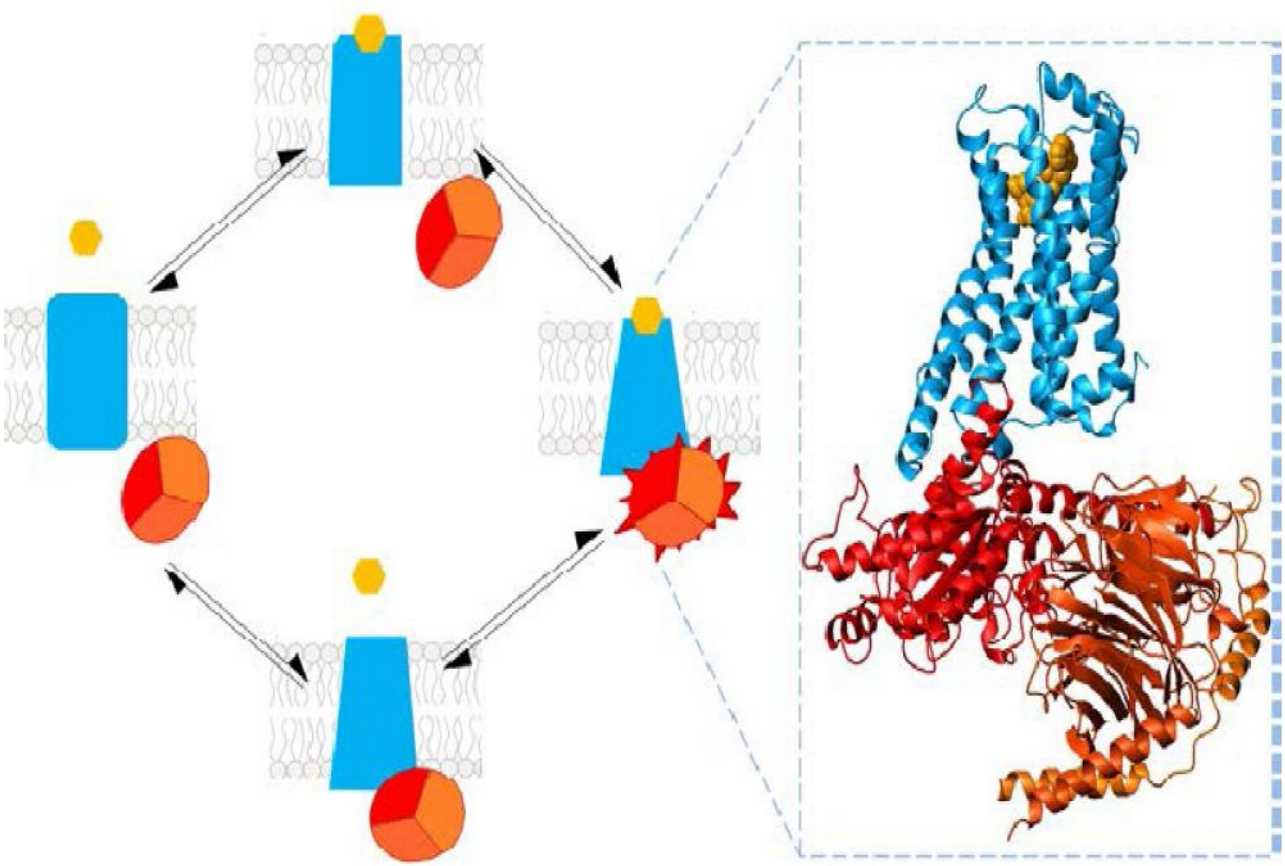


Fig. 11 A) Ternary complex model. **B)** Crystal structure of an active ternary complex. Drawn from the file 3SN6.pdb by Molmol (20).

Adapted from
http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2012/advanced.html

GPCRs have two dynamic structure conformations, an inactive structure conformation and active structure conformation.

Each GPCR shows affinity with a few numbers of ligands to grip. When this GPCR grips the ligand, GPCR transfers from the inactive conformation structure to the active conformation structure and its affinity with G-protein is increased. That means the ligand does not pass the cell membrane, but the signal is transferred from the outside to the inside of the cell through activation of the GPCRs. The activated GPCRs binds the C-terminus of the G-alpha protein thereby G-protein is activated. The receptor, ligand and G-protein form a ternary complex which is considered as an activated signaling unit. After forming the ternary complex numerous reactions inside the cell are started, and G-protein dissociates into G alpha protein and beta and gamma complex (21). The G alpha protein stimulates enzymes such as adenylate cyclase which transfers AMP to CAMP which acts as second messenger. The same active conformation of GPCR activates numerous G-proteins that lead to an amplification of the signal (22)

Hubbell and Khorana discovered that the helix six movements are significant for the activation of GPCRs, when the ligand binds to GPCRs at extracellular side, the helices open up at the intracellular side and create a binding site for the C terminus of G-protein. Tiny changes of GPCRs at the extracellular side by binding the ligand propagate into greater structural changes at the intracellular side of the GPCRs (20)

Kobilka and coworkers concluded that the C-terminus of G-alpha protein penetrates into this bouquet and binds to the GPCR. Through this binding to the GPCR the G-alpha protein is activated and forms a ternary complex.

The ternary complex formed by ligand, GPCR and of G-alpha protein is short-living, which allows the GPCR and ligand-complex to activate numerous G-alpha proteins and through that to amplify the signal Fig 12.

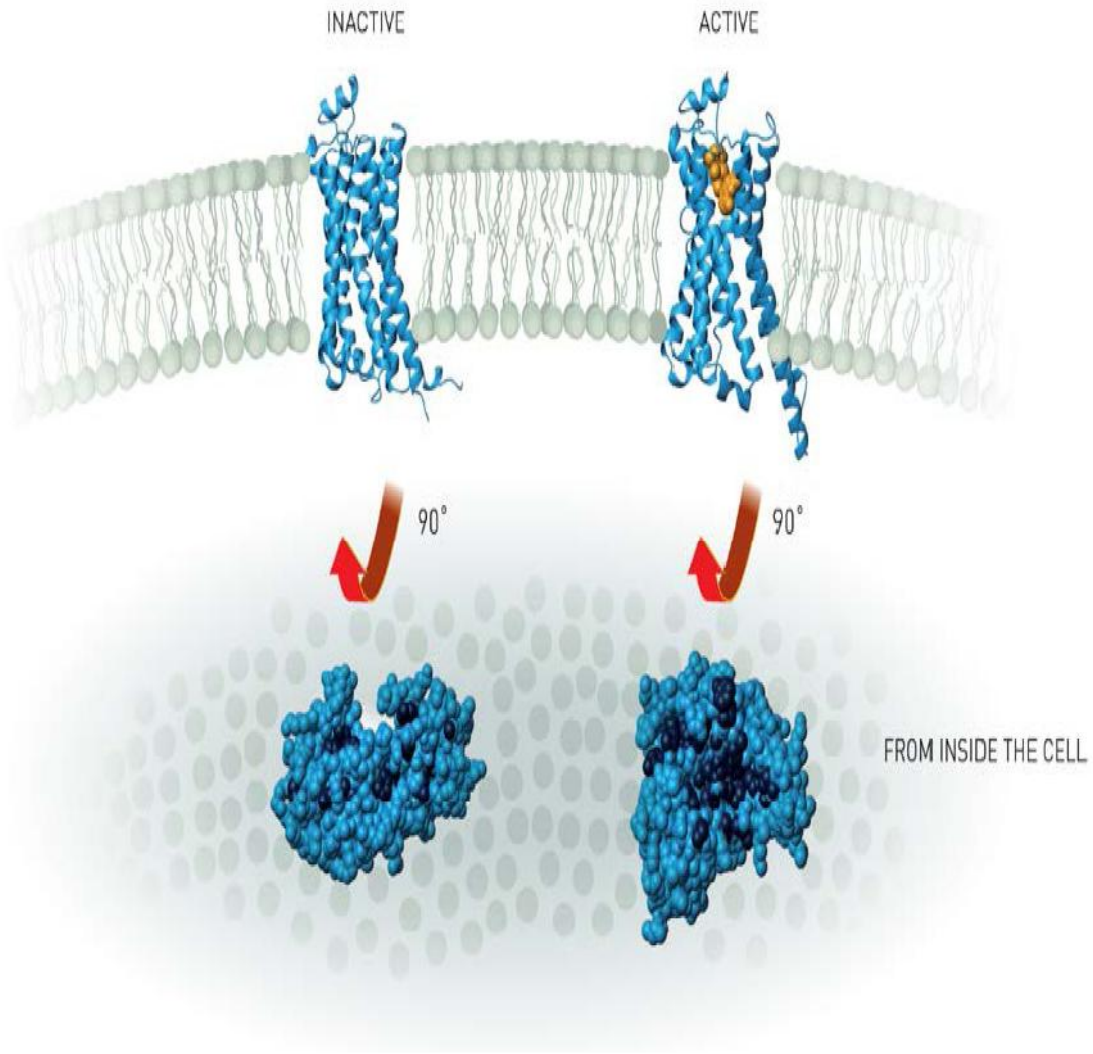


Fig. 12 Structural basis of the GPCR signaling mechanism.

Non-activated βAR (2BAR.pdb) is shown to the left, and activated βAR bound to ligand and G-protein (3SN6.pdb) to the right. At the top, the receptor in the membrane is drawn with blue ribbon that traces the backbone.

The vertical view shows the elongated helix 5 in active conformation structure which forms a large surface to contact the G-alpha protein.

The bottom view is from the inside of the cell membrane, with the receptor shown using a space-filling model with hydrophobic side chains in dark blue

Adapted from
http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2012/advanced.html

2.4.2 G-protein-dependent and independent signalling pathways

The same GPCR signal out of various pathways is relying on the identity of the ligand. The flexibility of the receptor and flexibility of its environment around make the receptor liable to bind with numerous ligands and thereby have various active or inactive forms.

Lefkowity proposed a G-protein independent pathway. That means the GPCR signal from the outside into inside of the cell could be also performed by other proteins not only by G-protein such as arrestins and others. The signaling pathway is chosen according to ligand identity.

In case of ligands GPCR response is probably followed by both G-protein dependent and G-protein independent signaling pathways.

The extracellular side of the receptor is significant for choosing the ligand and the ligand identity is significant for choosing or determining the signaling pathway.

2.4.3 Terminating the signal

1- By dissociation of the ligand.
2- By phosphorylation of the intracellular region, that makes the receptor less sensitive to the ligand (23), the phosphoraylation increases GPCRs affinity to arrestins (24)
3- By specific enzymes such as rhodopsin kinase (2)
4- By internalization through endocytosis and recycling of GPCRs (found by Kobika)(26)

2.4.4 Types of ligand according to biological response

1. ligand acts as agonist when it is binding to the receptorand activates the G-protein inside the cell

2. ligand acts as inverse agonist when it is binding to the receptor and stabilizes the inactive form

3. ligand acts as antagonist when it competes with the agonist and blocks the agonist-binding site (27).

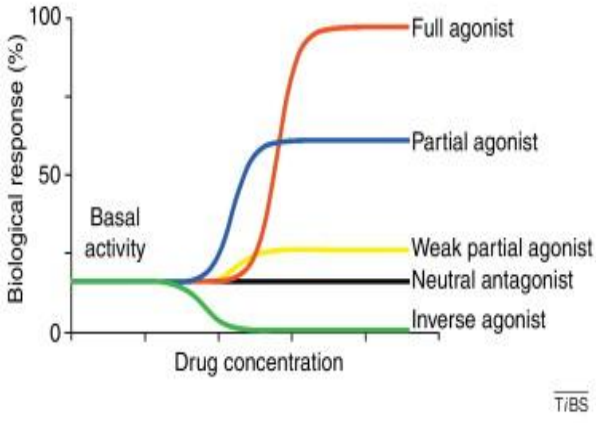


Fig. 13 types of ligand agonist, antagonist, and inverse agonist. Adapted from (28)

3. Olfactory receptors (ORs)

3.1 ORs genes as the largest gene known superfamily in GPCRs.

Odorants are detected by the large superfamily of odorant receptors (ORs). ORs are encoded by the largest known genes superfamily in GPCRs (Fig 14).

OR genes are unequally distributed among different species. For example elephants have more than 3000 ORs (15), catfish about 100 ORs (30) and humans around 900 ORs. OR genes were first discovered in mice, where around 1500 OR genes were found. In humans chromosome 11 bears the plurality of OR genes but the first OR gene was mapped on the chromosome 17.

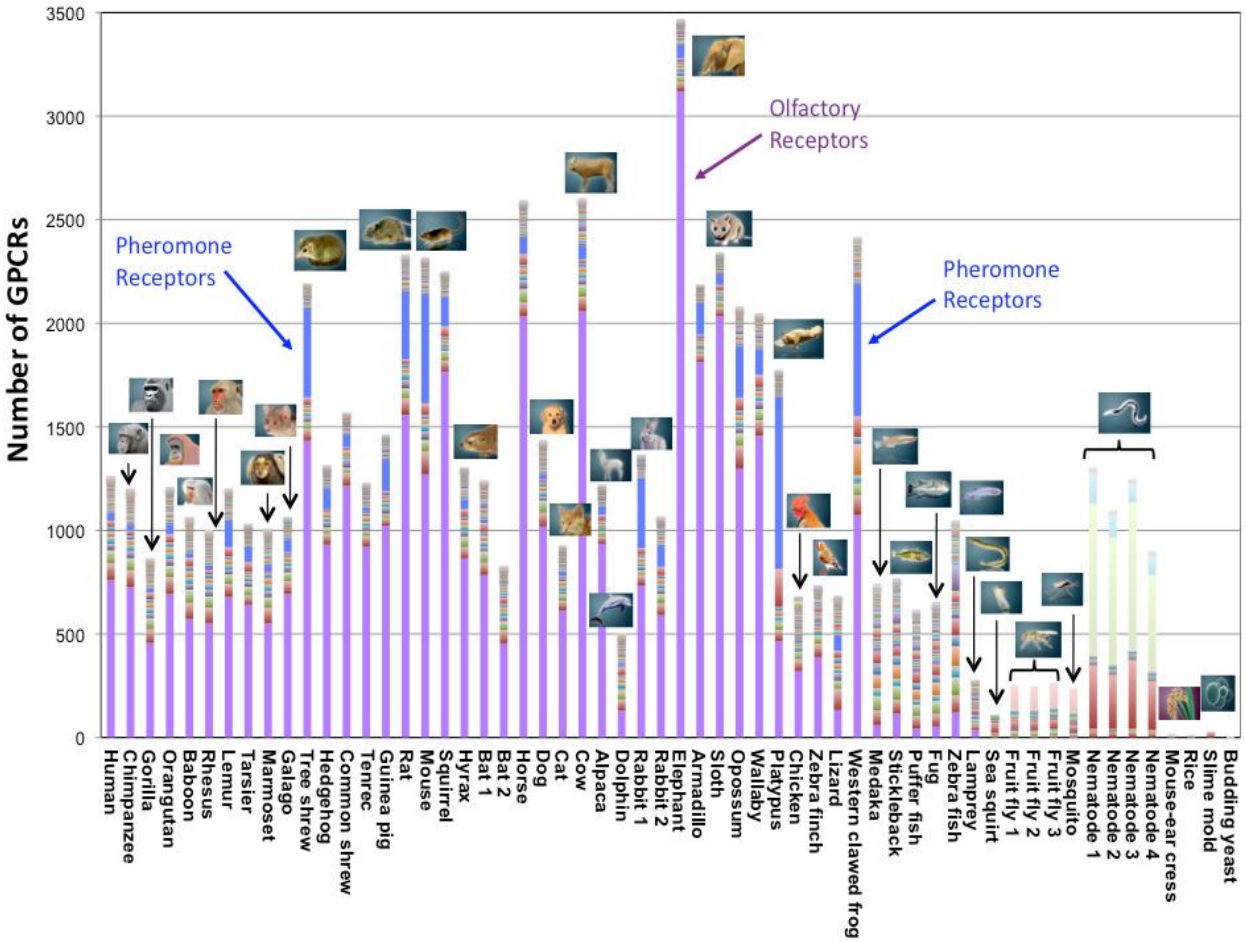


Fig 14 OR genes represent about 70% of GPCRs in 58 eukaryotes (colored in purple), and human OR about 60% of GPCRs. Elephants have the largest number of OR genes more than 3000, sea animals have rather small numbers of OR genes around one hundred. Adapted from (15)

About 1% of all the human genome are OR genes (29). The genes of ORs without introns (intervening sequence) thereby this facilitates their identification from the DNA (30). The human genome has around 900 OR genes, 391 intact genes and 466 pseudogenes (Adapted from HORDE, the human olfactory data explorer at 24.06.2015). Human olfactory receptor genes are found on around all chromosomes except two chromosomes (20 and Y) and arranged randomly in clusters (31) (Fig 15). Pseudo genes in humans represent about two-third of all OR genes, but in mice the amount of 20% of all OR genes, OR genes of mice are around 1500, which means the functional ORs in mice are three times larger than in humans, because of evolution in human genome or frameshift (32). Human OR genes are highly distributed on chromosome 11 and 1, around 42% of OR genes are found on these two chromosomes (Glusman et al. 2000), the rest is found mainly on chromosomes number 6, 9 and 14, and the other chromosomes have a few percent. Human OR genes are unequally distributed among 51 loci, one or more of the functional OR genes are located on 38 chromosomal loci and the other loci have pseudogenes. All human ORs families are composed of 172 subfamilies, the members of the same subfamily are 60% or more are identical in polypeptide sequence. Generally subfamilies are encoded by single chromosomal loci, and numerous loci encode one or a low number of subfamilies (31). There is a diversity of OR genes among different individuals, causing the variety of the sense of smell (32).

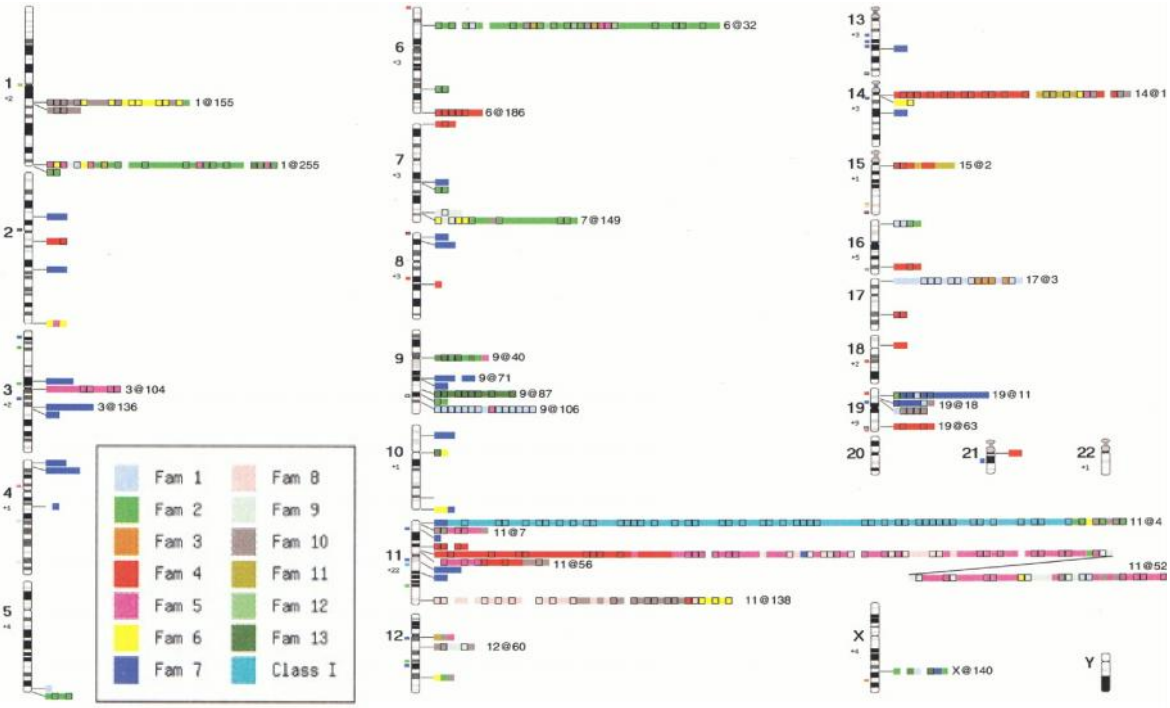


Fig. 15 Human OR genes are localized to 51 different chromosomal loci distributed on all chromosomes except 20 and Y, each color indicates on specific family. Adapted from (33)

3.2 ORs genes classification and nomenclature.

ORs are classified into two classes, class I and class II (Fig. 16 c). Class II is containing 14 families (from family 1 to family 14). Class I is containing 4 families (family 51, 52, 55 and 56) (Fig. 17). Class I ORs at the beginning are identified in fish, and perceive them water soluble odorants. Class II ORs are identified in mammalians, and perceive them airborne odorants.

There is variety of the sizes of each family for example family 2, 4, 5 and 7 stand for the largest families (Fig. 16 a) (33).

From the phylogenetic tree it is show that some families are related to each other as families 5 and 8, families 1 and 7, families 2 and 13, families 11 and 6, families 12 and 4, and families 51 and 52 Fig. 16 b.

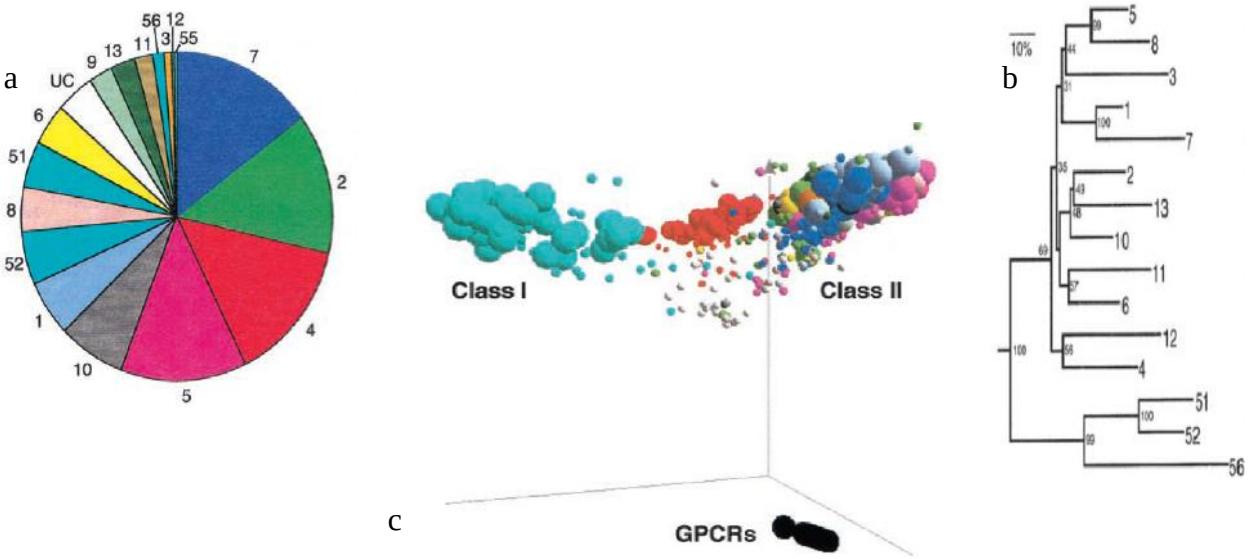


Fig. 16 a) Different families of ORs b) Phylogenetic tree of ORs

c) Principal component analysis (PCA) of all ORs class I and class II. Spheres represent ORs, small spheres denote OR pseudogenes. Adapted from (33)

At the beginning it was assumed that human ORs do not have any belonging to class I, but after that it was discovered that a large number of ORs from class I can be found at chromosome 11 of human, around 10% of the entire human ORs belong to class I (33).Families of ORs are defined by a 40% or more identical polypeptide sequences. Subfamilies of ORs are defined by a 60% or more identical polypeptide sequences (31). All of human ORs families are composed of 172 subfamilies, 94 subfamilies have only one OR, 73 subfamilies contain (from 2 to 6) different ORs, and 5 subfamilies contain (8 or 9) different ORs.

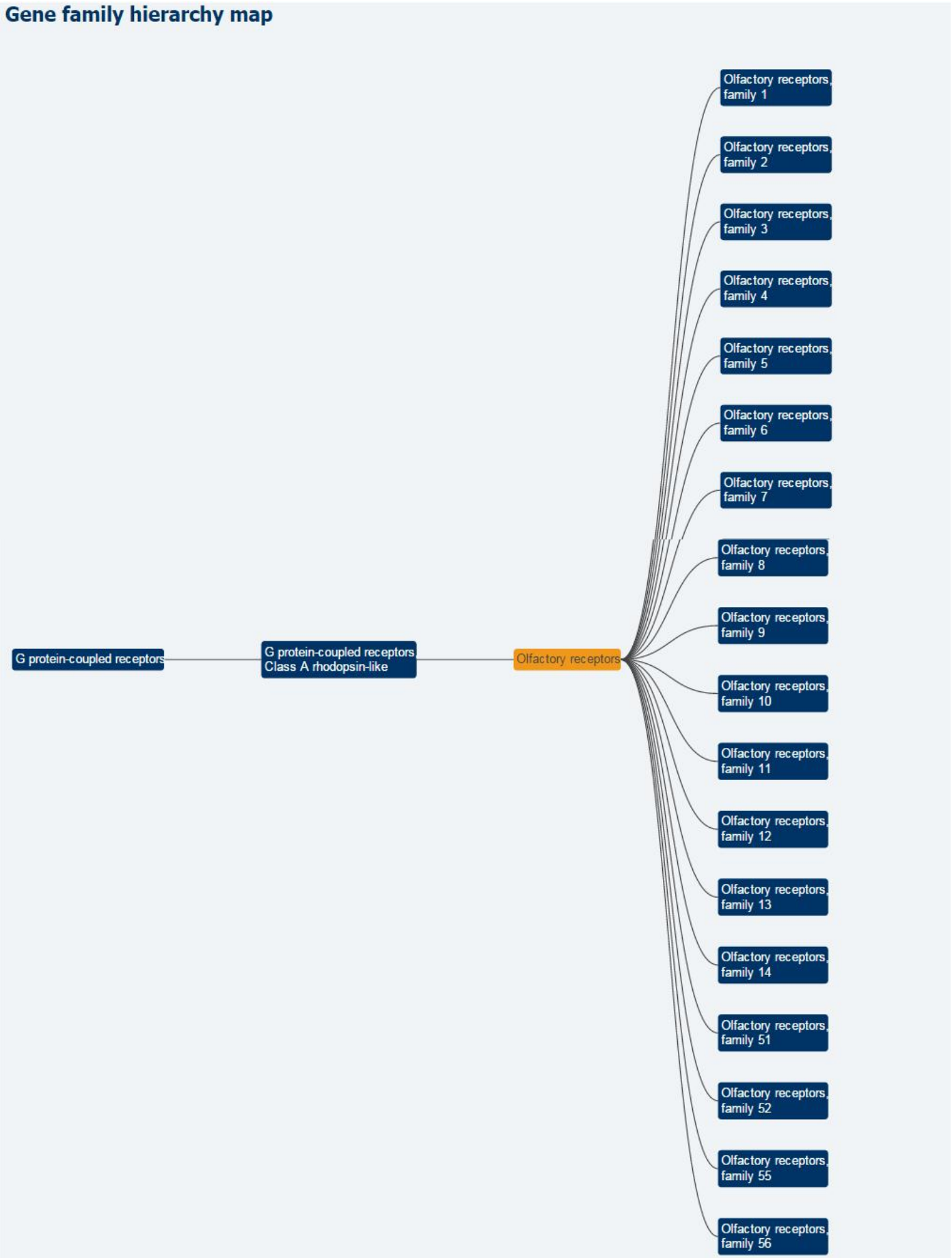


Fig. 17 ORs class II families (1-14), and class I families 51, 52, 55 and 56 adapted from www.genenames.org/cgi-bin/genefamilies/set/141 at 28.06.2015

ORs nomenclature

ORs nomenclature is based on family and subfamily classification.

