Structure Prediction of G-Protein Coupled Receptors

Thesis by

Václav Cvičeka

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended January 29, 2015)

To Diana

Acknowledgments

I am grateful to many for their support throughout my PhD journey. First and foremost, I would like to thank my advisor, Professor William Goddard, for his tireless enthusiasm in my research, for his encouragement, and for his innumerable ideas, which guided me into new scientific territories.

I would also like to thank Ravinder Abrol, for our long and productive discussions about GPCRs and science in general, as well as the Biogroup members, Soo-Kyung Kim, Fan Liu, Sijia Dong, Caitlin Scott, Andrea Kirkpatrick, and Adam Griffith, for our collaboration and great discussions on GPCR proteins. Working alongside such a talented group of researchers encouraged me to strive harder and made the journey all the more enjoyable.

I am also grateful to Saber Naserifar and Daniel Brooks for our great collaboration on the implementation of PQEQ, and to Andres Jaramillo-Botero, Robert Nielsen, Si-ping Han, Julius Su, Bartosz Trzaskowski, and other group members for the fantastic research (and office) community they provided.

I would also like to acknowledge the great contributors to my development as a researcher of my early group at Caltech: Professor Jim Eisenstein, Ghislain Granger, Bill Chickering, and Jing Xia. I will always cherish their effort in guiding me into the world of two-dimensional electron gases. Thanks are also due to the staff at Caltech, for their friendliness and helpfulness throughout the years I spent here.

Outside of Caltech, I am also grateful to Sadasivan Shankar from Intel, for giving me a taste of physics applications in industrial settings and for providing some powerful real-world motivation for continuing my research.

My group of close friends, which includes Paula Popescu, Dzhelil Rufat, and Dan Grin, have tremendously improved my quality of life, and I thank them for always being there when I needed them through the ups and downs of graduate school life.

Like many Czechs before me, I am thankful to Jára Cimrman, our fictional hero, for the inspiration and amusement he provided for me growing up. May we all live up to the standards he set for us.

Finally, I would like to thank my parents, Miluše and Václav, without whom none of what I have achieved so far would be possible, for the many sacrifices they made so that I could follow my

dreams. Together with my sister, Barbora Cvičková, they provided the most loving and supportive family anyone can have. Last but not least, a big thanks to my fiancée, Diana Negoescu, for her endless support and belief in my ability to succeed.

Abstract

G-protein coupled receptors (GPCRs) form a large family of proteins and are very important drug targets. They are membrane proteins, which makes computational prediction of their structure challenging. Homology modeling is further complicated by low sequence similarly of the GPCR superfamily.

In this dissertation, we analyze the conserved inter-helical contacts of recently solved crystal structures, and we develop a unified sequence-structural alignment of the GPCR superfamily. We use this method to align 817 human GPCRs, 399 of which are nonolfactory. This alignment can be used to generate high quality homology models for the 817 GPCRs.

To refine the provided GPCR homology models we developed the Trihelix sampling method. We use a multi-scale approach to simplify the problem by treating the transmembrane helices as rigid bodies. In contrast to Monte Carlo structure prediction methods, the Trihelix method does a complete local sampling using discretized coordinates for the transmembrane helices. We validate the method on existing structures and apply it to predict the structure of the lactate receptor, HCAR1. For this receptor, we also build extracellular loops by taking into account constraints from three disulfide bonds. Docking of lactate and 3,5-dihydroxybenzoic acid shows likely involvement of three Arg residues on different transmembrane helices in binding a single ligand molecule.

Protein structure prediction relies on accurate force fields. We next present an effort to improve the quality of charge assignment for large atomic models. In particular, we introduce the formalism of the polarizable charge equilibration scheme (PQEQ) and we describe its implementation in the molecular simulation package Lammps. PQEQ allows fast on the fly charge assignment even for reactive force fields.

Contents

Chapter 1

Introduction

Although living organisms seem to be infinitely complex, most of the diversity comes from the complex interactions of the simple building blocks, such as twenty different amino acids. A protein is a chain of amino acids, which fold into a compact structure due to the hydrophobic interaction with surrounding water molecules. Most proteins adopt a unique conformation, which depend on the order of amino acids in their sequence. The number of possible conformations grows exponentially with the length of the amino acid chain. The exponential complexity is the source of biological diversity, and at the same time is what makes it difficult to solve protein structures.

Proteins and molecules inside organisms can interact in many different ways. To organize multicellular life, nature creates many different compartments, such as cells and organelles. Communicating information across membranes between the compartments is the means by which multicellular organisms can function. G-Protein Coupled Receptors (GPCRs) are a large family of membrane proteins, which sense diverse molecules outside of the cell and transfer the information inside the cell where they regulate a number of signaling pathways. Since GPCRs are natural places to intercept signaling pathways they are very important drug targets. Predicting atomistic models for GPCRs would enable faster advances in drug discovery.

Structure prediction of GPCRs is a difficult problem because they are large proteins. Small proteins with about 50 residues can be folded using molecular dynamics, which simulates the natural dynamics of the protein [1]. Larger proteins with about 100 residues can often be solved with the Monte Carlo method combined with fragment assembly or other sampling techniques [2]. The space of the possible protein conformations scales exponentially with the size of the proteins. GPCRs are large proteins with more than 300 residues, thus the current structure prediction techniques are too slow given the presently available computational resources. To solve the structure of a large protein the only possibility is to start with a similar protein, the homology model, and to try to refine it with molecular dynamics or other sampling methods. The more we know about the type of proteins in question, the better we can tailor the sampling method for structure prediction.

Because proteins are small — only a few nanometers in diameter — they represent a challenge

to study. Many proteins have already been crystallized, thus their structure is already known to within atomic resolution. In order to find a unique structure, these proteins are removed from their native environment and crystallized at low temperature. In living organisms the proteins are often dynamic structures acting as machines: sometimes they have a well-defined structure, sometimes they fluctuate between multiple conformations, and sometimes they are completely disordered. Instead of a single snapshot, we can describe this dynamic structure as an ensemble of low lying energy conformations. Each conformation belongs to some functional state, e.g. active and inactive. When predicting new protein structures we pick the low energy model as the solution, and we also consider several higher energy models as candidates for functionally important states of the protein. Approaches based on energy function have been successful mainly for small globular proteins soluble in water.

Structure prediction of membrane proteins is in its infancy compared to the structure prediction of soluble proteins, because most of the membrane proteins are large as they need to span the 30 Å thick cell membrane. Also, there are significantly less known crystal structures of membrane proteins, which means less training examples for the computational approaches. Yarov et al. [3] developed a general membrane protein prediction method by generalizing the methods used for soluble proteins, but it was not accurate enough for drug discovery. The recent assessment of the GPCR structure prediction community, GPCRDock2013 [4], showed that approaches based on homology models are currently the most successful [5]. In order to go beyond homology modeling, we extend the ideas of the BiHelix method [6]. Our Trihelix method does complete local sampling of the GPCR structures, while using discretized coordinates for the transmembrane helices. We use physics tools to address the GPCR structure prediction. In particular, we use a multi-scale approach to simplify the problem by treating the transmembrane helices as rigid bodies. Furthermore, we use the mean field approximation in order to handle large number of conformations during the energy evaluation stages.

The objective of this thesis is to computationally predict GPCR structures, which can then be used in drug discovery. Since GPCRs are large proteins unsolvable by direct molecular dynamics, we first study the known GPCR structures to identify features which help to model new proteins. Based on this analysis we developed a sampling scheme to GPCR structure prediction. And finally we apply this technique to predict the structure of the HCAR1 protein.

In Chapter 2 we analyze known GPCR structures in order to identify the features specific to this protein superfamily. The analysis leads to a unified sequence-structural alignment of the GPCR superfamily which can then be used to enable structure prediction of such large proteins. Another result of this analysis is a list of residues which are critical for the receptor activation. Natural variants for amino acids at these locations dramatically influence the activation mechanism, and so these mutations can directly correspond to diseases.

To model a protein with an unknown structure, we need to first map it to a protein with a known structure. In Chapter 3 we construct an extension of this alignment to most other known GPCR sequences. The result is a sequence alignment of transmembrane regions of 817 proteins, and it is included in Appendix D. Using this alignment, an initial homology model for any GPCR can be quickly prepared by mapping the corresponding residues to the most similar available crystal structure.

The initial homology model has to be refined to describe the unique properties of the target protein. In Chapter 4, we describe the Trihelix sampling technique, which refines the initial homology model by complete local sampling of the helical conformations of the transmembrane bundle. Practical methods are limited by the available computational resources, thus we decrease the exponentially large space of conformations by sampling on a coarse grid and by taking mean field approximations at several stages of the protocol. We test the Trihelix method on known crystal structures.

In Chapter 5, we apply the Trihelix method to predict the structure of the HCAR1 protein, which has recently been connected to the regulation of pancreatic cancer. To complete the model, we build the extracellular loops using a new approach specific to GPCRs, which uses constraints from known disulfide bonds. Finally, we dock lactate and and 3,5-dihydroxybenzoic acid molecules to the protein model. The binding sites show a single ligand molecule interacting with three Arg residues, each on a different transmembrane helix, in agreement with available mutation data.

Both the structure prediction and ligand binding calculations rely on evaluation of energy of the models. An important component of the underlying calculations is the determination of the atomic charges. In the final Chapter 6, we describe an extension of the QEQ charge equilibration method, which includes polarization. We derive the relevant formulas and describe the implementation in the Lammps molecular dynamics code. Optimization of atomic parameters for this method is presented in [7]. This work is part of ongoing efforts to improve the energy models for molecular simulations.

Chapter 2

GPCR Fold and Structure Based Alignment

2.1 Signal Messengers

Signal transduction coordinates all life processes in multicellular organisms. All signals have to cross biological membranes, and therefore the membranes are a natural place to intercept biological pathways. G-Protein Coupled Receptors (GPCRs) are a large superfamily of membrane proteins facilitating information transfer across the cellular membrane.

In humans, GPCRs are sensitive to neurotransmitters, hormones, metabolites, and odors among other molecules [8]; rhodopsin is even sensitive to light thanks to the small molecule retinal. This means that proteins from this superfamily control a very diverse set of interactions in the human body. Different GPCRs regulate neural signaling in the brain, control blood pressure, influence metabolism, and are responsible for sensory function of vision, taste, and smell. Out of about 20,000 human proteins [9], there are about 800 human GPCRs [10], half of which are olfactory receptors. This leaves almost 400 GPCRs which can serve as drug targets. Currently, about 30-50% of all drugs and 20% of recently FDA approved drugs act via GPCRs [11], which makes the structure prediction of these protein very important for drug discovery.

There are multiple mechanisms of signal transduction via GPCRs. The best described pathway is the β_2 adrenoreceptor activation and Gs protein binding [12] (Fig. 2.1). First, an agonist binds to the extracellular part of the β_2 adrenoreceptor. While the binding causes relatively small conformational changes in the ligand pocket, it induces large conformational changes of the transmembrane helices. The movement of the helices then enables binding of the Gs heterotrimer (with subunits α, β, and γ) on the intracellular side. After Gs binding, the α subunit releases GDP. Then, GTP can bind to the α subunit, resulting in the separation of the α and $\beta\gamma$ subunits from the GPCR. After that, both subunits regulate their respective pathways until GTP is hydrolyzed to GDP and the Gs heterotrimer can be reassembled.

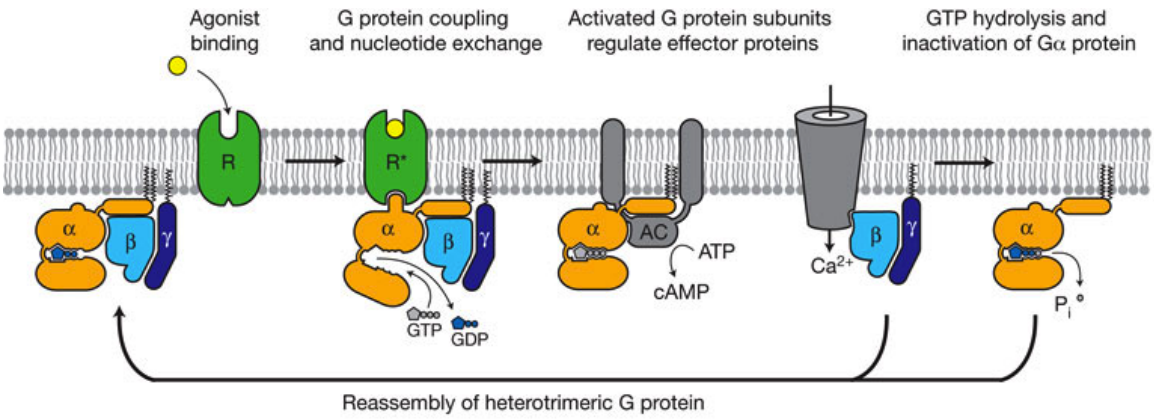


Figure 2.1: Typical mechanism of signal transduction by a GPCR. Source: [12].

This description of the activation mechanism is a result of a decade long effort on crystallization of the active state of a GPCR receptor. Membrane proteins are very difficult to crystalize compared to soluble proteins because folded proteins need to be removed from the membrane environment. Crystallization of GPCRs is even more difficult because the GPCR is not a rigid structure but rather a flexible and dynamic one. The degree of flexibility has been observed with fluorescence studies [13, 14] and long molecular dynamics (MD) simulations (up to 30 μs [15]). Therefore, the structure determination of a GPCR should not look for a single structure but rather the dynamic ensemble of structures low in the energy landscape.

2.2 Transmembrane Helices as Basis For Structure Prediction

The cellular membrane divides a GPCR protein into 3 regions: extracellular, membrane, and intracellular. The N-terminus and three extracellular loops (EC1, EC2, EC3) lie outside of the cell. Seven transmembrane (TM) helices (TM1-7) span the membrane. And inside the cell, there are three intracellular loops (IC1, IC2, IC3) together with the C-terminus, which is typically a shorter helix 8 (see Figure 2.2). The intracellular and extracellular loops have different lengths and admit different conformations among the known GPCR crystal structures. However, the fold of the transmembrane helices is very well conserved even for proteins with very small sequence similarity.

We use this common GPCR fold as a basis for structure prediction method for all GPCRs. GPCRs are large proteins with more than 300 residues, and thus difficult to model with the usual ab-initio protein modeling methods [17, 3]. Since the common GPCR fold is formed by alpha-helices, to first approximation, we can think of these as rigid bodies. In order to predict the structure of a novel GPCR we start from the most similar known structure, template, and then refine the structures with the necessary rigid body moves of the TM helices.

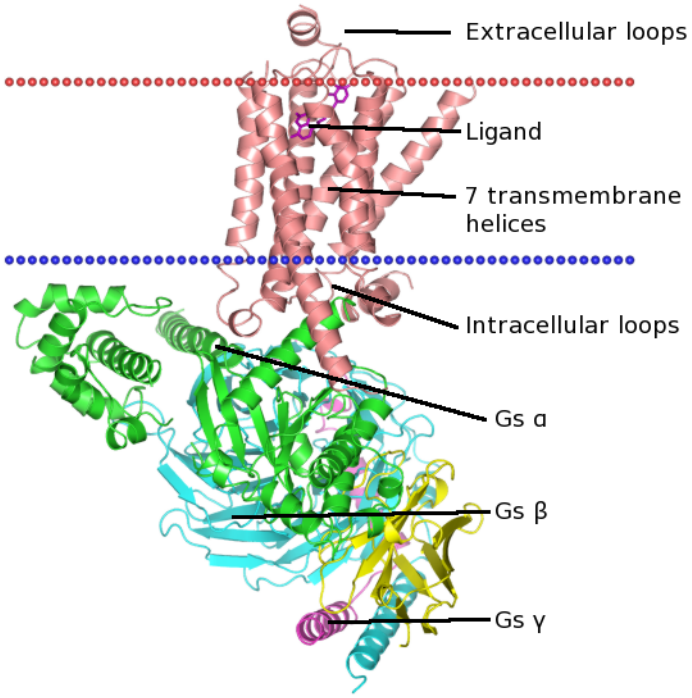


Figure 2.2: β_2 adrenoreceptor in complex with Gs protein (with subunits α, β, and γ). The membrane divides the receptor into the extracellular part (N-terminus, EC1-3), transmembrane helices, and the intracellular part (IC1-3, C-terminus). Image source: [16], structure: [12].

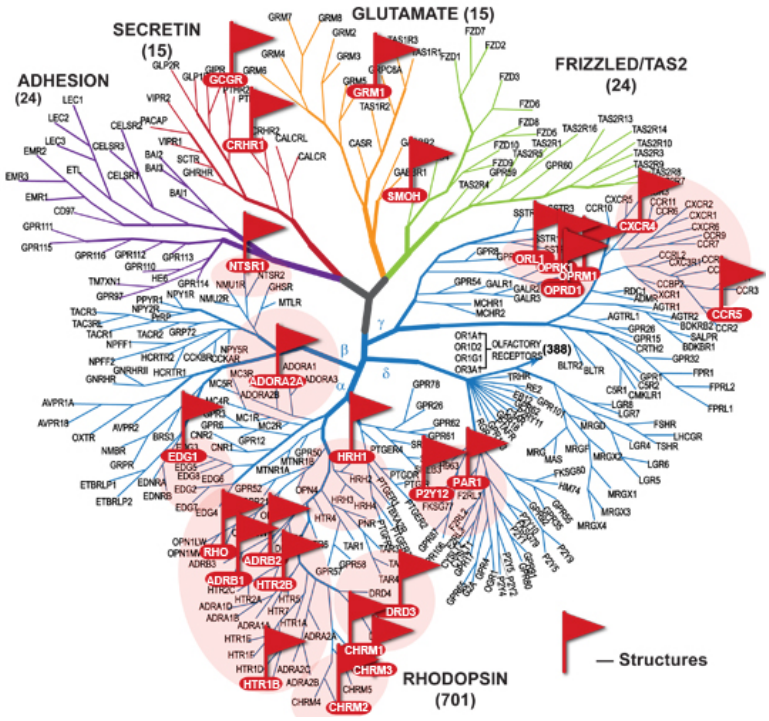


Figure 2.3: Phylogenetic tree of 342 human GPCRs by Fredrikson et al [18]. Red flags denote crystal structures solved prior to May 2014. The denoted groups can be assigned to the GPCR classes as follows: class A (rhodopsin-like receptors), class B (secretin and adhesion family), class C (glutamate receptors), and class F (frizzled and bitter taste receptors). Source: GPCR Network [10], a large collaboration trying to experimentally determine structure of many GPCRs.

2.3 Class A Alignment

Before we can model the TM bundle from a suitable template, we need to find the optimal sequence alignment. This turns out not to be an easy task for GPCRs because they are a very diverse class of receptors. There are many ways to classify the GPCR family, e.g., the International Union of Basic and Clinical Pharmacology classifies the receptors by known ligands [8], whereas sequence based classifications are used in [19, 20, 18, 21].

First, we use the usual classification of the GPCR superfamily into the classes A-F. In Chapter 3 we suggest a new classification based on TM similarity that is suitable for structure prediction. Fredrikson et al [18] performed detailed phylogenetic analysis further dividing classes A-F into branches: class A (rhodopsin-like receptors), class B (secretin and adhesion family), class C (glutamate receptors), and class F (frizzled and bitter taste receptors). Figure 2.3 shows the resulting phylogenetic tree of 342 nonolfactory human GPCR proteins. Classes D, E are part of the classification that includes other species, but contain no human proteins.

Sequence alignment is typically possible within each GPCR class separately, not because the whole sequences are similar enough, but because of several conserved motifs. In particular, the class

TM	.46	.47	.48	.49	**.50**	.51	.52	.53
1				G	**N**	x	x	V
2	L	x	x	x	**D**			
3				E/D	**R**	Y		
4					**W**			
5		F	x	x	**P**			
6		C	W	x	**P**	F	F	
7				N	**P**	x	x	Y

Table 2.1: Class A conserved motifs for each TM. x denotes a non-conserved residue. The numbers correspond to the Ballesteros-Weinstein numbering scheme [22], in which $n.50$ is the most conserved residue in TM n.

A proteins have highly conserved residues at specific positions, which are displayed in Table 2.1. These highly conserved residues define the $n.50$ residue Ballesteros-Weinstein numbering scheme [22] and are often involved in inter-helical hydrogen bonding or correspond to a proline kink in the alpha helix.