This is proposed the nomenclature schemes for OR genes.

ORnXmP

OR: means olfactory receptor or odorant receptor.

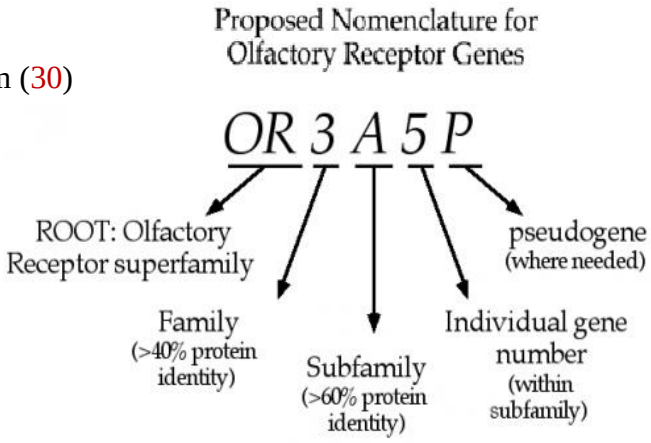
n: that is the number representing the family number, class II families are numbered beginning at 1 (presently 1-14), class I families are numbered beginning at 51 (presently 51-68) (30).Each family member has 40% or more identical in polypeptide sequences.

X: a single letter or more denoting a subfamily, 60% or more identical in polypeptide sequences.

m: that is the number representing an individual member (isoform)

P: Pseudogene in indicated

Example for ORs schemes Fig. 18 adapted from (30)



Five rules for nomenclature:

1- OR symbol for all species except mouse (*Or* for mouse)in italic symbols.

2- P symbol denotes a pseudogene where needed except mouse (*Ps* in mouse).

3- When a gene family or subfamily is known, the subfamily letter and gene number are included explicitly.

4- Human nomenclature applies for all species except mouse

5- The name of mRNA and protein (including mouse) by capital letters without italics or hyphens (30).

3.3 ORs genes distribution

Human OR genes are distributed at all chromosomes except chromosomes 20 and Y, Chromosomes 18 and 21 have only pseudogenes. All functional human OR gene are distributed in all human chromosomes except 18, 20, 21, and Y.

Based on the analysis of OR genes distribution as shown in the appendix

Chromosome 11 has the majority of functional ORs and a big number of class I functional OR genes from family 51, 52, and 56.

Chromosome 11 has functional OR genes from 12 different families.

Chromosomes 1 and X each one has 9 different families.

Families 8, 51, and 52 are found only at chromosome 11. Family 12 only 3 genes are found at chromosome 6.

3.4 ORs genes expression.

Odorant receptors are expressed in various different non-olfactory tissues not only in olfactory epithelium (Flegel et al. 2013). A lot of recent studies have described physiological functions of odorant receptors which can be found in other different non-olfactory tissues and play an important role in its physiological function. Example of this tissue enterchromaffin cells of the Gut (Braun et al. 2007), tongue, liver, heart, kidney, lungs, brain, spermatozoa (Veitinger et al. 2011), prostate epithelial tissue (Neuhaus et al. 2009), and skin keratinocytes which have odorant receptors (34), more than 15 odorant receptors that found in the nasal cavity are also found in the skin like as OR2AT4 which is stimulated by synthetic sandalwood odorants (Sandalore and Brahmanol) leading to an increase of the intracellular calcium concentration and thereby inducing wound-healing processes (35).

There are two studies that have described the functional role of ORs which are expressed in human spermatozoa through OR1D2 (Spehr et al. 2003) and in mouse spermatozoa through mOR23 (Fukuda et al. 2004). Around 50 ORs have been found in the testes of different species as human, dog, mouse and rat (34). The olfactory receptors in spermatozoa enable sperm cells to find their way directly to the unfertilized egg (Veitingel et al. 2012).

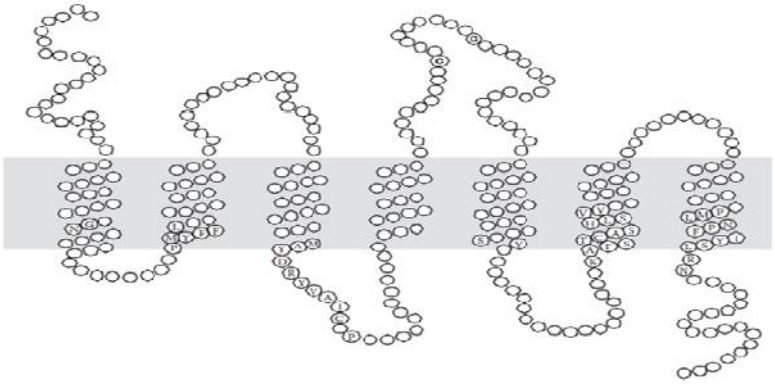
The ORs found in prostate are OR51E2, also named (prostate specific GPCR) or (PSGR). Activation of (PSGR) by odorant beta-ionone inhibits proliferation of prostate cancer cells (36). Olfactory chemosensory signaling occurs in the renal system. In studies it has been discovered that ORs, olfactory-related adenylate cyclase (AC3) and olfactory G protein are found in the renal tubules and macula densa cell which play an important role in glomerular filtration, reabsorption, or secretion rate and thereby have an effect on regulation of blood pressure. There are six ORs are found in the mouse kidney (Olfr78, Olfr90, Olfr1373, Olfr1392, Olfr1393, and Olfr NPTR6JSE50FPA). (37).

3.5 ORs 3D structure.

There is no X-ray crystallography structure of olfactory receptors because it is not possible to obtain crystals for ORs (38). Olfactory receptors have a characteristic structure of GPCRs, which are composed of a bundle of seven transmembran helices (7TM) spanning the cell membrane seven time. 7TM are linked by three intracellular loops (3CL) in the side of the cell cytoplasm and linked by three extracellular loops (3EL) in the side of extracellular membrane, with the carboxyl-terminus intracellular and amino-terminus extracellular. N-terminus and C-terminus are highly variable, extracellular loops are less conserved than intracellular loops. TM7 is highly conserved especially in the site close to the cytoplasm that is important to understanding the binding of G alpha protein (Fig. 19).

ORs belong to class A Rhodopsin-like family, which share two cysteines in first and second extracellular loops forming a disulfide bond, and has DRY motif in CL2 (Fig. 19) (39). The long (EL2) with an extra pair of cysteines is specific for ORs. The sequences in TM3, 4, and 5 are very variable and through these more variable regions are involved in odorant recognition and binding, these three TM3, 4, and 5 are predicted to face each other and compose a binding site pocket of odorants (Fig. 20).

Fig. 19 Typical structure of an odorant receptor shown the conserved motifs. Adapted from (39)



From Fig 19 and 21 there are many conserved motifs which are characteristic for ORs sequences such as in TM1 GN, in TM2 PMYF/LFL, in TM3 and CL2 MAYDRYVAIC, in TM5 SY, in TM6 KAFSTCA/GSHLSVV. These motifs are not present in other GPCRs which are responsible for interacting with G alpha protein and involved in signal transduction, in TM7 PMLNPFIYSLRN, in EL1 is cysteine, and in EL2 is three cysteine which form disulfide bond.

Based on their analysis of 200 ORs paralog sequences Pilpel and Lancent 1999 predicted 17 olfactory Complementarily Determining Residues (CDRs). These 17 residues were predicted to constitute a variable binding niche of odorants.

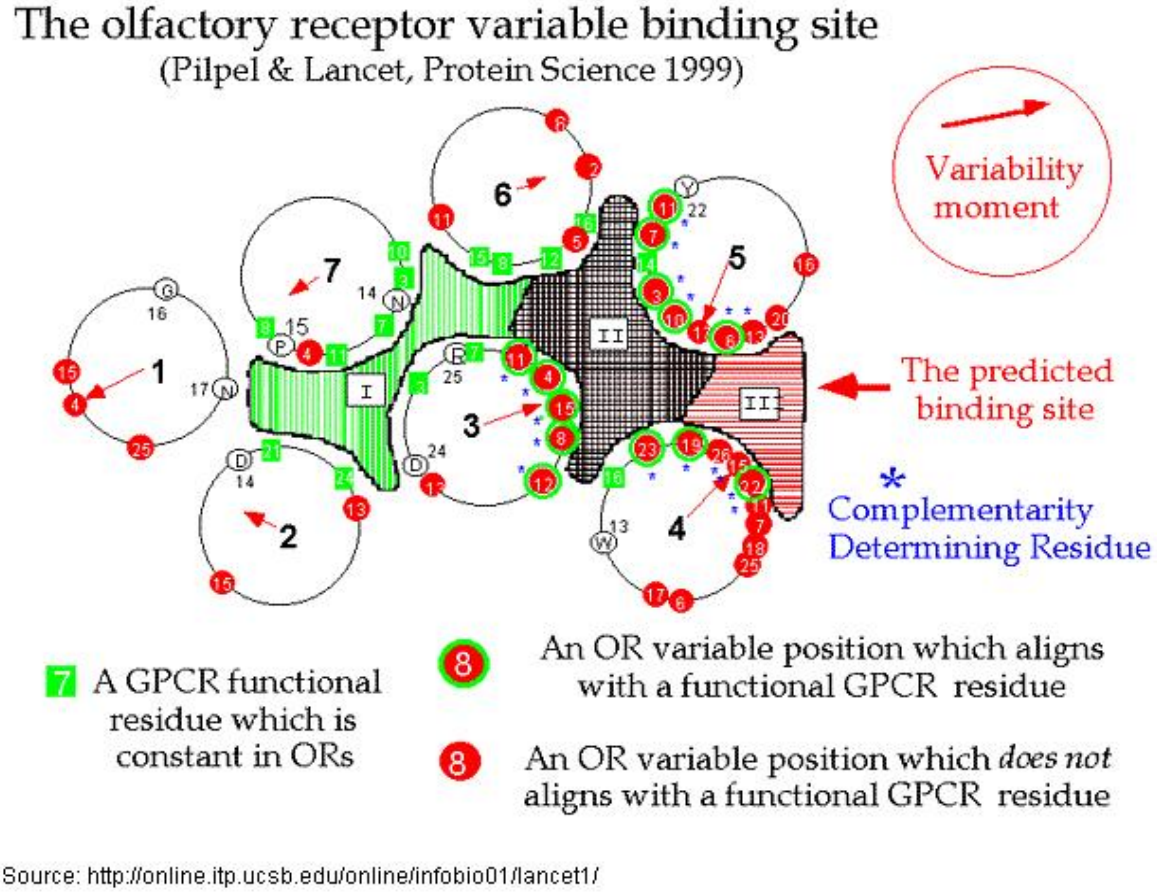


Fig. 20 the variable residue and the predicted binding site of ORs. Adapted from (40) and (41)

From the Fig. 20 the 17 predicted CDRs positions according to OR TM positions are TM3. 4, TM3. 8, TM3. 11,TM3. 12, TM3. 15, TM4. 11, TM4. 15, TM4. 19, TM4. 22, TM4. 23, TM4. 26, TM5. 3, TM5. 6, TM5. 7, TM5. 10, TM5. 11, and TM5. 13 (41). 12 of these 17 predicted CDRs are aligned with a functional GPCR residue (Fig. 20). The binding sites predictions were studied by Man et al. 2004 and Katada et al. 2005.

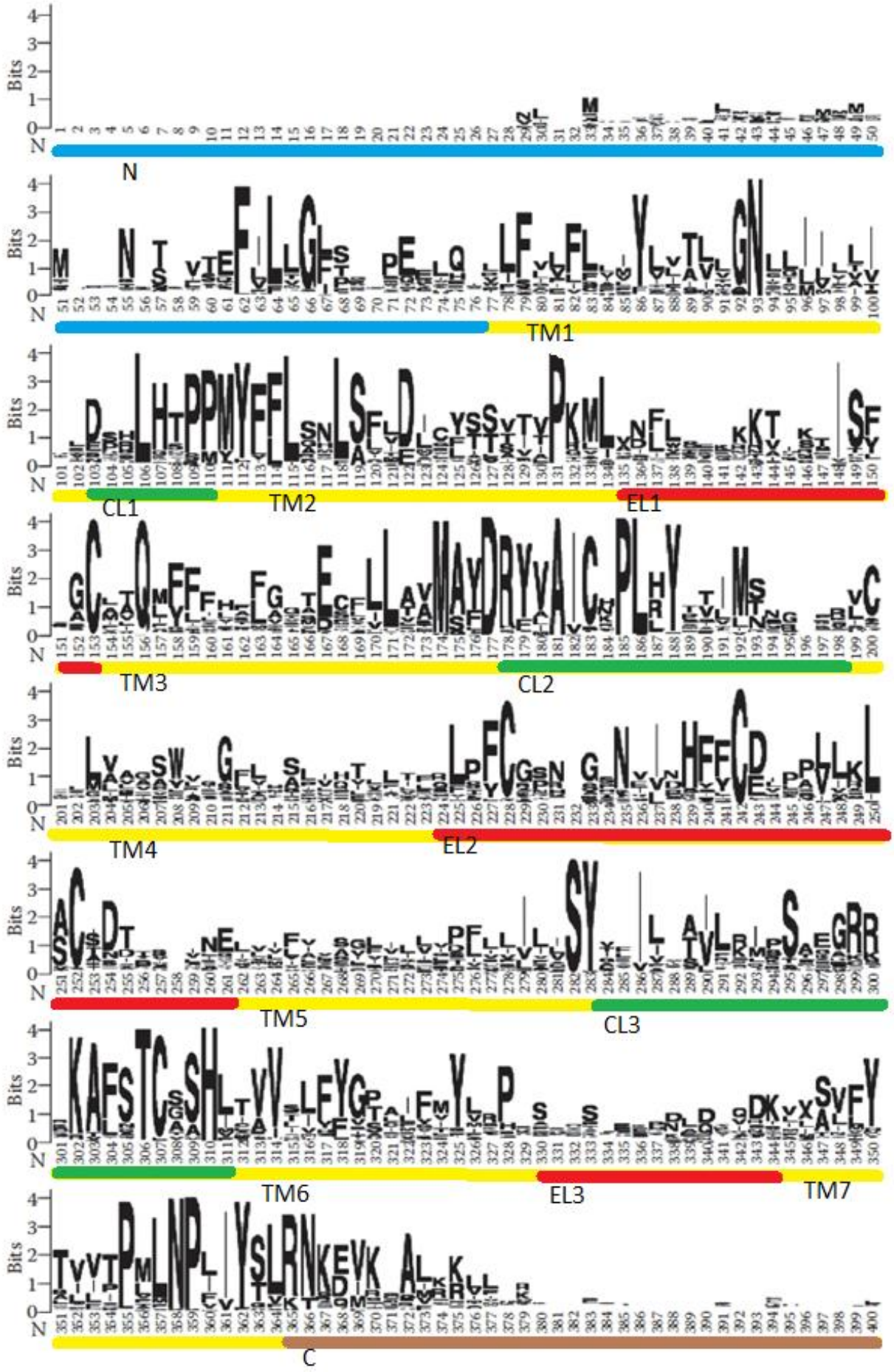


Fig. 21 Motif patterns found in the human ORs. Weblogo representation for 397 human OR amino acid sequences. Sequences were aligned using Clustalw. Adapted from (39)

3.6 ORs-odorant interaction.

The first step in the olfaction process is the solubilization of hydrophobic odorants from the air in the hydrophilic nasal mucus. This process is done by olfactory binding protein (OBP), which acts as a carrier for odorants. Until now it has not been detected if there any interaction between OBP and ORs. OBP belongs to the lipocalin family (42). OBPs are expressed in many other tissues e.g. testis and germ cells (Thomas et al. 1996), insulin-secreting beta cells, spleen (Blache et al. 1998), and heart (Drutel et al. 1995) (42). Odorants of typically chemical functional groups can possess completely distinct odours, at the same time as some odorants of different chemical functional groups can possess related odour impressions e.g. hydrogen cyanide and benzaldehyde can be discriminated like as almond. So it is not possible to predict the odour through its chemical structure or its functional group. On the other hand some antagonists are structurally related to the agonist, which means an antagonist might bind in the same binding site of an agonist, but does not activate the receptor. All of this means that there is no simple correlation between odours perception and its molecular structure (43). Odorant concentration can have an effect on its odour impression. Some chemicals have different odours at different concentrations e.g. indole which has a floral odour at low concentration and has putrid odour at high concentration. Other odorants have specific odour only of high concentration but are without odour at low concentration. Human olfactory system can discriminate numerous odorants even at very low concentration. One odor molecule can bind with several different ORs and one OR can bind with a few odor molecules (this point will be discussed in 3.8) (44).

Small numbers of ORs have matching odorant molecules that have been discovered. So most ORs are called orphans, which mean that their ligand odorants are not known. Only 10% of humans ORs have been deorphaned. That is because only few experimental studies have identified odorant-OR interaction. So further research in this field remains a challenge.

The recent study for the prediction of the interaction between human ORs and odor molecules is using a network protein procedure. This study can evaluate the selectivity and tendency of odorants to interact with ORs through developing an association score (AS) equation (44)

AS: an association score, the higher score is the more significant to OR-odorant interaction.

A: the number of compound for one OR.

$$AS = \left(\left(\frac{A}{B}\right) \times \left(\frac{A}{C}\right)\right) \times \left(\frac{B}{D}\right)$$

B: total number of molecules that have the same odor.

C: the number of ORs for the same odor.

D: total number of odorant molecules in the study.

List of ORs predicted to interact with molecules carrying the "anis" note.

Odor	ORs	Number of compounds (A)	AS
anis	OR1D2	2	1.59 10-4
anis	OR1D20	3	3.57 10-4
anis	OR1G1	5	9.93 10-4
anis	OR52D1	5	9.93 10-4
anis	OR5D18	4	6.36 -10-4
anis	OR6A2	2	1.59 10-4
Total	C = 6	B = 21	

doi:10.1371/journal.pone.0093037.t001

Table 1 Adapted from (44).

From the above table, OR1G1 and OR52D1 have the highest association score so these two ORs are the most significant association to the anis odor.
For more details see (44)

Another experimental trail to deorphan ORs is the coupling of ORs and ion channels for rapid visualization of odorants which can bind with these ORs. As mentioned before many ORs are still orphans because the deorphaning processes take a lot of time in conventional experimental methods.

To achieve a rapid response of odorants on ORs a group of researchers couple ORs with an ion channel (Kir 62 potassium), which is expressed in HEK293 cells. This way after the ligands bind with ORs leads to conformational changes in ORs. Thereby ion channels are opened and the resulting ion influx inside the cell can easily be recognized by using

membrane potential dye. The trial can be repeated with several odorants on the same OR and it can be determined which odorants have response on this specific OR. In this process ORs show high sensitivity and selectivity to the specific odorant molecules (45).

3.7 Signal transduction of olfaction

The olfactory sensory neurons (OSNs) are contacted directly through its cilia with the nasal mucosa. And as mentioned before ORs are found at the surface of the cilia of OSNs. OSNs are bipolar neurons with a single dendrite termination. Each OSN expresses one and only one type of ORs. A specific type of odorant can stimulate various types of ORs, and a specific OR responds to many types of odor molecules.

After the odorants binding to the ORs a cascade of reaction which is responsible for converting this binding into an electrical signals (46).

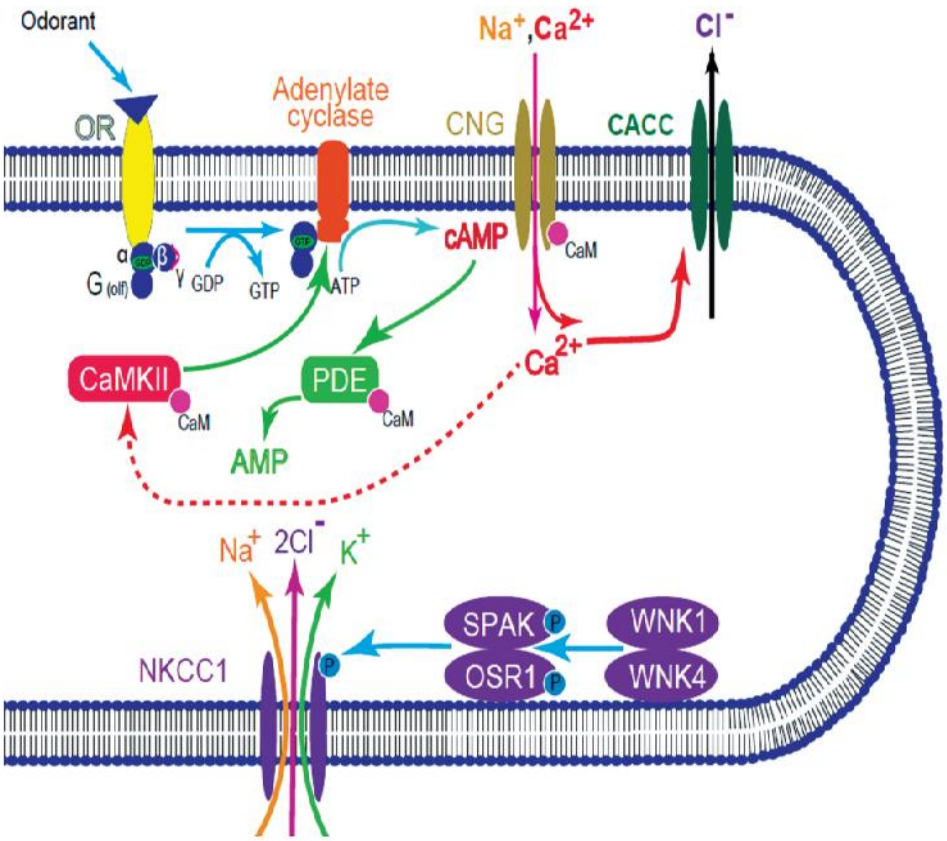


Fig. 22 Graphical representation of the signal amplification in OSN cilia. CNG, cyclic nucleotide-gated channel; CaM, calmodulin. Adapted from (46)

The perception of the odorant begins with a conformational activation of ORs. This conformational change which happens in ORs like in all GPCR displacements of helix 6 thereby allows the C terminus of G alpha protein to bind with ORs (see 2.4). The activated ORs catalyze the replacement of guanosine 5-diphosphate (GDP) by guanosine 5-triphosphate (GTP) on G-Protein thereby G alpha dissociates from the beta and gamma subunits complex and activates adenyl cyclase enzyme, which converts adenosine triphosphate (ATP) into cyclic adenosine monophosphate (cAMP) Fig 22 and 23 (46).

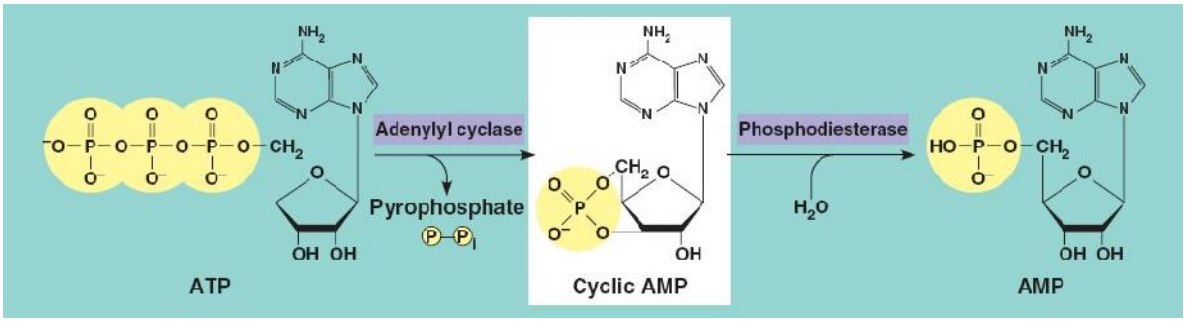


Fig. 23 formation of cyclic AMP and AMP

One active OR can activate numerous G alpha proteins thereby amplifying the action as mentioned in GPCRs. When intracellular concentration of (cAMP) is increased it acts as a second messenger and it can activate cyclic nucleotide-gated channels (CNG). These channels have an important role in chemoelectrical energy conversion in OSNs. By binding a cyclic nucleotide with CNG channels they are permeable to monovalent ions such as sodium, potassium, and calcium ions from the mucus into the ciliary cytoplasm, which leads to depolarization of the ciliary membrane.

Kleene and Gesteland reported the presence of calcium ions activating chloride ions efflux in OSNs through CACC and that leads to secondary depolarization and amplification of the first depolarization (46). Depolarization amplitudes rely on the amount and nature of the odor which are bound by the ORs.

The ion transporter Na+-K+-2Cl− cotransporter NKCC1 plays an important role in olfaction process, which is expressed in OSNs. NKCC1 allows the unidirectional cotransport of Na+, K+, and 2Cl− ions through the OSNs membrane from extracellular to cytoplasm of OSNs (Russell 2000) (Fig 22). The inhibition or lacking of NKCC1 in mice exhibits significant defects in perception of a complex odor mixture (47).

Signal termination can be by one of these mechanisms, first, inhibition of CNG channels by Calmodulin CaM. Second, calcium ion accumulation leads to activation of CaMKII by CaM thereby inhibiting activation of the receptors. Third is the conversion of cAMP into AMP by phosphodiesterase, which is activated by CaM (46) (Fig 22 and 23).

In some species, the olfactory signal transduction is done by two separate pathways , cAMP and Inositol Triphosphate (IP3), such as eugenol increases cAMP whereas lyral and lilial increase IP3 (Schid and Restrepo 1998). IP3 acts as second messenger in some mammals (40). The different signaling pathways are due to different types of G alpha protein which are coupled with ORs as shown in table 2.

Table 2 After G alpha protein has coupled with an OR it can activate the phospholipase C beta 2 instead of adenylate cyclase and produces IP3 instead of cAMP (48).

Effectors of heterotrimeric G proteins. The arrows indicate up- or down-regulation of the effector.

Subunit	Family	Main subtypes	Primary effector
α	α_s	$G_{\alpha s}$, $G_{\alpha olf}$	Adenylate cyclase ↑
	$\alpha_{i/o}$	$G_{\alpha i1}$, $G_{\alpha i2}$, $G_{\alpha i2}$	Adenylate cyclase ↓
		$G_{\alpha oA}$, $G_{\alpha_{oB}}$	K^+ channels ↑
		$G_{\alpha t1}$, $G_{\alpha_{t2}}$	Ca^{2+} channels ↓
		$G_{\alpha z}$	Cyclic GMP phosphodiesterase ↑
	αq	$G_{\alpha q}$, $G_{\alpha 11}$, $G_{\alpha 14}$,	Phospholipase C ↑
		$G_{\alpha 15}$, $G_{\alpha 16}$	
	$\alpha 12$	$G_{\alpha 12}$, $G_{\alpha 13}$	?
β	$\beta_{1\text{-}5\ (6?)}$	Different assemblies	Adenylate cyclase ↑/↓
		of β- and γ-subunits	Phsopholipase ↑
			Phosphatidylinositol 3-kinase ↑
γ	$\gamma_{1\text{-}11\ (12?)}$		Protein kinase C ↑
			Protein kinase D ↑
			GPCR kinases ↑
			Ca^{2+}, K^+ (and Na^+) channels

3.8 Combinatorial coding and olfactory pattern.

Detection of the odorants in a combinatorial receptor coding scheme means that each specific odorant can bind with numerous ORs and each OR is stimulated by numerous odorants. Different olfactory receptors which recognize the specific odorant might recognize different structure features of the odor molecules. This combinatorial coding acts as a way to increase the perceived odor space especially in the species which have small number of intact OR genes (49). There are only about 1000 OR genes in the genome and by combinatorial coding they can discriminate a huge number of different odor molecules. If the odor molecule is encoded by 3 olfactory receptors, the odor impesions that could be discriminated a huge number.

But recently many challenging studies provided many examples about specific odorants that are detected by highly specific ORs (Mathew et al. 2013. Liu et al. 2014) for example gambiae OR8 is specific for 1-octen-3ol and ionotropic receptor (IR92a) which specific for ammonia (49)

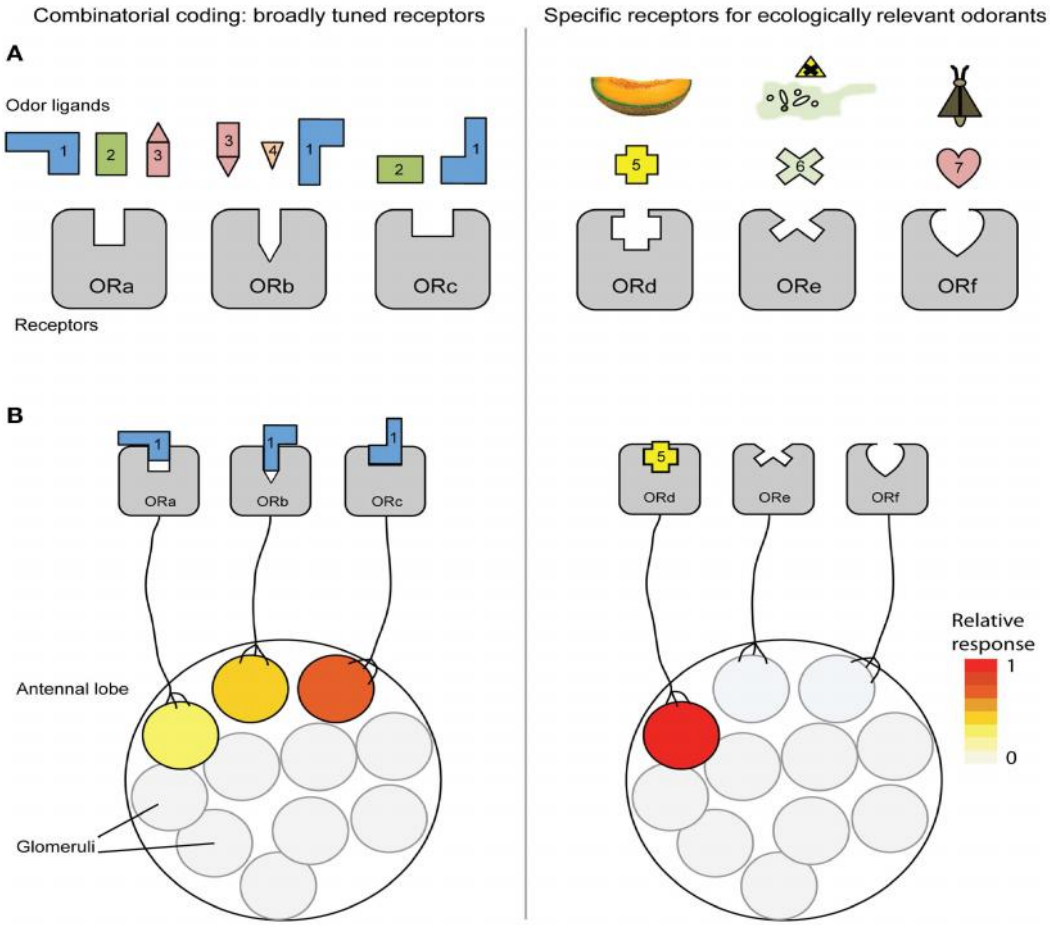


Fig 25 Combinatorial coding and specific receptors. Adapted from (48)

After stimulation of ORs by odor molecules, how are the odor molecules are coded in the brain to descriminate their odors? This question is chanallging and not completely understood. The recent scenario is after the odor molecule binds with multiple receptors or specific receptors (according to odor molecules features) multiple glomeruli or a single glomeruli will be activated (Fig. 25), thereby signals are transferred from the glomeruli to the brain. The brain will represent this signal in the form of a central pattern for this specific odor molecule Fig. 26 (50)

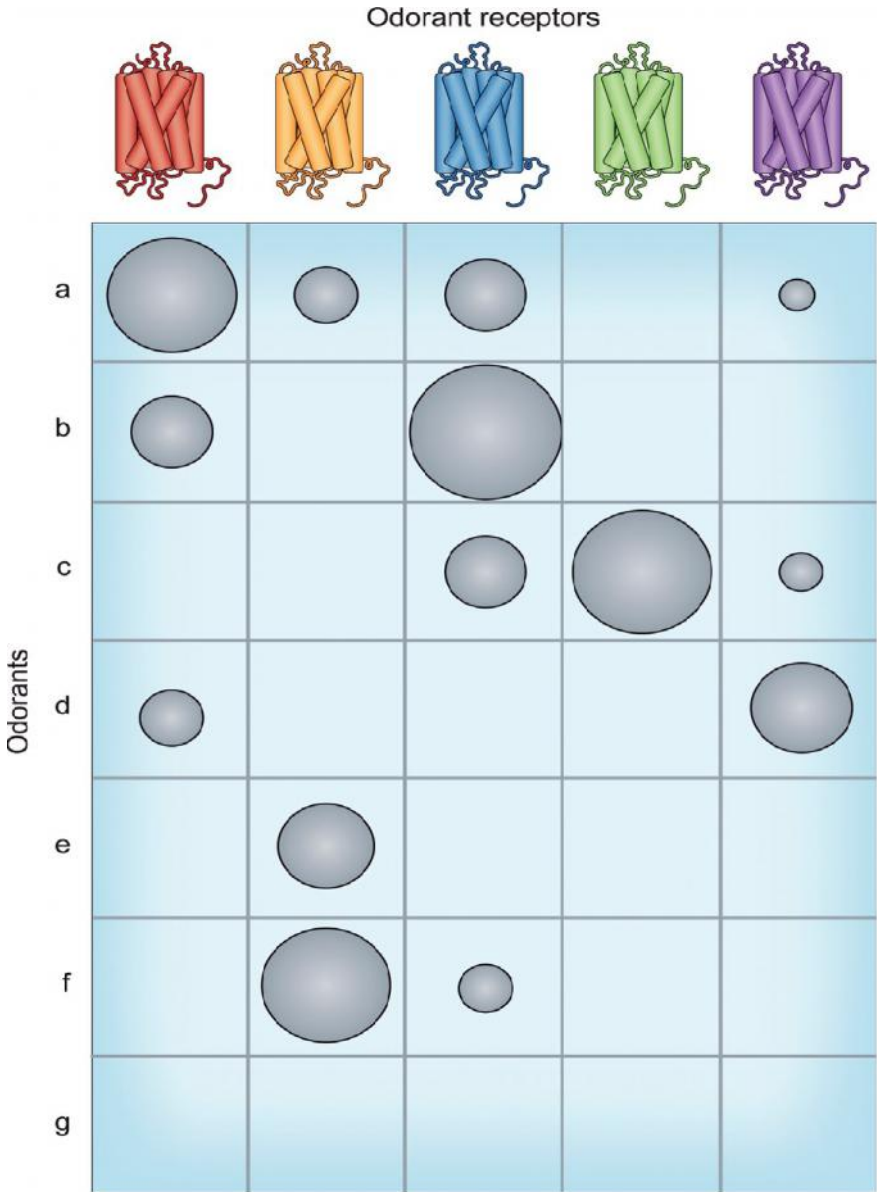


Fig. 26 The pattern of combinatorial receptor coding of the odorants. The size of the circle is relative to the intensity of the response of different ORs to the odor molecule. Adapted from (50)

4. Homology modeling

Homology modeling is also named comparative modeling or template-based modeling (TBM). Homology modeling is very important for determining the 3D protein structure and thereby understanding the protein function. To determine the protein 3D structure, there are many experimental methods like X-ray crystallography or NMR spectroscopy, but these experimental methods are time and cost consuming, and not suitable for all proteins. There is a huge gap between the numerous discovered protein sequences and their limited data of solved 3D structures. By using the homology modeling, we can solve this gap by using computational methods, which predict the 3D structure of a given protein sequence. The model depends on its sequence similarity to one or more proteins of known 3D structures (51). Homology modeling depends on evolutionary related proteins that share a similar structure (52). The query (or target protein) of which the 3D structure is unknown, can be modeled through the homology method with one or more known 3D structure template. Until now, no odorant receptor 3D structure has been experimentally determined, so the suitable way to determine the 3D structure of ORs is by homology modeling. ORs belong to GPCRs superfamily, class R (rhodopsin-like) and subclass delta according to GRAFS classification. Recently the 3D structures of a few receptors which belong to R class (rhodopsin-like) were solved; two of them belong to the delta subclass which may be suitable as a template because they belong to the same subclass of ORs. In this work we do homology modeling for OR1G1. Table 3 the sequence identities of OR1G1 and some of the available templates of GPCRs. Declares the sequence identity of all sequence, all TMs and of each helix separated. Adapted from (53)

GPCR	All	TMs	TM1	TM2	TM3	TM4	TM5	TM6	TM7
hOR1G1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$t\beta_1$ AR	*18.5*	*22.7*	*33.3*	*21.4*	*26.9*	*14.7*	*15.6*	*25.0*	*21.7*
hβ_2AR	17.3	21.9	25.9	28.6	26.9	14.7	9.4	21.4	26.1
hOPRK	*15.7*	*16.5*	*22.2*	*28.6*	*19.2*	*5.9*	*18.8*	*3.6*	*17.4*
rM_3MR	14.7	18.2	29.6	17.9	23.1	5.9	15.6	17.9	17.4
hNOP	14.7	17.0	25.9	28.6	19.2	5.9	9.4	3.6	26.1
hH_1HR	14.7	18.0	33.3	17.9	19.2	5.9	9.4	14.3	26.1
mOPRM	14.4	17.1	22.2	35.7	19.2	8.8	12.5	3.6	17.4
hM_2MR	14.4	17.0	29.6	17.9	19.2	8.8	15.6	10.7	17.4
mOPRD	14.4	15.5	18.5	35.7	23.1	2.9	6.3	0.0	21.7
hS1P1R	*14.1*	*19.1*	*14.8*	*32.1*	*15.4*	*14.7*	*12.5*	*17.9*	*26.1*
hAA_{2A}R	14.1	16.9	14.8	17.9	30.8	8.8	12.5	7.1	26.1
bRho	*13.7*	*17.6*	*14.8*	*28.6*	*30.8*	*11.8*	*12.5*	*7.1*	*17.4*
hD_3DR	13.7	15.0	22.2	14.3	11.5	8.8	3.1	14.3	30.4

General steps for homology modeling (54, 55)

1- One or more template determination.

2- Alignment of the query or the target sequence with one or more templates.

3- Building of the model.

4-Evaluation and refinement of the model.

We can repeat these general steps until we obtain the suitable model.

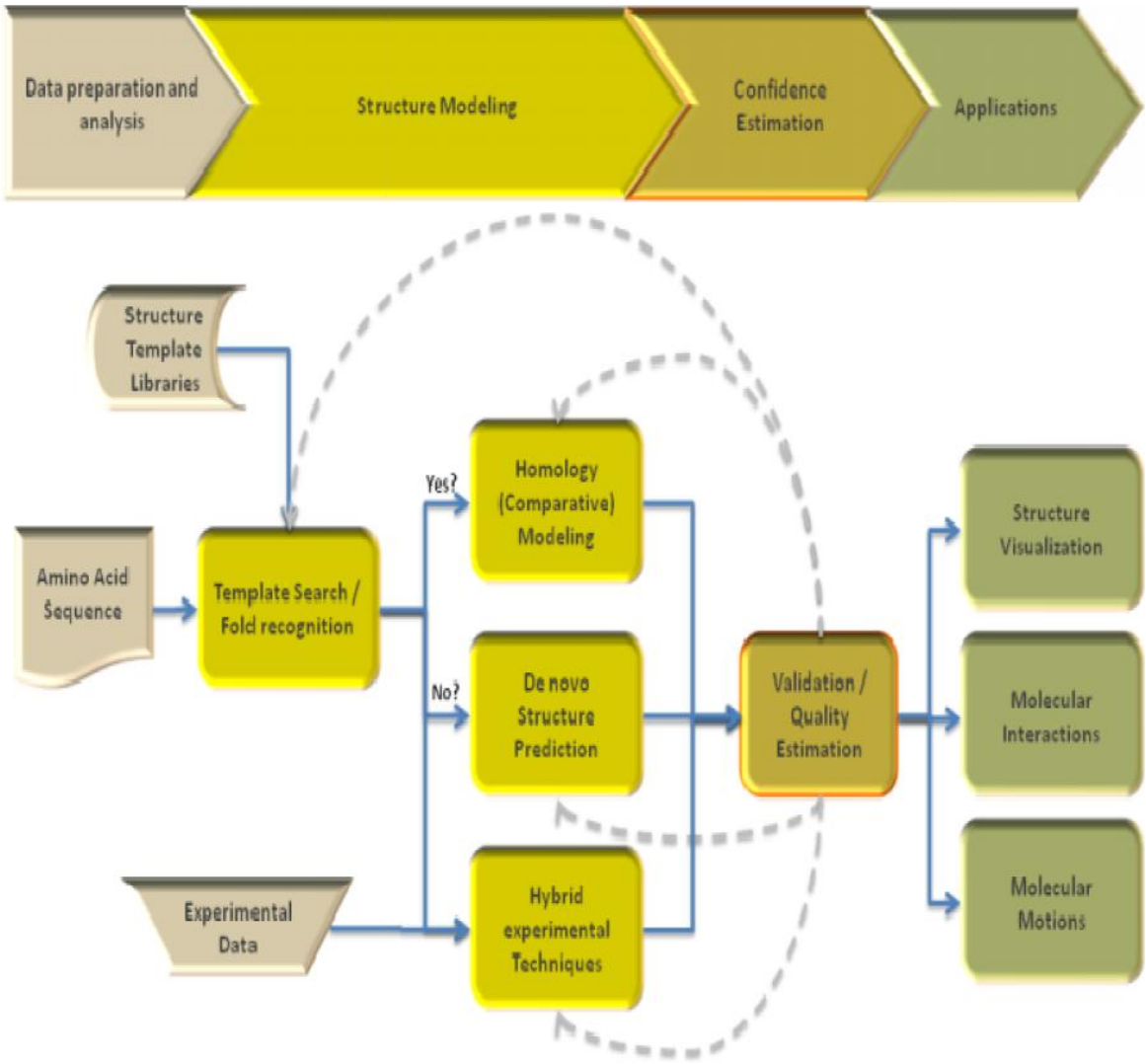


Fig. 27 Declares the steps for homology modeling coordinating with the experimental data which help to obtain the most accurate model.

Adapted from http://www.proteinmodelportal.org/?pid=101

As one can see in fig. 28 when sequence identity is more than 25% the sequence alignment is useful, but if it is less than 25% the use of multiple alignments is more suitable (56). The shorter sequence alignment is the more sequence identities are required.

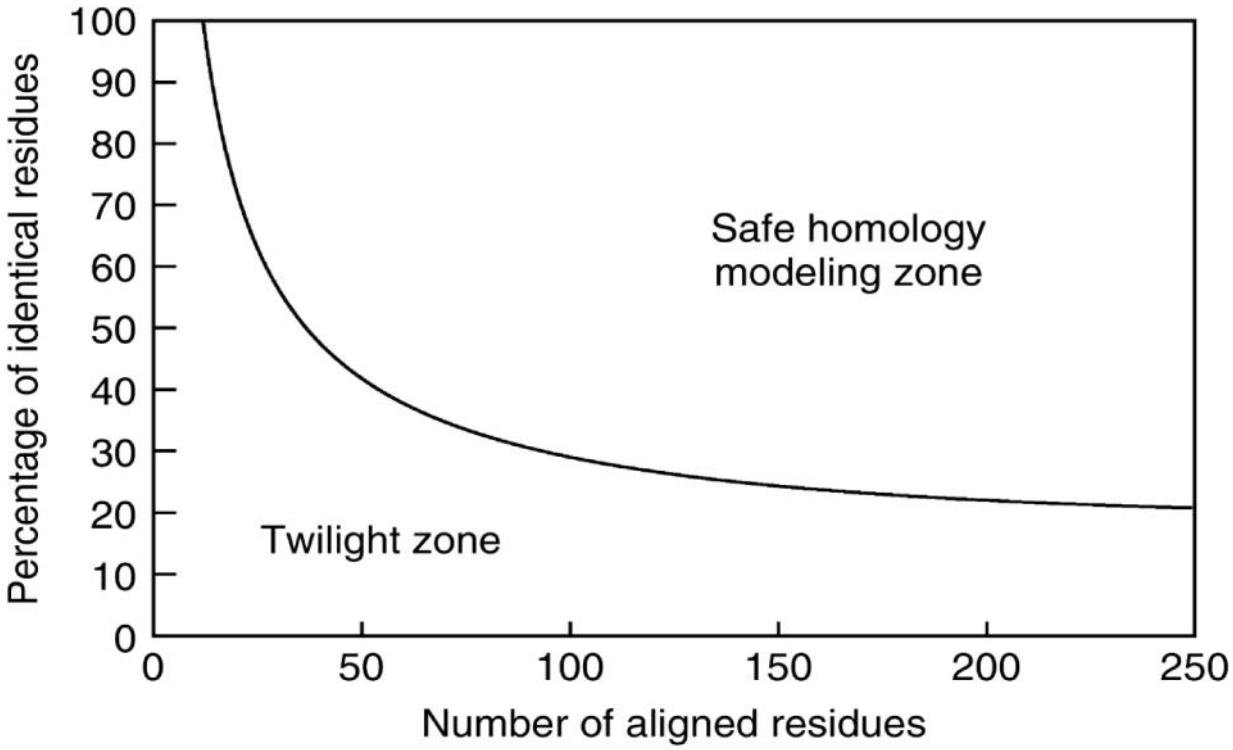


Fig. 28 The relationship between alignment length and sequence identity. Adapted from (51).

4.1 Human olfactory receptor OR1G1

From the nomenclature scheme this receptor belongs to class II family 1, subfamily G, member 1

Another synonyms for OR1G1 are

- Olfactory Receptor 17-209
- Olfactory Receptor OR17-8
- Olfactory Receptor 1G2
- OR17-209
- OR1G2
- Olfactory Receptor, Family 1, Subfamily G, Member 2
- Olfactory Receptor 1G1
- OR17-130

- **External Ids for OR1G1 Gene**

- HGNC: 8204
- Entrez Gene: 8390
- Ensembl: ENSG00000183024
- UniProtKB: P47890
- HORDE: OR1G1

Previous HGNC Symbols for OR1G1 Gene: OR1G2

OR1G1 Cytogenetic band: 17p13.3

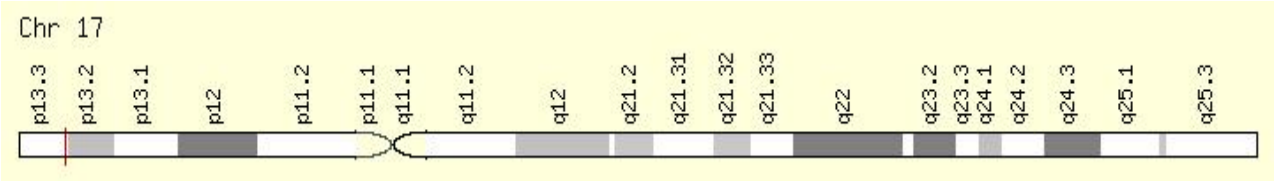


Fig 29 OR1G1 Position on Chromosome 17 p13.3 adapted from genecards.org

http://www.genecards.org/cgi-bin/carddisp.pl?gene=OR1G1

Molecular mass: 34924 Da

Sequence length: 313 amino acids

Amino acids sequences fasta format of OR1G1:

>sp|P47890|OR1G1_HUMAN Olfactory receptor 1G1 OS=Homo sapiens GN=OR1G1 PE=2 SV=2
MEGKNLTSISECFLLGFSEQLEEQKPLFGSFLFMYLVTVAGNLLIILVIITDTQLHTPMY
FFLANLSLADACFVSTTVPKMLANIQIQSQAISYSGCLLQLYFFMLFVMLEAFLLAVM
AYDCYVAICHPLHYILIMSPGLCIFLVSASWIMNALHSLLHTLLMNSLSFCANHEIPHF
FCDINPLLSLSCTDPFTNELVIFITGGLTGLICVLCLIISYTNVFSTILKIPSAQGKRKAFS
TCSSHLSVVSLFFGTSFCVDFSSPSTHSAQKDTVASVMYTVVTPMLNPFIYSLRNQEIK
SSLRKLIWVRKIHSP

FASTA format started by (>) thereby the code and the name of the protein after that the amino acid sequences. These sequences from uniprot server.
http://www.uniprot.org/uniprot/

OR1G1 gene expression:

OR1G1 gene is expressed not only in Olfactory Epithelium but also in Testis and Bladder Fig. 30

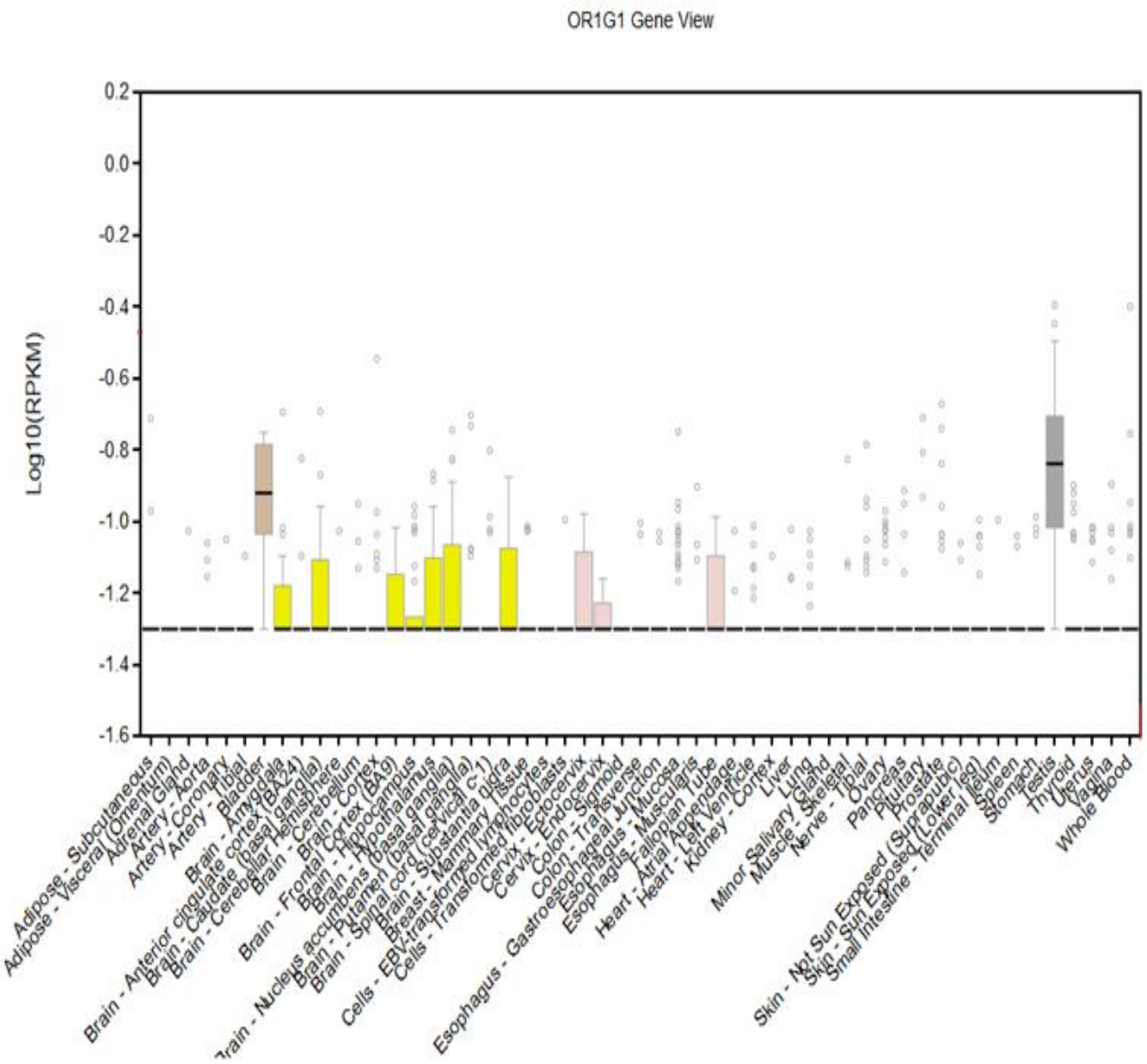


Fig. 30 OR1G1 gene excerption at different tissue. Adapted from

http://www.gtexportal.org/home/gene/OR1G1

Ligand (Odorants) which can binding with OR1G1:

(+/-)-citronellal its odour is citrus / 1-decanol its odour is polished
1-dodecanol its odour is raw carrot / 1-heptanol its odour is green
1-hexanol its odour is green / 1-nonanol its odour is green
2-ethyl-1-hexanol its odour is rose/2-isobutyl-3-methoxypyrazine its odour is green
2-methyl pyrazine its odour is roasted nuts / 2-nonano its odour is lplastic
2-nonanone its odour is green /2-octano its odour is fatty
2-undecanone its odour is floral / 3-hydroxybutan-2-one its odour is buttery
3-nonanone its odour is green / 3-octanol its odour is earthy
4-octanol its odour is soapy / 9-decen-1-ol its odour is waxy
acetophenone its odour is vegetal / benzaldehyde its odour is almond
benzothiazo its odour is lrubbery / beta-ionone its odour is violet
camphor its odour is camphor / capric acid its odour is waxy
cinnamaldehyde its odour is cinnamon / citral its odour is lemon
coumarin its odour is woodruff / decanal its odour is waxy
ethyl butyrate its odour is strawberry / ethyldecanoate its odour is grape
ethylisobutyrate its odour is fruity / ethyl nonanoate its odour is waxy
ethyl octanoate its odour is waxy / ethyl vanillin its odour is vanilla
eugenyl acetate its odour is spicy / floralozone its odour is ocean
gamma-decalactone its odour is peach / geraniol its odour is rose
guaiacol its odour is smoky / hedione its odour is floral
heptanal its odour is fatty / hexanal its odour is green apple
isoamyl acetate its odour is overripe banana / jasmonyl its odour is jasmine
lauric aldehyde its odour is waxy / limonene its odour is lemon
lyral its odour is floral / maltol its odour is caramel
maltyl isobutyrate its odour is fruity / manzanate its odour is apple
Menthol its odour is mint / methyl decanoate its odour is floral
Methyl nonanoate its odour is fruity / methyl octanoate its odour is green
nonanal its odour is orange / nonanoic acid its odour is coriander
octanal its odour is watermelon / octanol its odour is soap
phenyl methanol its odour is green / piperonyl acetone its odour is floral
pyrazine its odour is nutty / pyridin its odour is scallop
quinoline its odour is coal tar / safrole its odour is spicy

S-methyl thiobutanoate its odour is cheesy / thiazol its odour is meaty
thymol its odour is spicy / trans-anethol its odour is anise
tridecanal its odour is perfum-like / vanilline its odour is vanilla (57)
-santalol its odour is Sandalwood-Like Odorant (58)

4.2 Descriptions of some programs, servers and websites for modeling and docking.

4.2.1 The amino acid sequences for the target protein can be derived from:

1- Uniprot server, in which the protein amino acid sequences and some information about the protein function can be found. http://www.uniprot.org

2- NCBI server, in which genomic DNA, transcript RNA and amino acid sequences of protein can be found. On this server BLAST search tools can be used. http://www.ncbi.nlm.nih.gov

4.2.2 Secondary structure prediction servers:

1- TMHMM (Krogh et al. 2001) and HMMTOP (Tusnady and Simon 2001), these two websites firstly predict the localisation of the TM helices and topology of the protein based on a Hidden Markov Model (HMM). Mostly there is no difference between the results of the two websites. In this work TMHMM is used (Fig. 31). http://www.cbs.dtu.dk/services/TMHMM

2- GPCR tools, the benefits of this server are that it not only constructs a snake plot of the protein but also a helix box which can illustrate some information about the binding bucket (Fig. 32). http://tools.gpcr.org/visualise/protein

3- CFSSP in this server the Chou and Fasman algorithm is used.