Figure 2.4 shows the structural alignment of the class A structures corresponding to the class A sequence alignment. All the structures are in the same fold, as the backbone root mean square deviation (RMSD) of the alignment is typically between 1 and 3 Å (see Figure B.2). The RMSD is defined as:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_{1i} - \mathbf{r}_{2i}|^2},$$

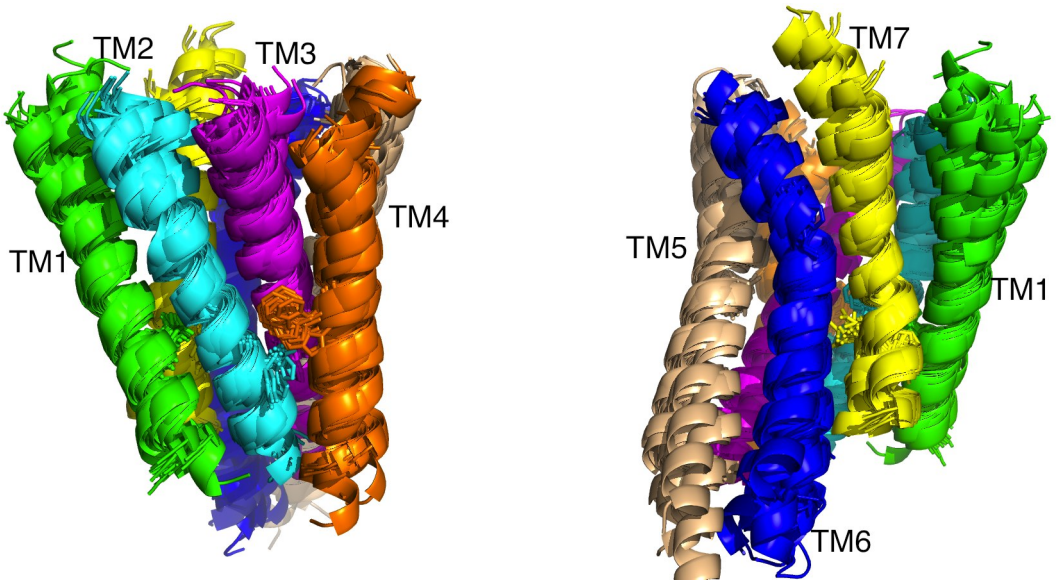
where the sum runs over corresponding pairs of atoms at positions $\mathbf{r}_{1i}$ and $\mathbf{r}_{2i}$. For comparing similar folds we include only backbone atoms from each amino acid: N, C_α, C, O.

Furthermore, the correspondence of the residues in this alignment is very good, since the corresponding residues overlap, as can be seen in Figure 2.5. The figure shows the residues involved in hydrogen bonding networks, which are most conserved in class A. Thus, within class A we see the structural biology paradigm at work:

$$\text{Sequence} \sim \text{Structure} \sim \text{Function}$$

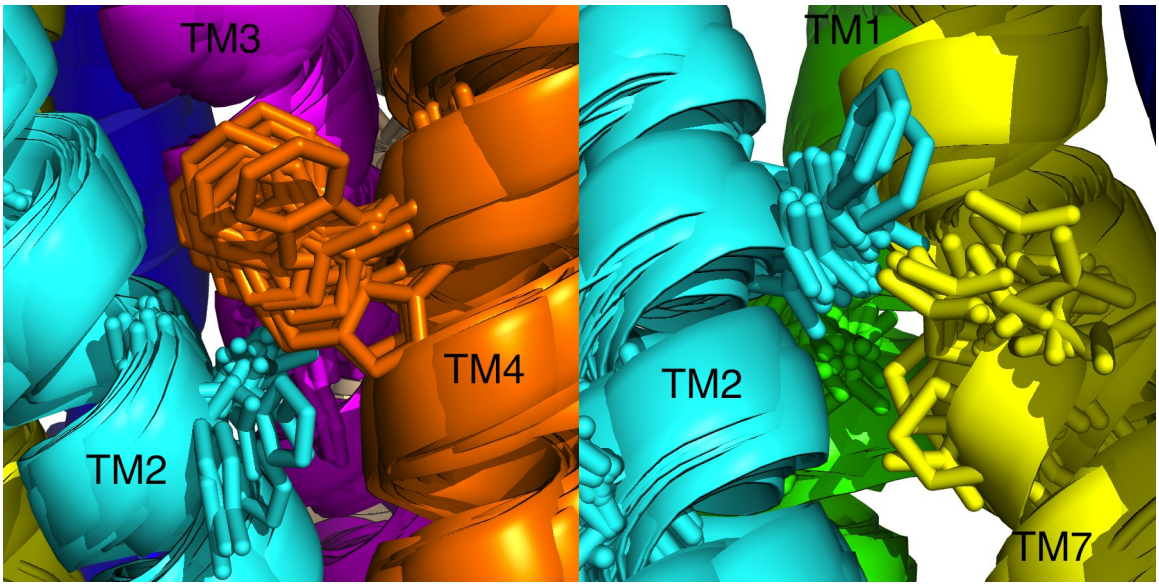
Within class A, the sequence identity between the corresponding transmembrane regions is 15-40% (Figure B.3). However, the sequence identity between class A and GPCR proteins from other classes is so low, typically 5-15%, that one cannot simply align the sequences between GPCRs from different classes. At present, there are 5 non-class A GPCR structures available. When these structures are properly aligned to class A, it is clear that they are in the same fold as class A GPCRs (see Figure 2.6), thus leading to a disagreement with the structural biology paradigm within the GPCR superfamily:

$$\text{Sequence} \nsim \text{Structure} \sim \text{Function}$$



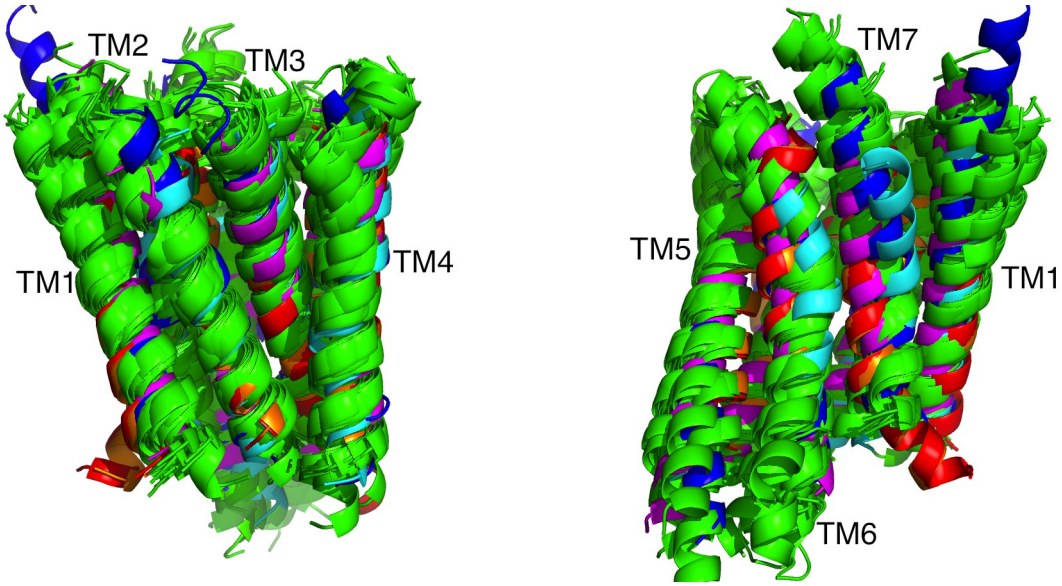
(a) View at TM 1, 2, 3, 4 (b) View at TM 5, 6, 7, 1

Figure 2.4: 20 different class A structures aligned to β_2. The helices overlap each other very well, showing how well-defined the class A TM fold is. Each TM has a different color. Loops are omitted since they do not share the same fold. Detailed view is shown in Fig. 2.5. Rendered using PyMOL [23].



(a) W4.50 ↔ S/N/T2.45 ↔ S/N/T3.42 (b) N1.50 ↔ D2.50 ↔ N7.49 (water-mediated)

Figure 2.5: Detailed view into Fig. 2.4 showing conserved motifs in class A GPCRs. Even though 20 different crystal structures are shown, the conserved residues have a very similar position.



(a) View at TM 1, 2, 3, 4

(b) View at TM 5, 6, 7, 1

Figure 2.6: All available structures aligned to β_2. 20 different class A structures are colored green. CRF1 is cyan, GLR is blue (both class B). MGLU1 is orange, MGLU5 is red (both class C). SMO is magenta (class F). The corresponding helices overlap each other quite well, showing how well defined the general GPCR TM fold is.

In the next section, we construct the optimal sequence alignment, which corresponds to the Figure 2.6.

One way to align two GPCR structures with small sequence identity is to use a fully structural alignment. Structural alignment typically works iteratively, removing from the alignment those atoms which were too far in the previous rounds. This works relatively well, but the exact sequence pairing is not uniquely defined and depends on the cutoff parameter. In many cases it leaves an ambiguity of ± 4 residues (1 helical turn).

In order to remove this ambiguity and to see which alignment would minimize the RMSD of the full TM bundle, we start from the approximate structural alignment, and try all nearby sequence alignments (± 1 helical turn on each helix). But minimizing only RMSD does not necessarily lead to an optimal alignment. For several cases in class A, we find TM alignments which have slightly lower RMSD than the alignment conserving the Ballesteros-Weinstein residues. The reason for this is that the extracellular ends of the helices sometimes have significantly different tilt, so the most tilted helix can dominate the RMSD measure. To avoid these issues with the RMSD measure, we instead analyze conserved inter-helical contacts. In the following section, we obtain the alignment, which leads to the structural alignment of all available GPCR proteins on Figure 2.6.

Class	Ligand	Protein Name	Short	Gene	PDB Id Inactive	PDB Id Active	Uniprot	Ref.
Aα	opsin	bovine rhodopsin	RHO	RHO	1GZM.A	3PQR.A	P08100	[25, 26]
Aα	aminergic	turkey β_1 adrenergic receptor	β_1AR	ADRB1	2VT4.B		P08588	[27]
Aα	aminergic	human β_2 adrenergic receptor	β_2AR	ADRB2	2RH1.A	3SN6.R	P07550	[28, 12]
Aα	aminergic	human dopamine D3 receptor	D3	DRD3	3PBL.A		P35462	[29]
Aα	aminergic	human histamine H1 receptor	H1	HRH1	3RZE.A		P35367	[30]
Aα	aminergic	human muscarinic acetylcholine receptor M2	M2	CHRM2	3UON.A	4MQS.A	P08172	[31, 32]
Aα	aminergic	rat muscarinic acetylcholine receptor M3	M3	CHRM3	4DAJ.A		P20309	[33]
Aα	aminergic	human serotonin receptor 1B	5HT1B	HTR1B	4IAR.A		P28222	[34]
Aα	aminergic	human serotonin receptor 2B	5HT2B	HTR2B	4NC3.A		P41595	[35]
Aα	nucleoside	human adenosine receptor A2a	A2A	ADORA2A	3EML.A	3QAK.A	P29274	[36, 37]
Aα	lipid	human sphingosine 1-phosphate receptor 1	S1P1	S1PR1	3V2Y.A		P21453	[38]
Aβ	peptide	rat neurotensin receptor type 1	NTS1	NTSR1		4GRV.A	P30989	[39]
Aγ	peptide	human C-X-C chemokine receptor 4	CXCR4	CXCR4	3ODU.A		P61073	[40]
Aγ	peptide	human C-C chemokine receptor 5	CCR5	CCR5	4MBS.A		P51681	[41]
Aγ	peptide	human κ-opioid receptor	κOR	OPRK1	4DJH.A		P41145	[42]
Aγ	peptide	mouse μ-opioid receptor	μOR	OPRM1	4DKL.A		P35372	[43]
Aγ	peptide	human nociceptin/orphanin FQ opioid receptor	NOP	OPRL1	4EA3.A		P41146	[44]
Aγ	peptide	mouse δ-opioid receptor	δOR	OPRD1	4EJ4.A		P41143	[45]
Aδ	peptide	human proteinase-activated receptor 1	PAR1	F2R	3VW7.A		P25116	[46]
Aδ	nucleoside	human P2Y purinoreceptor 12	P2Y12	P2RY12	4NTJ.A		Q9H244	[47]
B	peptide	human corticotropin-releasing factor receptor 1	CRF1	CRHR1	4K5Y.B		P34998	[48]
B	peptide	human glucagon receptor	GLR	GCGR	4L6R.A		P47871	[49]
C	peptide	human metabotropic glutamate receptor 1	MGLU1	GRM1	4OR2.A		Q13255	[50]
C	peptide	human metabotropic glutamate receptor 5	MGLU5	GRM5	4OO9.A		P41594	[51]
Frizzled	peptide	human smoothened homolog	SMO	SMO	4N4W.A		Q99835	[52, 53]

Table 2.2: List of GPCRs with available crystal structure. When multiple structures were available, then the one with the highest resolution or the one with least deformed TM helices is listed.

2.4 Alignment Based on Common Contacts

In class A, the analysis of inter-helical interactions typically focuses on hydrogen bonds because many of the hydrogen bonding residues are highly conserved. However, the sequence is not conserved between different classes, and therefore we look at conserved contacts instead. We use the definition of a inter-helical contact as in [24]: any two heavy atoms from different TMs that are closer than the sum of their van der Waals radii plus 0.6 Å.

All known crystal structures are listed in Table 2.2. Many of the transmembrane helices contain bends, and sometimes the termination of the helices is not well defined. We define the extent of each transmembrane helix as the residues positioned in the membrane (as placed by the Orientations of Proteins in Membranes (OPM) database [16]) extended until the end of the alpha helix by the DSSP secondary structure determination [54]. The helices were manually inspected and only a few manual corrections were needed. The final TM lengths are displayed in Table B.1.

Once the TM regions are defined, we analyze the number of inter-helical contacts of all known structures (listed in Table 2.2). Each structure has about 200 contacts (see Table B.1). For class A, the alignment preserving Ballesteros-Weinstein (BW) numbering maximizes the number of common contacts between any two structures. For Classes B, C, and F the BW numbering is not defined, because the sequence is not conserved. Starting with an approximate initial structural alignment, we try all possible adjustments to BW 50 residues on each helix and count the number of contacts common with any of the class A structures.

Table 2.3 shows the best alignments. Class B (CRF1, GLR) alignment agrees with the suggested

	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Common contacts
CRF1	**L134**	**F162**	**L213**	**W236**	**V279**	**L329**	**S360**	2212
	L134	F162	L213	W236	V279	I325	A363	2084
	L134	F162	L213	W236	V279	I325	S360	2081
GLR	**L156**	**F184**	**L249**	**W272**	**A314**	**V364**	**A397**	1972
	L156	F184	L249	G273	A314	V364	A397	1913
	L156	F184	L249	M276	A314	V364	A397	1876
MGLU1	**T607**	**I638**	**I682**	**I714**	**L763**	**A800**	**L827**	2017
	T607	I638	I682	S711	L763	A800	L827	2012
	T607	I638	I682	S711	C767	A800	L827	1873
MGLU5	T594	I625	A669	F698	L750	A787	L814	1974
	T594	**I625**	**A669**	**I701**	**L750**	**A787**	**L814**	1954
	T594	I625	A669	F698	L750	I784	L814	1820
SMO	**T245**	**F274**	**W339**	**W365**	**V411**	**I465**	**S533**	2358
	T245	F274	W339	W365	V411	C469	G529	2311
	T245	F274	W339	W365	V411	C469	S533	2283
	...							
	T245	F274	W339	W365	V411	S468	I530	1827

Table 2.3: This table shows the selection process for assigning BW .50 residues to non class A proteins. Shifting BW .50 residue on each helix renumbers the relative BW numbers, effectively changing the labels of contacts observed in these proteins. Subsequently, the number of common contacts each structure shares with the class A structures is different for different BW residue assignments. The rightmost column shows the cumulative number of contact occurrences among the 24 class A structures. The BW assignment with the highest number of contacts is selected (except for MGLU5). The selected alignment is in bold.

alignment in [49], which was obtained by an iterative structural alignment. Similarly, we chose class C (MGLU1, MGLU5) alignment, which agrees with the suggested alignment in [50]. For MGLU5, only the second highest scoring alignment was chosen, so that the alignment is consistent with the MGLU1. This choice was checked manually and the corresponding residues are in a more similar position in the selected alignment.

For the SMO receptor our procedure obtains different sequence alignment than the iterative structural alignment [52]. The proposed alignment is shown on the last line on Table 2.3, and it differs from reference [52] in helix 6 and 7. Helix 7 of the SMO receptor does not have any proline residue, and so it is missing the kink that is typical for class A. There are many inter-helical contacts in the extracellular part of the TM7, and so the chosen alignment gives a good spatial correspondence for the larger part of helix 7.

This alignment was used to overlay the structures on Fig. 2.6. The RMSD of the TM bundle alignment between classes is about 2.5-3.5A, from which we reach the surprising conclusion that the class B, C and, F proteins share the same fold as class A. Thus considering all GPCRs, we have an example of proteins with similar structure and function, but dissimilar sequence:

$$\text{Sequence} \not\sim \text{Structure} \sim \text{Function}$$

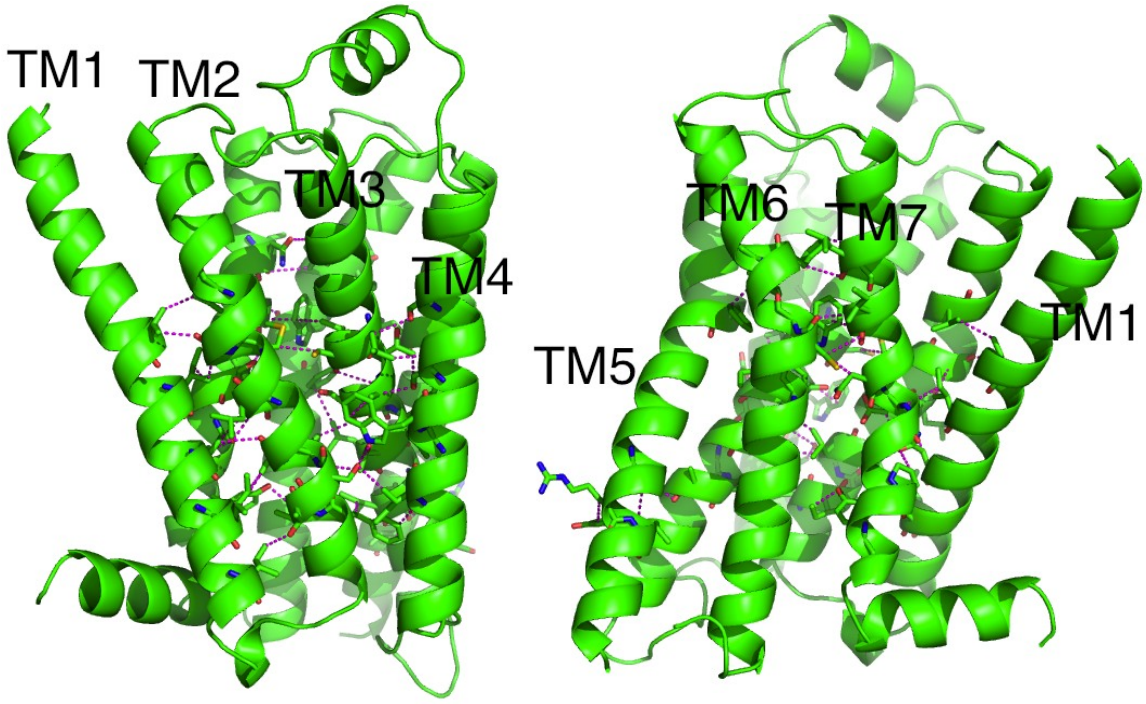


Figure 2.7: β_2 crystal structure in two side views. Inter-helical contacts conserved in class A are represented as dotted lines in magenta. The residues forming them are displayed using stick representation. The detailed list of the contacts is in Fig. 2.8.

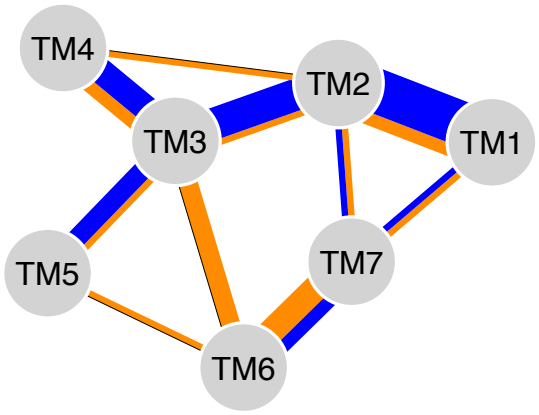
It may be possible to gain new insights into the class A fold by comparing it to the other classes.

Chapter 3 uses this alignment of the known crystal structures and generalizes it to the alignment for most of the GPCR sequences by anchoring each subfamily to the known alignment.

2.5 Conserved Contacts

Venkatakrishnan et al. [24] pointed out that, besides several hydrogen bonds, a large number of inter-helical contacts are conserved for most structures in class A. Considering any inter-helical contacts, and not just hydrogen bonds, is better for comparing structures across the GPCR classes, since the chemical nature of the contacts does not have to be conserved. For a diverse set of proteins, the residues can mutate from polar to hydrophobic while the contact of the nearby residues is preserved. In the previous section we used common contacts as a way to compare structures across the GPCR classes. In this section we extend the comparison to find common contacts that possibly define or constrain the GPCR fold. This analysis provides additional insights that are useful for modeling large proteins.

The inter-helical contacts, which are present in almost all studied class A structures, are displayed in Fig. 2.7 and listed in Figure 2.8. This list is very similar to the contacts found by [24], but there are minor differences caused by using a different set of crystal structures. More interesting is it to



TM2-TM3:	2.42-3.45, 2.42-3.46, 2.45-3.42, 2.46-3.39, 2.46-3.42, 2.53-3.35
TM3-TM4:	3.34-4.53, 3.34-4.57, 3.37-4.53, 3.38-4.50, 3.38-4.53, 3.41-4.49
TM3-TM5:	3.47-5.57, 3.51-5.57, 3.51-5.60, 3.51-5.61
TM3-TM6:	3.36-6.48, 3.40-6.44, 3.43-6.44
TM2-TM4:	2.45-4.46
TM5-TM6:	5.47-6.52
TM6-TM7:	6.47-7.41, 6.47-7.42, 6.47-7.45, 6.48-7.42, 6.51-7.38, 6.51-7.39
TM2-TM7:	2.43-7.53, 2.50-7.46
TM1-TM2:	1.43-2.54, 1.43-2.58, 1.46-2.54, 1.50-2.47, 1.50-2.50, 1.50-2.51, 1.53-2.47, 1.54-2.47, 1.57-2.44
TM1-TM7:	1.46-7.47, 1.50-7.50

Figure 2.8: Diagram of inter-helical contacts present in at least 23 out of 24 studied class A structures. The contacts common to all classes are in blue, and contacts present only in class A in orange.

compare the list of chemically unspecific contacts to the conserved hydrogen bonds. Within class A, one often focuses on two conserved networks of hydrogen bonds, which were shown on Fig 2.5: W4.50 ↔ S/N/T2.45 ↔ S/N/T3.42 and N1.50 ↔ D2.50 ↔ N7.49.