4- APSSP advanced protein secondary structure protein server

5- COILS in this program amino acid sequence are compared to a database of known parallel two-stranded coiled-coils.

6- SOPMA Self-Optimized Prediction Method with Alignment, it operates with the homology method of Levin et al. 1986.

There are a lot of other programs like the ADS-TM filter, DLP-SUM, COR, PREDATOR, PHD, HTMSRAP, JPRED, MARCOIL, Net SurfP, PORTER, Predictprotein and SOSUL

4.2.3 Template libraries

The PDB files for templates can be obtained from template libraries which are available through the world wide Protein Data Bank. Until now there are three organized Protein Data Banks:

1- USA Protein Data Bank RCSB PDB http://www.rcsb.org

2- Europe Protein Data Bank PDBe http://www.ebi.ac.uk/pdbe

3- Japan Protein Data Bank PDBj http://pdbj.org

4.2.4 Servers for genetic protein information

1- Genecards website, at this website many information about any gene can be found. Like other names for the gene, distribution of the gene at the chromosomes, length of the sequence and molecular mass. http://www.genecards.org

2- Gtexportal website offers data about expression of the gene in different tissues. http://www.gtexportal.org

4.2.5 Server for template search:

1- NCBI server Through this server BLAST search tools can be found. http://www.ncbi.nlm.nih.gov

2- PSI-BLAST, alignment search tools can be used to determine the suitable template. http://www.ebi.ac.uk/Tools/sss/psiblast

3- HHpred server depends in its prediction on HMM-HMM comparison. This server is very user-friendly and its results are reported to be more accurate than those of BLAST and PSI-BLAST. Thus it offers a template of higher probability to be homologous. (Fig. 35) http://toolkit.tuebingen.mpg.de/hhpred

4- The MOE program offers the Homology Searching tool. By using the target amino acid sequence, the suitable template is determined according to E-values and Z-scores, the smallest E-value and the biggest Z-score is the best template. (Fig. 34)

4.2.6 Conserved sequences determination server:

1- The Evolutionary Trace server is helpful to determine the conserved residues of all 7TMs in all olfactory receptors and in all rhodopsin-like superfamily receptors. Color scale is from red to violet, red is highly conserved and more important, violet is highly varied and less important for alignment (Fig. 36). http://mammoth.bcm.tmc.edu/ETserver.html

2- The Consurf server is used to determine the conserved residues of a specific protein. In this work this server was used to determine the conserved residues of OR1G1 (Fig. 33) http://consurf.tau.ac.il

4.2.7 Sequence alignment server:

Structural alignment and coordinate assignment are the most important step. After determination of 7TMs sequences through secondary structure prediction, it is necessary to select the template (or several) and to determine the highly conserved sequences. The following step is the sequence alignment. A lot of programs can be used for sequence alignment. Some of them are used for pairwise alignment such as AlignMe, G-PAS, LALIGN, NW-align, Satsuma and UGENE. Another group is designed for multiple sequence alignment such as T-coffee, ClustalW, DECIPHER, MSAP robs, MAFFT and GLP robs. Firstly, in this work I used TM coffee, which is specialized on GPCRs. I did multiple sequence alignment for some olfactory receptors, which I selected randomly to determine the conserved residue. (Fig. 37) http://tcoffee.crg.cat/apps/tcoffee/do:tmcoffee

As second step is I did sequence alignment for OR1G1 and the template which I had selected before. This step can be done by MOE program and the alignment can be edited manually (Fig. 41), or it can be done by TM-coffee and the alignment can be edited by using any editing program as Jalview http://www.jalview.org

4.2.8 The building of the homology model

MOE program (59) (the Molecular Operating Environment) is one package which consists of a set of applications and methodology development platforms. MOE integrates visualization, application development and simulation. It supports homology modeling of protein, evaluates the 3D protein structure, and docking of small molecule and proteins. It is developed by

Chemical Computing Group (CCG). This program is used in this work. http://www.chemcomp.com

Other working facilities

MODELER (60) http://salilab.org/modeler

SWISS Model (61): http://www.expasy.org/swissmod/SWISS-MODEL.html

SYBYL: http://www.tripos.com . WHAT IF: http://www.cmbi.kun.nl/swift/servers

The CPHModels Server: http://www.cbs.dtu.dk/services/CPHmodels

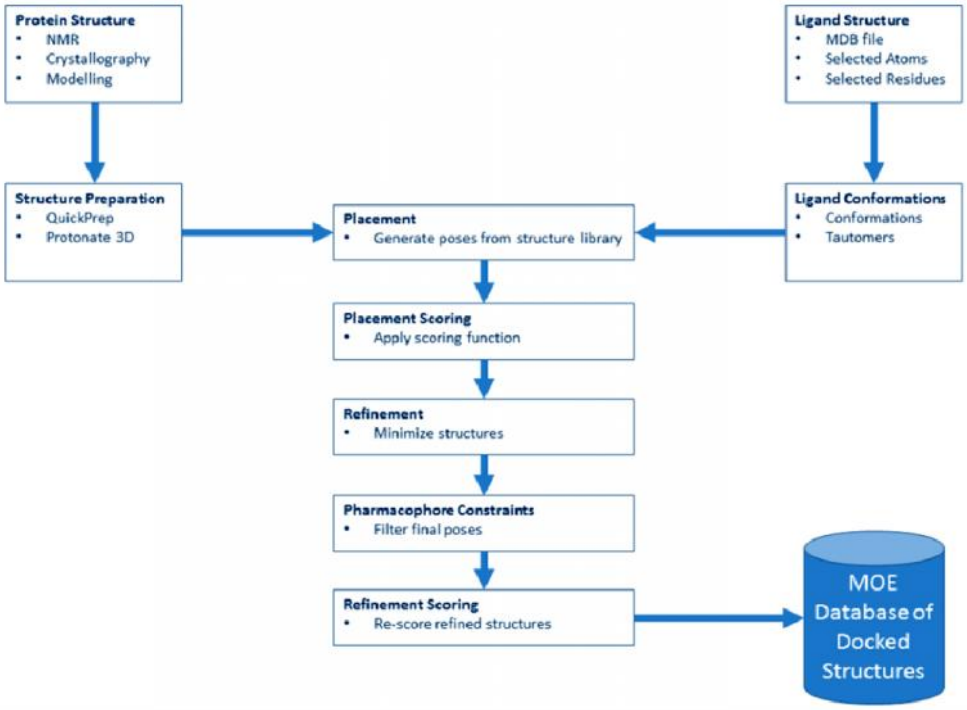
3D Jigsaw: http://www.bmm.icnet.uk/~3djigsaw. SDSC1: http://cl.sdsc.edu/hm.html

EsyPred3D: http://www.fundp.ac.be/urbm/bioinfo/esypred. InsightII: http://www.msi.com

COMPOSER: http://www.tripos.com/sciTech/inSilicoDisc/bioInformatics/matchmaker.html

4.2.9 Docking programs.

Docking attempts to find the best matching between a ligand and a receptor. Computational processes are used to calculate optimal binding geometries, and determine the orientation with minimizing the total energy of the complex which is formed. Many programs are used for docking, here are mentioned some of them. A) MOE as mentioned before in 4.2.8. (59). It is an application, which searches for optimal binding configurations between small-to medium-sized ligand molecules. In MOE the receptor is treated as rigid and the ligand is treated as fixable. Before running dock, firstly prepare the receptor and the ligand, then setup the Dock Protocol. Fig. 31 Methodology overview of the docking by using MOE program (59).



By the MOE there are four different protocols use different docking parameters:

1) Virtual Screening, London dG is used for rescoring 1. This protocol is used in the present work. 2) Rigid Receptor. 3) Induced Fit. 4) Pose Rescoring.

In rescoring stages, lower scores indicate more optimal poses. In MOE there are five Scoring functions:

1) ASE Scoring

2) Affinity dG Scoring

$$\Delta G = C_{\mathrm{hb}} f_{\mathrm{hb}} + C_{\mathrm{ion}} f_{\mathrm{ion}} + C_{\mathrm{mlig}} f_{\mathrm{mlig}} + C_{\mathrm{hh}} f_{\mathrm{hh}} + C_{\mathrm{hp}} f_{\mathrm{hp}} + C_{\mathrm{aa}} f_{\mathrm{aa}}$$

3) Alpha HB Scoring

4) London dG Scoring, it estimates the free energy of binding of the ligand from a given pose.

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{HB} f_{HB} + \sum_{m-lig} c_M f_M + \sum_{atoms\ i} \Delta D_i$$

$$\Delta D_i = c_i R_i^3 \left\{ \iiint_{u \notin A \cup B} |u|^{-6} du - \iiint_{u \notin B} |u|^{-6} du \right\}$$

5) GBVI/WSA dG Scoring for Rescoring 2.

$$\Delta G \approx c + \alpha \left[\frac{2}{3} (\Delta E_{coul} + \Delta E_{sol}) + \Delta E_{vdw} + \beta \Delta SA_{weighted} \right]$$

B) AutoDock http://autodock.scripps.edu/

AutoDock and AutoDock Vina are automated tools for docking. They are freely distributed under GPL open source license, so they are widely-used

C) RosettaDOCK http://rosie.rosettacommons.org/

D) FlexX https://www.biosolveit.de/FlexX/

Receptor is treated as rigid. Firstly the ligand is broken up into fragments, then those fragments are docked into the receptor finally the ligand is reassembled in low energy conformations. E) GOLD http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx

4.3 Secondary structure prediction of OR1G1

The secondary structure of the proteins give us a simple description of where the transmembrane helices and beta sheets are found. Through this structure it can be predicted where the disulfide bonds are found, which are formed between cystines residues. By viewing horizontally of the secondary structure, it can be predicted where the binding buckets could be found (Fig. 33). The secondary structure plays an essential role in protein folding and acts as intermediate step on the way to 3D structure prediction (62). Different computational programs or servers can be used for secondary structure prediction. Here we chose the TMHMM Server v. 2.0 to predict the transmembrane helices in OR1G1 http://www.cbs.dtu.dk/services/TMHMM/

TMHMM result

HELP with output formats

```
# sp|P47890|OR1G1_HUMAN Length: 313
# sp|P47890|OR1G1_HUMAN Number of predicted TMHs:  7
# sp|P47890|OR1G1_HUMAN Exp number of AAs in TMHs: 145.15909
# sp|P47890|OR1G1_HUMAN Exp number, first 60 AAs:  23.14109
# sp|P47890|OR1G1_HUMAN Total prob of N-in:        0.17916
# sp|P47890|OR1G1_HUMAN POSSIBLE N-term signal sequence
sp|P47890|OR1G1_HUMAN    TMHMM2.0        outside      1    26
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix     27    49
sp|P47890|OR1G1_HUMAN    TMHMM2.0        inside      50    60
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix     61    83
sp|P47890|OR1G1_HUMAN    TMHMM2.0        outside     84    97
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix     98   120
sp|P47890|OR1G1_HUMAN    TMHMM2.0        inside     121   131
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix    132   154
sp|P47890|OR1G1_HUMAN    TMHMM2.0        outside    155   196
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix    197   219
sp|P47890|OR1G1_HUMAN    TMHMM2.0        inside     220   239
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix    240   262
sp|P47890|OR1G1_HUMAN    TMHMM2.0        outside    263   271
sp|P47890|OR1G1_HUMAN    TMHMM2.0        TMhelix    272   294
sp|P47890|OR1G1_HUMAN    TMHMM2.0        inside     295   313
```

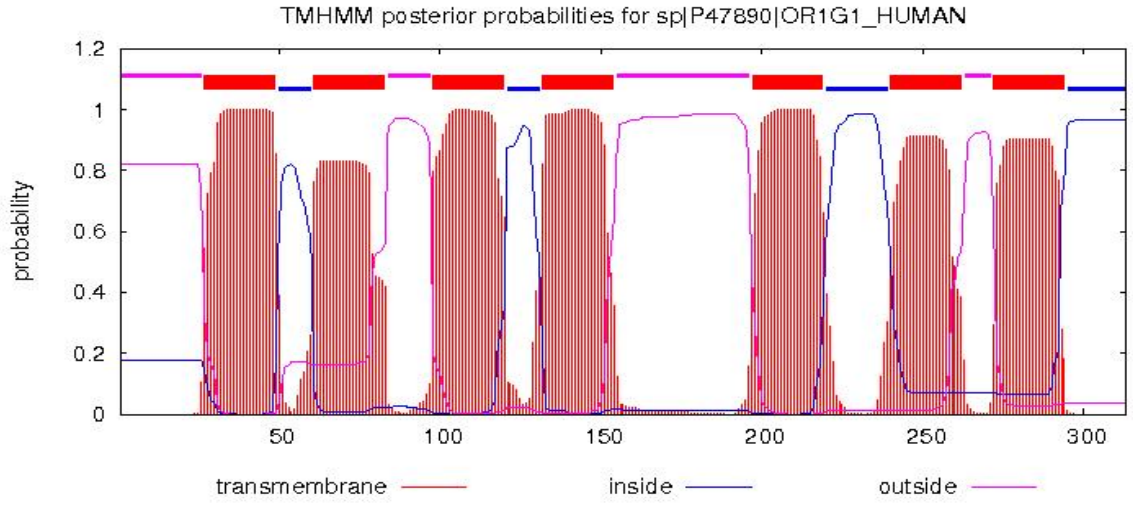


Fig. 32 Secondary structure of OR1G1 using TMHMM server

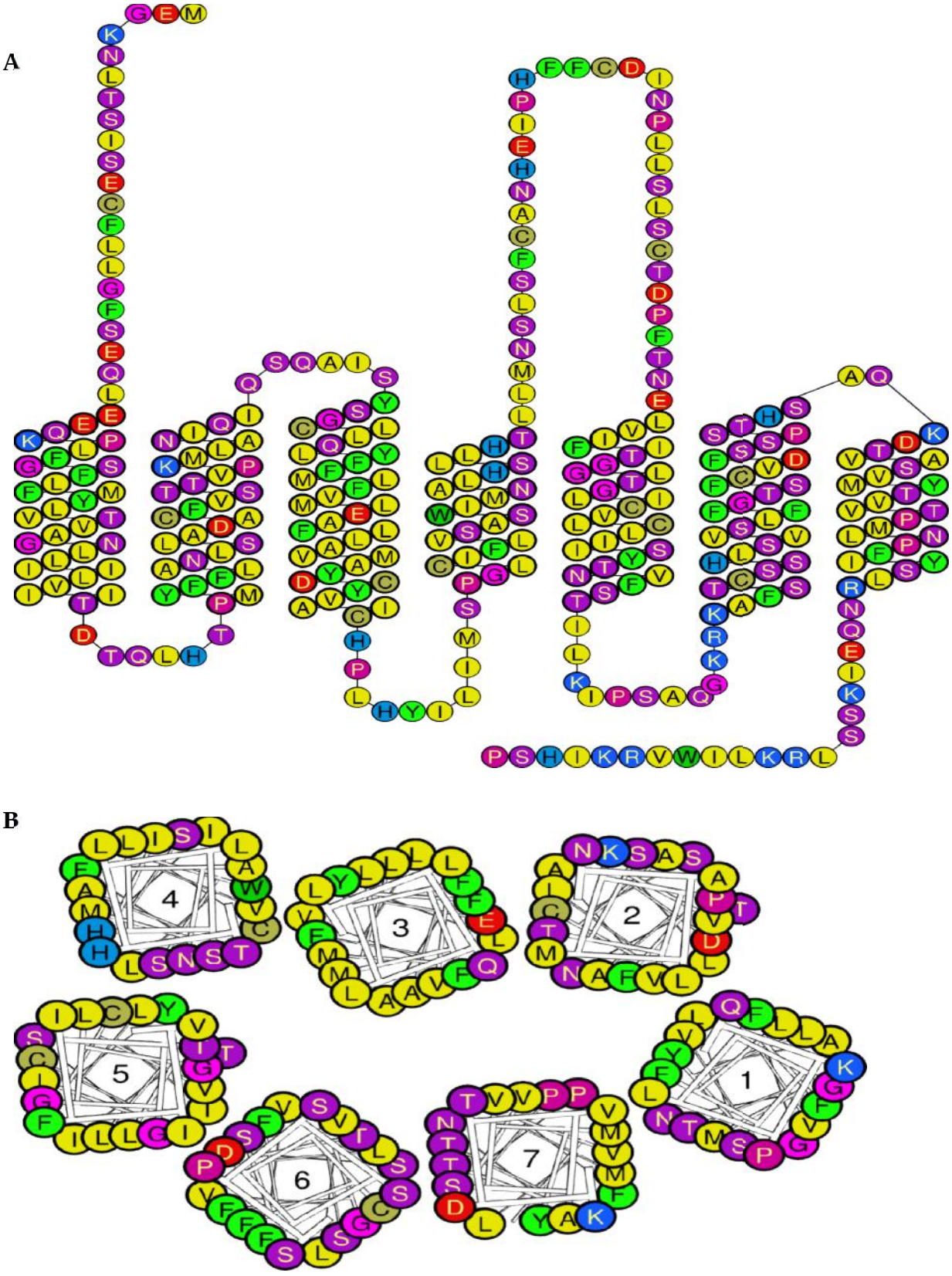


Fig. 33 Another computational tools for secondary structure prediction. A) Is the snake plot. B) Is the helix box for OR1G1. Through these it can be predicted where the binding buckets could be found. The sequences are colored according to their properties. Adapted from http://tools.gpcr.org/visualise/protein/or1g1_human

4.4 Conserved sequence motifs in OR1G1

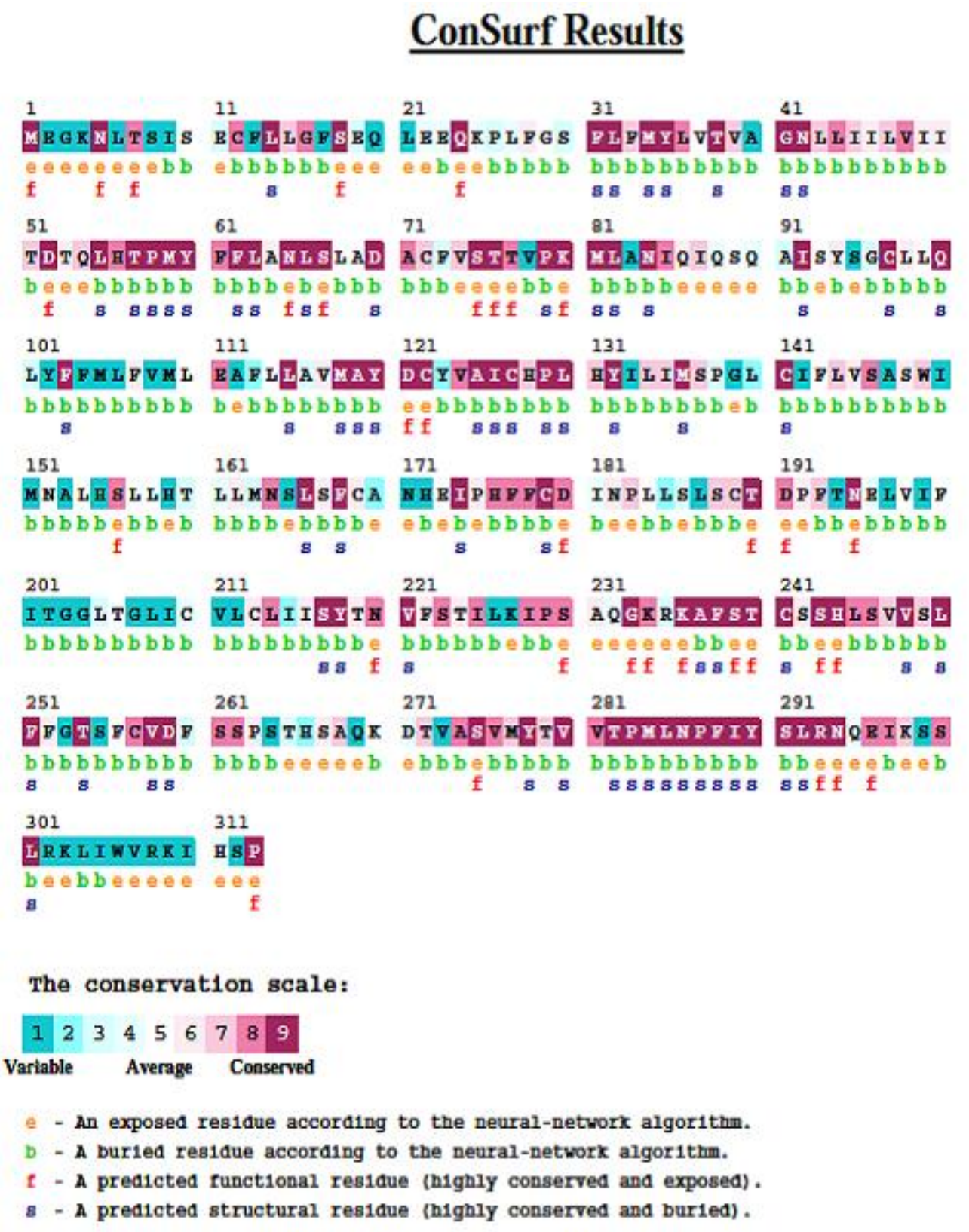


Fig. 34 Conserved sequences at OR1G1 by consurf results server. This server for the identification of functional regions in proteins http://consurf.tau.ac.il

4.5 Template and templates selection.

The first important step for homology modeling is the template selection. To select the most suitable template, many factors should be taken into account, like the sequence similarity, the length of template in comparison to target, the presence of the highly conserved residues of the target, the resolution and R-factor of a crystallographic structure of the template. There are different methods to select the suitable template (as discussed before in 4.2.5). One of these methods is a measure of compatibility between the sequence of target and the structure of template by calculating The PROSAII Z-score (63). PROSAII Z-score is sufficiently accurate to choose the most suitable template (64). In MOE server one can select the template by using homology searching, using a generalized version of the FASTA. Then the selection can be narrowed by calculating the E-Value and the Z-Score. The biggest Z-Score shows the best template family.

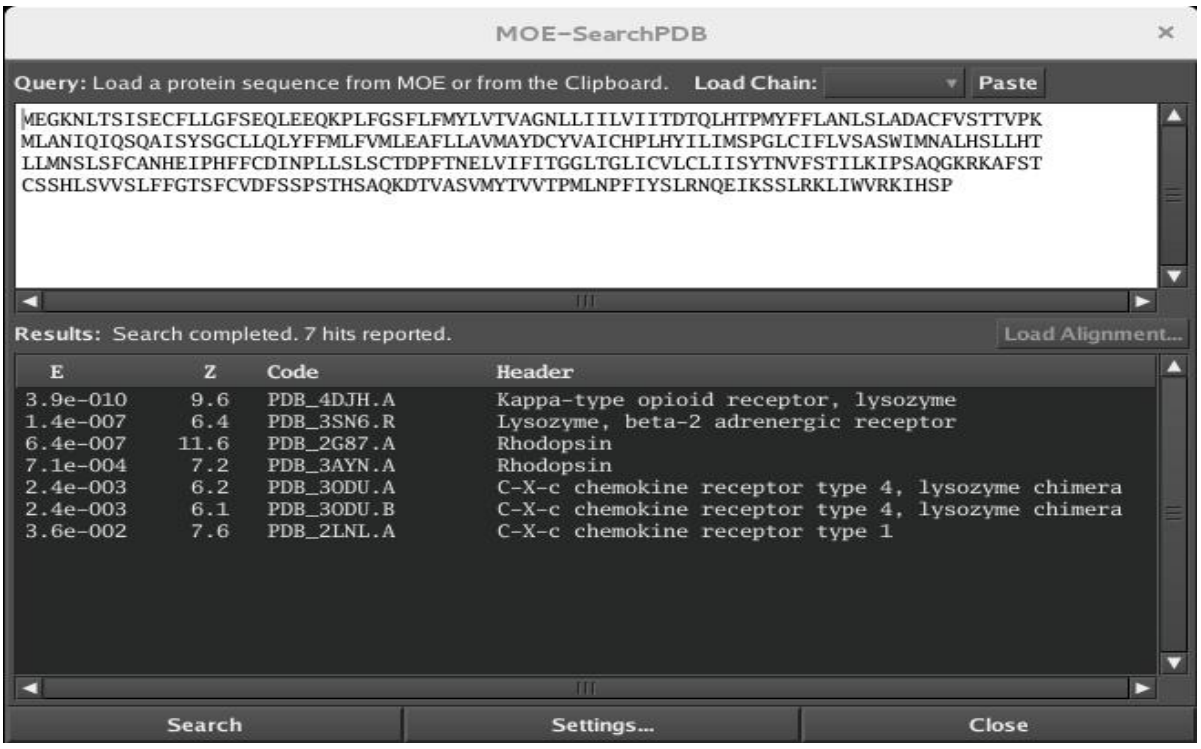


Fig.35 MOE-search PDB, the biggest Z-Score is 2G87. A (Rhodopsin) as a template

Based on the analysis of fig 35 and 36, the most suitable template for homology modeling of OR1G1 is Rhodopsin. By using MOE the biggest Z-Score for 2G87 which is 11.6 can be seen. By using hhpred server one can see 1U19 as a suitable template with probability 100%, in addition it features a good resolution 2.2 A, and a length of 349 aa which is close to the target length of 313 aa.