Fig. 2.8 shows that the network 4-2-3 is well conserved across classes. In particular, the contact 2.45-3.42 is present in all classes, and there are many conserved contacts in its immediate vicinity, such as 2.42-3.45, 2.42-3.46, and 2.46-3.42. The contact between TM 3 and 4 is also well conserved, as the highly conserved bulky residue W4.50 leans on A3.38.

The network 1-2-7 also has many conserved contacts, including N1.50-D2.50, but interactions with N7.49 are not conserved. Even In class A, N7.49 interacts with the other residues of this hydrogen bonding network indirectly through a water molecule. Still, the side chain packing in the regions where the helices are close together is important for structural stability.

Any mutations to the conserved contacts can cause structural stability issues for the protein. Thus, naturally occurring mutations of the residues involved in the conserved contacts could be direct causes of diseases. In the next section we discuss several diseases and disorders caused by naturally occurring single nucleotide polymorphisms (SNP). The exact position of each SNP is important since it determines whether the SNP causes some structural defects or whether it is likely to be benign. The positions of many SNPs are known from genetic studies, and by using the global GPCR alignment, which is derived in the next chapter, we can get the BW numbering of each SNP. The BW numbers of each SNP can then be compared against the list of conserved contacts to estimate their importance.

We cannot determine with confidence the conserved contacts in classes B, C, and F yet, because there are at most two structures in each of these classes. Nevertheless, we start the analysis with the available structures, so that we at least reduce the number of residues that can be more structurally

important. There are two structures available in classes B, C and only one in class F. Figures 2.9a, 2.9b, and 2.9c show the inter-helical contacts found in classes B, C, and F. The lists show interactions which should be considered first when one studies the structure or function of these proteins.

Even though the sequence similarity between the classes is low, there are structural similarities. Especially important are the similarities on the intracellular side of the GPCRs, since the same G-proteins bind to all GPCR classes. Let us now compare class A to the other classes. The blue color in Fig. 2.8 denotes the contacts common to all classes, and orange denotes contacts specific to class A. Only one contact, 6.51-7.39, is present in all of class A structures (active and inactive), but not in the structures of the other classes. Furthermore, the interactions of TMs 1-5 are more conserved across all classes, but the TM 6 and 7 contacts are very class A specific. It is possible that during the GPCR assembly the helices 1-5 form some intermediate partially folded state before helices 6 and 7 are fully present in the membrane. This might be the reason why the contacts between helices 1-5 are more similar across the classes. Another possibility is that the motion of TM6, which is critical for activation, is different for classes B, C, and F.

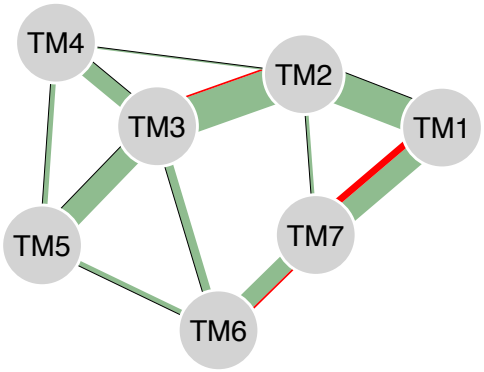
To summarize, in this section we identified mutations which likely cause structural changes in proteins. We next study mutations that cause functional problems.

2.6 Inter-helical Contacts Involved in Activation

In this section we show that the class A activation mechanism relies critically on a small number of residues. If any of these residues is mutated, the free energy balance between the active and inactive states is modified, and the receptor is likely to become either constitutively active or inactive, which is often the cause of a disease. Comparing the common contacts among different proteins, as we did in the previous section, is not straightforward because many of the sequence differences are random. Focusing on the difference between active and inactive structures of the same protein makes the significance of the individual residues much clearer. There are 3 active-inactive structure pairs available: RHO, β_2AR, and M2. The active structure of A2A is only partially active, and for NTS1act the inactive structure is not available.

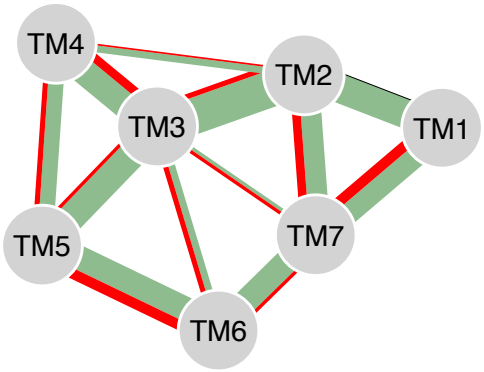
For rhodopsin, which is the most studied, the main signature of activation is the breaking of the salt-bridge R3.50↔E6.30 and the forming of the salt-bridge K5.66↔E6.30. Instead of keeping track of hydrogen bonds only, the analysis of contacts allows us to study more general changes during the activation. Figure 2.10 shows the common changes in contacts among all the active-inactive structure pairs.

Most of the changes occur for TM 6, since the intracellular end of helix 6 undergoes the largest movement upon activation. However, TM 7 also shows a large number of systematic changes, as it breaks a contact with TM1 and creates new contacts with TM 2 and 3. The residues 3.43 and



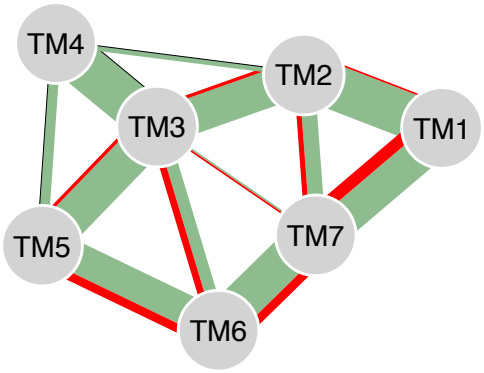
TM2-TM3: 2.63-3.25
TM6-TM7: 6.47-7.38
TM1-TM7: 1.43-7.42, 1.46-7.44, 1.50-7.49, 1.56-7.54

(a) **Class B**: CRF1, GLR



TM2-TM3: 2.48-3.38, 2.52-3.38, 2.59-3.27
TM3-TM4: 3.33-4.53, 3.37-4.50, 3.41-4.46, 3.44-4.49, 3.45-4.42
TM3-TM5: 3.33-5.47, 3.37-5.50
TM3-TM6: 3.40-6.41, 3.43-6.37, 3.46-6.33
TM3-TM7: 3.32-7.38, 3.40-7.45
TM2-TM4: 2.42-4.43
TM4-TM5: 4.49-5.46, 4.53-5.43, 4.53-5.47
TM5-TM6: 5.37-6.56, 5.41-6.56, 5.44-6.48, 5.44-6.51, 5.48-6.48, 5.59-6.42
TM6-TM7: 6.55-7.34, 6.55-7.37
TM2-TM7: 2.46-7.54, 2.50-7.54, 2.54-7.43, 2.54-7.47, 2.60-7.38
TM1-TM7: 1.36-7.36, 1.52-7.58, 1.53-7.58, 1.56-7.58, 1.56-7.61, 1.57-7.58

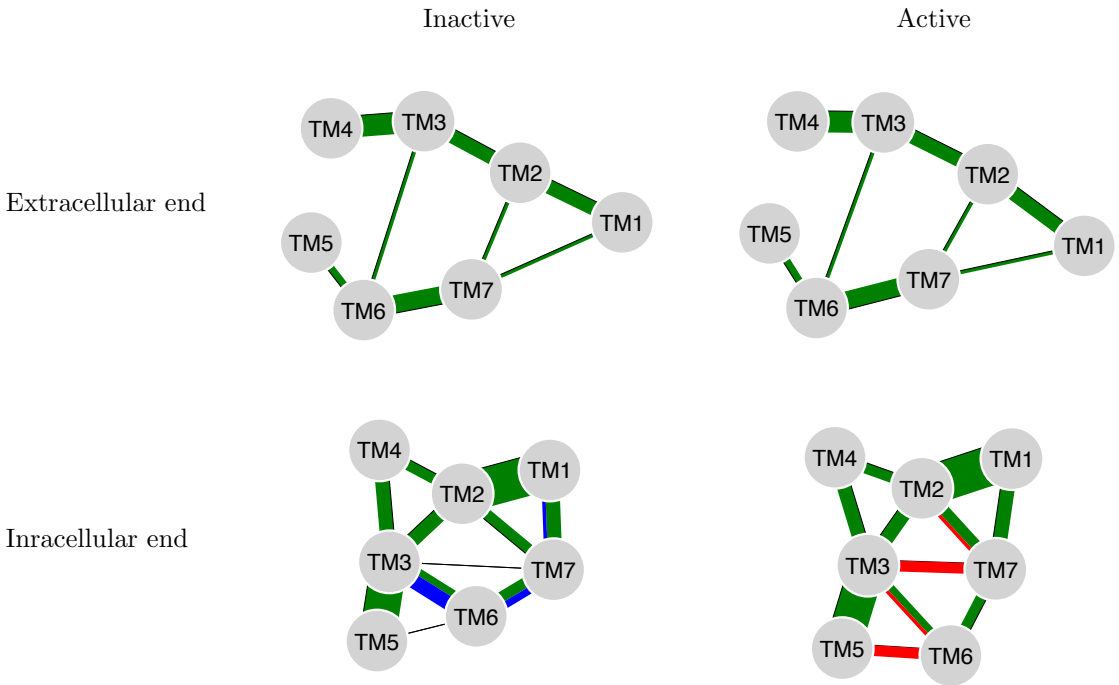
(b) **Class C**: MGLU1, MGLU2



TM2-TM3: 2.50-3.36, 2.57-3.36
TM3-TM5: 3.43-5.50, 3.50-5.62
TM3-TM6: 3.43-6.45, 3.50-6.38, 3.50-6.42, 3.53-6.34
TM3-TM7: 3.46-7.49
TM5-TM6: 5.44-6.59, 5.47-6.51, 5.50-6.48, 5.58-6.45, 5.62-6.42
TM6-TM7: 6.33-7.52, 6.36-7.51, 6.40-7.51, 6.47-7.40, 6.50-7.40
TM2-TM7: 2.40-7.52, 2.50-7.42, 2.57-7.42
TM1-TM2: 1.36-2.60
TM1-TM7: 1.31-7.32, 1.35-7.32, 1.35-7.35, 1.38-7.39, 1.52-7.53, 1.56-7.53, 1.57-7.52

(c) **Class F**: SMO

Figure 2.9: Diagram of interhelical contacts present in classes B, C, and F. The width of the line connecting two TMs is proportional to the number of contacts present in all structures from the given class. The list (emphasized in red) shows the contacts not present in any another available structure.



TM3-TM6:	3.43-6.40, 3.43-6.41, 3.46-6.37, 3.50-6.37	3.50-6.40
TM3-TM7:		3.43-7.49, 3.43-7.53, 3.46-7.53
TM5-TM6:		5.55-6.41, 5.58-6.40, 5.62-6.37
TM6-TM7:	6.36-7.53, 6.40-7.49	
TM1-TM7:	1.53-7.53	
TM2-TM7:		2.43-7.54

Figure 2.10: Diagram of interhelical contacts changed upon activation. Width of the green lines is proportional to the number of contacts common to all six structures (RHO, β_2AR, M2, and their active structures). Blue shows the contacts present only in inactive structures, but in no inactive structures; and red shows the opposite. The upper diagrams show contacts in the extracellular half of the membrane. We see that there is no systematic change common to the class A receptors in the conformation of the extracellular half of the TMs. This is not obvious, because there are conformational changes accompanying ligand binding. All the systematic changes, which enable G protein binding, occur in the intracellular half of the TMs. The list only contains 16 different residues, but it has 15 different contacts. Thus many of the residues switch partners upon activation.

3.46 occur in the list of conserved contacts in both active and inactive structures, therefore the conformational changes around these residues seem to be very important for the conformational changes during activation.

It has been experimentally shown that single amino acid mutations can have a dramatic effect on GPCR activity. For example, the mutation T3.46A makes the receptor CB1 inactive, while the mutations T3.46I and L3.43A make it constitutively active [55, 56]. These mutations are introduced by experimentalists, but the exact list of contacts on Figure 2.10 is useful for explaining certain natural variants as well. Below we provide several examples. We scanned the Uniprot database [57] for naturally occurring mutations and converted the residue numbering to the BW scheme using our alignment.

For example, the natural variants R3.50C and R3.50L cause the vasopressin V2 receptor to be constitutively active. This causes 'nephrogenic syndrome of inappropriate antidiuresis', which presents itself as an inability to excrete a free water load, resulting in low sodium levels [58]. The mutations of R3.50 clearly interfere with arginines's ability to form hydrogen bonds, and so they disrupt the activation mechanism.

Similarly the natural variant H2.43R in Parathyroid hormone receptor causes its constitutive activity. This mutation of class B receptor causes 'Jansen metaphyseal chondrodysplasia', which is characterized by short-limbed dwarfism [59]. Since the same G proteins couple to different GPCR classes, we can expect the same or similar structural signatures of activation in class B as in class A.

For both of these examples the mutations were shown to cause constitutive activity. However, the Uniprot database also shows many natural mutations, for which the effect is unknown.

For example, we predict that the mutation M3.43T of the G-protein coupled receptor 56 will influence its activation, because the residue 3.43 has to switch contact residues during activation. This adhesion GPCR is involved in cell adhesion as well as in cell to cell interactions, and regulates the migration of neural precursor cells; thus the mutation likely has serious consequences. No databases of single nucleotide polymorphisms contain any information about this mutation (we checked the TinyGRAP [60] and NAVA [61] databases), therefore this is a true prediction.

Another example is the T6.36P mutation of the D(1B) dopamine receptor. This is a class A receptor and it influences the activity of adenylyl cyclase. Again, we predict that the T6.36P mutation makes the receptor either constitutively active or fully inactive.

Finally, let us consider natural variants of the HCAR1 receptor, whose structure we study in the next chapter. The receptor has four known naturally occurring mutations [62]:

H43Q	no effect	IC1
A110V	reduced activity	3.47, close to 3.46
S172L	reduced activity	EC2
D253H	reduced activity	EC3

In this case, we can attribute the 3.47V mutation to influencing the activation mechanism due to indirect interaction with the 3.46 residue. The mutations S172L and D253H likely influence the sensitivity to the endogenous ligand of HCAR1, lactate, by changing the loop structure of the receptor.

To summarize, in this section we identified specific residues which are critical for GPCR activation. We illustrated the importance of these residues by finding diseases that are caused by mutations on these residues. The list of residues directly involved in activation only contains 16 residues, which is only about 5% of the residues in the transmembrane domain.

Chapter 3

Extension of Alignment to All GPCRs

3.1 Low Sequence Similarity Between Classes

In Section 2.4 we described how to construct a sequence alignment for all the known GPCR crystal structures. We had to use structural information to find the relation between the different classes, as the sequence similarity across the classes is very low. In this section we study in detail the differences between the structures.

Since most of the proteins are in class A, we projected the Ballesteros-Weinstein numbering [22] to the other classes. In this scheme, the most conserved residue for each TM is denoted n.50 (n = 1 to 7 denotes the TM). We use this numbering even for classes B, C, and F, in which different residues might be most conserved on some of the TMs. For example, class B Wootten numbering [63] labels n.50 residues which are most conserved within this class only.

Figure 3.1 shows the alignment of the TM3 regions for all the studied crystal structures. We can see that the DRY motif at positions 3.49-3.51 is highly conserved within class A (in all but the last 5 sequences), and even when there are mutations only similar amino acids occur: ERY, DRF. In classes B, C, and F the DRY motif is not conserved at all.

Proline residues often cause a helix kink, so they are structurally important for deciding which structure should be used as template for modeling a new protein. In Figure 3.1, prolines are highlighted in red. Only MGLU5 has a proline in a central region of TM3, but in this case, the shape of TM3 is very similar to MGLU1, which does have the corresponding proline.

The consensus sequence for TM3 mostly agrees with class A residues, because most of the crystal structures are from the class A. Interestingly the most conserved residue across all classes is Cys3.25, which forms a disulfide bond to the extracellular loop EC2. This bond is important for the stability of the protein, and if one of the cysteine residues forming this bond is mutated, the protein does not even assemble in the membrane [64].

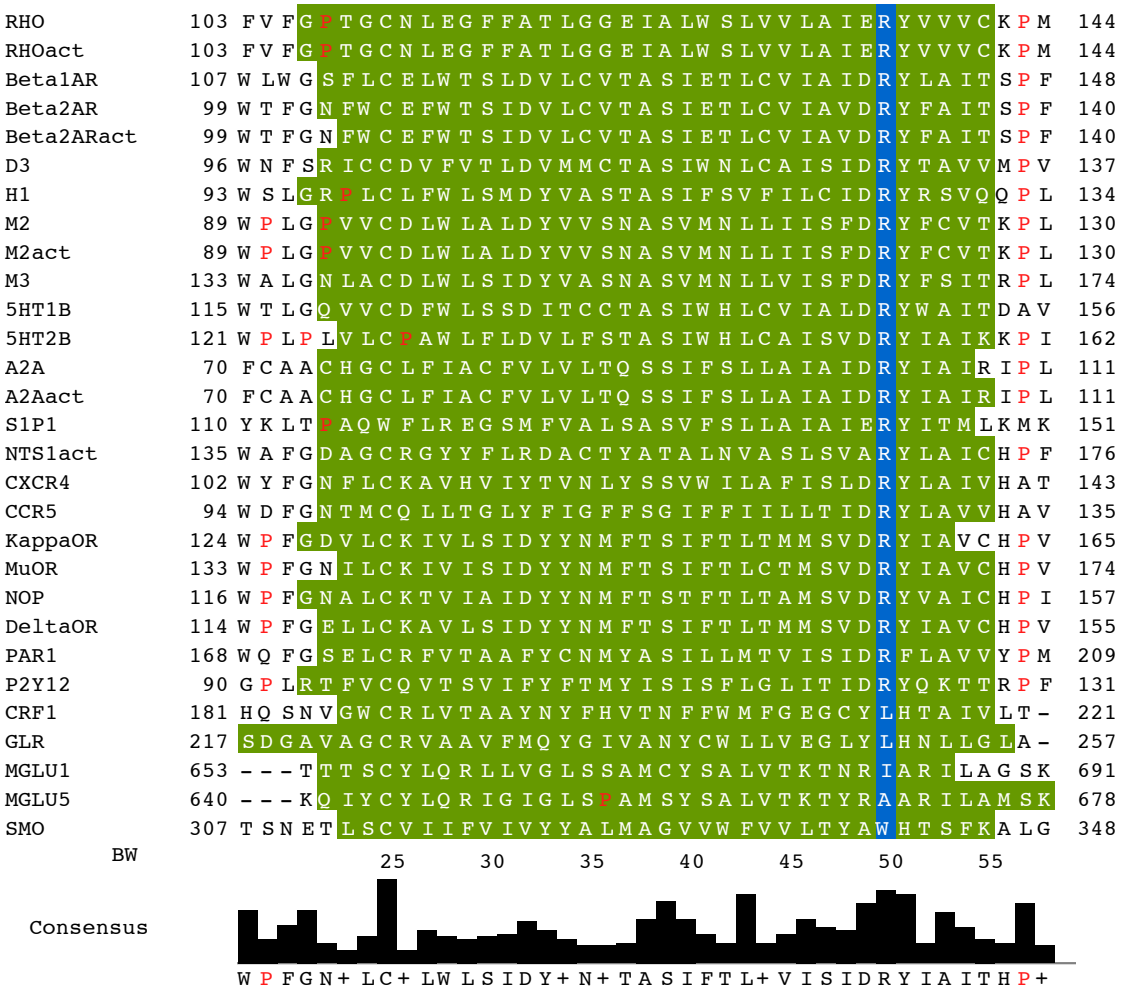


Figure 3.1: TM 3 alignment for the crystal structures. The blue column labels the n.50 residues in BW numbering and the green color denotes the extent of the TM regions, which is very similar across all classes. Prolines are highlighted in red. The last 5 sequences belong to classes B, C, and F, and the rest to class A. There is very low sequence similarity between the different classes.

The alignment of other TMs is shown on Figures figs. C.1 to C.7 in Appendix C.

3.2 Source of GPCR Sequences

Fredriksson [18] performed the first detailed phylogenetic analysis and classified 342 nonolfactory human GPCR sequences into five main groups: rhodopsin, secretin, adhesion, glutamate, and frizzled/taste2. The phylogenetic tree based on this analysis is displayed on Figure 2.3.

The complete list of GPCRs is still not set and changes as new analyses of the human genome are completed. Therefore we extended the list of candidate GPCRs with the proteins from the following resources: [10, 8, 57]. Altogether we collected 836 candidate sequences, which can be categorized in the following classes:

88	Aα	16	B	5	Vomeronasal
33	Aβ	22	C	25	Taste2
57	Aγ	11	F	11	Other
58	Aδ	33	Adhesion	8	Pseudogene
51	A-other	418	Olfactory		

In this work, we label the rhodopsin group as class A. Since this is the largest group, we keep Fredriksson's subdivision to the 4 subgroups: $\alpha, \beta, \gamma, \delta$. Class A sequences, which were not included in Fredriksson's list, were labeled *A-other* in the present work. Class B labels the secretin group, class C labels the glutamate group, and class F labels the frizzled group. The adhesion group has sequences, which at the TM regions are similar to class B. The olfactory and vomeronasal receptors are similar to class A. The main deviation from the classification of Fredriksson is observing that the *Taste2* group is more similar to class A than to class F, as will be argued in the following section.

Sequences which were collected from sources without classification were assigned to a class by running the Basic Local Alignment Search Tool (BLAST) and searching for the most similar proteins. A curious case is the protein GPR157 (Uniprot ID Q5UAW9), which is most similar to class B, but its TM1 has a gap in the alignment to class B. However, the TM1 aligns well to the class A TM1, so this protein is a hybrid between these two classes.

We identified 11 *Other* proteins (they are listed in Appendix D) for which the BLAST search does not return any sequences with already assigned classes. To identify a good alignment for such individual sequences, one could search for similar sequences in different species, and then average over the sequence variation when trying to find a good alignment. However, this has to be done for each protein separately, so we skip it in our general analysis.

Finally, some of the candidate sequences were ruled to be pseudogenes, because they align well with some known GPCR, but miss one or more TM regions.

3.3 Extension of Alignment to All Known GPCR Sequences

The approximate positions of the TM regions are already annotated in the Uniprot database as predicted by the TMHMM program [65]. The prediction is quite noisy, and even for similar proteins which align well, it typically differs by 5-8 residues and sometimes even misclassifies a TM. However, the overall trend clearly determines the approximate TM location and allows us to judge the quality of the alignment of several proteins. If there are gaps in the TM regions, the alignment cannot be used to successfully create homology models.

First, we try to align all 828 sequences of the GPCR superfamily using a multiple sequence alignment program Clustal Omega [66]. However, the overall sequence conservation is very low, and so the resulting alignment has many large gaps, and even some highly conserved residues end up aligned incorrectly. In order to avoid this problem, we try to align only class A sequences (705, including olfactory), and again the resulting alignment has large gaps even in the TM region. It seems that the large variability of the loop region is what confuses the alignment algorithm.

Fortunately, sequences within individual subgroups can be aligned using Clustal Omega without large gaps in the TM regions. We take these individual subgroup alignments and fix them into a *profile*. We then align any two profiles to see how similar the two groups are.

A profile alignment of Aα to each of the other class A groups (Aβ, Aγ, Aδ) has no gaps in the TM regions, and also gives correct alignment of the BW .50 residues.

The multiple sequence alignment of the group A-other showed gaps in the TM regions for several proteins (Uniprot IDs: Q96P67, Q8TDU6, Q16570, Q86SM8, Q9NS66, Q9NS67, P60893, Q86SM5), so a separate profile is created for these sequences. Both profiles of the A-other proteins aligned to Aα without any gaps in the TM region. The profile of olfactory receptors aligned well to Aα as well. The Vomeronasal and Taste2 groups were more problematic, and are discussed in the following section.

The profile of the adhesion group aligns well to class B and class B is aligned to class A using the structural analysis, as discussed before. Aligning profiles of Aα and B does not yield meaningful alignment, because the TM regions are offset and there are many gaps in the TM regions. Similarly, aligning classes A and C or A and F does not yield meaningful alignment.

Once the alignment is fixed, the TM lengths for new proteins are taken as the average TM lengths from the available structures in the same class. For example, for TAS1 proteins the predicted TM length is the average TM lengths of GMR1 and GMR5; and for TAS2 the average TM lengths of 15-20 class A structures.

Appendix D has the complete listings of all the proteins and their alignment suitable for building homology models or starting sampling of the TM bundle.

			Offset										
Group 1	Group 2	TM	-5	-4	-3	-2	-1	0	1	2	3	4	5
Vomeronasal	A-alpha	TM1	0.21	0.20	0.23	0.26	0.29	0.33	0.23	0.22	0.24	0.25	0.23
Vomeronasal	A-alpha	TM2	0.21	0.24	0.20	0.27	0.26	0.35	0.23	0.22	0.27	0.23	0.27
Vomeronasal	A-alpha	TM3	0.22	0.20	0.21	0.21	0.16	0.36	0.16	0.20	0.21	0.14	0.17
Vomeronasal	A-alpha	TM4	0.23	0.23	0.23	0.25	0.25	0.33	0.22	0.23	0.21	0.27	0.19
Vomeronasal	A-alpha	TM5	0.20	0.24	0.20	0.22	0.28	0.25	0.21	0.28	0.24	0.21	0.19
Vomeronasal	A-beta	TM5	0.25	0.23	0.22	0.22	0.27	0.28	0.20	0.25	0.24	0.23	0.20
Vomeronasal	A-gamma	TM5	0.22	0.23	0.23	0.23	0.26	0.30	0.21	0.24	0.23	0.22	0.20
Vomeronasal	A-delta	TM5	0.21	0.23	0.23	0.23	0.25	0.30	0.22	0.22	0.23	0.23	0.18
Vomeronasal	B	TM5	0.25	0.26	0.23	0.22	0.25	0.31	0.29	0.26	0.24	0.27	0.25
Vomeronasal	C	TM5	0.25	0.25	0.22	0.24	0.24	0.27	0.26	0.24	0.26	0.16	0.15
Vomeronasal	F	TM5	0.22	0.25	0.24	0.28	0.17	0.24	0.27	0.28	0.25	0.23	0.22
Vomeronasal	A-alpha	TM6	0.24	0.27	0.28	0.23	0.26	0.39	0.25	0.21	0.26	0.23	0.21
Vomeronasal	A-alpha	TM7	0.20	0.21	0.19	0.22	0.24	0.31	0.19	0.21	0.20	0.23	0.16

Table 3.1: Testing the robustness of the alignment of the Vomeronasal receptors with the other groups. The table shows similarity between TMs averaged over all pairs of sequences formed from group 1 and group 2. Red denotes high similarity, blue low smilarity.

3.4 Bitter Taste and Vomeronasal Receptors

The profile of the vomeronasal group aligns better with class Aα compared to classes B and C, but there is still a gap of length 2 near the center of TM5. We remove the gap in such a way that the residue which aligned with 5.50 stays fixed. To check that this is indeed the best alignment we explore small changes in the alignment by shifting individual TM by -5 – +5 residues. In Table 3.1 we can see that for TMs 1 to 4, the current alignment gives the highest sequence similarity with Aα, so the alignment of these TMs is correct. However, for TM5, the alignment shifted by -1 or +2 residues gives higher similarity with Aα. Nevertheless, the similarity with groups Aβ, Aγ, Aδ, and B is the highest for the current alignment. We therefore keep the current choice.

We performed similar analysis for the Taste2 receptor, for which adjustments had to be made. The profile alignment of Taste2 with Aα has some gaps, but it is still the best alignment (i.e., it has the fewest gaps) compared to aligning to classes other than class A. TM3 has two gaps in the alignment: Gap of length 4 in the middle of TM3, and gap of length 5 at the DRY motive. As the first iteration we kept the alignment fixed on residue 3.50, then we computed the similarity to other groups for -5 – +5 residue shifts. The shift by +3 residues gives better similarity and so it was kept. See Table 3.2 for the computed similarities after the shift has been made. All class A subclasses favor this new choice, as the highest similarity has offset 0. Class B would favor shift by 2 residues, but the similarity is less than 30%.

TM4 has low sequence similarity, and in particular highly conserved Trp is not present in Taste2. Again as a starting point we keep the alignment at 4.50, but later had to adjust it by 4 residues. Table 3.2 shows the similarity after this shift has been made. For TM4 the similarity is only slightly higher at the new best offset than at nearby offsets.

Taste2 TM6 showed partially conserved motif IYFLS, with S being aligned with P6.50, which

			Offset										
Group 1	Group 2	TM	-5	-4	-3	-2	-1	0	1	2	3	4	5
Taste2	A-alpha	TM1	0.23	0.24	0.27	0.31	0.30	0.34	0.28	0.27	0.29	0.28	0.23
Taste2	A-beta	TM1	0.25	0.26	0.29	0.31	0.26	0.37	0.28	0.25	0.32	0.29	0.23
Taste2	A-gamma	TM1	0.25	0.25	0.31	0.28	0.27	0.39	0.28	0.25	0.31	0.29	0.23
Taste2	A-delta	TM1	0.24	0.24	0.28	0.25	0.27	0.36	0.26	0.24	0.30	0.27	0.25
Taste2	B	TM1	0.24	0.30	0.31	0.23	0.26	0.29	0.28	0.25	0.26	0.29	0.23
Taste2	C	TM1	0.26	0.22	0.25	0.25	0.27	0.36	0.27	0.26	0.29	0.26	0.27
Taste2	F	TM1	0.20	0.16	0.16	0.17	0.27	0.25	0.22	0.20	0.25	0.22	0.23
Taste2	A-alpha	TM2	0.27	0.23	0.23	0.28	0.25	0.39	0.27	0.27	0.22	0.27	0.27
Taste2	A-beta	TM2	0.30	0.23	0.26	0.27	0.23	0.38	0.27	0.28	0.22	0.26	0.29
Taste2	A-gamma	TM2	0.29	0.22	0.22	0.30	0.22	0.35	0.26	0.27	0.21	0.24	0.23
Taste2	A-delta	TM2	0.28	0.23	0.24	0.30	0.23	0.33	0.26	0.27	0.23	0.26	0.24
Taste2	B	TM2	0.23	0.23	0.14	0.22	0.23	0.33	0.23	0.22	0.27	0.25	0.29
Taste2	C	TM2	0.19	0.23	0.21	0.20	0.33	0.23	0.24	0.25	0.29	0.30	0.24
Taste2	F	TM2	0.20	0.19	0.18	0.23	0.22	0.31	0.25	0.28	0.20	0.25	0.27
Taste2	A-alpha	TM3	0.21	0.21	0.20	0.22	0.18	0.30	0.19	0.23	0.25	0.19	0.14
Taste2	A-beta	TM3	0.21	0.22	0.20	0.22	0.18	0.31	0.21	0.20	0.24	0.19	0.16
Taste2	A-gamma	TM3	0.20	0.23	0.21	0.21	0.20	0.36	0.22	0.22	0.25	0.22	0.18
Taste2	A-delta	TM3	0.19	0.23	0.20	0.21	0.17	0.33	0.22	0.21	0.24	0.20	0.17
Taste2	B	TM3	0.17	0.20	0.20	0.17	0.24	0.26	0.21	0.29	0.25	0.19	0.19
Taste2	C	TM3	0.16	0.22	0.17	0.24	0.24	0.18	0.17	0.23	0.22	0.23	0.20
Taste2	F	TM3	0.21	0.22	0.25	0.22	0.21	0.28	0.28	0.28	0.24	0.27	0.18
Taste2	A-alpha	TM4	0.30	0.31	0.30	0.29	0.28	0.33	0.28	0.28	0.23	0.26	0.24
Taste2	A-beta	TM4	0.32	0.28	0.28	0.27	0.28	0.34	0.29	0.29	0.26	0.30	0.25
Taste2	A-gamma	TM4	0.27	0.31	0.28	0.31	0.30	0.31	0.28	0.29	0.27	0.27	0.26
Taste2	A-delta	TM4	0.28	0.29	0.30	0.32	0.32	0.31	0.28	0.29	0.28	0.28	0.26
Taste2	B	TM4	0.20	0.21	0.27	0.25	0.27	0.29	0.26	0.25	0.20	0.23	0.22
Taste2	C	TM4	0.32	0.30	0.30	0.31	0.32	0.33	0.34	0.31	0.32	0.31	0.29
Taste2	F	TM4	0.14	0.26	0.25	0.26	0.20	0.23	0.26	0.27	0.23	0.26	0.29
Taste2	A-alpha	TM5	0.25	0.30	0.23	0.24	0.28	0.32	0.27	0.25	0.23	0.27	0.24
Taste2	A-beta	TM5	0.27	0.28	0.23	0.26	0.25	0.35	0.27	0.22	0.24	0.25	0.26
Taste2	A-gamma	TM5	0.26	0.29	0.23	0.24	0.28	0.37	0.26	0.23	0.25	0.25	0.25
Taste2	A-delta	TM5	0.23	0.28	0.24	0.24	0.27	0.37	0.26	0.24	0.24	0.26	0.25
Taste2	B	TM5	0.23	0.33	0.32	0.27	0.27	0.32	0.30	0.27	0.35	0.35	0.25
Taste2	C	TM5	0.28	0.28	0.25	0.25	0.25	0.23	0.27	0.20	0.23	0.26	0.21
Taste2	F	TM5	0.21	0.29	0.24	0.29	0.26	0.29	0.20	0.28	0.32	0.25	0.20
Taste2	A-alpha	TM6	0.23	0.25	0.29	0.23	0.22	0.33	0.24	0.24	0.29	0.34	0.24
Taste2	A-beta	TM6	0.24	0.31	0.29	0.25	0.25	0.35	0.23	0.24	0.28	0.33	0.27
Taste2	A-gamma	TM6	0.24	0.26	0.29	0.24	0.24	0.35	0.24	0.25	0.31	0.35	0.26
Taste2	A-delta	TM6	0.24	0.25	0.29	0.23	0.23	0.33	0.24	0.25	0.29	0.33	0.25
Taste2	B	TM6	0.25	0.21	0.31	0.29	0.24	0.37	0.31	0.22	0.30	0.32	0.22
Taste2	C	TM6	0.24	0.27	0.19	0.23	0.23	0.30	0.27	0.22	0.29	0.32	0.27
Taste2	F	TM6	0.17	0.30	0.34	0.22	0.27	0.32	0.27	0.20	0.23	0.23	0.24
Taste2	A-alpha	TM7	0.21	0.23	0.22	0.25	0.24	0.27	0.18	0.20	0.17	0.25	0.16
Taste2	A-beta	TM7	0.19	0.23	0.23	0.26	0.26	0.29	0.20	0.23	0.17	0.23	0.16
Taste2	A-gamma	TM7	0.18	0.24	0.22	0.26	0.24	0.29	0.18	0.21	0.14	0.25	0.14
Taste2	A-delta	TM7	0.20	0.26	0.23	0.26	0.22	0.26	0.18	0.20	0.19	0.24	0.16
Taste2	B	TM7	0.22	0.24	0.22	0.22	0.27	0.22	0.17	0.20	0.20	0.19	0.15
Taste2	C	TM7	0.25	0.26	0.24	0.25	0.26	0.22	0.24	0.22	0.22	0.20	0.21
Taste2	F	TM7	0.23	0.18	0.20	0.23	0.24	0.26	0.29	0.23	0.22	0.20	0.19

Table 3.2: Testing the robustness of the alignment of the Taste2 receptors with the other groups. The table shows similarity between TMs averaged over all pairs of sequences formed from group 1 and group 2. Red denotes high similarity, blue low smilarity.

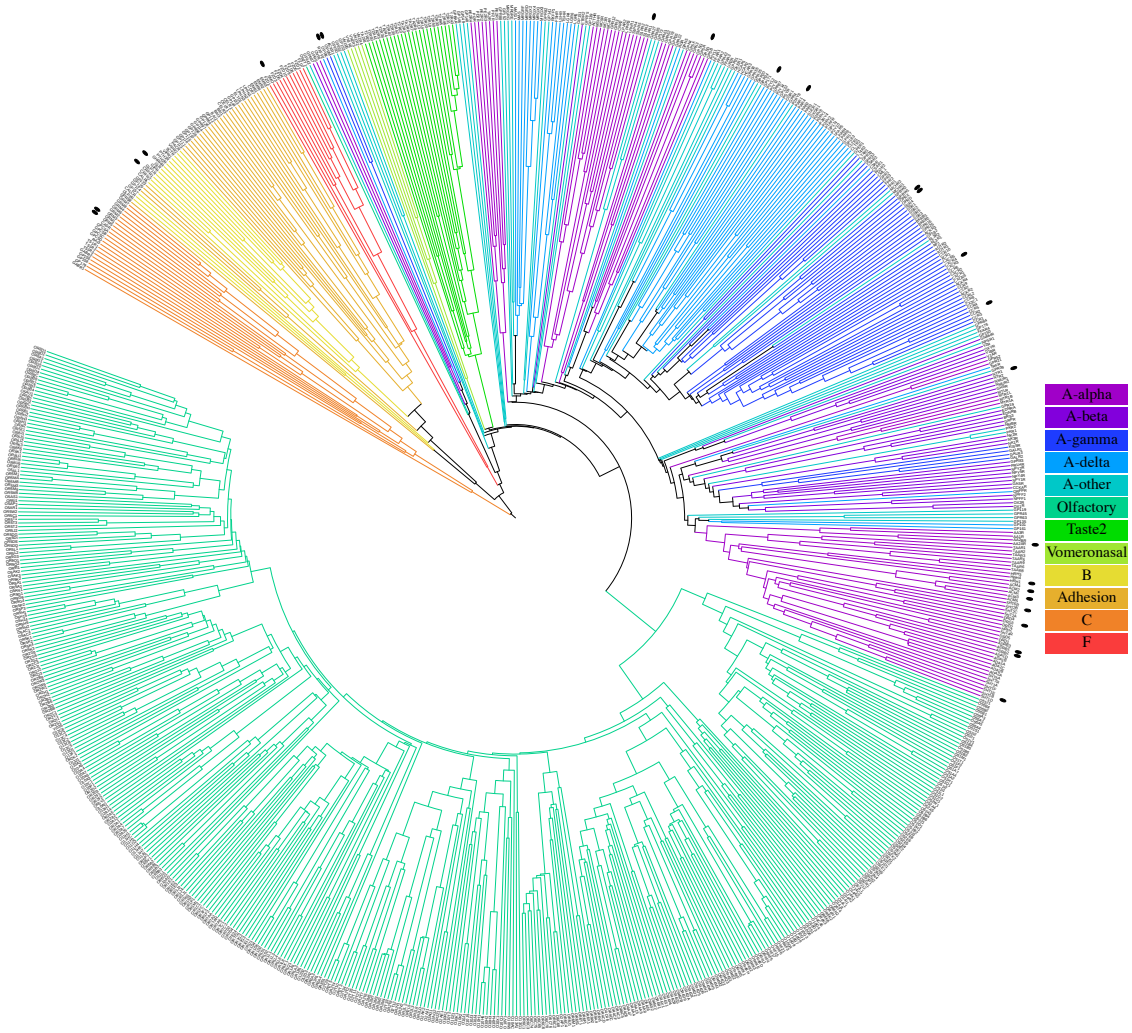


Figure 3.2: Phylogenetic tree based on TM similarity only (loops were ignored). Color coding denotes the GPCR class. Proteins with known crystal structure are emphasized with a dot. The tree was visualized using the Iterative Tree of Live toolkit [67].

we kept as an initial try. This choice is kept in Table 3.2. However, we see that an offset of +4 residues, which correspond to one turn shift (the motif IYFLS aligns Ile with P6.50), also gives high similarity. Based solely on sequence similarity we cannot distinguish which alignment is better, and therefore both cases should be considered when building homology models.

3.5 Phylogenetic Tree

Based on the constructed alignment, we can compute the similarity for each of the two sequences using the weights from the BLOSUM62 matrix, and the similarities can be used as a distance metric to cluster the proteins. We used the unweighted pair-group clustering algorithm (implemented in Jalview [68]), which iteratively extends clusters by finding a non-member sequence with the lowest

average dissimilarity over the cluster members. Figure 3.2 shows the phylogenetic tree constructed by this clustering algorithm.

The branches near the root of the tree are very sensible: First class C separates, then class B and adhesion proteins branch off, and then class F. The rest of the tree contains class A-like sequences. Except for several outliers, the first major branches to separate are the sensory receptors: Vomeronasal, Taste2, and Olfactory. In the olfactory branch, the first split separates the fish-like receptors (families 51-56) from the tetrapod-like receptors (families 1-13). The subdivision of the rest of the class A family does not exactly follow the $\alpha - \delta$ subclasses, but it is close.

Near the leaves (i.e., for closely related proteins), the displayed tree might not be the best classification, since the computation of similarity ignored loops. For related proteins, it is advantageous to align loops as well, since loops often interact with ligands, and therefore can determine receptor specificity. However, for very dissimilar proteins present in the GPCR super-family, aligning the loops does not provide any useful information.

3.6 Loop Alignment

In such a large protein family loops can be very diverse, and so a meaningful alignment of the loop regions could only be done for smaller subgroups. The intracellular side of the GPCRs interacts with the G-proteins, and thus we expect the intracellular loops to be similar among the proteins. Extracellular loops adapt each protein to its particular function and thus are more variable.

In many cases there are several disulfide bonds in the extracellular loops keeping the loops in a compact conformation. If we can find a conserved disulfide link, it significantly limits the possible conformations of the loops. In particular, there is a highly conserved disulfide bond between TM3 and loop EC2. This bond is important even for folding the protein in the membrane. For example, in HCAR1 [64] mutating cysteines of this conserved bond prevent the protein from being assembled in the membrane. However, mutating other cysteines in the loop regions, which are probably involved in separate disulfide bonds, only interrupts the protein sensitivity to the ligand. The constraints coming from conserved or experimentally observed disulfide bonds should be used in constructing the loops.