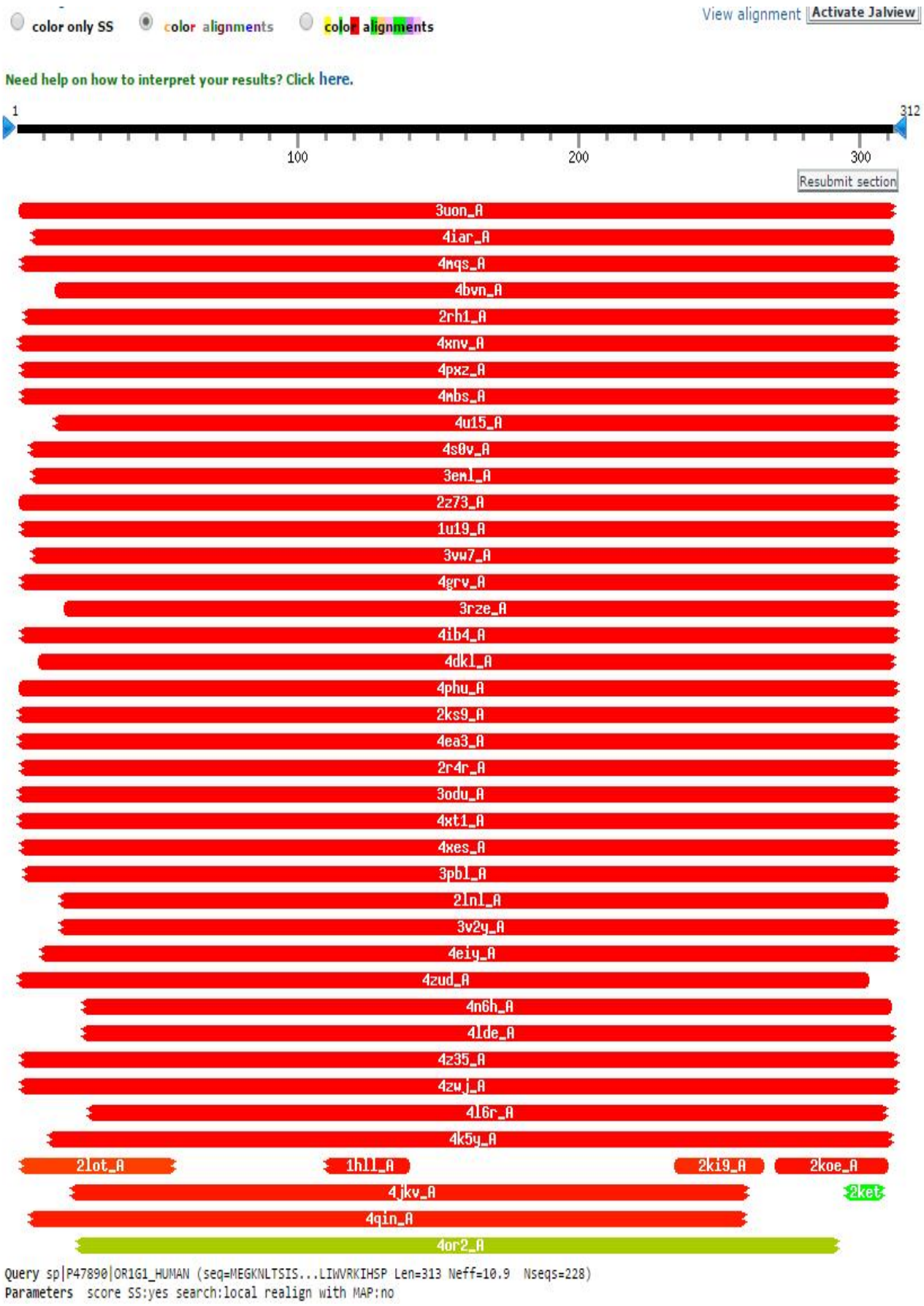


Fig. 36 Template searching using hhpred server http://toolkit.tuebingen.mpg.de/hhpred

Table 4 Probability of each template to OR1G1 using hhpred server.

Query sp|P47890|OR1G1_HUMAN (seq=MEGKNLTSIS...LIWVRKIHSP Len=313 Neff=10.9 Nseqs=228)
Parameters score SS:yes search:local realign with MAP:no

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	3uon_A Human M2 muscarinic ace	100.0	5.2E-44	1.4E-48	317.7	14.2	299	1-312	1-462 (467)
2	4iar_A Chimera protein of huma	100.0	1.9E-45	5.2E-50	320.7	2.3	291	6-312	2-401 (401)
3	4mqs_A Muscarinic acetylcholin	100.0	3.4E-44	9.3E-49	307.4	10.0	300	2-313	18-339 (351)
4	4bvn_A Beta-1 adrenergic recep	100.0	5.3E-45	1.5E-49	307.9	3.1	290	14-313	1-306 (315)
5	2rh1_A Beta-2-adrenergic recep	100.0	7.3E-44	2E-48	319.5	8.7	301	3-313	19-482 (500)
6	4xnv_A P2Y purinoceptor 1, rub	100.0	1.7E-43	4.6E-48	310.5	10.3	304	1-313	23-394 (421)
7	4pxz_A P2Y purinoceptor 12, so	100.0	5.4E-44	1.5E-48	317.2	5.5	300	2-313	12-435 (466)
8	4mbs_A Chimera protein of C-C	100.0	8.3E-43	2.3E-47	305.1	11.5	301	2-313	10-372 (414)
9	4u15_A Muscarinic acetylcholin	100.0	6E-43	1.7E-47	306.6	9.2	288	14-313	2-406 (418)
10	4s0v_A Human orexin receptor t	100.0	6.8E-42	1.9E-46	310.7	11.6	300	5-313	41-549 (559)
11	3eml_A Human adenosine A2A rec	100.0	1.2E-42	3.4E-47	310.7	4.8	302	6-313	3-472 (488)
12	2z73_A Rhodopsin; visual pigme	100.0	6.3E-42	1.8E-46	302.9	9.3	303	1-313	1-342 (448)
13	1u19_A Rhodopsin; G protein-co	100.0	2.3E-42	6.4E-47	295.7	6.1	303	2-313	13-329 (349)
14	3vw7_A Proteinase-activated re	100.0	4.3E-42	1.2E-46	306.7	6.9	301	6-313	2-470 (484)
15	4grv_A Neurotensin receptor ty	100.0	7E-43	2E-47	313.8	1.4	304	2-313	10-492 (510)
16	3rze_A Histamine H1 receptor,	100.0	4.3E-42	1.2E-46	304.2	6.3	285	17-313	1-448 (452)
17	4ib4_A Chimera protein of huma	100.0	7.7E-42	2.1E-46	300.5	7.4	299	2-313	7-418 (430)
18	4dkl_A MU-type opioid receptor	100.0	2.9E-41	8E-46	299.8	10.7	292	8-312	1-460 (464)
19	4phu_A Free fatty acid recepto	100.0	2.9E-41	8.1E-46	301.9	10.4	297	1-313	1-473 (491)
20	2ks9_A Substance-P receptor; w	100.0	9.1E-42	2.5E-46	293.8	5.5	302	1-313	9-327 (364)
21	4ea3_A Fusion protein of nocic	100.0	1.2E-40	3.3E-45	293.4	12.7	302	1-313	100-415 (434)
22	2r4r_A Beta-2 adrenergic recep	100.0	3.2E-43	9E-48	302.8	-4.7	302	2-313	11-347 (365)
23	3odu_A C-X-C chemokine recepto	100.0	2.3E-42	6.3E-47	309.9	0.3	298	1-313	26-498 (502)
24	4xt1_A HHRF3, G-protein couple	100.0	1.2E-40	3.5E-45	286.6	11.0	292	1-313	9-321 (362)
25	4xes_A Neurotensin receptor ty	100.0	9.3E-41	2.6E-45	302.1	10.0	304	2-313	10-359 (541)
26	3pbl_A D(3) dopamine receptor,	100.0	3E-41	8.4E-46	301.1	6.1	291	3-313	17-477 (481)
27	2lnl_A C-X-C chemokine recepto	100.0	1.3E-41	3.6E-46	284.7	2.2	282	16-310	3-296 (296)
28	3v2y_A Sphingosine 1-phosphate	100.0	1.2E-39	3.2E-44	293.6	13.5	277	16-313	55-498 (520)
29	4eiy_A Adenosine receptor A2A/	100.0	4.6E-42	1.3E-46	303.2	-3.6	299	9-313	16-431 (447)
30	4zud_A Chimera protein of solu	100.0	2E-40	5.7E-45	289.9	5.8	296	1-303	94-410 (410)
31	4n6h_A D-OR-1, DOR-1, soluble	100.0	4.3E-39	1.2E-43	281.9	11.5	277	24-311	125-414 (414)
32	4lde_A Lysozyme, beta-2 adrene	100.0	3.9E-40	1.1E-44	292.9	4.4	279	24-312	182-467 (469)
33	4z35_A Lysophosphatidic acid r	100.0	2.6E-39	7.2E-44	287.3	8.3	293	2-313	27-442 (464)
34	4zwj_A Chimera protein of huma	100.0	1.3E-35	3.6E-40	282.5	9.8	303	2-313	172-488 (906)
35	4l6r_A Soluble cytochrome B562	100.0	4.9E-32	1.4E-36	237.6	-0.4	256	26-309	124-402 (425)
36	4k5y_A Corticotropin-releasing	99.6	8.1E-17	2.2E-21	139.9	3.3	272	12-311	2-438 (441)
37	2koe_A Human cannabinoid recep	98.7	5.1E-09	1.4E-13	58.6	1.5	40	270-310	1-40 (40)
38	1hll_A Alpha-2A adrenergic rec	98.4	1.8E-08	4.9E-13	53.3	-1.9	31	110-140	2-32 (32)
39	4jkv_A Soluble cytochrome B562	97.9	2.4E-05	6.7E-10	69.3	6.6	229	20-260	143-387 (475)
40	4qin_A Smoothened homolog/solu	97.6	0.00067	1.9E-08	59.5	10.6	239	5-259	22-380 (468)
41	2ki9_A Cannabinoid receptor 2;	96.4	0.0011	2.9E-08	34.7	1.0	33	234-266	1-33 (33)
42	2lot_A Apelin receptor; membra	94.7	0.00016	4.4E-09	44.4	-7.3	55	2-56	6-60 (64)
43	4or2_A Soluble cytochrome B562	80.0	32	0.00089	29.6	15.1	243	22-292	116-367 (389)
44	2ket_A Cathelicidin-6; antimic	69.8	2.7	7.5E-05	19.4	1.2	13	295-308	4-16 (27)
45	2rdd_B UPF0092 membrane protei	37.2	32	0.0009	18.1	2.2	33	205-237	4-36 (37)
46	3twe_A Alpha4H; unknown functi	31.6	23	0.00064	16.3	0.9	16	295-312	11-26 (27)
47	4oo9_A Metabotropic glutamate	27.6	3.6E+02	0.01	23.7	9.9	238	27-292	14-420 (444)

4.5 Multiple sequence alignment, MSA.

4.5.1 Evolutionary Trace server ET.

In the beginning, I used Evolutionary Trace server to determine the highly conserved residues in each helix in all rhodopsin-like family and in all ORs (65). The members of the input superfamily are all rhodopsin-like (465 sequences) and those of the input subfamily are all ORs (2279 sequences). After the submission I got the following report which helped me with the manual adjustment of the MSA.

GPCR Results

Your input superfamily is All_Rhodopsin
Your input subfamily is All_Olfactory
Both were traced with the real-valued ET (Mihalek, I., et. al. (2004). J. Mol. Bio. 336(5): 1265-82.)
ET mapped onto the sequence
Hover your mouse pointer over the amino acid symbol to see the residue number (numbering follows PDB 1f88) Click on the amino acid letter and the corresponding residue will be highlighted on the structure as space fill / spheres

```
SUPERCLASS TM1 QFSMLAAYMFLLIMLGFPINFLTLYVTVQ
  SUBCLASS TM1 QFSMLAAYMFLLIMLGFPINFLTLYVTVQ

SUPERCLASS TM2 LNYILLNLAVADLFMVFGGFTTTLYT
  SUBCLASS TM2 LNYILLNLAVADLFMVFGGFTTTLYT

SUPERCLASS TM3 TGCNLEGFFATLGGEIALWSLVVLAIERYVVVCK
  SUBCLASS TM3 TGCNLEGFFATLGGEIALWSLVVLAIERYVVVCK

SUPERCLASS TM4 AIMGVAFTWVMALACAAPPLVGW
  SUBCLASS TM4 AIMGVAFTWVMALACAAPPLVGW

SUPERCLASS TM5 SFVIYMFVVHFIIPLIVIFFCYGQLVFTV
  SUBCLASS TM5 SFVIYMFVVHFIIPLIVIFFCYGQLVFTV

SUPERCLASS TM6 VTRMVIIMVIAFLICWLPYAGVAFYIFT
  SUBCLASS TM6 VTRMVIIMVIAFLICWLPYAGVAFYIFT

SUPERCLASS TM7 IFMTIPAFFAKTSAVYNPVIYIMMNK
  SUBCLASS TM7 IFMTIPAFFAKTSAVYNPVIYIMMNK
```

ET color scale

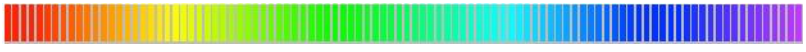
More important  Less important

Fig. 37 Evolutionary Trace results are mapped onto Rhodopsin receptor (PDBID 1F88) and Beta2-adrenergic receptor (PDBID 2RH1).
Paggenerated on: Wed Sep 30 21:05:53 2015
http://mammoth.bcm.tmc.edu/ETserver.html

4.5.2 MSA for randomly selected ORs.

Based on analysis of this alignment and according to Ballesteros-Weinstein numbering, it was concluded that: N of conserved GN at TM1 is 1.50 (66), GN is conserved at 90% within hORs and 99% within mouse ORs (67,68). D of conserved LSxxD at TM2 is 2.50, R or C which after conserved MAYD/E and before the conserved Y at TM3 is 3.50, conserved W at TM4 is 4.50, Y of conserved SY at TM5 is 5.58, H of conserved KAFSTCxSH at TM6 is 6.40, and P of conserved NPxxY at TM7 is 7.50. All of these motifs are often conserved within most of ORs.

Fig. 38 multiple sequence alignment of some of ORs by TM-coffee

```
HOR1G1_HUMAN    ------------------------------------------------------------------------
HOR2AT4         MDA---------------T---ACNESVDGSPVFYLLGIPSLPETFFLPVFFIFLLFYLLILMGNA
HOR1A1          MRE---------------N-----NQS--STLEFILLGVTGQQEQ-EDFFYILFLFIYPITLIGNL
HOR1A2          MKK---------------E-----NQS--FNLDFILLGVTSQQEQ-NNVFFVIFLCIYPITLTGNL
HOR2A1          MGE-----------------------NQT--MVTEFLLLGFLLGPRI-QMLLFGLFSLFYIFTLLGNG
HOR2A14         MEG-----------------------NKT--WITDITLPRFQVGPAL-EILLCGLFSAFYTLTLLGNG
HOR3A1          MQPE------------SGA-----NGT--VIAEFILLGLLEAPGL-QPVVFVLFLFAYLVTVRGNL
HOR4A16         MR----------------------PSS--NVTEFVLLGLTQDPDV-KKTLFVMFLLIYIVTMVGNL
HOR5A1          MSIT------------KAW-----NSS--SVTMFILLGFTDHPEL-QALLFVTFLGIYLTTLAWNL
HOR6B1          MEL---------------E-----NQT--RVTKFILVGFPGSLSM-RAAMFLIFLVAYILTVAENV
HOR7A17         MEP---------------E-----NDT--GISEFVLLGLSEEPEL-QPFLFGLFLSMYLVTVLGNL
HOR8B2          MLA---------------R-----NNS--LVTEFILAGLTDHPEF-RQPLFFLFLVIYIVTMVGNL
HOR9G1          MQR---------------S-----NH---TVTEFILLGFTTDPGM-QLGLFVVFLGVYSLTVVGNS
HOR10A6         MER---------------Q-----NQS--CVVEFILLGFSNYPEL-QGQLFVAFLVIYLVTLIGNA
HOR11H1         MCPLTLQVTGLMNVSEPNS-----SFA--FVNEFILQGFSCEWTI-QIFLFSLFTTTYALTITGNG
HOR12D2         ML----------------------NTT--SVTEFLLLGVTDIQEL-QPFLFVVFLTIYFISVTGNG
HOR13C2         MEW---------------E-----NHT--ILVEFFLKGLSGHPRL-ELLFFVLIFIMYVVILLGNG
HOR51A2         MSI---------------I-----NTSYVEITTFFLVGMPGLEYA-HIWISIPICSMYLIAILGNG
HOR52B2         MSH---------------T-----NVTIFHPAVFVLPGIPGLEAY-HIWLSIPLCLIYITAVLGNS
HOR52A1         MSI---------------S-----NITVYMPSVLTLVGIPGLESV-QCWIGIPFCAIYLIAMIGNS
HOR56A1         MIQ---------------PMASPSNSSTVPVSEFLLTCFPNFQSW-QHWLSLPLSLLFLLAMGANT

HOR1G1_HUMAN    ----------------FFLANLSLADACFVSTTVPKMLANIQIQSQAISYSGCLLQLYFFMLFVM
HOR2AT4         LILVAVVAEPSLHKPMYFFLINLSTLDILFTTTTVPKMLSLFLLGDRFLSFSSCLLQMYLFQSFTC
HOR1A1          LIVLAICSDVRLHNPMYFLLANLSLVDIFFSSVTIPKMLANHLLGSKSISFGGCLTQMYFMIALGN
HOR1A2          LIILAICADIRLHNPMYFLLANLSLVDIIFSSVTIPKVLANHLLGSKFISFGGCLMQMYFMIALAK
HOR2A1          AILGLISLDSRLHTPMYFFLSHLAVVDIAYTRNTVPQMLANLLHPAKPISFAGCMTQTFLCLSFGH
HOR2A14         VIFGIICLDCKLHTPMYFFLSHLAIVDISYASNYVPKMLTNLMNQESTISFFPCIMQTFLYLAFAH
HOR3A1          SILAAVLVEPKLHTPMYFFLGNLSVLDVGCISVTVPSMLSRLLSRKRAVPCGACLTQLFFFHLFVG
HOR4A16         LIWVTTIGSPSLGSLMYFFLAYLSLMDAIYSTAMSPKLMIDLLCDKIAISLSACMGQLFIEHLLGG
HOR5A1          ALIFLIRGDTHLHTPMYFFLSNLSFIDICYSSAVAPNMLTDFFWEQKTISFVGCAAQFFFFVGMGL
HOR6B1          IIILLVLQNRPLHKPMYFFLANLSFLETWYISVTVPKLLFSFWSVNNSISFTLCMIQLYFFIALMC
HOR7A17         LIILATISDSHLHTPMYFFLSNLSFADICFISTTIPKMLINIQTQSRVITYAGCITQMCFFVLFGG
HOR8B2          GLITLFGLNSHLHTPMYYFLFNLSFIDLCYSSVFTPKMLMNFVSKKNIISNVGCMTRLFFFLFFVI
HOR9G1          TLIVLICNDSCLHTPMYFFTGNLSFLDLWYSSVYTPKILVTCISEDKSISFAGCLCQFFFSAGLAY
HOR10A6         IIIVIVSLDQSLHVPMYLFLLNLSVVDLSFSAVIMPEMLVVLSTEKTTISFGGCFAQMYFILLFGG
HOR11H1         AIAFVLWCDRRLHTPMYMFLGNFSFLEIWYVSSTVPKMLVNFLSEKKNISFAGCFLQFYFFFSLGT
HOR12D2         AVLMIVISDPRLHSLMYFFLGNLSYLDICYSTVTLPKMLQNFLSTHKAISFLGCISQLHFFHSLGS
HOR13C2         TLILISILDPHLHTPMYFFLGNLSFLDICYTTTSIPSTLVSFLSERKTISLSGCAVQMFLGLAMGT
HOR51A2         TILFIIKTEPSLHGPMYYFLSMLAMSDLGLSLSSLPTVLSIFLFNAPETSSSACFAQEFFIHGFSV
HOR52B2         ILIVVIVMERNLHVPMYFFLSMLAVMDILLSTTTVPKALAIFWLQAHNIAFDACVTQGFFVHMMFV
HOR52A1         LLLSIIKSERSLHEPLYIFLGMLGATDIALASSIMPKMLGIFWFNVPEIYFDSCLLQMWFIHTLQG
HOR56A1         TLLITIQLEASLHQPLYYLLSLLSLLDIVLCLTVIPKVLAIFWYDLRSISFPACFLQMFIMNSFLP
                                  :   :.  :        *  :              *  :  :   :
```

```
HOR1G1_HUMAN    LEAFLLAVMAYDCYVAICHPLHYILIMSPGLCIFLVSASWIMNALHSLLHTL-LMNSLSFCANHEI
HOR2AT4         SEAFILVVMAYDRYVAICHPLHYPVLMNPQTNATLAASAWLTALLLPIPAVV-RTSQMAYNSIAYI
HOR1A1          TDSYILAAMAYDRAVAISRPLHYTTIMSPRSCIWLIAGSWVIGNANALPHTL-LTASLSFCGNQEV
HOR1A2          ADSYTLAAMAYDRAVAISCPLHYTTIMSPRSCILLIAGSWVIGNTSALPHTL-LTASLSFCGNQEV
HOR2A1          SECLLLVLMSYDRYVAICHPLRYSVIMTWRVCITLAVTSWTCGSLLALAHVV-LILRLPFSGPHEI
HOR2A14         VECLILVVMSYDRYADICHPLRYNSLMSWRVCTVLAVASWVFSFLLALVPLV-LILSLPFCGPHEI
HOR3A1          VDCFLLTAMAYDRFLAICRPLTYSTRMSQTVQRMLVAASWACAFTNALTHTV-AMSTLNFCGPNVI
HOR4A16         AEVFLLVVMAYDRYVAISKPLHYLNIMNRLVCILLLVVAMIGGFVHSVVQIV-FLYSLPICGPNVI
HOR5A1          SECLLLTAMAYDRYAAISSPLLYPTIMTQGLCTRMVVGAYVGGFLSSLIQAS-SIFRLHFCGPNII
HOR6B1          TECVLLAAMAYDRYVAICRPLHYPTIMSHGLCFRLALGSWAIGFGISLAKIY-FISCLSFCGPNVI
HOR7A17         LDSLLLAVMAYDRFVAICHPLHYTVIMNPRLCGLLVLASWMIAALNSLSQSL-MVLWLSFCTDLEI
HOR8B2          SECYMLTSMAYDRYVAICNPLLYKVTMSHQVCSMLTFAAYIMGLAGATAHTG-CMLRLTFCSANII
HOR9G1          SECYLLAAVAYDRYVAISKPLLYAQAMSIKLCALLVAVSYCGGFINSSIITK-KTFSFNFCRENII
HOR10A6         AECFLLGAMAYDRFAAICHPLNYQMIMNKGVFMKLIIFSWALGFMLGTVQTS-WVSSFPFCGLNEI
HOR11H1         SECLLLTVMAFDQYLAICRPLLYPNIMTGHLYAKLVILCWVCGFLWFLIPIV-LISQMPFCGPNII
HOR12D2         TESMLFAVMAFDLSVAICKPLRYTVIMNPQLCTQMAITIWVIGFFHALLHSV-MTSRLNFCGSNRI
HOR13C2         TECVLLGMMAFDRYVAICNPLRYPIIMSKDAYVPMAAGSWIIGAVNSAVQSV-FVVQLPFCRNNII
HOR51A2         LESSVLLIMSFDRFLAIHNPLRYTSILTTVRVAQIGIVFSFKSMLLVLPFPF-TLRSLRYCKKNQL
HOR52B2         GESAILLAMAFDRFVAICAPLRYTTVLTWPVVGRIALAVITRSFCIIFPVIF-LLKRLPFCLTNIV
HOR52A1         IESGILVAMALDRYVAICYPLRHANIFTHQLVIQIGTMVVLRAAILVAPCLVLIKCRFQFYHTTVI
HOR56A1         MESCTFMVMAYDRYVAICHPLRYPSIITNQFVAKASVFIVVRNALLTAPIPI-LTSLLHYCGENVI
                 :   :  :: *    *  ** :   :.                          :       :

HOR1G1_HUMAN    PHFFCDINPLLSLSCTDPFTNELVIFITGGLTGLICVLCLIISYTNVFST-ILKIPSAQGKRKAFS
HOR2AT4         YHCFCDHLAVVQASCSDTTPQTLMGFCIAMVVSFLPLLLVLLSYVHILAS-VLRISSLEGRAKAFS
HOR1A1          ANFYCDITPLLKLSCSDIHFHVKMMY-LGVGIFSVPLLCIIVSYIRVFST-VFQVPSTKGVLKAFS
HOR1A2          ANFYCDIMPLLKLSCSDVHFNVKMMY-LGVGVFSLPLLCIIVSYVQVFST-VFQVPSTKSLFKAFC
HOR2A1          NHFFCEILSVLRLACADTWLNQVVIFAACVFFLVGPPSLVLVSYSHILAA-ILRIQSGEGRRKAFS
HOR2A14         NHFFCEILSVLKLACADTWLNQVVIFAACVFILVGPLCLVLVSYLRILAA-ILRIQSGEGRRKAFS
HOR3A1          NHFYCDLPQLFQLSCSSTQLNELLLFAVGFIMAGTPMALIVISYIHVAAA-VLRIRSVEGRKKAFS
HOR4A16         DHSVCDMYPLLELLCLDTYFIGLTVVANGGIICMVIFTFLLISCGVILNF--LKTYSQEERHKALP
HOR5A1          NHFFCDLPPVLALSCSDTFLSQVVNFLVVVTVGGTSFLQLLISYGYIVSA-VLKIPSAEGRWKACN
HOR6B1          NHFFCDISPVLNLSCTDMSITELVDFILALVIFLFPLFITVLSYGCILAT-ILCMPT--GKQKAFS
HOR7A17         PHFFCELNQVIHLACSDTFLNDMGMYFAAGLLAGGPLVGILCSYSKIVSS-IRAISSAQGKYKAFS
HOR8B2          NHYLCDILPLLQLSCTSTYVNEVVVLIVVGTNITVPSCTILISYVFIVTS-ILHIKSTQGRSKAFS
HOR9G1          DDFFCDLLPLVELACGEKGGYKIMMYFLLASNVICPAVLILASYLFIITS-VLRISSSKGYLKAFS
HOR10A6         NHISCETPAVLELACADTFLFEIYAFTGTFLIILVPFLLILLSYIRVLFA-ILKMPSTTGRQKAFS
HOR11H1         DHVVCDPGPRFALDCVSAPRIQLFCYTLSSLVIFGNFLFIIGSYTLVLKA-MLGMPSSTGRHKAFS
HOR12D2         HHFLCDIKPLLKLACGNTELNQWLLSTVTGTIAMGPFFLTLLSYFYIITYLFFKTRSCSMLCKALS
HOR13C2         NHFTCEILAVMKLACADISDNEFIMLVATTLFILTPLLLIIVSYTLIIVS-IFKISSSEGRSKASS
HOR51A2         SHSYCLHQDVMKLACSDNRIDVIYGFFGAL-CLMVDFILIAVSYTLILKT-VPGIASKKEELKALN
HOR52B2         PHSYCEHIGVARLACADITVNIWYGFSVPIVMVILDVILIAVSYSLILRA-VFRLPSQDARHKALS
HOR52A1         SHSYCEHMAIVKLAAANVQVNKIYGLFVAFTVAGFDLTFITLSYIQIFIT-VFRLPQKEARFKAFN
HOR56A1         ENCICANLSVSRLSCDNFTLNRIYQFVAGWTLLGSDLFLIFLSYTFILRA-VLRFKAEGAAVKALS
                 .  *         . .                          *   :              **

HOR1G1_HUMAN    TCSSHLSVVSLFFGTSFCVDFSSPS--THSA------QKDTVASVMYTVVTPMLNPFIYSLRNQEI
HOR2AT4         TCSSHLLVVGTYYSSIAIAYVAYRA--DLPL------DFHIMGNVVYAILTPILNPLIYTLRNRDV
HOR1A1          TCGSHLTVVSLYYGTVMGTYFRPLT--NYS-------LKDAVITVMYTAVTPMLNPFIYSLRNRDM
HOR1A2          TCGSHLTVVFLYYGTTMGMYFRPLT--SYS-------PKDAVITVMYVAVTPALNPFIYSLRNWDM
HOR2A1          TCSSHLCVVGLFFGSAIIMYMAPKS--RHPE------EQQKVFFLFYSFFNPTLNPLIYSLRNGEV
HOR2A14         TCSSHLCVVGLFFGSAIVTYMAPKS--RHPE------EQQKVLSLFYSLFNPMLNPLIYSLRNAEV
HOR3A1          TCGSHLTVVAIFYGSGIFNYMRLGS--TKLS------DKDKAVGIFNTVINPMLNPIIYSFRNPDV
HOR4A16         TCISHIIVVALVFVPCIFMYVRPV--SNFPF---------DKLMTVFYSIITLMLNPLIYSLRQSEM
HOR5A1          TCASHLMVVTLLFGTALFVYLRPSS--SYLL------GRDKVVSVFYSLVIPMLNPLIYSLRNKEI
HOR6B1          TCASHLVVVTIFYSAIIFMYARPRV--IHAF------NMNKIISIFYAIVTPSLNPFIYCLRNREV
HOR7A17         TCASHLSVVSLFCCTGLGVYLTSAA--THNS------HTSATASVMYTVATPMLNPFIYSLRNKDI
HOR8B2          TCSSHVIALSLFFGSAAFMYIKYSS--G-SM------EQGKVSSVFYTNVVPMLNPLIYSLRNKDV
HOR9G1          TCSSHLTSVTLYYGSILYIYALPRS--SYSF------DMDKIVSTFYTVVFPMLNLMIYSLRNKDV
HOR10A6         TCAAHLTSVTLFYGTASMTYLQPKS--GYSP------ETKKVMSLSYSLLTPLLNLLIYSLRNSEM
HOR11H1         TCGSHLAVVSLCYSSLMVMYVSPGL--GHST------GMQKIETLFYAMVTPLFNPLIYSLQNKEI
HOR12D2         TCASHFMVVILFYAPVLFTYIHPAL--ESFM------DQDRIVAIMYTVVTPVLNPLIYTLRNKEV
HOR13C2         TCSAHLTVVIIFYGTILFMYMKPKS--KETLNSDDLDATDKIISMFYGVMTPMMNPLIYSLRNKDV
HOR51A2         TCVSHICAVIIFYLPIINLAVVHRFAGHVSP------LINVLMANVLLLVPPLMKPIVYCVKTKQI
HOR52B2         TCGSHLCVILMFYVPSFFTLLTHHFGRNIPQ------HVHILLANLYVAVPPMLNPIVYGVKTKQI
HOR52A1         TCIAHICVFLQFYLLAFFSFFTHRFGSHISP------YIHILFSSIYLLVPPFLNPLVYGAKTTQI
HOR56A1         TCGSHFILILFFSTILLVVVLTNVARKKVPM------DILILLNVLHHLIPPALNPIVYGVRTKEI

                ** :*.  .                                          :: ::*  :  ::
```