In fact, the longer loops are expected to be quite flexible at body temperature, as they are often not resolved in the crystal structures. Furthermore, one can expect the loop conformations to be influenced by crystal packing. We compared how loops differ between several different crystal packings for proteins for which crystal structures in multiple crystal packing are available. In the case of A2A, we indeed observed two different conformations of the EC2 loop in different crystal packings. However, all loops of β_2AR and RHO seem to have loops of very similar conformations in the different crystal structures.

3.7 Size of Helix Movements

By careful analysis of the inter-helical contacts we constructed a unique alignment between all the GPCR protein, from which new homology models can be derived. For structure prediction we would like to know how far the homology models are from the target structure. In this section we compare the variability of the TM bundles, when we treat the individual TM helices as rigid bodies. In particular we are interested in how large moves should be sampled, in order to effectively model the target protein starting from the homology model.

The Figure 3.3 shows the observed move sizes. Each pair of known structures was first aligned together, then each helix of the first protein was individually aligned to the corresponding helix of the second protein and the size of the move was measured. The center of mass translation was broken down into the direction along the helical axis and a direction perpendicular to it. The 'tilt of axis' measures how much axis 1 had to be rotated to axis 2. And finally the 'rotation around axis' measures the necessary rotation around the axis to map the corresponding atoms to each other.

From Figure 3.3 we see that the maximal move sizes, which need to be considered, get smaller as the similarity of the TM sequence increases. If we are predicting a structure starting from a homology model with higher than 50% similarity, then we only need to consider translating the helices up to 1.5 Å in any direction, tilting them up to 10, and rotating around their axis by 40. This is a very useful bound for refining homology models.

The same comparison can be applied to one protein in multiple conformations. The red points in Fig. 3.3 show the magnitude of rigid body moves underwent during activation for the 3 available pairs of active-inactive structures. Activation involves mainly the movement of TMs 5, 6, and 7. The computation of the move sizes ignores the bending of TM6 during activation, so it should be understood as an approximate description only.

3.8 Conclusion

We constructed the sequence alignment of the transmembrane regions for most known human GPCRs, which is a promising starting point for structure prediction. For a protein in question one can build homology models based on any of the available templates by mutating the corresponding amino acids. One of the most visible structural errors in the predicted structures comes from misalignment of prolines, since prolines often cause kinks in the long helical TM regions. Such errors can be avoided by choosing a template with the least misaligned prolines. However, most of the differences among the proteins cannot be resolved only by using a suitable template, and one needs to refine the structure of the initial homology models. GPCRs are too large (> 300 residues) for exploring larger conformational changes using molecular dynamics. In the next chapter, we

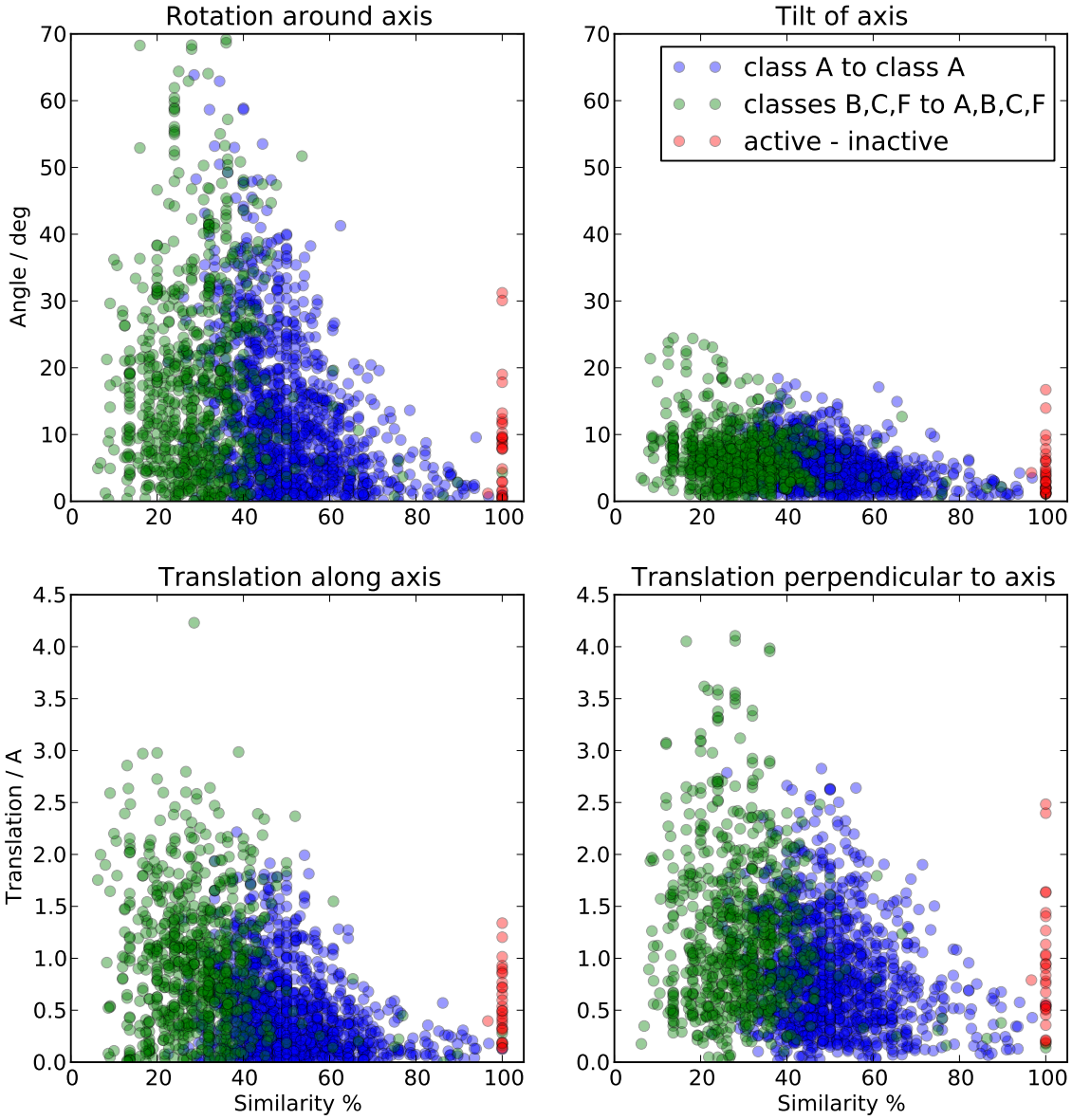


Figure 3.3: Magnitude of the rigid body moves of the helices necessary to map one structure to another. All TMs 1-7 from all available structure pairs were compared. The coordinates systems is defined in the text. The maximal observed deviation is approximately proportional to the sequence similarity of the two compared TMs, and it follows the same trend within class A (blue points) and across the GPCF superfamily (green points). The red points, which correspond to the active-inactive structure pairs, show rigid body moves caused by receptor activation.

explore an alternative sampling scheme which refines the initial homology models.

Chapter 4

Trihelix Sampling Method

4.1 GPCR Structure Prediction

We begin this chapter by summarizing possible approaches to protein folding. As discussed in the introduction, the number of possible protein conformations scales exponentially with the length of the proteins. Small proteins with about 50 residues can sometimes be folded using molecular dynamics (MD) by directly simulating the protein folding process in time. Some fast folding proteins fold on microsecond timescale, but most natural proteins fold at millisecond timescale or longer. For example, the folding time for ubiquitin, which has 76 residues, is about 3ms. Piana et al. [1] simulated in total 8 ms of ubiquitin MD and observed 2 folding and 8 unfolding events. Reaching millisecond timescale requires state of the art supercomputers, and for large proteins we would likely need to simulate even longer MD. Another timescale that comes into play for large proteins is the rate of protein synthesis, which is only about 10 residues per second. Thus the N-terminus has several more seconds to assemble, while the protein is still being synthesized. This is relevant for GPCRs because they are inserted into the membrane helix by helix: first TM1, then TM2 and 3, then TM4 and TM5, and finally TM6 and TM7. It is possible that the ensemble of helices 1-5 creates intermediate states, which are important to the later formation of the whole 7 TM bundle. Simulating the complete membrane insertion would requite 10 s or longer of MD, which is out of the reach of current computers.

Instead of simulating the physical dynamics directly, one can take shortcuts. Larger proteins with about 100 residues can often be solved with the Monte Carlo method combined with fragment assembly or other sampling techniques [2]. The most successful methods rely on assembly of fragments, which are shapes of short amino chains extracted from the PDB crystal structure database. Since the number of solved structures in the database is increasing — it is almost 100,000 currently — the library of possible motives is more and more comprehensive. Unfortunately, there are only a few solved membrane structures available, and therefore only few motives specific to membrane proteins have been observed. A particularly successful Monte Carlo structure prediction toolkit

is Rosetta [69]. Besides the amino acids fragments, Rosetta also collects distributions of atomic distances and torsion angles from the PDB database, which are then converted into energy terms. The potentials cannot be well calibrated to the protein-membrane interface, however, because there are only few membrane proteins in the PDB database. The Rosetta community attempted to implement de novo membrane protein structure prediction with a program called Membrane AbInitio [3, 70, 17], but the effort was discontinued. The lack of known structures makes predicting new ones more challenging.

In the next section we discuss possible approaches to protein structure refinement, and in the rest of this chapter we discuss the Bihelix and Trihelix sampling methods, which use several approximations to simplify the GPCR structure prediction problem.

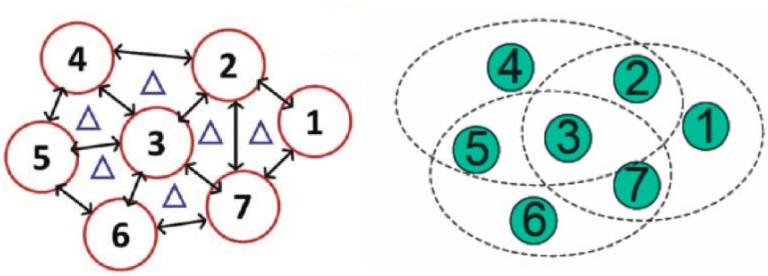
4.2 Structure Refinement

To solve the structure of a large protein, it is best to start with a structure of a related protein, and then try to refine it with MD or other sampling methods. Indeed, the current most successful GPCR structure prediction methods are based on homology modeling [5]. As discussed in Section 2.3, the crystal structures solved since 2008 are very similar to each other in the TM region. The TM RMSD is ∼2.5 Å - 3.5 Å. Thus homology models built with the correct sequence alignment should have TM RMSD in a similar range. Nevertheless, the refinement of the initial model below 2 Å remains challenging. First, let us consider the simulation timescales that would guarantee refinement with MD.

There are many MD simulations of the β_2 receptor which can help us estimate how long it takes to change the conformation. Rosenbaum et al. [15] performed 30μs long MD of the β_2 active state. At about 11μs from the start, the protein went from the active state to the inactive one. More recently in 2014, Kohlhoff et al. [71] performed an exhaustive exploration of the activations pathways of β_2 using large Google computational resources. The cumulative time of their simulations was 2ms, which can serve as an estimate of the timescale necessary to find conformational changes in a GPCR. Refining a homology template using MD would likely take an order of magnitude more time.

A more successful refinement protocol is the Fast Relax algorithm [72] from the Rosetta macromolecular modeling suite. The algorithm cycles through side chain repacking and minimization, while gradually increasing the Van der Waals repulsion until its full value. This algorithm is typically used on soluble proteins up to 100 residues in length. Since the typical GPCR protein with loops (∼350 residues) is too large, and the fold of the loops is often unknown, we try this approach to refine the TM bundle without the loops (∼200 residues).

The Fast Relax protocol samples large changes in the side chain packing by decreasing the Van der Waals repulsion. After several iterations the low energy structures start to have large RMSD



(a) Bihelix

(b) Quadhelix

Figure 4.1: Separation of the 7 TM bundle into interacting parts for Bihelix. Images from [6].

relative to the X-ray reference. The missing loops, which would normally constraint the positions of the helices, allow separation of the helices as the Van der Waals repulsion is returned to its full value. In order to avoid this problem, we add constraints to the starting helix positions. The constraints do not help in discriminating the correct structure because either the constraints are too weak to affect the energy, or they are too strong and do not allow any significant movement.

Fast Relax introduces dramatic changes to the protein structure and in order for the minimization to not disassemble the protein, it is important that there is only one amino acid chain, which cannot break. Furthermore, for small proteins the strong hydrophobic forces keep the protein from disassembling. If we do not allow steps which disassemble the TM bundle, then the loops can be ignored. In the next section, we consider a very different approach where only a small subset of coordinates is sampled.

4.3 Complete Sampling Approach

Before the year 2008, the only solved GPCR structure was rhodopsin. At that time, the structural similarities of rhodopsin to other GPCR proteins could only be inferred from sequence comparisons. Because the sequence similarity of proteins in the GPCR superfamily was very low, it was not clear how large the differences between different GPCR structures were.

In order to refine homology models, Abrol et al. [73, 74] developed the Bihelix sampling method. This method ignores loops of the helical bundle, and only samples the rotation of each helix around its axis in 30° increments. The total number of conformations is too large to build directly, so a mean field approach is used to approximate the energy of the seven helix bundle from pairwise interactions energies between neighboring helices. Fig. 4.1a shows the 12 helical interactions in the seven helix bundle. For each pair of interacting helices, the backbone atoms are first rotated around the helical axes to the desired angles θ_i, θ_j; side-chains are repacked using a side-chain packing algorithm; then the structure is minimized in the desired force field; and finally the energy of the helix pair is split

Sidechain placement	Scream		Scwrl		Rosetta	
Force field	Dreiding		Amber		Rosetta	
	Bihelix	Combihelix	Bihelix	Combihelix	Bihelix	Combihelix
A2A	1	1	168	1	1	1
β_2	2	1	—	1	1	1
PAR1	1	1	2	1	1	1
P2Y12	2	1	133	1	1	1
RHO	1	4	1	2	1	1

Table 4.1: Test if the Bihelix method can select the correct structure when starting to sample from a crystal structure. The table shows the rank of the all zero structure after the Bihelix stage and after the Combihelix stage; 1 means that the all zero structure has the lowest energy, and so was correctly identified. The energies were evaluated using different combinations of side-chain placement algorithms and force fields: Scream [75], Scwrl [76], Dreiding [77], Amber [78], Rosetta [69].

into the energy of the i-th helix E_i^{ij}, energy of the j-th helix E_j^{ij}, and their interaction energy E_{i-j}^{ij}. The energy of the whole bundle is then approximated as:

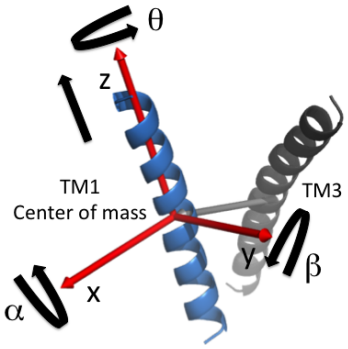
$$\begin{aligned} E(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7) =& E_{1-2}^{12}(\theta_1,\theta_2) + \frac{1}{2}E_1^{12}(\theta_1,\theta_2) + \frac{1}{4}E_2^{12}(\theta_1,\theta_2) + \\ &+ E_{1-7}^{17}(\theta_1,\theta_7) + \frac{1}{2}E_1^{17}(\theta_1,\theta_7) + \frac{1}{4}E_7^{17}(\theta_1,\theta_7) + \\ &+ E_{2-3}^{23}(\theta_2,\theta_3) + \frac{1}{4}E_2^{23}(\theta_2,\theta_3) + \frac{1}{5}E_3^{23}(\theta_2,\theta_3) + \\ &+ \dots \text{ (9 other pairwise helix interactions).} \end{aligned} \tag{4.1}$$

In this manner the mean field energy for all $12^7 = 35 \cdot 10^6$ conformations is computed from only $12 \cdot 12^2 = 1728$ constructed structures.

To test whether the mean field energy can distinguish the native conformation, we start the Bihelix sampling for several proteins with known structure. The reference structure is the native conformation, so the desired solution has all θ_i angles 0. Table 4.1 shows the performance of the Bihelix sampling with several side-chain packing algorithms and several different force fields. For example, the Scream side-chain packing together with the Dreiding force field is able to identify the native structure as the first or second lowest energy from all $35 \cdot 10^6$ conformations.

After the approximate Bihelix stage, we select the 2000 lowest energy conformations for a more accurate energy evaluation: we explicitly build the bundles, repack the side-chains, and evaluate the bundle energies. The rank of the all zero conformation is shown in the Cobihelix column of the Table 4.1. We see that Combihelix recovers the native structure in all cases except for rhodopsin. The selected rhodopsin crystal structure has a covalently bound molecule retinal, so the rhodopsin structure without the retinal molecule might be different. In summary, the two stage energy evaluation (Bihelix and Combihelix) is able to resolve the native structure with the used force fields.

We described this algorithm in detail because it is very uncommon in the protein modeling



(a) Coordinate system for each helix.

θ	-15°	0°	+15°
α, β	-7°	0°	+7°
z	-2 Å	0 Å	+2 Å

(b) Step sizes in the present Trihelix runs.

Figure 4.2

community. Bihelix performs complete sampling on a small set of discretized coordinates. The complete sampling is done in several stages: first, approximate energy is used for all conformations, and second, a more accurate energy is evaluated for a smaller set of structures. In order for this scheme to work, the coordinate grid has to be large enough, or the energy function will not be able to distinguish the correct structure as the lowest energy structure. On the other hand, if the grid is too coarse, it is likely that the correct structure is the trivial all zero conformation. The Bihelix sampling was a novel way to study GPCR proteins when only the rhodopsin structure was available. At present good sequence alignments and several GPCR structures are commonplace, thus the initial homology models are better quality than the resolution of the Bihelix sampling.

In later development, Bray [6] refined the Bihelix sampling to a finer grid of points and also allowed for tilts of the helices. The larger computational cost was resolved by adding an intermediate mean field approximation, where only the low scoring conformations for each 4 helix bundle are kept (see Quadhelix diagram in Fig. 4.1b). This modified approach selects near native states, when the helix shapes from the target protein are provided. However, this approach did not manage to refine a homology model without prior knowledge of the helix shapes of the target protein. Thus, as far as structure prediction, the original homology model is better than the output of the modified Bihelix sampling.

In the next section we describe our algorithm for GPCR structure refinement, which uses the same ideas of complete sampling on finite grid, but samples more degrees of freedom with more relevant step sizes.

4.4 Trihelix Sampling

A detailed analysis of the GPCR crystal structures suggests that the rigid body manipulations of the individual helices are suitable degrees of freedom, if one considers translations as well (Figure

3.3). Thus, in addition to the 3 rotational degrees of freedom for each helix, we add translations of the helix center of mass. To limit the computational requirements, we only consider the translations along the helical axis, which makes 4 degrees of freedom per helix. The coordinate system for each helix is shown in Fig. 4.2. Sampling 3 points per degree of freedom gives $3^4 = 81$ conformations per helix, and $81^7 = 2.3 \times 10^{13}$ conformations for the whole bundle.

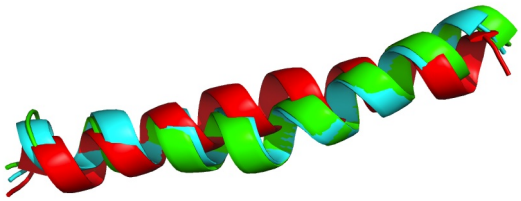
Bihelix stage. First, we explicitly build models for 12 helix pairs for all $81 * 81$ conformations (78,732 in total) and evaluate their energies. Based on these energies, we use the mean field approximation as in Equation 4.1 to compute the energy of all the conformations of the four helical bundle TM2-3-6-7. We keep the lowest lying 10^6 conformations and compute the mean field energy of all five helical bundles TM2-3-5-6-7 for each TM5 conformation. We again keep the lowest lying 10^6 conformations and add all conformations of TM4 to form bundle of 6 helices TM2-3-4-5-6-7. Finally, for the lowest lying 10^6 conformations we add all conformations of TM1, to form the complete seven helix bundle.

Then, starting from the lowest energy bundle conformations, we pick 20 conformations for each helix.