4.5.3 Previous multiple sequence alignment of OR1G1 with some templates

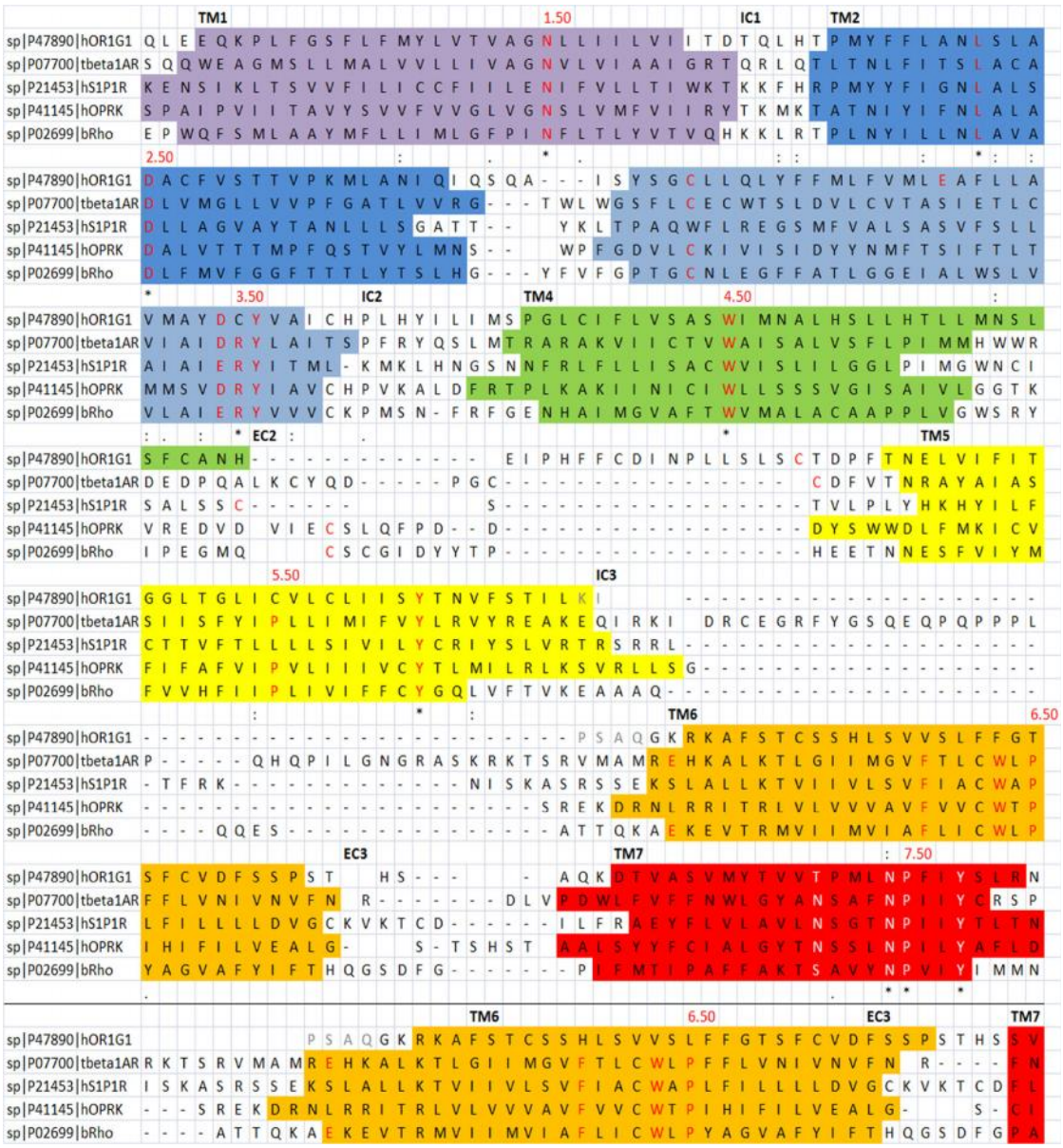


Fig. 39 The multiple sequence alignment of OR1G1 with four templates. The predicted helices are colored, each TM region is shown with a different color. The extracellular and intracellular regions are shown with gaps.

According to Ballesteros-Weinstein numbering:

TM1.50 is N, which highly conserved in rhodopsin-like family

TM2.50 is D. TM3.50 is Y. TM4.50 is W

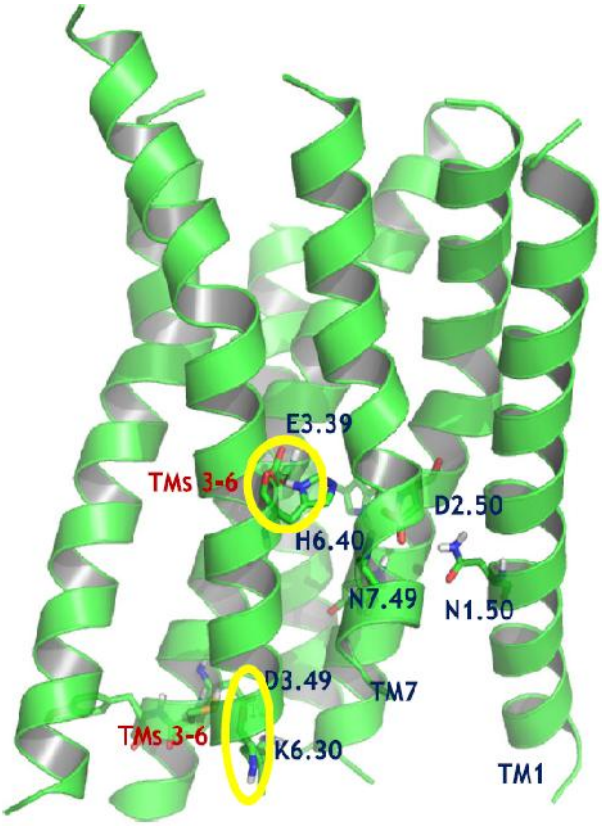
TM5.50 is P in non olfactory and C in OR1G1, which before Y5.58

TM6.50 is P in non olfactory and T in OR1G1

TM7.50 is p as shown. Adapted from (53)

	TM3				3.25																						3.50			
sp|P47890|OR1G1	S	Y	S	G	C	L	L	Q	L	Y	F	F	M	L	F	V	M	L	E	A	F	L	L	A	V	M	A	Y	D	C	Y	V	A
sp|O76100|OR7AA	T	Y	A	G	C	I	T	Q	M	C	F	F	L	L	F	V	G	L	D	N	F	L	L	T	V	M	A	Y	D	R	F	V	A
sp|Q15622|OR7A5	T	Y	I	A	C	L	M	Q	M	Y	F	F	I	L	F	A	G	F	E	N	F	L	L	S	V	M	A	Y	D	R	F	V	A
sp|O43749|OR1F1	S	F	C	G	C	L	T	Q	M	Y	F	V	F	M	F	V	D	M	D	N	F	L	L	A	V	M	A	Y	D	H	F	V	A
sp|O14581|OR7AH	T	Y	A	G	C	I	T	Q	M	C	F	F	V	L	F	G	G	L	D	S	L	L	L	A	V	M	A	Y	D	R	F	V	A
sp|Q8NHA8|OR1FC	S	Y	P	G	C	L	A	Q	M	Y	F	C	M	M	F	A	N	M	D	N	F	L	L	T	V	M	A	Y	D	R	Y	V	A
sp|Q96RA2|OR7D2	S	Y	M	D	C	L	T	Q	V	Y	F	S	M	F	F	P	I	L	D	T	L	L	L	T	V	M	A	Y	D	R	F	V	A
sp|Q8NG98|OR7D4	S	Y	M	G	C	L	T	Q	V	Y	F	L	M	M	F	A	G	M	D	T	F	L	L	A	V	M	A	Y	D	R	F	V	A
sp|P30953|OR1E1	P	Y	A	D	C	L	T	Q	M	Y	F	F	L	L	F	G	D	L	E	S	F	L	L	V	A	M	A	Y	D	R	Y	V	A
sp|Q15619|OR1C1	S	F	A	G	C	L	T	Q	L	F	F	F	V	S	F	V	N	M	D	S	L	L	L	C	V	M	A	Y	D	R	Y	V	A
sp|Q8NGA1|OR1M1	S	Y	P	C	C	L	I	Q	M	Y	F	F	H	F	F	G	I	V	D	S	V	I	I	A	M	M	A	Y	D	R	F	V	A
	TM6																					6.50											
sp|P47890|OR1G1	G	K	R	K	A	F	S	T	C	S	S	H	L	S	V	V	S	L	F	F	G	T	S	F	C	V	D	F	S	S	P	S	T
sp|O76100|OR7AA	G	K	Y	K	A	F	S	T	C	A	S	H	L	S	V	V	S	L	F	Y	G	T	C	L	G	V	Y	L	S	S	A	A	T
sp|Q15622|OR7A5	G	K	Y	K	A	F	S	T	C	A	S	H	L	S	V	V	S	L	F	Y	G	A	I	L	G	V	Y	L	S	S	A	A	T
sp|O43749|OR1F1	G	R	W	K	A	F	S	T	C	G	S	H	L	A	V	V	L	L	F	Y	S	T	I	I	A	V	Y	F	N	P	L	S	S
sp|O14581|OR7AH	G	K	Y	K	A	F	S	T	C	A	S	H	L	S	V	V	S	L	F	C	C	T	G	L	G	V	Y	L	T	S	A	A	T
sp|Q8NHA8|OR1FC	G	K	L	K	A	F	S	T	C	G	S	H	L	A	L	V	I	L	F	Y	G	A	I	T	G	V	Y	M	S	P	L	S	N
sp|Q96RA2|OR7D2	G	K	Q	K	A	L	S	T	C	G	S	H	L	S	V	V	S	L	F	Y	G	T	G	I	G	V	H	F	T	S	A	V	T
sp|Q8NG98|OR7D4	G	K	Y	K	A	F	S	T	C	G	S	H	L	C	V	V	S	L	F	Y	G	T	G	L	G	V	Y	L	S	S	A	V	T
sp|P30953|OR1E1	G	I	C	K	A	F	S	T	C	G	S	H	L	S	V	V	S	L	F	Y	G	T	V	I	G	L	Y	L	C	S	S	A	N
sp|Q15619|OR1C1	G	K	Q	R	A	V	S	T	C	S	C	H	L	S	V	V	V	L	F	Y	G	T	A	I	A	V	Y	F	S	P	S	S	P
sp|Q8NGA1|OR1M1	G	R	K	K	A	F	S	T	C	S	S	H	L	S	V	V	A	L	F	Y	G	T	T	I	G	V	Y	L	C	P	S	S	V

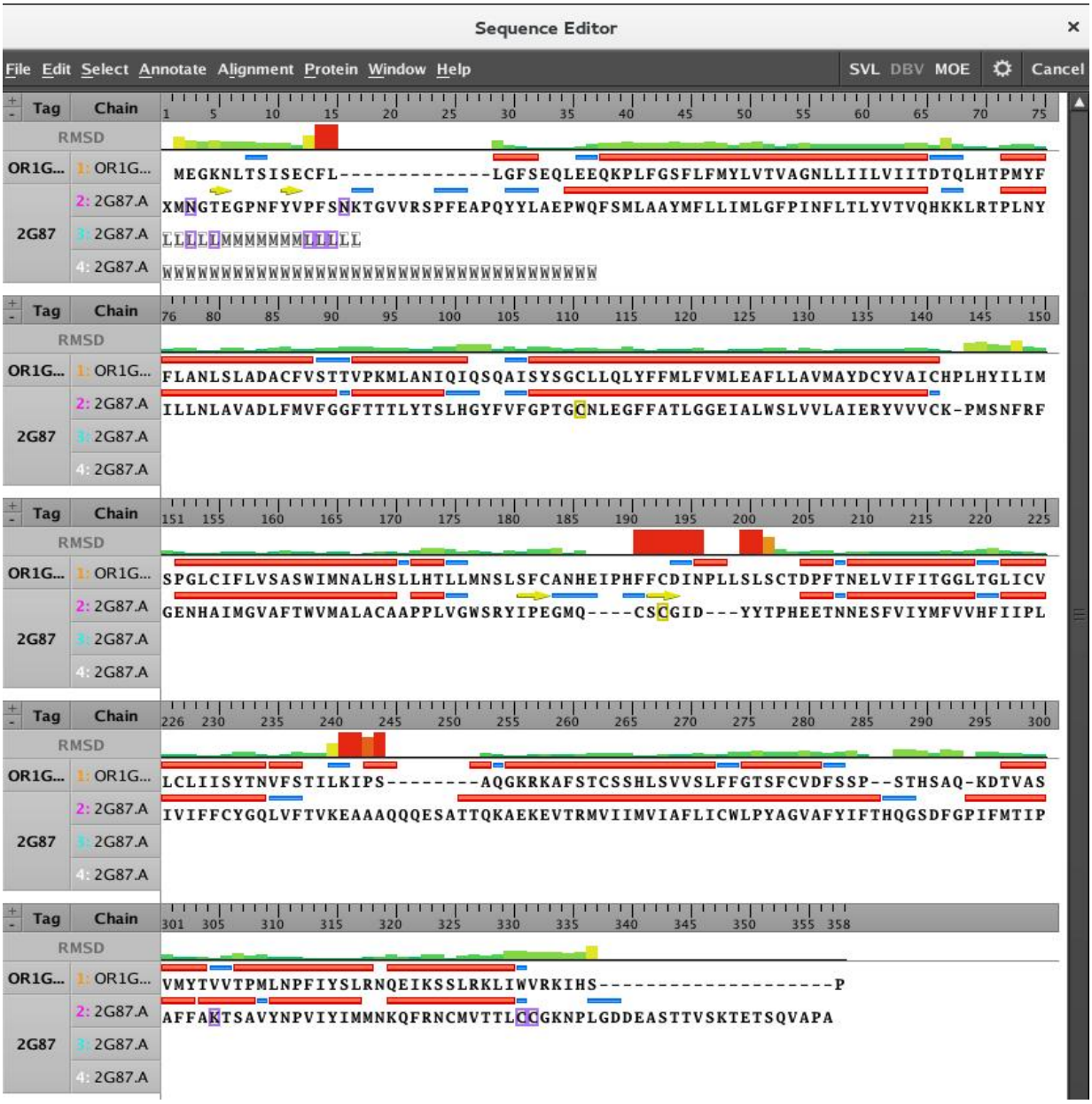
Fig. 40 TM3 and TM6 show multiple alignments of some of ORs. Adapted from (53). Based on the analysis of (Fig. 38 and 40) C97 at OR1G1 (or 3.25 according to Ballesteros-Weinstein numbering), I asserted that this residue is conserved at most ORs, which are responsible to form a disulfide bridge with C 179 at EL2 of OR1G1. Another disulfide bridge within EL2 is between C169 and C189, which was characterized by mass spectrometry in hOR1D2 (53.69). Conserved D/E 3.39 at TM3 and conserved H 6.40 at TM6 are predicted to form an ionic bridge. Conserved D 3.49 and conserved K 6.30 might also form an ionic bridge. Fig. 41 Ionic bridges predicted between TM3 and TM6. Adapted from (53)



4.5.4 Alignment of OR1G1 and 2G87.A using MOE/SE

I edited the alignment of OR1G1 and 2G87.A at Sequence Editor of MOE. In the alignment, I created GPCR constraints by this command SE/Alignment/Constraints/GPCR. Then I did manual editing by adding gaps or removing gaps, with a focus on the conserved residues at each helix. In the alignment the helices does not contain any gap, only at the loops.(70) Example for these conserved residues: N at TM1, D at TM2, C at TM3, W at TM4, Y at TM5, T at TM6 of OR1G1 aligns with P at TM6 of 2G87 (as discussed in 4.5.3) and P at TM7 (Fig. 42).

Fig. 42 Sequence Editor Panel for alignment of OR1G1 and 2G87.A



4.6 Building of the homology model

To construct the model for OR1G1, one can use ab-initio or homology modeling methods. The two methods depend on the known template (or templates) (71,72). The alignment between the target protein and the template is the most important step as discussed before. After the alignment with the suitable template, one can use any program (discussed in 4.2.8) for building the model. Here in this work I used MOE program for building a OR1G1 model by using 2G87.A as a template (Fig. 43)

The modeling by MOE comprises four steps:

1- Initial partial geometry specification.

Through this step an initial partial geometry for the query sequence from the template (or templates) is copied. Conserved residues between the query and template are copied as well as heavy-atoms, and disulfide bound which is formed between C 97 at OR1G1 and C 179 at EL2 are copied, too. In this step only the backbone of the query are copied.

2- Insertion and deletions.

After first step still some residues are without coordinates left. These residues are one of the following resedues:

A- Residues in loops, which can be added through insertion in the model according to the template.

B- Outgaps residues which are located after the N-terminus or before the C-terminus of the template.

C- Residues in the deletion-region in this case must be modeled as indels, which are copied from high-resolution chains. These chains are taken from Protein Data Bank (73). When no segment meets through RMSD criteria, the indels are permitted to back off (74).

3- Loop selection and side chain packing.

Once the insertion and deletions are finished, loops are modeled randomly. The contact energy function is analyzed for each loop, in this step the atoms which previously have been modeled and the ligand which is bound to the template, must be taken into account. Then a Boltzmann-weighted choice can be done according to the energy of each loop. After the loops are modeled, the server chooses the best side chain conformations from a library of high-resolution PDB.

4- Final model selection and refinement.

The choice of the final model is dependent on the best-scoring model, according to different available choices:

A- According to RMSD of each model.

B- According to the electrostatic solvation energy based on Generalized Born/Volume Integral (GB/VI) (75).

C- According to a knowledge-based residue packing quality function, which depends on solvent accessibility for each residue, possibility of residue to form hydrogen bonds, and the ratio between polar versus non-polar contacting atoms in its environment.

D- According to the effective atomic contact energy (ACE) (76,77)

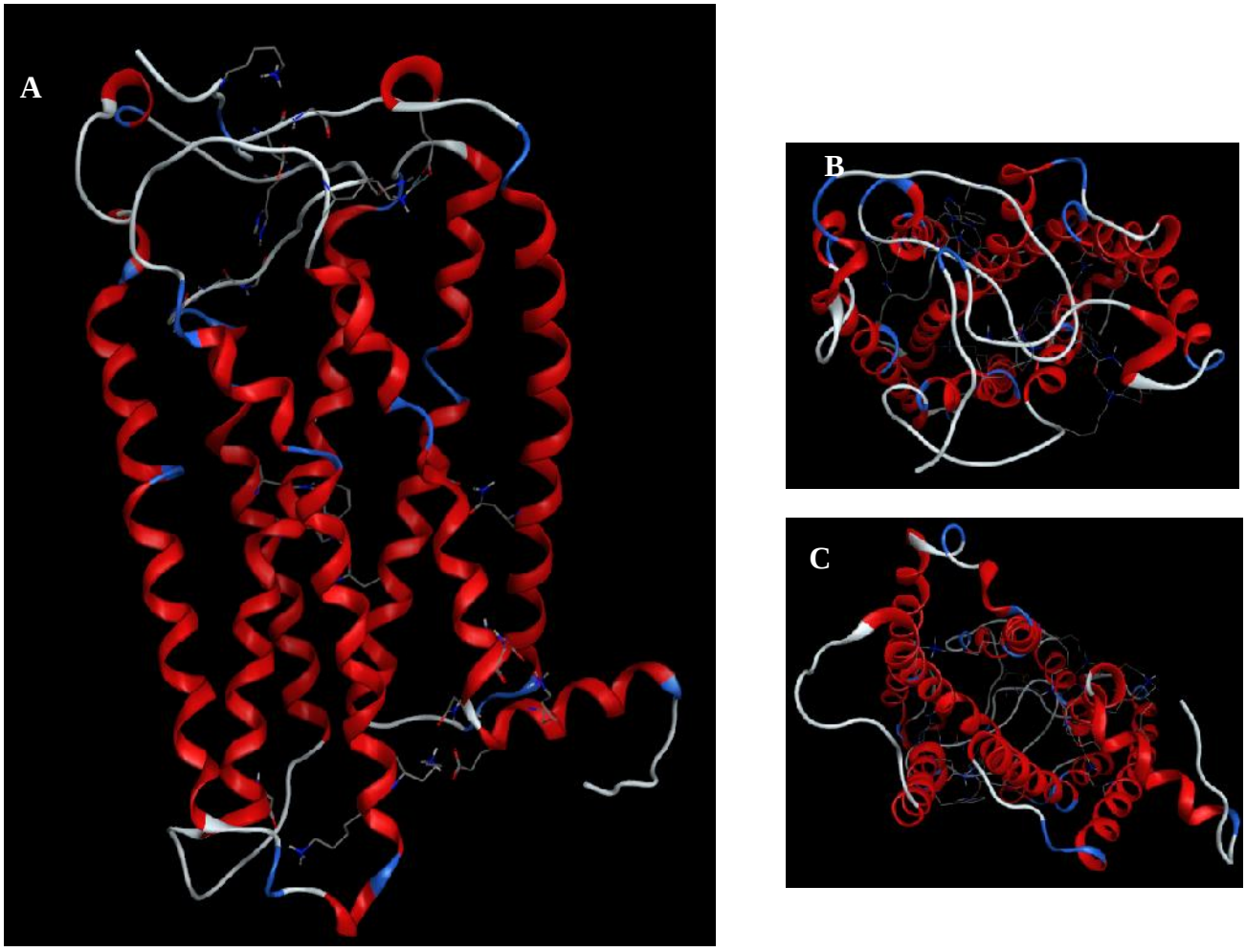


Fig. 43 3D structure of OR1G1, A) vertical view of OR1G1 describes 7TMs, intracellular and extracellular loops, and some of hydrogen bonds between the helices. B) Horizontal view from extracellular site, describes EL2 which acts as lid for binding bucket. C) Horizontal view from intracellular site, describes the cleft for binding of the receptor with alpha G-Protein.

4.7 Evaluation and refinement of the model

Homology model may contain errors because of two main reasons

1- The percentage of sequence identity between the template and the model is not high enough.

2-There are some errors in the templates

So it is important to evaluate the correctness of overall fold/ structure, errors of localized regions and stereochemical parameters which includes geometries, bond lengths, bond angles, peptide bond and side-chain planarity, torsion angels of the main-chain and the side-chain of the protein.