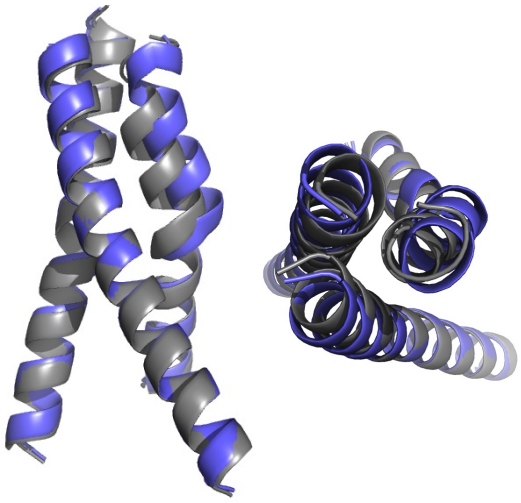
Trihelix stage. The side-chain placement, when only two helices are present, is often not accurate, and so to obtain better side-chain placements within a reasonable time, we build all bundles of the interacting three helices. With the limited set of 20 conformations per helix, we explicitly build all possible helical triplets ($6 * 20^3 = 48000$ conformations). The selected conformations allow 20^7 different bundles. We evaluate the full bundle energies based on the mean field approximation from the triplet energies (Eq. 4.2), and keep the 2000 lowest energy conformations for the full bundle stage.

$$
\begin{aligned}
E(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7) = & E^{127}_{1-2}(\theta_1,\theta_2,\theta_7) + E^{127}_{1-7}(\theta_1,\theta_2,\theta_7) + \frac{1}{2}E^{127}_{2-7}(\theta_1,\theta_2,\theta_7) \\
& + E^{127}_{1}(\theta_1,\theta_2,\theta_7) + \frac{1}{3}E^{127}_{2}(\theta_1,\theta_2,\theta_7) + \frac{1}{3}E^{127}_{7}(\theta_1,\theta_2,\theta_7) \\
& + \frac{1}{2}E^{237}_{2-3}(\theta_2,\theta_3,\theta_7) + \frac{1}{2}E^{237}_{2-7}(\theta_2,\theta_3,\theta_7) + \frac{1}{2}E^{237}_{3-7}(\theta_2,\theta_3,\theta_7) \\
& + \frac{1}{3}E^{237}_{2}(\theta_2,\theta_3,\theta_7) + \frac{1}{5}E^{237}_{3}(\theta_2,\theta_3,\theta_7) + \frac{1}{3}E^{237}_{7}(\theta_2,\theta_3,\theta_7) \\
& + \ldots \text{ (4 other trihelix bundles)}
\end{aligned}
\tag{4.2}
$$

Full bundle stage. We build the 2000 structures explicitly and evaluate their energies. These structures give the spectrum of low lying conformations and the lowest energy structure is the predicted model.



(a) TM1 of the PAR1 crystal structure (green) serves as reference helix with a proline kink. The homology model of TM1 based on β_2 template (red) with no proline kink minimizes to the blue structure with an almost exact proline kink.



(b) Clash in a trihelix bundle (grey) is resolved by change in backbone torsion (blue), not by moving the helices apart.

Figure 4.3: Examples of the efficiency of torsional minimization.

4.4.1 Implementation

This protocol uses rigid body manipulations, sidechain placement, and minimization. We use PyRosetta [79, 80] as a convenient way to perform all the steps in memory. We use the Talaris 2013 scoring function, and set the weight of the Lazaridis-Karplus solvation energy to 0, since most of the TM bundle is exposed to the lipid membrane and not water. Energy is then reported in REU (Rosetta energy units), which approximately equal kcal/mol.

Energy for each conformation is evaluated in the following way: first, the backbone atoms of each helix are moved as rigid bodies to the desired conformation using numpy [81]; then, the sidechains are repacked using Rosetta's default repack algorithm [82]; after that, sidechains are minimized using the Davidon-Fletcher-Powell minimizer ('dfpmin_armijo_nonmonotone'); finally, the whole structure is minimized using the same minimizer. Before the final minimization, we add harmonic constraints on the CA atoms to their positions after the rigid body move.

Since the Rosetta's sidechain repack algorithm is a Monte Carlo annealing with random outcomes, we repeat this procedure 3 times, and keep the lowest achieved energy.

4.4.2 Torsional Minimization

The Bihelix method has difficulty getting the helix shape correct. Particularly problematic are helical kinks and irregular turns near proline residues, because Cartesian minimization, which is used in Bihelix, resolves the clashes locally. We use torsional coordinates, where the atom positions are determined by torsion angles relatively to the bonded atoms. The minimizer implemented in

the torsional coordinates can resolve some local clashes with moves that have a global effect. Figure 4.3a shows a successful minimization of a helix with one proline kink. Furthermore, when using the torsional minimizer with the nearby helices present, the helix shapes mold to fit their local environment (example shown in Fig. 4.3b).

In the absence of loops, the torsional minimizer sometimes makes large moves, separating the helices and bending them away from each other. To supplement the role of loops and to tie the bundle together, we add harmonic constraints to the CA atoms. The constraints keep balance between helix flexibility and staying on the grid of the coordinates we are using to describe the helix orientation. On the one hand, if the constraints are too high, the helix shapes are not allowed to adjust. On the other hand, if the constraints are too low, the moves during minimization can be larger than the coordinate grid size.

4.5 Testing the Trihelix Protocol

In the following section, we present a series of simulations designed to characterize the Trihelix sampling method. In particular, we explore the coordinate grid size and the validity of the energy model. The sampling is complete in the sense that all coordinates are considered at the Bihelix level, but some low energy structures might still be missed if the coordinates are removed in the Bihelix stage. We study in detail two cases: human adenosine receptor A2a (A2A) and human P2Y purinoreceptor 12 (P2Y12). The A2A receptor is modeled by refining a homology template based on another class Aα receptor, β_2, which has TM sequence similarity 51% (identity 36%); the P2Y12 receptor is refined from the PAR1 receptor, which has TM sequence similarity 49% (identity 25%). P2Y12 and PAR1 are the only receptors with known structure in the class Aδ, which is the class of interest in the next chapter because it has the HCAR1 receptor in it.

The suitable translation and rotation magnitudes for refining homology models with 50% similarity were analyzed in the previous chapter and can be found in Fig. 3.3. The ideal rigid body moves, which optimally map the helices of the homology model to the crystal structure, involve all three rotations and all three translations. In order to limit the computational cost we do not sample helix translations in the membrane plane, but only the translation along the helical axis. The selected move sizes are shown in Fig. 4.2.

The sampling of 3 steps for 4 degrees of freedom for all 7 helices takes approximately 200 CPU hours at the Bihelix stage, 300 CPU hours at the Trihelix stage, and 200 CPU hours at the full bundle level, in total about 700 CPU hours per protein. The computation is composed of evaluating energies of many independent structures, thus scales to any number of processors.

Next we present an analysis of 4 types of simulations, gradually working towards the prediction of one protein from another. First, to identify the resolution of the energy function, we try to refine the

crystal structure itself. Next, we run two intermediate stages by using some degree of information from the correct crystal structure. Finally, we perform full prediction as if the modeled protein was currently unknown.

The many RMSD numbers in the following analysis should be compared to the RMSD of the initial homology models for A2A and P2Y12: **1.85 Å** and **2.23 Å**. We try to refine the homology models, i.e., the goal is to decrease the RMSD.

4.5.1 Recovery of the Crystal Structure

At first, we start the Trihelix sampling from the A2A crystal structure itself. Figure 4.4 shows that the starting structure is correctly identified as the lowest energy structure. We call the starting structure *all-zero*, since all the coordinates are 0. The gap to the second lowest energy state is 1.2 REU. The scatter plot in Fig. 4.4 is similar to the energy *funnel* plot often seen in Monte Carlo protein folding, but it is not as dense at the bottom, because we sample on discrete grid of coordinates. Here the funnel means that only small errors in structure are made if the energy does not distinguish the correct lowest energy state.

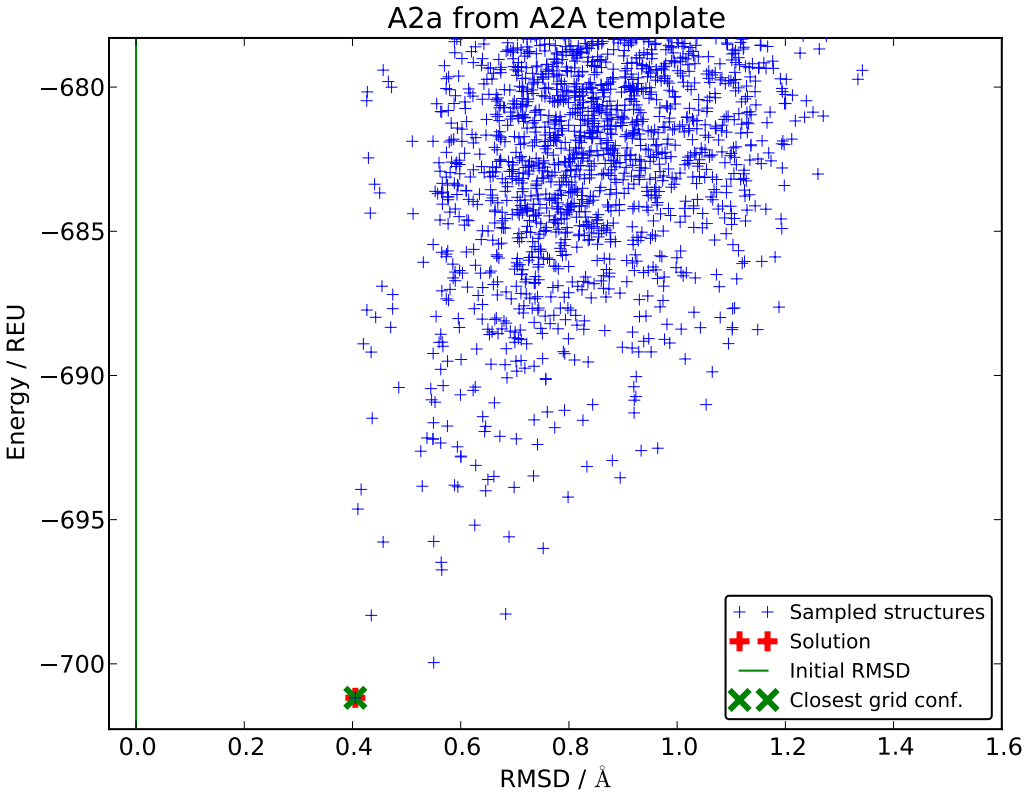
In the case of A2A reconstruction, the method finds the correct lowest energy structure. This is not a trivial result, since this structure had to pass through both the Bihelix and Trihelix approximations.

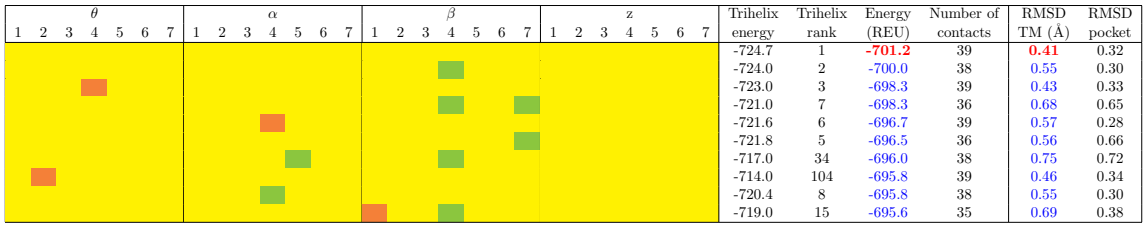
Fig. 4.5 shows an analogous simulation for P1Y12, where the Trihelix sampling starts from the correct P2Y12 crystal structure. In this case, the lowest energy structure has RMSD 0.77 Å, but the all-zero structure has the 8th lowest energy and RMSD 0.42 Å (after minimization). Nevertheless, RMSD below 1Å is still very low. These two runs suggest that the energy model can be used to identify structures similar to the crystal structure. In fact, for these two examples, the Trihelix mean field energy identifies the same lowest energy structure as the full bundle energy.

4.5.2 Sampling with the Correct Helix Shape

In the previous subsection we have shown that if the helices are at the right positions and if they have the right shape, the energy model can distinguish the native structure or find a structure very close to the native one. Below, we separately analyze the importance of the correct helical shape and of the correct helical positions. First, we run the Trihelix sampling on a hybrid template that is formed by the correct A2A helices that are aligned to the orientation of helices in β_2.

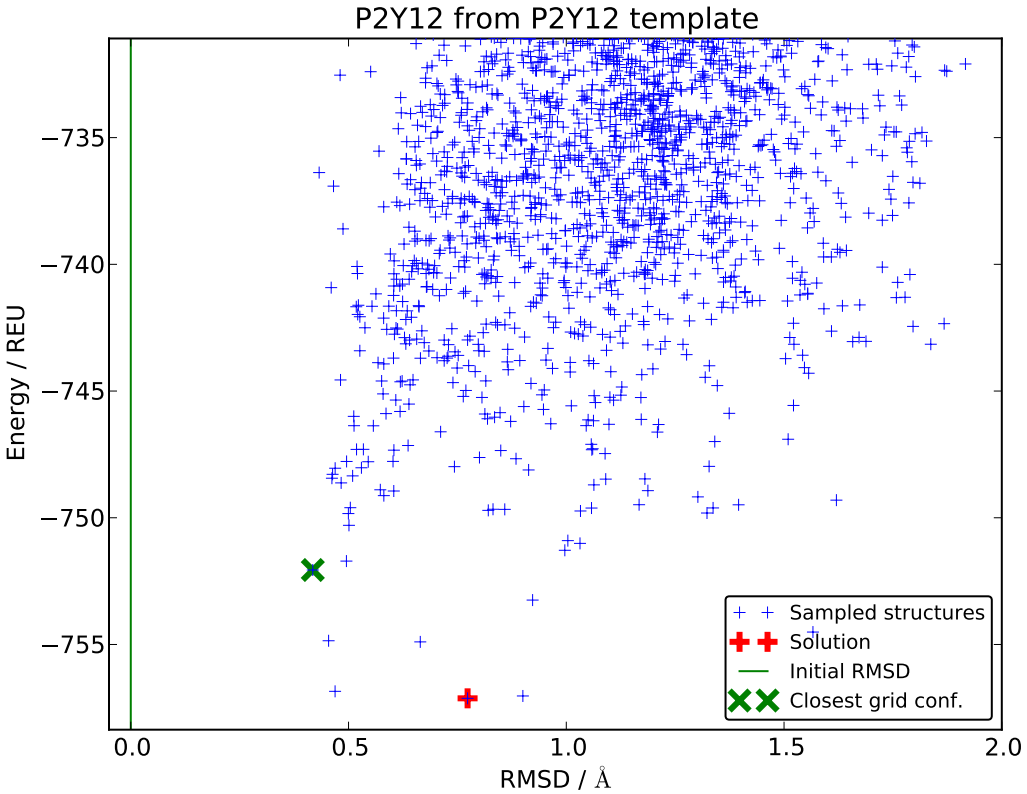
The starting hybrid template has RMSD 1.49 Å (from the A2A crystal structure). Figure 4.6 shows that the energy does not select the structure with the lowest RMSD from the sampled set, but finds a structure with RMSD 1.52 Å, which is a similar RMSD to the starting one. This number should not be compared to 0, but to the best possible RMSD given the used restricted set





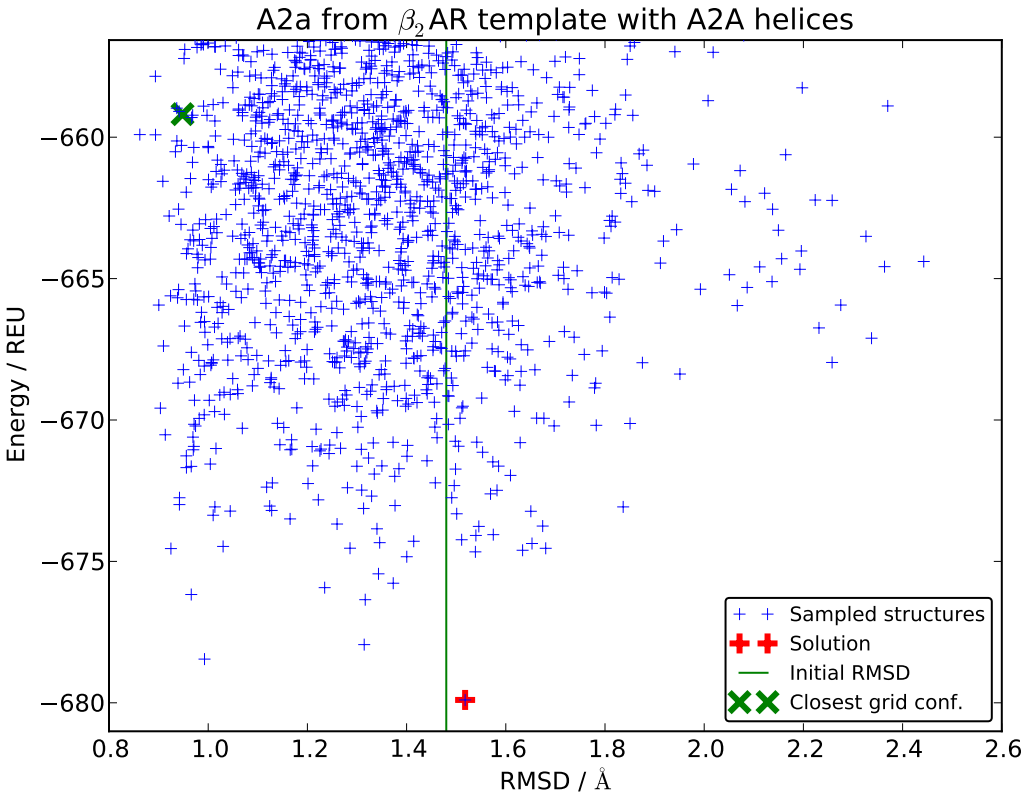
Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
-724.7	1	**-701.2**	39	**0.41**	0.32
-724.0	2	-700.0	38	0.55	0.30
-723.0	3	-698.3	39	0.43	0.33
-721.0	7	-698.3	36	0.68	0.65
-721.6	6	-696.7	39	0.57	0.28
-721.8	5	-696.5	36	0.56	0.66
-717.0	34	-696.0	38	0.75	0.72
-714.0	104	-695.8	39	0.46	0.34
-720.4	8	-695.8	38	0.55	0.30
-719.0	15	-695.6	35	0.69	0.38

Figure 4.4: *Trihelix results for the A2A crystal structure when starting from itself.* **Scatter plot:** The blue points denote 2600 structures from the Trihelix run. The energy is in REU (Rosetta Energy Units, approximately kcal/mol) and the RMSD is after structure minimization. The red + highlights the lowest energy structure. The vertical line shows the RMSD of the initial homology model (before minimization, here at 0). The green × shows the all-zero structure (after minimization, so its RMSD is > 0). **Table:** The table lists the 10 lowest energy conformations. The colors denote values of the sampled coordinates according to the color key in Fig. 4.2. The 'number of contacts' column is the number of class A common contacts present in this structure (the conserved contacts are listed in Fig. 2.8). The last two columns show RMSD to the correct crystal structure. RMSD of the binding pocket is discussed in Section 4.5.5.



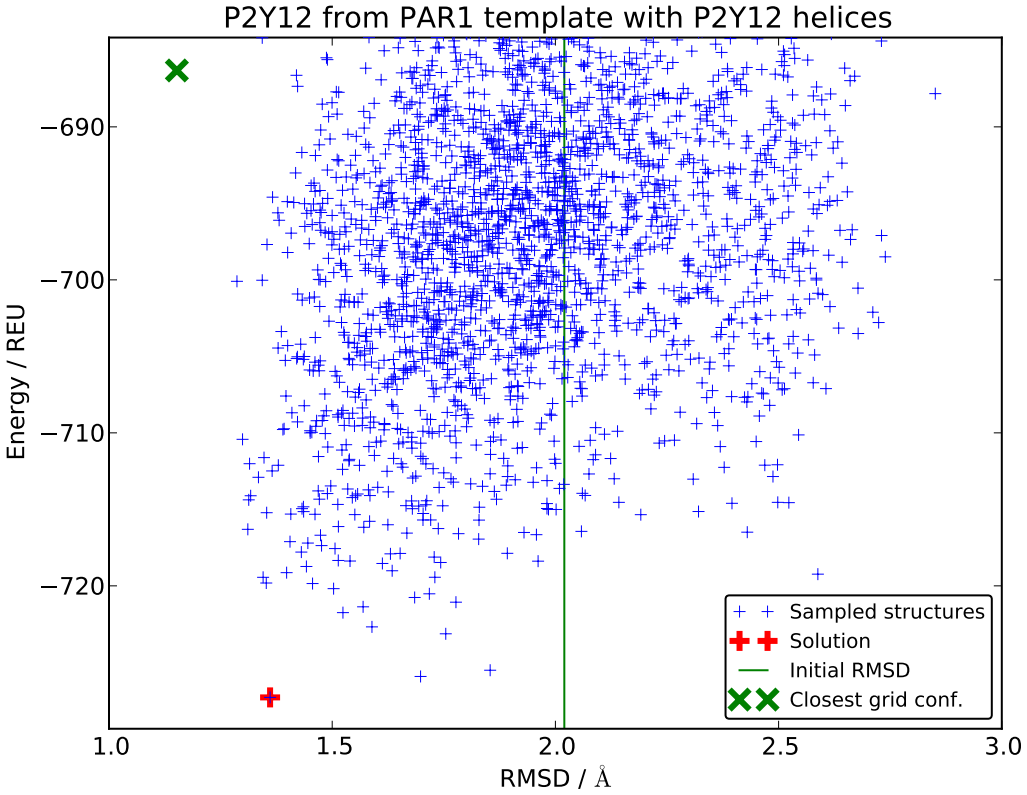
θ 1 2 3 4 5 6 7	α 1 2 3 4 5 6 7	β 1 2 3 4 5 6 7	z 1 2 3 4 5 6 7	Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
				-801.2	1	**-757.1**	39	**0.77**	0.64
				-791.7	35	-757.0	39	0.90	0.82
				-790.4	59	-756.8	39	0.47	0.42
				-792.8	26	-754.9	39	0.66	0.53
				-793.5	18	-754.9	39	0.45	0.35
				-793.6	17	-754.5	38	1.57	1.69
				-789.6	75	-753.2	38	0.92	0.89
				-790.9	51	-752.1	39	0.42	0.37
				-789.1	92	-751.7	39	0.50	0.42
				-791.2	43	-751.3	38	1.00	0.96

Figure 4.5: *Trihelix results for the P2Y12 crystal structure when starting from itself.* Same legend as for Fig. 4.4. The all-zero conformation has the 8th lowest energy.