4.7.1 Stereochemistry

After the homology modeling steps have been finished, we should inspect the final model through Phi-Psi plot, or also known as a Ramachandran Plot SE/Protein/Geometry/Phi-Psi Plot

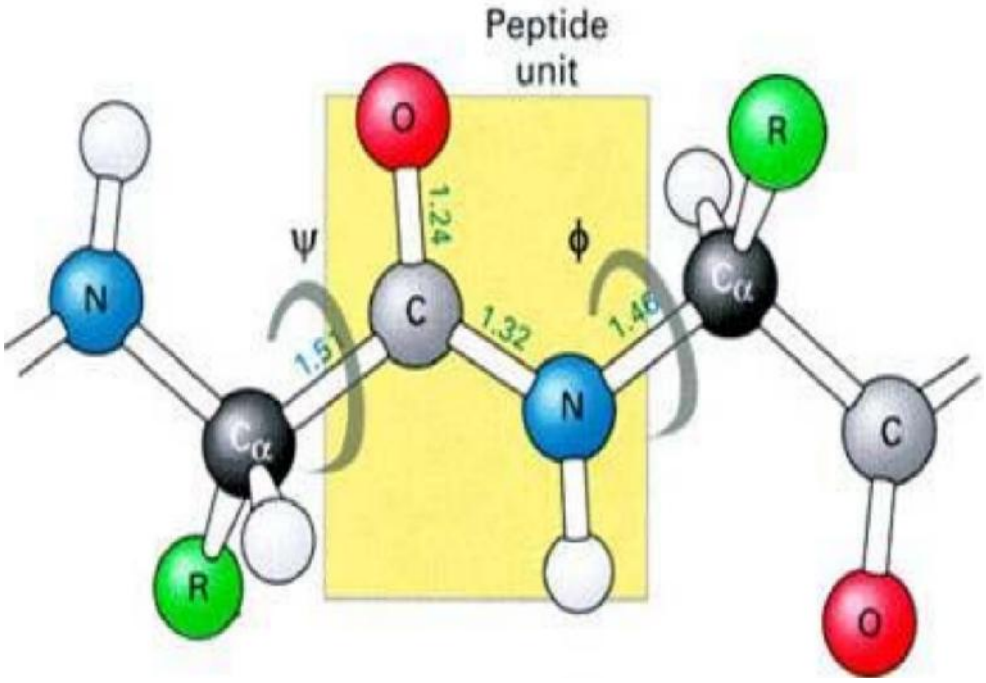


Fig. 44 Phi is the dihedral angle showing as clockwise rotations around the N-C alpha bond, Psi is the dihedral angle showing as clockwise rotations around the C alpha-C bond. Adapted from (78).

Ramachandran plot

Ramachandran Plot is a diagram to visualize backbone dihedral angles phi against psi. Ramachandran plot studied the most favourable combination of Phi and Psi dihedral angles in isolated dipeptides at first. Based on the analysis of Fig. 45, in the Ramachandran plot there are three areas representing the combination of the Phi and Psi dihedral angles. Inside the green area is the core area, representing the most favoured angle combination. Inside red area is the allowed region. White area or disallowed region represents residues having unfavourable Phi and Psi dihedral angles (79). Glycine residues are not restricted to any particular region in Ramachandran plot because the glycine residues have no a side chain (80). To get an ideal model, one should try to have more than 90% of the residues inside the green area or the core area. In my work there are four residues lying in disallowed region. Ser 188, which lies in EL2 region, Asp 271, which lies at EL3, Val 307, and Lys 309 which lie at the C-terminus. All of the four residues which lie in disallowed region are not affected on the binding site of the odorant molecules. So the result of the modeling can be accepted (Fig. 45).

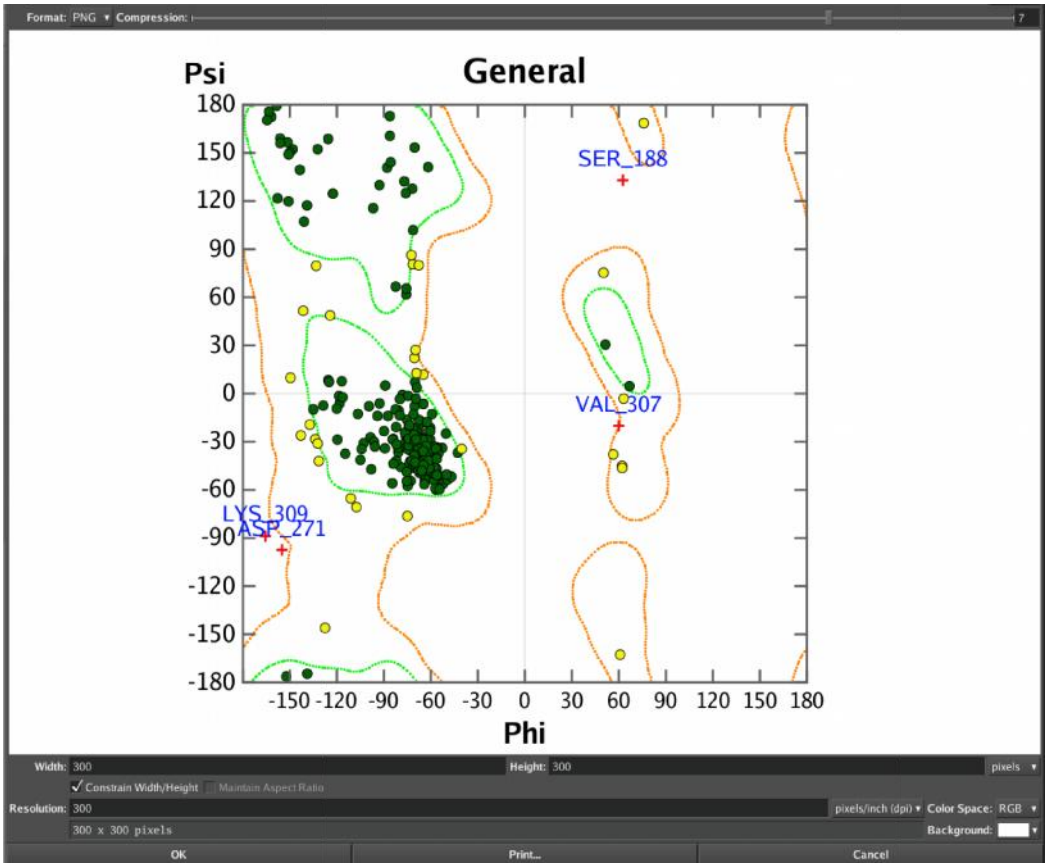


Fig. 45 Phi-Psi Plot for OR1G1 using MOE.

4.8 Comparisons between Rhodopsin and OR1G1

Rhodopsin can be considered as a good template for ORs because the ligand (retinal) binds deeply inside the rhodopsin receptor without any access to the aqueous phase (Fig. 9, 10). The binding site of retinal with rhodopsin is predicted to be similar to the binding site of odorants with ORs (81), so rhodopsin is predicted to be the most suitable template for many of ORs modeling and docking with the odorant molecules, more than the other available crystal structures of GPCRs (81).

Fig. 46 Render the final model using MOE. OR1G1 in orange color and 2G87 in magenta color show that the model is reasonably close to the X-ray structure of 2G87 for most of the residues. The major differences are near the gap areas only at the loops.

5 Predicting of binding site and docking.

5.1 Some of previous studies of the ORs binding site

Up to now many studies have been performed to predict the binding site of olfactory receptors, one of these in 2003 by Orna M., Yoav G., and Doron L.. They have been working on human–mouse comparisons for the amino acid sequences of their ORs, based on the analyses of 1441 OR sequences from human and mouse. They assumed that the functional site would be highly conserved among orthologous pairs, but less conserved among pairs of paralogous. At the end of their analyses they have identified 22 of the predicting binding sites (table 5) (82).

Table 5 The OR binding sites by human–mouse comparisons. Adapted from (82)

OR segment position	Alignment position	Other GPCR	GPCR amino acid
TM2 13	86	Human endothelin-1 receptor precursor (ET-A)	Y:129
TM3 4	115	Rat muscarinic m_1 receptor	L:102
TM3 7	118	Rat muscarinic m_1 receptor	D:105
TM3 8	119	Rat muscarinic m_3 receptor	Y:148
TM3 11	122	Human dopamine D_3 receptor	C:114
TM3 12	123	Rat muscarinic m_1 receptor	N:110
TM3 15	126	Rat muscarinic m_1 receptor	V:113
TM3 16	127	NA	
TM4 12	167	Bovine rhodopsin	A:164
TM4 16	171	Human dopamine D_2 receptor	S:267
TM4 19	174	Rat muscarinic m_3 receptor	P:201
EL2-1	193	cholecystokinin type B (CCKB) receptor	Q:204
EL2 2	196	cholecystokinin type B (CCKB) receptor	H:207
TM5 2	214	Human α2A adrenergic receptor	V:197
TM5 6	218	Human α2A adrenergic receptor	C:201
TM5 9	221	Human α2A adrenergic receptor	S:204
TM5 10	222	Rat $5HT_{2A}$ serotonin receptor	F:243
TM6 12	288	Bovine rhodopsin	F:261
TM6 15	291	Rat type-1B angiotensin II receptor	S:252
TM7 5	321	Human neurokinin-1 (substance P) receptor	I:290
TM7 6	322	Human dopamine D_3 receptor	T:369
TM7 9	325	Rat muscarinic m_1 receptor	C:407

Another study used a combination of three methods, cell biology, calcium imaging, and molecular modeling. Based on the result analyses it was concluded that some residues which are involved in the binding cavity are not changed regardless of the change of the ligand. By the application of this principle to the OR1G1 it was found that residues for F104, M105, F168, I181 and F260, all that of these residues are hydrophobic, and T202 and T279 both can form hydrogen bond with the ligand as seen in Fig. 47

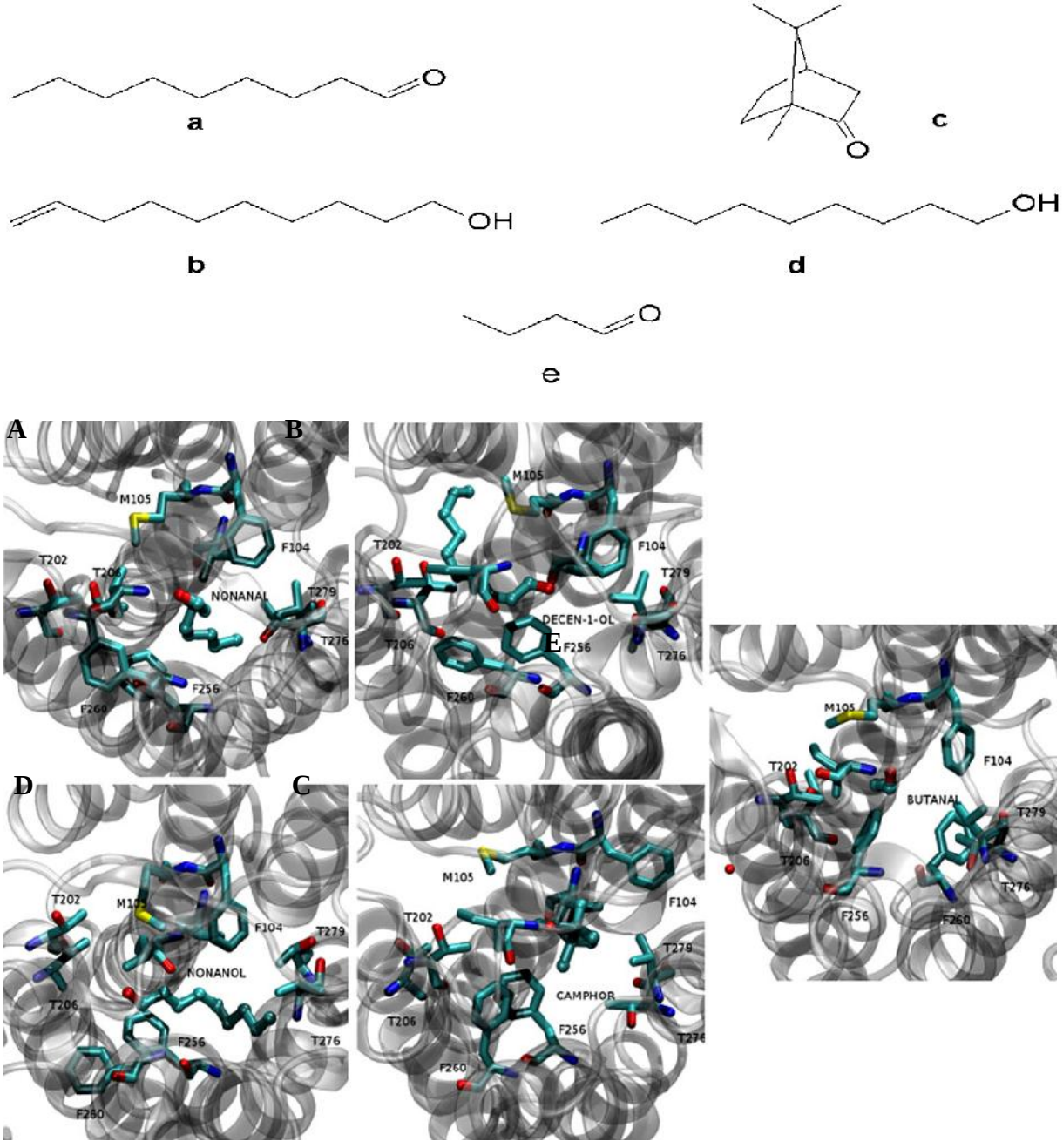


Fig. 47 Chemical structures and the binding site of OR1G1with A) nonanal, B) decanol, C) camphor, D) nonanol, E) butanal. Adapted from(83)

Other author used DarwinDock for docking five odornts, decan-1-ol, nonanol, nonanal, camphor, and butanal with OR1G1. Based on the analysis of these docking calculation it was found that residues of F104 (3.32), M105 (3.33), V108 (3.36), T202 (5.42), T206 (5.46), F256 (6.52), F260 (6.56) and T279 (7.42) are very essential residues for the binding bucket (53) (Fig. 48). These residues or other which close to 3.32, 3.33, 3.36, 5.42, 5.46, 6.52, 6.56 and 7.42 are involved in many previous docking trail for the odorant molecules with ORs (84,85). Docking of 1-amylbutyrate for hOR2AG1 model by using rhodopsin as template found residues A104 3.32, V260 6.48, S263 6.51, S264 6.52, and T279 7.42 are responsible to form the binding bucket (86). Docking of dicarboxylic acids for mouse olfactory receptor MOR42-3 model by using rhodopsin as template found residues I112 3.33, V113 3.37, N117 3.41, V202 5.38, V206 5.43, R207 5.43, N210 5.46, and T259 6.52 are responsible to form the binding bucket (87). Docking of S-(-)citronellal and S-(-)citronellol for human OR1A1 model by using rhodopsin as template found residues of G108 3.36, N109 3.37, S112 3.40, N155 4.56, and I205 5.46 are responsible to form the binding bucket (88). Docking of S-(-)citronellal and S-(-)citronellol for human OR1A2 model by using rhodopsin as template found residues of A108 3.36, K109 3.37, S112 3.40, S155 4.56, and V205 5.46 are responsible to form the binding bucket (88).

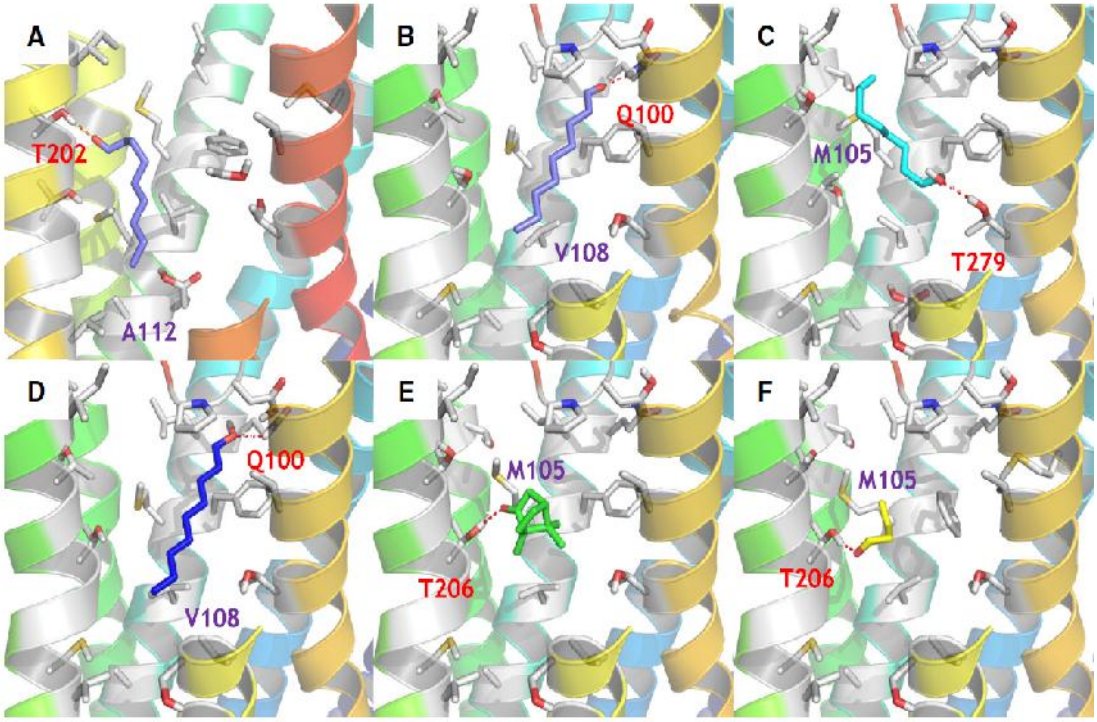


Fig. 48 Docking of five odorants for OR1G1 A,B) nonanal, C) 9-decan-1-ol, D) 1-nonanol, E) camphor , F) n–butanal. . Adapted from(53)

5.2 Docking of menthol as agonist for OR1G1 using MOE (59)

5.2.1 Build the menthol molecule and saving in MOE database

To build the menthol molecule, one can use this command MOE/RHS/Builder or MOE/Edit/Build/Molecule. Then save it in MOE database (menthol.mdb)

5.2.2 Preparing of the menthol molecule for docking

In MOE there are some of automatic and manual application to process the menthol structure in a MOE database:

1. Database washing

2. Correction of bonding patterns

3. Protonation and tautomeric state

4. Partial charges

5. 3D conformation

5.2.3 Preparing of the OR1G1 for docking

For preparing the receptor, one can do this step manually, or by using the Ligx panel provides an automated process in the following order:

1. Correct structural issues with structure preparation.

2. Protonate the structure with protonate 3D.

3. Delete unbound water.

4. Set tethers of different strengths on the receptor, ligand and solvent atoms.

5. Fix atoms beyond a cut off distance from the active site.

6. Minimize the structure.

5.2.4 Calculate possible active sites in the OR1G1 by using site finder panel

In my work I used site finder tools of MOE to determine the suitable site for the docking. Based on the analysis of the result of site finder and comparing them with some of the previous studies, predicted that site number 7 is the most suitable site (Fig 49).

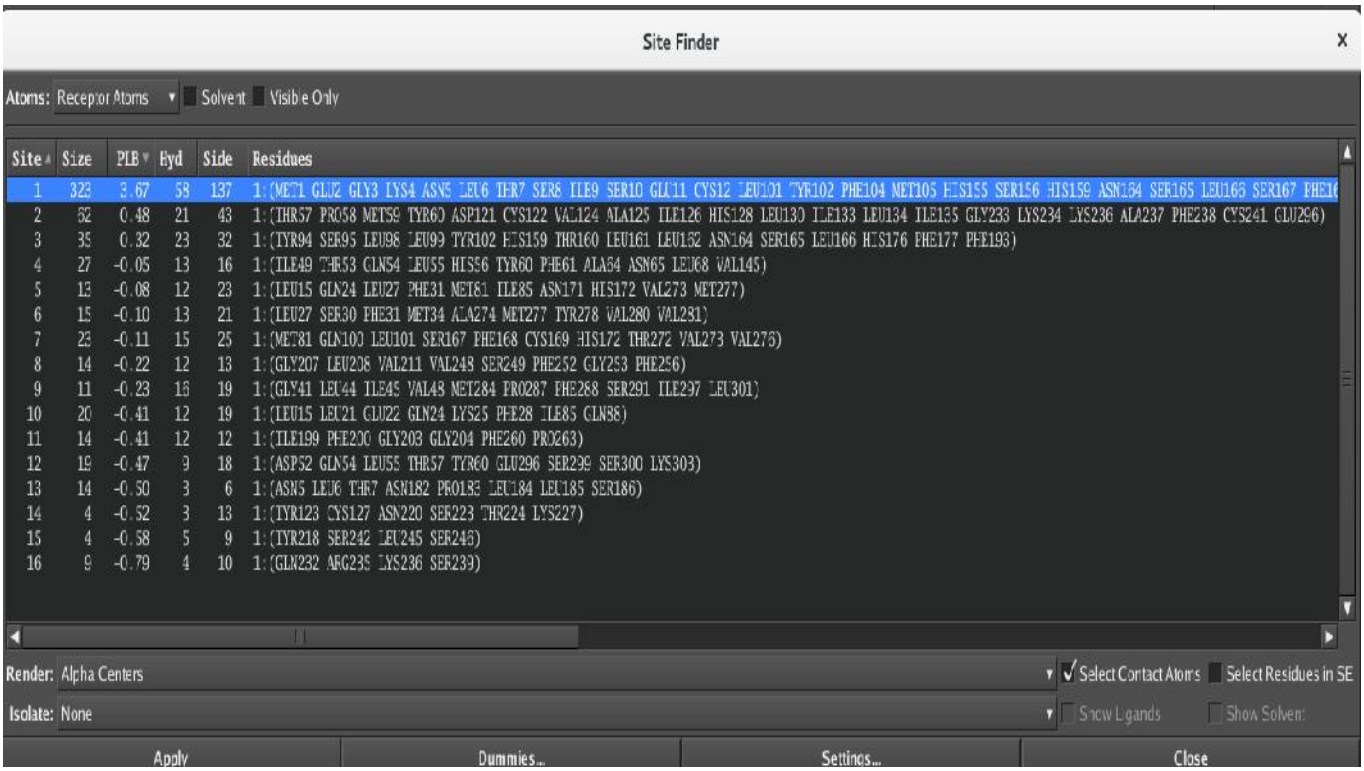


Fig. 49 Site finder panel for OR1G1, the residues of each site are shown.

The residue Val 273 (7.36) can form the hydrogen bund with the oxygen molecule of the menthol. The residues Thr 272 (7.35), Val 273 (7.36), Val 276 (7.39), Met 277 (7.40) and Leu 101 (3.29) form the binding site for binding of menthol as agonist for OR1G1. The residues Phe 168, Cys 169 and His 172, which are found at EL2, are participated in the formation of the lid for the binding site (Fig. 50, 51)

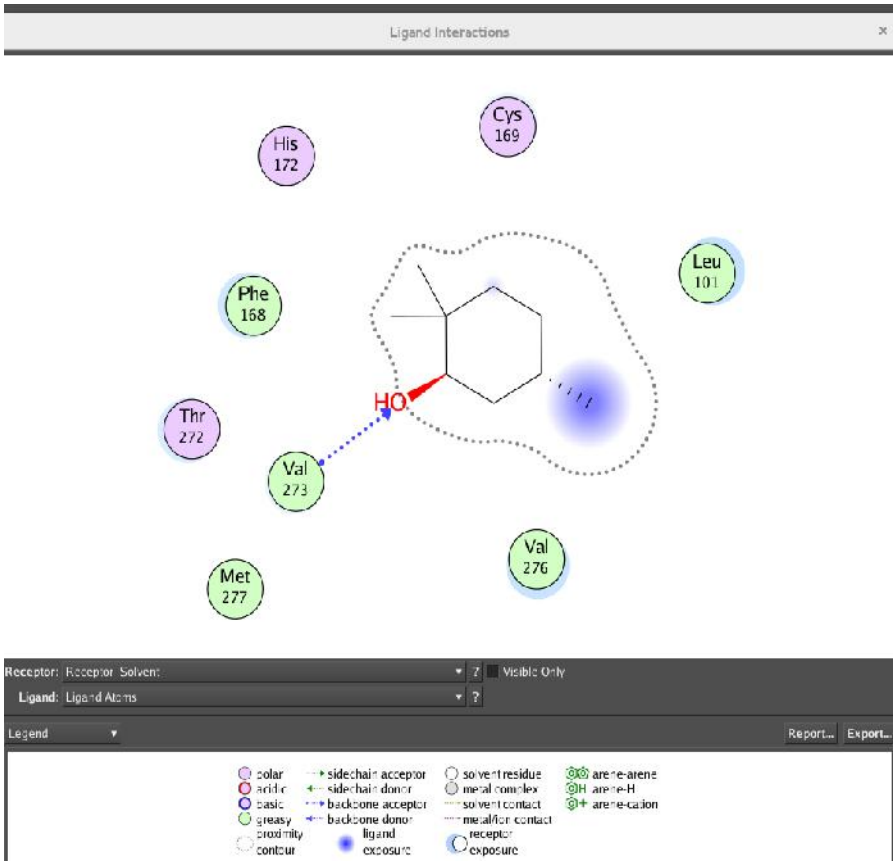


Fig. 50 2D diagram of the menthol molecules and the OR1G1 and the schematic representation of the binding site residues, with the important interaction between the menthol as agonist and the binding site of the OR1G1.

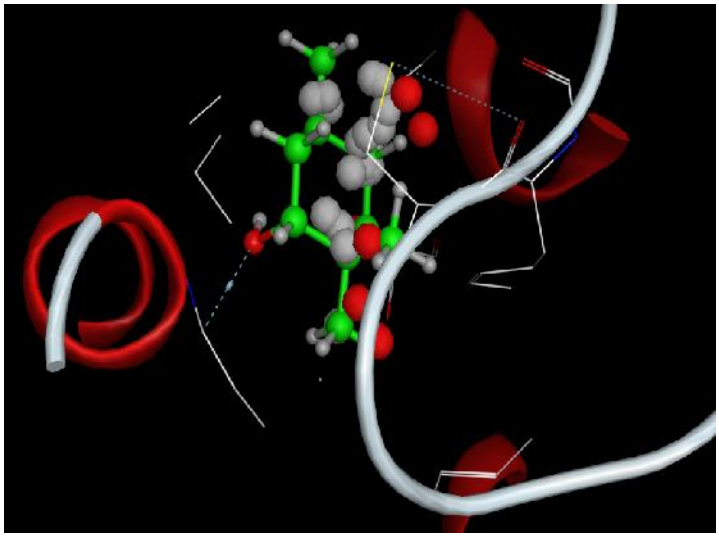


Fig. 51 3D structure of OR1G1 with menthol clarifies the hydrogen bund between oxygen of menthol and the residue Val 273 (7.36), helices 7 and 3 which form the binding site, and EL2 which form the lid to prevent exit of the ligand. Menthol in green color, helices in red, hydrogen bund in cyan, and EL2 in white.