θ 1 2 3 4 5 6 7	α 1 2 3 4 5 6 7	β 1 2 3 4 5 6 7	z 1 2 3 4 5 6 7	Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
				-697.9	24	**-679.9**	37	**1.52**	1.31
				-701.2	3	-678.5	39	0.99	0.72
				-701.5	2	-677.9	38	1.31	0.89
				-695.3	111	-676.4	37	1.32	1.08
				-699.0	12	-676.2	39	0.97	0.85
				-697.4	28	-675.9	38	1.23	0.83
				-691.9	578	-675.8	38	1.37	1.38
				-693.4	275	-675.4	38	1.34	0.91
				-694.1	210	-674.8	37	1.40	1.26
				-693.1	326	-674.7	33	1.54	1.56
						...			
						-659.2	37	**0.95**	0.76

Figure 4.6: *Trihelix results for A2A when starting from a hybrid structure of A2A helices aligned to the β_2 crystal structure.* **Scatter plot:** The blue points denote 2600 structures from the Trihelix run. The energy is in REU and the RMSD is after structure minimization. The red + highlights the lowest energy structure. The vertical line shows the RMSD of the initial homology model (1.49 Å). The green × was added after the Trihelix run and corresponds to the conformation, with ideal moves rounded to the nearest grid coordinate. **Table:** The table lists the 10 lowest energy conformations and the added rounded conformation. The colors denote values of the sampled coordinates according to the color key in Fig. 4.2. 'The number of contacts' column is the number of class A common contacts present in this structure (the conserved contacts are listed in Fig. 2.8). The last two columns show RMSD to the correct crystal structure. RMSD of the binding pocket is discussed in Section 4.5.5.



θ 1 2 3 4 5 6 7	α 1 2 3 4 5 6 7	β 1 2 3 4 5 6 7	z 1 2 3 4 5 6 7	Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
				-758.3	86	**-727.3**	36	**1.36**	1.88
				-759.7	33	-725.9	36	1.70	1.77
				-751.5	1747	-725.5	38	1.85	2.38
				-765.2	3	-723.1	36	1.75	1.92
				-756.7	194	-722.7	38	1.59	2.24
				-751.4	1839	-721.8	38	1.52	1.78
				-751.3	1877	-721.4	39	1.57	1.94
				-752.5	1180	-721.1	38	1.78	2.25
				-765.2	2	-720.8	35	1.68	1.66
				-750.7	2313	-720.5	39	1.72	2.30
						...			
						-686.3	36	**1.15**	1.12

Figure 4.7: *Trihelix results for P2Y12 when starting from a hybrid structure of P2Y12 helices aligned to the PAR1 crystal structure.* Same legend as for Fig. 4.6. The starting RMSD is 2.02.

of coordinates. To find a close conformation on the sampled grid of coordinates, we rounded the ideal helical moves to the nearest grid point. This closest grid point is shown as a green × and is specified in the bottom line of the table (Fig. 4.6). The grid point was not found among the low energy structures from Trihelix, so we had to add it in the analysis step. However, the energy of this structure is relatively high compared to other sampled structures, therefore it is not surprising that the Trihelix energy did not pick it up.

An analogous simulation for P2Y12 starts at RMSD 2.02 Å and the energy selects structure with 1.36 Å, which is a large improvement (Fig. 4.7). The rounded grid conformation has RMSD 1.15 Å, smaller still, but its energy is higher.

In both cases, the full bundle stage was necessary to identify the lowest energy model. The solution was only 24th and 36th, respectively, at the Trihelix level. Both models have accuracy near RMSD 1.5 Å.

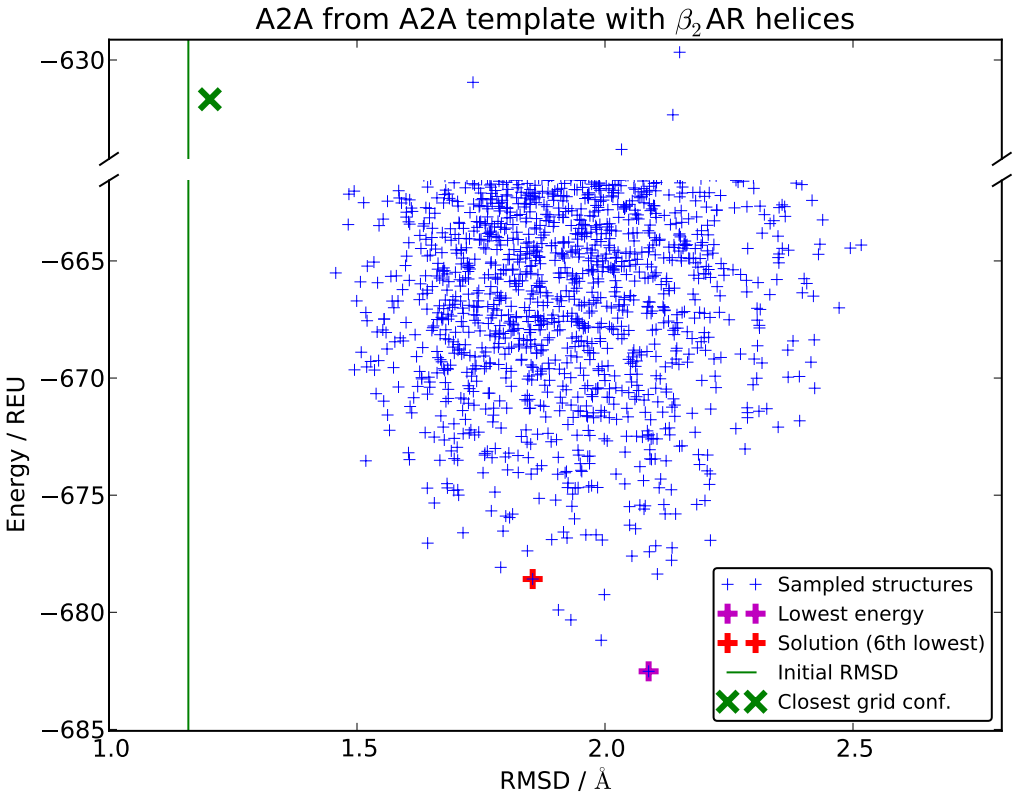
4.5.3 Sampling with the Correct Helix Positions

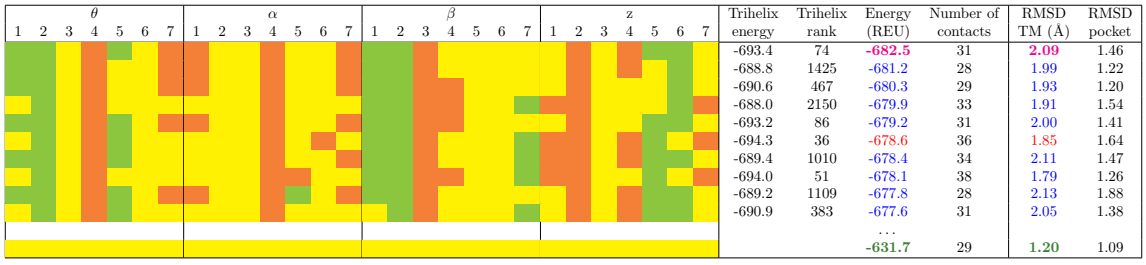
Next, we consider a hybrid structure formed with the homology (incorrect) helices placed in the crystal (correct) orientations. In this case, the starting structure (RMSD $\sim$ 1 Å) cannot be improved by the rigid body moves. We want to see whether the energy function selects this structure, or if the helical positions are inconsistent with the helical shapes as a result of clashes.

Figure 4.8 shows the result for A2A. The initial RMSD is 1.16 Å (1.20 Å after minimization), but this structure has very high energy. The lowest energy structure has RMSD 2.09 Å, which is surprisingly high. For this structure, we identify all the inter-helical contacts and compare them to the 40 conserved contacts in class A, which are listed in Fig. 2.8. This structure has only 31 out of them. Thus we reject the structure and consider models which are higher in energy. Only the 6th model has more than 35 contacts, so we select it as the solution. Its RMSD is 1.85 Å. To select this structure we described *a GPCR health heuristic, which rejects models with less than 35 inter-helical conserved contacts.*

Figure 4.9 shows the result for P2Y12. The starting structure has RMSD 0.97 Å, and the Trihelix sampling identifies a model with RMSD 1.75 Å. This model has 35 inter-helical contacts, therefore it passes the GPCR health heuristic.

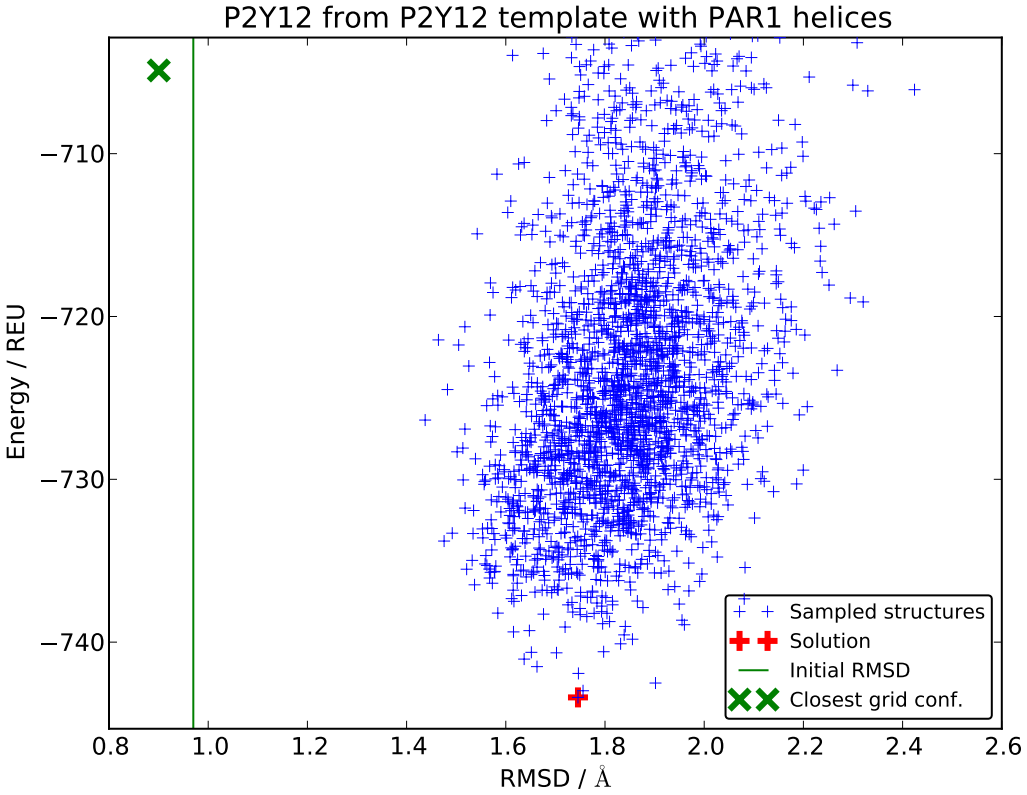
In both studied cases, the initial structure has a low RMSD $\sim$ 1 Å because we placed the helices at the optimal orientations, but the energy function does not select it. Instead, structures with RMSD around 1.8 Å are selected (after using the GPCR health heuristic). This shows that removing the information about the exact helical shapes is a major obstacle for structure refinement, and without the correct helical shapes we cannot expect to improve homology models below about 1.8 Å.





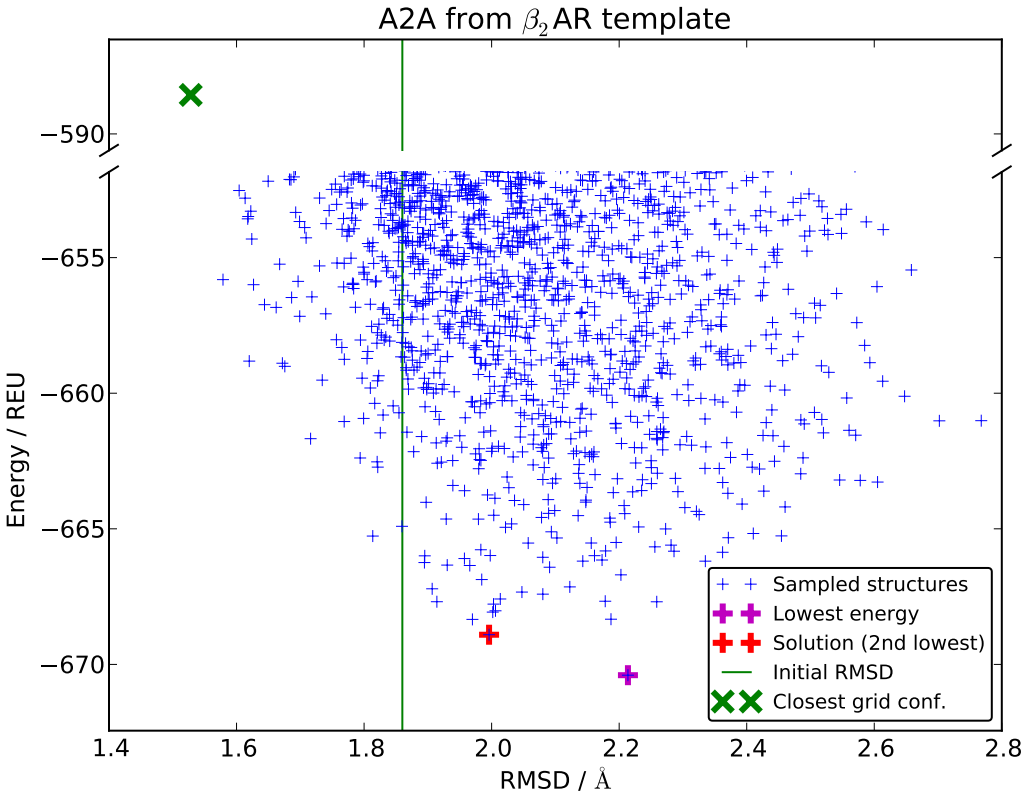
Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
-693.4	74	**-682.5**	31	**2.09**	1.46
-688.8	1425	-681.2	28	1.99	1.22
-690.6	467	-680.3	29	1.93	1.20
-688.0	2150	-679.9	33	1.91	1.54
-693.2	86	-679.2	31	2.00	1.41
-694.3	36	-678.6	36	1.85	1.64
-689.4	1010	-678.4	34	2.11	1.47
-694.0	51	-678.1	38	1.79	1.26
-689.2	1109	-677.8	28	2.13	1.88
-690.9	383	-677.6	31	2.05	1.38
		...			
		-631.7	29	**1.20**	1.09

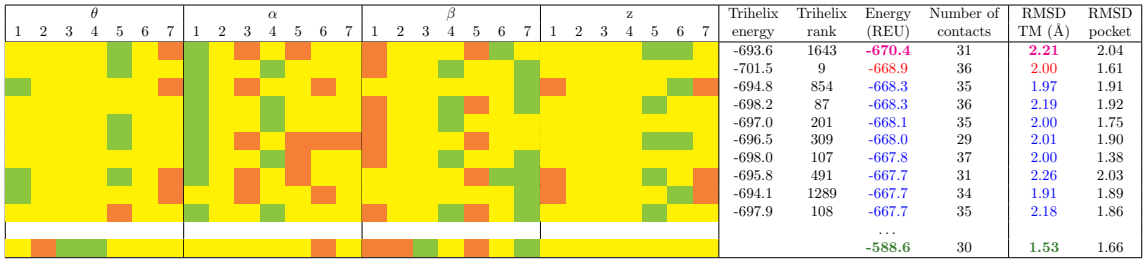
Figure 4.8: *Trihelix results for A2A when starting from a hybrid structure of β_2 helices aligned to the A2A crystal structure.* Same legend as for Fig. 4.6. The starting RMSD is 1.16. The 5 lowest energy models are rejected because they have less than 35 class A conserved contacts. Only the 6th lowest energy model satisfies this criterion.



θ 1 2 3 4 5 6 7	α 1 2 3 4 5 6 7	β 1 2 3 4 5 6 7	z 1 2 3 4 5 6 7	Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
				-773.7	479	**-743.4**	35	**1.75**	1.82
				-769.8	2356	-743.0	34	1.76	1.70
				-771.1	1467	-742.5	33	1.90	1.28
				-775.5	186	-741.9	35	1.75	1.71
				-776.6	97	-741.5	36	1.66	1.29
				-775.6	175	-741.0	39	1.64	1.42
				-773.7	490	-740.7	35	1.70	1.10
				-771.6	1179	-740.6	37	1.65	1.46
				-778.8	26	-740.6	35	1.80	1.25
				-773.4	555	-740.1	34	1.83	1.42
						...			
						-704.9	33	**0.90**	0.69

Figure 4.9: *Trihelix results for P2Y12 when starting from a hybrid structure of PAR1 helices aligned to the P2Y12 crystal structure.* Same legend as for Fig. 4.6. The starting RMSD is 0.97.





Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
-693.6	1643	**-670.4**	31	**2.21**	2.04
-701.5	9	-668.9	36	2.00	1.61
-694.8	854	-668.3	35	1.97	1.91
-698.2	87	-668.3	36	2.19	1.92
-697.0	201	-668.1	35	2.00	1.75
-696.5	309	-668.0	29	2.01	1.90
-698.0	107	-667.8	37	2.00	1.38
-695.8	491	-667.7	31	2.26	2.03
-694.1	1289	-667.7	34	1.91	1.89
-697.9	108	-667.7	35	2.18	1.86
		...			
		-588.6	30	**1.53**	1.66

Figure 4.10: *Trihelix results for A2A when starting from the homology model based on β_2.* Same legend as for Fig. 4.6. The starting RMSD is 1.85. The lowest energy model is rejected because it has less than 35 class A conserved contacts. The 2nd lowest energy model has 36 contacts, so it is selected as the solution.

4.5.4 Prediction from Homology Model

Finally, we move to the full prediction of the A2A and P2Y12 helical bundles. First, we study the Trihelix sampling of the A2A homology model based on the β_2 template without using any information about A2A other than its sequence. Figure 4.10 shows that the starting template has RMSD 1.86 Å. The lowest energy model has RMSD 2.21 Å, but it only has 31 conserved contacts contacts, so we discard it using the GPCR health heuristic. The second lowest energy structure with RMSD 2.00 has 36 conserved contacts, so we keep it as the solution. There are conformations with lower RMSD — for example the conformation with rounded coordinates has 1.53 Å — but they are

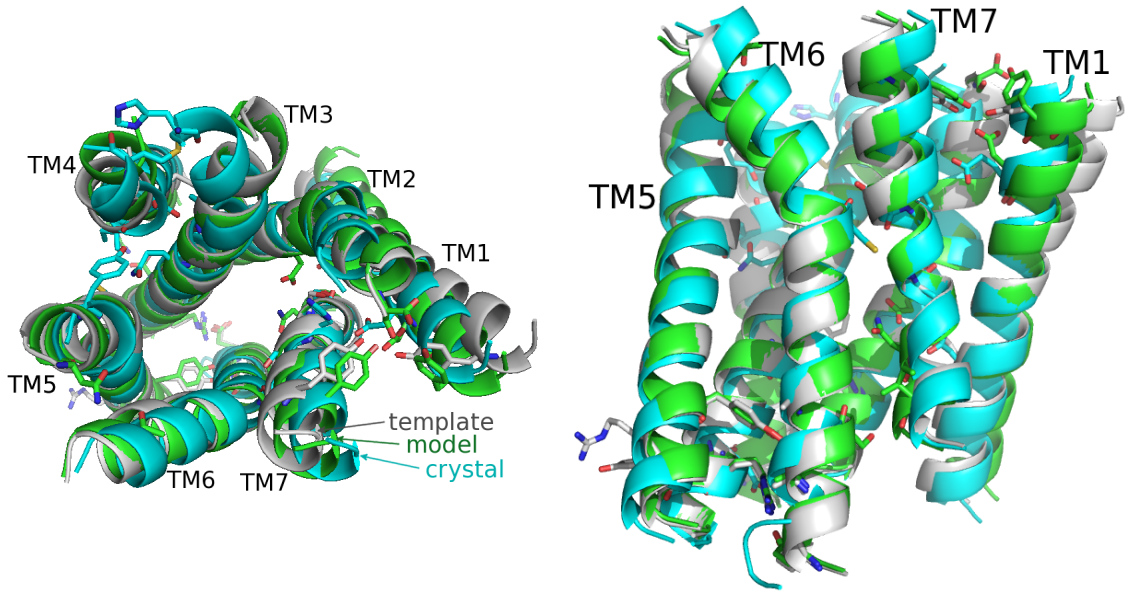


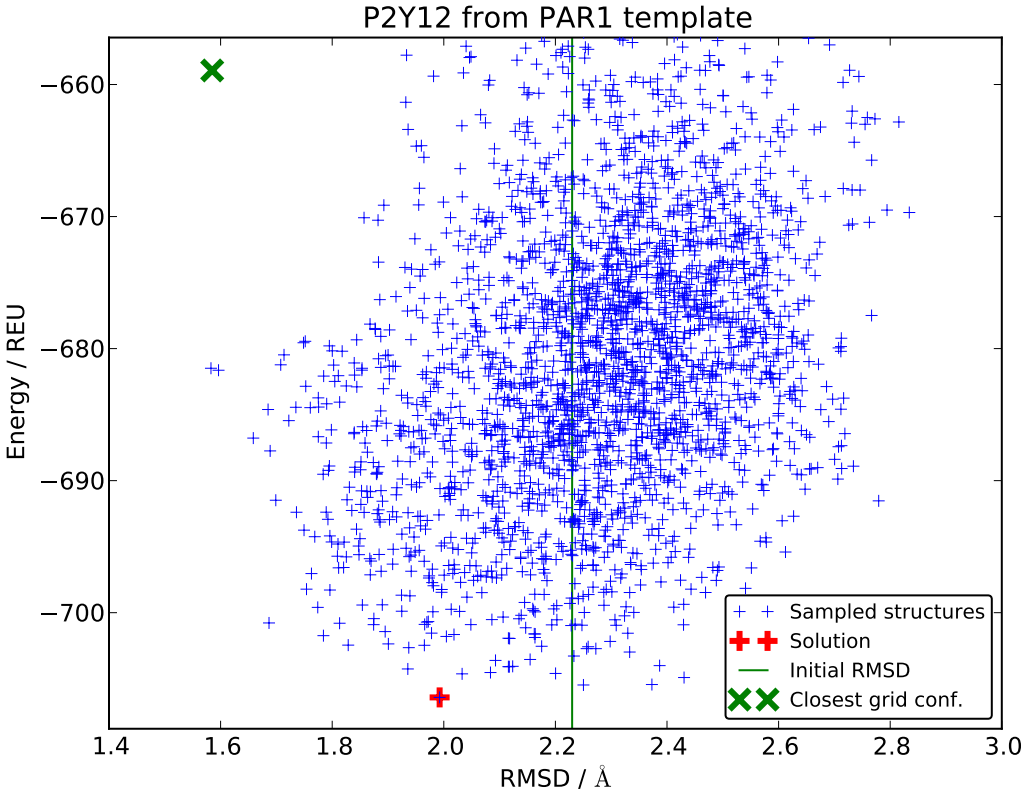
Figure 4.11: The predicted A2A structure (green) based on the β_2 homology model (grey). The A2A crystal structure is shown in blue. Residues forming the inter-helical hydrogen bond are displayed in stick representation.

not low in energy.