6 Summary

In this work the 3D structure of OR1G1 was studied. Until now no crystal structure for any olfactory receptor was discovered, so the only way to study the structure and the function of ORs through computational methods, which including the homology modeling techniques, which depends on evolutionary related proteins. Each specific odorant can bind with numerous ORs and each OR is stimulated by numerous odorants. This is known as a combinatorial receptor coding scheme. The functions of ORs not only smell the odorant molecules but also have another physiological function such as inducing wound-healing processes through OR2AT4 which is found in the skin, the olfactory receptors in spermatozoa enable sperm cells to find their way directly on unfertilized egg, and activation of (PSGR) by odorant beta-ionone inhibits proliferation of prostate cancer cells. Based on the importance of the physiological function of the ORs, many previous studies for the 3D structures of ORs have been done. Olfactory receptors have a characteristic structure of (GPCRs), which are composed of a bundle of seven transmembran helices (7TM) spanning the cell membrane seven time. 7TM are linked by three intracellular loops (3CL) in the side of the cell cytoplasm and linked by three extracellular loops (3EL) in the side of extracellular membrane, with the carboxyl-terminus intracellular and amino-terminus extracellular. ORs belong to Class A Rhodopsin-like family. For the study of the structure of ORs highly conserved residues must be determined by the multiple sequence alignment. Based on the analysis of selected ORs it can be concluded that: N of conserved GN at TM1 is 1.50. D of conserved LSxxD at TM2 is 2.50, R or C which after conserved MAYD/E and before the conserved Y at TM3 is 3.50, conserved W at TM4 is 4.50, Y of conserved SY at TM5 is 5.58, H of conserved KAFSTCxSH at TM6 is 6.40, and P of conserved NPxxY at TM7 is 7.50. The important step for homology modeling is the accurate template selection, and the accurate alignment. According to sequence similarity, length of template in comparison to target, presence of the highly conserved residues of the target, resolution and R-factor of a crystallographic structure of the template we can chose rhodopsin as a template. After building the model of OR1G1, it can be evaluated using Ramachandran plot. To have an ideal model, one has to try to have more than 90% of the residues inside the core area of Ramachandran plot. Using the site finder tool of MOE one can predict the binding site. From a previous study concluded residues or other which are close to 3.32, 3.33, 3.36, 5.42, 5.46, 6.52, 6.56 and 7.42 are involved in many docking for the odorant molecules with ORs.

By using MOE for docking OR1G1 with menthol it was found that: The residues Thr 272 (7.35), Val 273 (7.36), Val 276 (7.39), Met 277 (7.40) and Leu 101 (3.29) form the binding site. The residues Phe 168, Cys 169 and His 172, which are found at EL2, are participated in the formation of the lid for the binding site.

References

1. Huart, C., Rombaux, P. & Hummel, T. (2013). Plasticity of the Human Olfactory System: The Olfactory Bulb. Molecules, 18 (9), 11586-11600.
2. Sultan-Styne K., Toledo R., Walker C., Kallkopf A., Ribak C. E. & Guthrie K. M. (2009). Long-term survival of olfactory sensory neurons after target depletion. J Comp Neurol, 515 (6), 696-710.
3. Bushdid, C., Magnasco, M. O., Vosshall, L. B. & Keller, A. (2014). Humans Can Discriminate More than 1 Trillion Olfactory Stimuli. Science, 343 (6177), 1370-1372.
4. Pifferi, S., Menini, A. & Kuurahashi, T. (2010). Signal transduction in vertebrate olfactory cilia. A. Menini, The Neurobiology of Olfaction. Boca Raton (FL): CRC Press, Chapter 8.
5. Vaidehi, N., Floriano, W. B.,Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zaamanakos, G. & Goddard, W. A. (2002). Prediction of structure and function of G proteincoupled receptors. Proc Natl Acad. Sci., 99 (20), 12622-12627.
6. Fredriksson R., Schiöth H. B. (2005). The repertoire of G-protein-coupled receptors in fully sequenced genomes. Mol Pharmacol, 67 (5), 1414-1425.
7. Pang,G., Xie, J., Chen, Q. & Hu, Z. (2012). How functional foods play critical roles in human health. Food Science and Human Wellness, 26-60
8. Joost, P. & Methner, A. (2002). Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands. Genome Biol, 3 (11), 1-16.
9. Foord, S. M., Bonner, T. I., Neubig, R. R., Rosser, E. M., Pin, J., Davenport, A. P., Spedding, M., & Harmar, A. J. (2005). International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List. Pharmacological Reviews, 57 (2), 279-288.
10. Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Hugh, R. & Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. Nature Reviews Drug Discovery, 12 (1), 25-34
11. Krishnan, A., Almen, M. S., Fredriksson, R. & Schiöth, H. B. (2012). The Origin of GPCRs: Identification of Mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in Fungi. public library of science ONE, 7 (1), e29817
12. Fredriksson, R., Lagerström, M. C., Lundin, L. & Schiöth, H. B. (2003). The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Molecular Pharmacology, 63 (6), 1256-1272.

13. Lagerström, M. C. & Schiöth, H. B. (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. Nature Reviews Drug Discovery, 7, 339-357.
14. Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Hugh, R. & Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. Nature Reviews Drug Discovery, 12, 25-34.
15. Suwa, M., Sugihara, M. & Ono, Y. (2011). Functional and Structural Overview of G-Protein-Coupled Receptors Comprehensively Obtained from Genome Sequences. Pharmaceuticals, *4*, 652-664
16. Fredriksson, R., Lagerström, M. C., Lundin, L. & Schiöth, H. B. (2003). The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Mol Pharmacol, 63, 1256-1272.
17. Zhou, E., Melcher, K. & Xu, E. (2012). Structure and activation of rhodopsin. Acta Pharmacologica Sinica, 33, 291-299.
18. Janz, J. M., Fay, J. F. & Farrens, D.L. (2003). Stability of Dark State Rhodopsin Is Mediated by a Conserved Ion Pair in Intradiscal Loop. The Journal of Biological Chemistry, 278, 16982-16991.
19. Luttrell, L. M. & Lefkowitz, R. J. (2002). The role of β-arrestins in the termination and transduction of G-protein-coupled receptor signals. Journal of Cell Science, 115, 455-465.
20. Scientific Background on the Nobel Prize in Chemistry 2012 Studies of g-protein–coupled receptors. from the royal swedish academy of sciences
21. Digby, G. J., Lober, R. M., Sethi, P. R. & Lambert, N. A. (2006). Some G protein heterotrimers physically dissociate in living cells. Proc Natl Acad. Sci., 103 (47), 17789-17794.
22. Pedersen S., and Ross E. (1982). Functional reconstitution of beta-adrenergic receptors and the stimulatory GTP-binding protein of adenylate cyclase. Proc. NatL Acad. Sci. USA, 79, 7228-7232 Biochemistry.
23. Ferguson, S. S. (2001). Evolving Concepts in G Protein-Coupled Receptor Endocytosis: The Role in Receptor Desensitization and Signaling. Pharmacological, 53 (1), 1-24.
24. Wilden U., Hall S. & Kühn H. (1986). Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kDa protein of rod outer segments. Proc. Nai. Acad. Sci. USA, 83, 1174-1178, Biochemistry

25. Weller M., Virmaux N. & Mandel P. (1975). Light-stimulated phosphorylation of rhodopsin in the retina: the presence of a protein kinase that is specific for photobleached rhodopsin. Proc. Nat. Acad. Sci. USA, 72 (1) 381-385.
26. Zastrow, M. von & Kobilka, B. K. (1994). Antagonist-dependent and -independent Steps in the Mechanism of Adrenergic Receptor Internalization. the journal of biological chemistry, 269, (28), 18448-18452.
27. Urban, J. D., Clarke, W. P., Zastrow, M. von, Nichols, D. E., Kobilka, B., Weinstein, H., Javitch, J. A., Roth, B. L., Christopoulos, A. , Xeston, P. M., Miller, K. J., Spedding, M., & Mailman, R. B. (2007). Functional Selectivity and Classical Concepts of Quantitative Pharmacology. Journal for Electrochemistry and Plating Technology, 320 (1), 1-13.
28-Tate, C. G. (2012). A crystal clear solution for determining G-protein-coupled receptor structures. Trends in Biochemical Sciences 915, 1-10.
29. Lane, R. P., Cutforth, T., & Trask, B. J. (2001). Genomic analysis of orthologous mouse and human olfactory receptor loci. Proc Natl Acad. Sci., 98 (13), 7390-7395.
30. Glusman, G., Bahar, A., Sharon, D., Pilpel, Y., White, J. & Lancet, D. (2000). The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mammalian Genome, 11, 1016-1023.
31. Malnic, B., Godfrey, P. A. & Buck, L. B. (2004). The human olfactory receptor gene family. Proc Natl Acad. Sci., 101 (8), 2584-2589.
32. Knape, K., Beyer, A., Stary, A., Buchbauer, G. & Wolschann, P. (2008). Genomics of selected human odorant receptors. Monatsh Chem, 139, 1537-1544.
33. Glusman, G., Yanai, I., Rubin, I., & Lancet, D. (2001). The Complete Human Olfactory Subgenome. Genome Res, 11, 685-702.
34. Flegel, C., Manteniotis, S., Osthold, S., Hatt, H. & Gisselmann, G. (2013). Expression Profile of Ectopic Olfactory Receptors Determined by Deep Sequencing. public library of science ONE, 8 (2), e55368
35. Busse, D., Kudella, P., Grüning, N.M., Gisselmann, G., Ständer, S., Luger, T., Jacobsen, F., Steinsträßer, L., Paus, R., Gkogkolou, P., Böhm, M., Hatt, H. & Benecke, H. (2014). A Synthetic Sandalwood Odorant Induces Wound-Healing Processes in Human Keratinocytes via the Olfactory Receptor OR2AT4. Journal of Investigative Dermatology, 134, 2823-2832.

36. Neuhaus, E. M., Zhang, W., Gelis, L., Deng, Y.,Noldus, J., & Hatt, H. (2009). Activation of an Olfactory Receptor Inhibits Proliferation of Prostate Cancer Cells. J Biol Chem, 284(24), 16218-16225.
37. Pluznick, J. L.,Zou, D.J., Zhang, X., Yan, Q., Rodriguez-Gil, D. J.,Eisner, C., Wells, E., Greer, C. A., Wang, T., Firestein, S., Schnermann, J. & Caplan, M. J. (2009). Functional expression of the olfactory signaling system in the kidney. Proc Natl Acad. Sci., 106 (6), 2059-2064.
38. Hall, S.E., Floriano,W. B., Vaidehe, N. & Goddard III, W.A. (2004). Predicted 3-D Structures for Mouse I7 and Rat I7 Olfactory Receptors and Comparison of Predicted Odor Recognition Profiles with Experiment. Chem. Senses, 29, 595-616
39. Malnic B, Gonzalez-Kristeller D., and Gutiyama L. (2010). Odorant Receptors. Menini, A., The Neurobiology of Olfaction. Boca Raton (FL), CRC Press, Chapter 7, 181-202.
40. Leffingwell Reports, May, 2002, Vol. 2 (1).
41. Pilpel, Y., & Lancet, D. (1999). The variable and conserved interfaces of modeled olfactory receptor proteins. Protein Science, 8 (5), 969-977
42. Matarayyo, V., Zsürger, N., Guillemot, J. C., Clot-Faybesse, O., Botto, J. M., Dal Farra, C., Crowe, M., Demaille, J., Vincent, J.P., Mazella, J., & Ronin, C. (2002). Porcine Odorant-binding Protein Selectively Binds to a Human Olfactory Receptor. Chem. Senses, 27 (8), 691-701.
43. Triller, A., Boulden, E. A., Churchill, A., Hatt, H., Englund, J., Spehr, M. & Sell, CS. (2008). Odorant-receptor interactions and odor percept: a chemical perspective. Chem Biodivers, 5(6), 862-886.
44. Audouze, K., Tromelin, A., Marie Le Bon, A., Belloir, C., Petersen, R. K., Kristiansen, K., Brunak, S. & Taboureau, O. (2014). Identification of Odorant-Receptor Interactions by Global Mapping of the Human Odorome. The public library of science ONE, 9 (4), e93037, 1-12.
45. Oh, E. H., Lee, S. H., Ko, H. J., Lim, J. H. & Park, T. H. (2015). Coupling of olfactory receptor and ion channel for rapid and sensitive visualization of odorant response. Acta Biomaterialia, 22, 1–7.
46. Pervez, N., Ham, H. G. & Kim, S. (2014). Interplay of Signaling Molecules in Olfactory Sensory Neuron toward Signal Amplification. Hanyang Med Rev, 34 (3), 137-142.

47. Haering C., Kanageswaran N., Bouvain P., Scholz P., Altmüller J., Becker C., Gisselmann G., Wäring-Bischof J.& Hatt H. (2015). Ion transporter NKCC1, modulator of neurogenesis in murine olfactory neurons. J Biol. Chem., 290 (15), 9767-9779.
48. Persuy, M. A., Sanz, G., Tromelin, A. & Thomas-Danguin, T. (2014). Mamalian olfactory receptors: molecular mechanisms of odorant detection 3D modeling and structure-activity relationships. Progress in molecular biology and translational science, 1-36.
49. Andersson, M. N., Löfstedt, C. & Newcomb, R. D. (2015). Insect olfaction and the evolution of receptor tunig. Frontiers in Ecologyand Evolution, 3 (53), 1-14.
50. DeMaria, S. & Ngai, J. (2010). The cell biology of smell. The journal of cell biology, 191 (3), 443-452.
51. Bourne, P. E. & Weissig, H. (2003). Homology modeling. In E. Krieger, S. B. Nabuurs & G. Vriend. Structural Bioinformatics, 507-521.
52. Lesk, A. M. & C. H. Chothia C. (1986). The response of protein structures to amino-acids sequence changes. Philos. Trans. R. Soc. Lond. B Biol. Sci., 317, 345-356.
53. Kim, S. K. & Goddard, W. A. (2014). Predicted 3D structures of olfactory receptors with details of odorant binding to OR1G1. J Comput Aided Mol, 28 (12), 1175-1190.
54. Sanchez, R. & Sali, A. (1997). Advances in comparative protein-structure modelling. Curr Opin Struct Biol, 206-214.
55. Marti-Renom, M. A., Stuart, A. C., Fisher, A., Sanchez, R., Melo, F. & Sali, A. (2000). comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct., 29, 291-325.
56. Sauder, J. M., Arthur, J. W. & Dunbrack, R. L. (2000). Large-Scale Comparison of Protein Sequence Alignment Algorithms With Structure Alignments. PROTEINS: Structure, Function, and Genetics, 40, 6-22.
57. De March, C. A., Ryu, S. E., Sicard, G., Moon, C., & Golebiowski, J. (2015). Structure-odour relationships reviewed in the postgenomic era. Flavour Fragr. J., 30 (5), 342-361.
58. de March, C. A., Baldovini, N., Lebon, A. M. & Golebiowski, J. (2014). The human OR1G1 olfactory receptor is involved in the combinatorial code of □-santalol. 7th JFIC - GIFC, Turin, Italy, May 5th-6th
59. Molecular Operating Environment (MOE version 2015) Chemical Computing Group, Inc. Montreal, Quebec, Canada. http://www.chemcomp.com
60. Sali, A. & Blundell, TL. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol., 234, 770-815.

61. Guex N, Peitsch MC. SWISS-MODEL. (1997). An envionment for comparative protein modeling. Electroph., 18, 2714-2723. http://www.expasy.org/swissmod/SWISS-MODEL.html
62. Alirezaee, M., Dehzangi, A. & Mansoori, E. (2012). Predicting the secondary structure of proteins by cascading neural networks. International Journal of Artificial Intelligence & Applications (IJAIA), 3, (6), 39-47.
63. Sippl M.L. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins, 17, 355-362.
64. Pearson,W.R. (1996). Effective protein sequence comparison. Methods Enzymol, 266, 227-258.
65. Madabushi, S., Yao, H., Marsh, M., Kristensen, D. M., Philippi, A., Sowa, M. E. & Lichtarge, O. (2002). Structural clusters of evolutionary trace residues are statistically significant and common in proteins. J. Mol. Biol, 316, 139-154.
66. Ballesteros, J. A., Weinstein, H. (1995). Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Stuart CS, Ed. Methods in Neurosciences. Academic Press, 366-428.
67. Zozulya, S., Echeverri, F., Nguyen, T. (2001). The human olfactory receptor repertoire. GenomeBiol, 2, 1-12.
68. Zhang, X. & Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. Nat Neurosci, 5, 124-133.
69. Cook, B. L., Steuerwald, D., Kaiser, L., Graveland-Bikker, J., Vanberghem, M., Berke, A. P., Herlihy, K.,Pick, H., Vogel, H.& Zhang, S. (2009). Large-scale production and study of a synthetic G proteincoupled receptor: Human olfactory receptor 17-4. Proc Natl Acad. Sci. USA, 106, 11925-11930.
70. Wheatley, M., Wootten, D., Conner, M. T., Simms, J., Kendrick, R., Logan, R.T., Poyner, D.R.& Barwell, J. (2012). Lifting the lid on GPCRs: the role of extracellular loops. Br J Pharmacol, 165, 1688-1703.
71. Yarnitzky, T., Levit, A. & Niv, M. Y. (2010). Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. Curr Opin Drug Discov Devel, 13, 317-325.
72. Charlier, L., Topin, J., de March, C. A., Lai, P. C., Crasto, C. J. & Golebiowski, J. (2013). Molecular Modelling of Odorant/Olfactory Receptor Complexes. In: Crasto CJ, Ed. Olfactory Receptors. Humana Press, 53-65.

73. Berman, H.M., Westbrook,J., Feng, Z. Gilliland, G., Bhat,T.N., Weissig, H., Shindyalov, I.N.& Bourne, P.E. The Protein Data Bank. Nucleic Acids Research, 28, 235-242.
74. Fechteler T., Dengler, U., Schomberg, D. (1995). Prediction of protein Three-Dimensional Structures in Insertion and Deletion Regions: A Procedure for searching Databases of Representative Protein Fragments UsingGeometric Scoring Criteria. J. Mol. Biol., 253, 114-131.
75. Labute, P. (2008). The Generalized Born / Volume Integral (GB/VI)Implicit Solvent Model: Estimation oft he Free Energy of Hydration Using London Dispersion Instead of Atomic Surface Area. J. Comp. Chem., 29, 1693-1698.
76. Zhang, C., Vasmatzis, G., Cornette, J.L.& DeLisi C. (1997). Determination of Atomic Desolvation Energies from the Structures of Crystalized Proteins. J. Mol. Biol., 267, 707-726.
77. Miyazawa, S., Jerrigan, R.L. (1985). Estimation of Effective Inter residue Contact Energies from Protein Crystal Structures: Quasi-chmical Approximation. Macromolecules, 18, 534-552.
78. Purushe, S., Anbalagan, S. K. & Grinstein, G. (2011). Development of an Interactive Ramachandran Plot in Weave. Computer society, 232-236.
79. Morris,A.L., MacArthur, M.W., Hutchinson, E G. & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. Protein: Structure, Function, and Genetics, 12 (4), 345-364.
80. Ho,B. K.& Brasseur, R. (2005). The Ramachandran plots of glycine and pre-proline. BMC Structural Biology, 5-14.
81. Park, J. H., Morizumi, T., Li, Y., Hong, J. E., Pai, E. F., Hofmann, K. P., Choe, H.W. & Ernst, O. P. (2013). Opsin, a Structural Model for Olfactory Receptors? Angew. Chem. Int. Ed Engl., 52, 11021-11024.
82. Man, O., Gilad, Y. & Lancet, D. (2004). Prediction of the odorant binding site of olfactory receptor proteins by human–mouse comparisons. Protein Science, 13 (1), 240-254.
83. Charlier, Topin, L., Ronin, C., Kim, S. K., Goddard W. A. 3rd, Efremov, R. & Golebiowski, J. (2012). How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. Cell. Mol. Life Sci., 69, 4205-4213.
84. Gelis, L., Wolf, S., Hatt, H., Neuhaus, E. M. & Gerwert, K. (2012). Prediction of a ligand-binding niche within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology modeling. Angew Chem Int Ed Engl., 51, 1274-1278.

85. Schmiedeberg, K., Shirokova, E., Weber, H. P., Schilling, B., Meyerhof, W. & Krautwurst, D. (2007). Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. J Struct Biol., 159, 400-412.
86. Abrol, R., Bray, J. K. & Goddard, W. A. 3rd. (2012). Bihelix: towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins, 80 (2), 505-518.
87. Charlier, L., Topin, J., Ronin, C., Kim, S. K., Goddard, W. A. 3rd, Efremov, R. & Golebiowski, J. (2012). How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. Cell Mol Life Sci, 69 (24), 4205-4213.
88. Sanz, G., Schlegel, C., Pernollet, J. C. & Briand, L. (2005). Comparision of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidnece for antagonism. Chem Senses, 30 (1), 69-80.

Appendix

OR genes distribution

These data are collected from:

(ORDB) https://senselab.med.yale.edu/OrDB/info/humanorseqanal.htm

(HGNC) www.genenames.org/cgi-bin/genefamilies/set/141 at 28.06.2015

Chromosome 1

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
C1	AJ1		F5	BF1	Y1				K1	L1		G1	A2				
	AK1		F16	AV1	N2				Z1				A16				
	B11		F29	AT1	K6				T2				C36				
	C1			AX1	P1				R2				I1				
	C3			AY1	N1				J5				K1				
	C4				F1				J1								
	G2				K2				K2								
	G3				K3				J6								
	G6								J3								
	I1								J4								
	L2								X1								
	L3																
	L5																
	L8																
	L13																
	M2																
	M3																
	M4																
	M5																
	M7																
	T1																
	T2																
	T3																
	T4																
	T5																
	T6																
	T7																
	T8																

	T10															
	T11															
	T12															
	T27															
	T29															
	T33															
	T34															
	T35															
	W3															
	W5															

Chromosome 2

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
					B2	E102											
					B3												

Chromosome 3

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
F10				K1													
				H2													
				H6													
				K2													
				AC1													
				AC2													
				H1													
				H8													
				H14													
				H15													
				K3													
				K4													

Chromosome 4

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
				L5													

Chromosome 5

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
E6	Y1		F3														
	V1																
	V2																
	V3																

Chromosome 6

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
F12	B6		F14	U1					C1	A1	D3		J1				
	B1		F16	V1					C2		D2						
	H3		F10								D1						
	B9		F15														
	H2																
	H1																
	B3																
	J3																
	B2																
	W1																
	J2																
	B8																
	I2																
	H7																
	A4																
	J1																

Chromosome 7

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
	A10				V1			A4	AC1								
	A4				W1			A2									
	A1				B1												
	AE1																
	F1																
	A6																
	F2																
	A5																
	A7																
	A2																
	A12																
	A14																
	A25																
	A42																

Chromosome 8

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
	V2		F3														
			F21														

Chromosome 9

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
J4	K2			C1								C3					
N3	S2											C4					
L8												C5					
K1												C8					
L3												C9					
L6												C2					
J1												F1					
B1												J1					
L4												C7					

Q1												D1					
N2																	
L1																	
J2																	
N1																	
J5																	

Chromosome 10

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
												A1					

Chromosome 11

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
S2	AG1		D5	P3	A1		D4	G4	G4					M1	L1	E1	A1
S1	D3		A4	M10	B1		D1	Q1	G7					E1	J3		A6
L18.02	D2		A5	T2	M1		K1	G5	V1					A7	E4		A4
L35.03	AT4		P4	A1	T1		J3	I1	A5					I2	E5		B2
	AG2		C15	A2	Q1		K5	G1	D4					B2	N4		B4
			X1	B2	X1		A1	G9	A4					B4	N2		A3
			C12	AS1	A2		B12	Q2	A3					S1	B4		A5
			C13	B13			D2		S1					G1	A1		B1
			F17	M3			K3		G8					A4	E8		
			B1	I1			H2		Q1					L1	H1		
			X2	M8			I2		A6					A2	K1		
			A15	AK3			H3		A1					H1	D1		
			D9	F1			H1		G9					Q1	E2		
			D6	G3			B8		AG1					G2	N1		
			C6	T3			B4		A2					F2	E6		
			C16	M11			J1		D3					E2	B2		
			A16	P2			B3		G6					D1	K2		
			A47	M9			B2		W1					I1	E1		
			C46	D16			G1							T1	P1		
			C45	D18			G2							B5	R1		
			C3	L1			G5							B6	B6		
			C5	AR1			J2							F1	I1		
			C11	M1			S1							J1	I2		
			D10	AP2			U1							V1	N5		

			D11	B16			U8							L2		
			S1	B17			U9							A1		
			S2	D14										A5		
				M11										M1		
				J2										N5		
				D13										W1		
				AN1										Z1		
				AK2												
				AU1												
				B2												
				AL1												
				AN1												
				B3												
				B12												
				B21												
				L2												
				R1												
				T1												
				W2												

Chromosome 12

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
AP1					C1			K2	A7								
					C2				AD1								
					C3				P1								
					C4												
					C6												
					C65												
					C68												
					C70												
					C74												
					C75												
					C76												

Chromosome 13

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
E5																	

Chromosome 14

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
	F3		K2	AU1	J2				G3	H4							
			K3		S1				G2	G2							
			K5		J1					H6							
			N5							H2							
			L1							H7							
			K13							H12							
			K15														
			K17														
			N2														
			E2														
			K14														
			M1														
			K1														
			E1														
			Q2														
			Q3														

Chromosome 15

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
			N4														
			N2														
			F15														
			F12														
			F6														
			F4														
			M2														

Chromosome 16

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
F1	C1		S1														
F2			C3														

Chromosome 17

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
E2		A2	D1														
G1		A4	D2														
D4		A1															
A1		A3															
E5																	
D2																	
A2																	
D5																	
E1																	
E3																	

Chromosome 19

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
I1	Z1		F19			D2			H1								
M1			F18			C2			H2								
			F17			G1			H4								
						G3			H5								
						C1			H3								
						A5											
						A17											
						A10											
						A2											
						G2											
						D4											
						E24											
						G2											

Chromosome 22

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
										H1							

Chromosome X

Family 1	Family 2	Family 3	Family 4	Family 5	Family 6	Family 7	Family 8	Family 9	Family 10	Family 11	Family 12	Family 13	Family 14	Family 51	Family 52	Family 55	Family 56
F11	L2		Q3	H1	C2	E120	G1					H1			A2		
E7	K1			AC2	C1	E7	G2					E2					
	M4			B3	C3												
	D1																
	A19																
	M1																
	A20																
	M2																
	A21																
	L1																

Curriculum Vitae

Personal data:

Name: Girgis Noser
Date of birth: 06/10/1977.
Place of birth and nationality: Aswan, Egypt- Egyptian.
Tel.No..: 0043-6767928956
E-Mail: girgis77@yahoo.com.

Education:

2014-actual: nostrification student- faculty of pharmacy- Vienna Uni.
May, 2013 Strategic human resources management - Course
May, 2009 How to negotiate a profitable and successful close - Course.
Sep. 1996- Nov. 2001: Bachelor degree in Pharmaceutical sciences-Assiut Uni. Egypt.

Work experiences:

10/2004 - 07/2013: Executive medical representative – Astra Zeneca – Egypt.

11/2001- 9/2004: Pharmacist.

Personal skills:

Languages:
Arabic: Mother tongue. English: Very good. German: Good

Computer:
MC Office, MOE

Others:
Driving license, category B.
Photography.