The selected structure is shown in Figure 4.11, where we can see that the overall position of TM 1 and 7 improved. On the other hand, TM 2 and 3 undergo a rotation which puts their ends at an incorrect orientation. Such misalignment accumulates a lot of cost into the RMSD measure. In this Trihelix run the initial RMSD increased from 1.86 to 2.00, but residues contributing most to the RMSD are not important for binding. In the next section we point out that the binding pocket actually improved slightly. Another characteristic in which the structure improved is the number of hydrogen bonds, which increased from 8 to 11 (there are 15–18 for a typical GPCR crystal structure). The behavior of polar residues, which can form hydrogen bonds, is important because they cannot make polar contacts with the membrane environment.

The last Trihelix run we discuss is the prediction of P2Y12 from the PAR1 template. Figure 4.12 shows that the lowest energy structure improved the RMSD from 2.23 Åto 1.99 Å. This structure has 35 conserved inter-helical contacts, which passes the health test. Again, some of the sampled structures have lower RMSD than this, but they score too high in energy.

Figure 4.13 shows the lowest energy model for P2Y12. TM5 undergoes the most significant improvement, as it moves about halfway to its correct position. TM 6 improves as well, but it does not make a large tilt, which would decrease the RMSD even more. The reason this move is not found is because of the deformed TM6 of the PAR1 template, which has two neighboring small amino acids, Gly and Pro, while P2Y12 only has Pro. Even without this move, the number of inter-helical hydrogen bonds increased from 10 in the homology model to 14.



θ 1 2 3 4 5 6 7	α 1 2 3 4 5 6 7	β 1 2 3 4 5 6 7	z 1 2 3 4 5 6 7	Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts	RMSD TM (Å)	RMSD pocket
				-736.9	73	**-706.4**	35	**1.99**	2.43
				-731.8	626	-705.5	32	2.25	2.64
				-741.0	11	-705.4	33	2.37	2.96
				-734.1	262	-704.9	35	2.43	2.90
				-733.3	343	-704.6	38	2.03	2.58
				-739.7	21	-704.6	33	2.20	2.50
				-730.5	1075	-704.5	36	2.08	2.64
				-732.0	582	-704.3	32	1.94	2.44
				-734.7	203	-704.1	35	2.12	2.35
				-740.0	19	-704.1	34	2.12	2.56
						...			
						-658.9	32	**1.58**	1.66

Figure 4.12: *Trihelix results for P2Y12 when starting from the homology model based on PAR1.* Same legend as for Fig. 4.6. The starting RMSD is 2.23.

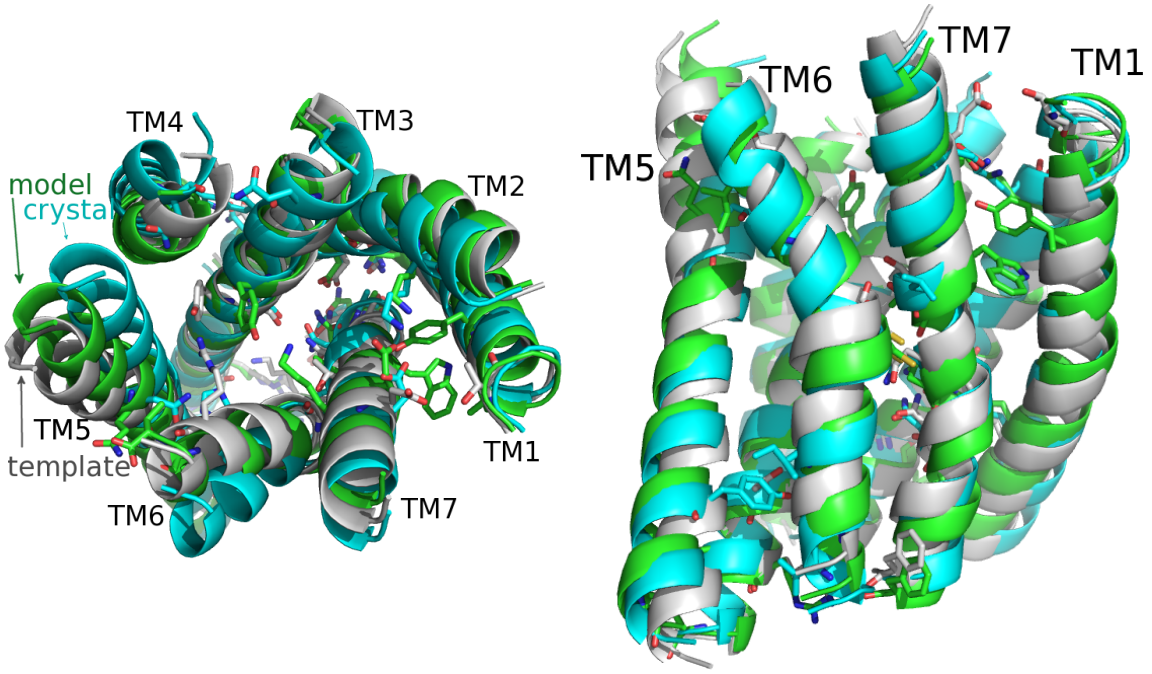


Figure 4.13: The predicted P2Y12 structure (green) based on the PAR1 homology model (grey). The P2Y12 crystal structure is shown in blue. Residues forming inter-helical hydrogen bond are displayed in the stick representation.

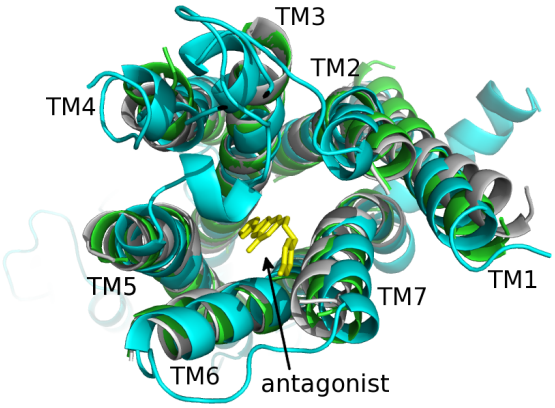
4.5.5 Binding Pocket RMSD

The most important structural feature for drug discovery is the shape of the binding pocket. Any improvements of the binding site result in highly improved docking outcomes. Let us define the binding pocket RMSD as the backbone RMSD of residues which are close (< 5 Å) to the ligand in the crystal structure. Figure 4.14 shows the ligand placement in the two studied cases.

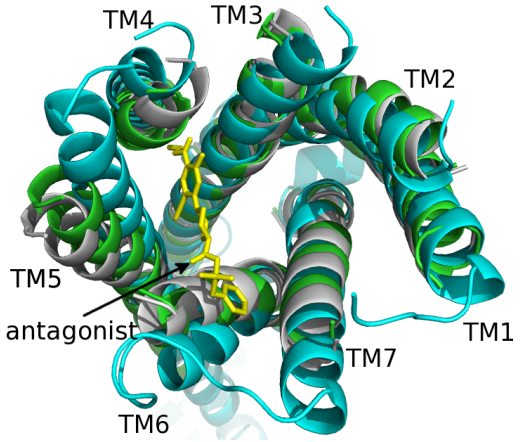
For A2A the binding pocket RMSD improved slightly from 1.64 Å to 1.61 Å. However, for P2Y12 there was a significant improvement from 2.82 Å to 2.43 Å. In both cases, the RMSD improvement in the binding pocket is better than the RMSD improvement of the full transmembrane bundle. This is because much of the contribution to the RMSD measure comes from the ends of helices that are most misaligned. If the loops were present, they would constraint the helical ends and the RMSD would be a more suitable measure.

4.6 Discussion

The Trihelix method improves the RMSD of a homology model for P2Y12 from 2.23 Å to 1.99 Å. The same method does not improve the RMSD of a homology model for A2A, as the RMSD increased from 1.86Å to 2.00Å. This suggests that the transmembrane bundle similarity below 2 Å is very good and is difficult to refine without the prior knowledge of the correct helical shapes.



(a) Orientation of the high-affinity subtype-selective antagonist ZM241385 in the A2A binding pocket [36].



(b) Orientation of the reversible antagonist AZD1283 in the P2Y12 binding pocket [47].

Figure 4.14: Improvements in the binding pocket.

As we suggested, the RMSD of the TM bundle may not be the best measure to compare the performance of the structure refinement. In both cases, the RMSD of the binding pocket decreased: for P2Y12 significantly from 2.82 Å to 2.43Å, and for A2A slightly from 1.64 Å to 1.61 Å.

The TM5 of the P2Y12 model moves approximately halfway from the starting position to the desired crystal position (shown in Fig. 4.13). This suggests that we should consider either increasing the grid size to allow large moves or increasing the number of grid points. Increasing the grid size makes impossible minor adjustments of helical positions so that the neighboring helices fit together. On the other hand, increasing the number of grid points creates too many conformations at the early stages of the protocol. When there are too many conformations, the errors in the Bihelix mean field energy cause selection of incorrect low energy conformations.

There is a fine balance between the coordinate grid steps and resolution of the energy function. Increasing the step size makes the crystal structure reconstruction more robust, but it makes it less likely to capture a good structure when making a refinement from a homology model. Decreasing the step size overwhelms the resolution of the energy function, resulting in an inability to select the correct structure. In fact, the all-zero conformation in the P2Y12 crystal structure reconstruction (Fig. 4.5) is not the lowest energy one. Decreasing the step size further would only make this worse.

In comparison to Monte Carlo protein refinement, the Trihelix energy vs. RMSD plot is not a funnel plot because only a few structures near the target are sampled. This makes it harder to test whether a good structure has been found. The energy gap separating the lowest energy model is not very high in most of the runs. Instead of relying solely on the energy function to select the solution, we use a GPCR health heuristic, which only keeps the models with 35 or more conserved inter-helical contacts.

The structure refinement starts from a homology model with approximate helix shapes and approximate helix orientations. From the two approximations, the helix shapes are more important for being able to select the correct structures using the energy function (Section 4.5.3). This is because the homology helix orientations are consistent with each other, but the homology helix shapes do not have to be consistent with the sequence. In particular, Pro and Gly residues cause kinks and irregular turns, which can cause large clashes with the nearby helices in the homology models. Any clashes at the Bihelix stage are problematic, since the conformation with clashes are removed from further considerations.

In the present work, we ignore the role of the loops. Since the loops are missing, minimization sometimes moves the helices apart to avoid clashes. To prevent this breakdown of the helical bundle during minimization we add small constraints on the backbone positions. The strength of the constraints is tuned to avoid the unwanted large moves, while allowing some flexibility of the helices.

4.7 Conclusion

It remains a challenge to refine the homology models below 2 Å because the helix shapes are unknown. The helices have to be consistent with each other, similar to a lock and a key. A similar situation occurs in other computational problems involving proteins, such as docking: if the correct protein backbone and the correct ligand conformation are known, the docking algorithms can often reliably find the binding site and recover the side-chains. However, if the correct backbone conformation or the correct ligand conformation is not known, the problem becomes much more difficult.

One way to significantly improve the effectivity of the structure refinement is to make sure there are no misaligned prolines. In some cases, this can be achieved by combining more templates with each other depending on the sequence of the studied protein.

We developed and characterized Trihelix, a new method for GPCR structure refinement. This method uses mean field approximations to completely sample a large number of coarse degrees of freedom. We have successfully demonstrated a refinement of the P2Y12 homology model and explored the limits of validity of the method. In the next chapter, we apply this method to predict the lactate receptor structure.

Chapter 5

Structure Prediction of the Lactate Receptor HCAR1

5.1 Receptor Family and Expression in Cancer

In this chapter we apply the Trihelix sampling method to model the TM bundle of the hydroxy-carboxylic acid receptor 1 (HCAR1, also lactate receptor). We complete the extracellular part of the model by generating and refining many possible loop conformations. Modeling of moderately long loops is made possible by constraints coming from three expected disulfide bonds. Finally, we dock the lactate molecule and a new higher affinity ligand to study possible binding sites. We propose a binding site which agrees with the available mutation data and thus provides a validation of the receptor model. A high accuracy model is needed to search for new antagonists that block the receptor activation.

Interest in the HCAR1 receptor is rising because it is expressed in many types of cancer. A recent study [83] shows that the lactate receptor HCAR1 is expressed at high levels in 95% of pancreatic cancer patients, and that its expression levels correlate with cancer growth and metastasis. Silencing the HCAR1 expression leads to a dramatic decrease in the tumor growth. The receptor is important for cancer survival in the low oxygen, high lactate microenvironment of tumors. A small molecule antagonist could block this receptor and slow down cancer growth.

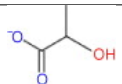
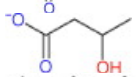
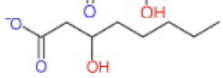
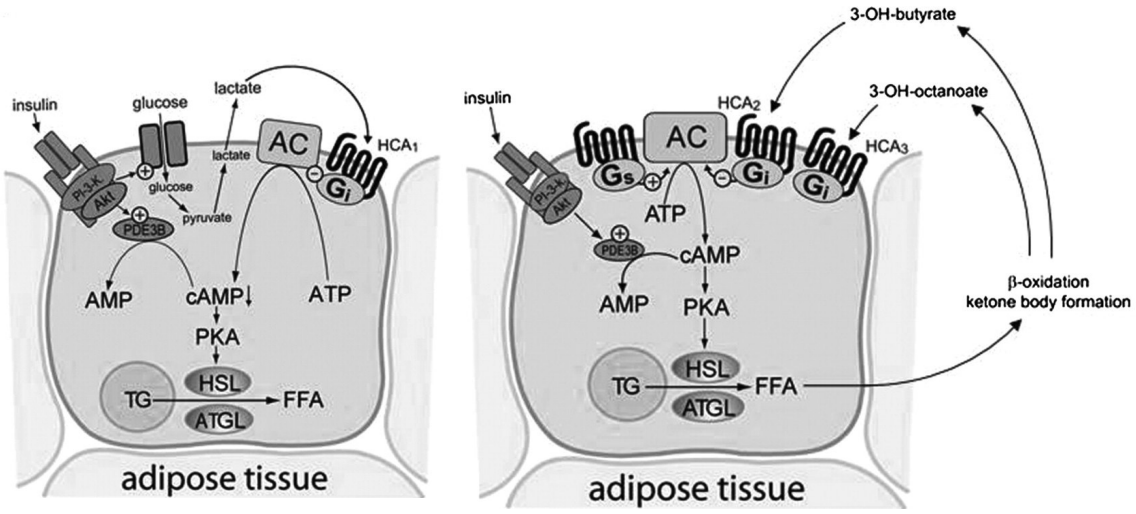
Endogenous Ligand	Structure	HCAR1	HCAR2	HCAR3	Reference
2-hydroxy-propanoate (lactate)		1500, 3000	Inactive	Inactive	[84, 85]
3-hydroxy-butyrate		Inactive	770	>25,000	[86]
3-hydroxy-octanoate		Inactive	Inactive	8	[87]

Table 5.1: Structures of endogenous hydroxy-carboxylic acids acting on HCAR1, HCAR2, and HCAR3. Half maximal effective concentration values (EC50, in μM) are determined from a downstream signaling pathway. Source [88].



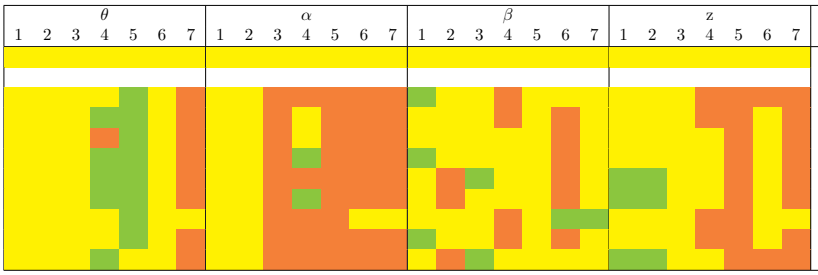
(a) HCA1

(b) HCA2 and HCA3

Figure 5.1: Functions of HCA1, HCA2, and HCA3. (a) The lactate receptor HCA1 mediates the short-term anabolic effects of insulin on adipocytes and thereby helps store energy after feeding. (b) HCA2 and HCA3 receptors are involved in the long-term regulation of lipolytic activity. In situations of increased β-oxidation rates (e.g., during starvation), 3-hydroxy-butyrate and 3-hydroxy-octanoate plasma levels are increased and result in the inhibitory regulation of lipolysis via HCA2 and HCA3 receptors, respectively, in the form of a negative feedback loop. Therefore, HCA2 and HCA3 receptors help preserve energy stores during starvation. Source: [91]. AC, adenylyl cyclase; TG, triglycerides; HSL, hormone-sensitive lipase; ATGL, adipocyte triglyceride lipase; FFA, free fatty acids; PKA, cAMP-regulated protein kinase.

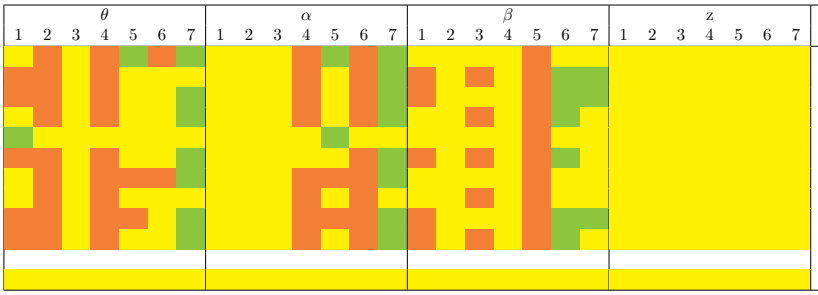
The lactate receptor is part of the hydroxy-carboxylic acid receptor family, whose endogenous ligands are intermediates of energy metabolism [88, 89, 90, 91]. These ligands (shown in Table 5.1) are small hydroxy-carboxylic acid compounds, which are very similar to each other, yet the receptors are specific to their respective endogenous ligands. The high specificity of these receptors to the set of similar molecules suggests that a high accuracy model is needed if we want to find new drug candidates.

Besides its involvement in cancer, the lactate receptor could be a direct drug target for dyslipidemia. At present, nicotinic acid, which targets HCAR2, serves as a dyslipidemic drug. However, besides fat tissues, HCAR2 is also expressed in skin, and so the nicotinic acid often has a serious side effect of flushing (redness). An agonist biding to HCAR1 could treat dyslipidemia without the side-effects, because HCAR1 is mostly expressed in fat tissue. The differences in function of these receptors are outlined in Figure 5.1. In particular, HCAR1 helps store energy after feeding, and HCA2 and HCA3 help save energy during starvation.



Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts
		-691.7	33
		...	
-700.0	1989	-690.3	31
-700.4	1752	-687.9	31
-702.0	904	-686.8	32
-703.3	506	-686.7	32
-705.6	177	-686.5	29
-707.4	65	-685.9	31
-699.8	2200	-685.4	32
-704.8	265	-684.3	29
-711.9	1	-684.3	30

(a) Trihelix sampling of HCAR1 starting from the P2Y12 template fails to find a good candidate structure. The all-zero structure was dropped at the bi-helix and Trihelix stages, but it actually has a lower energy than the lowest energy Trihelix prediction. Furthermore, the sampled structures have 32 or less contacts from the conserved contacts list of 40 contacts, which indicates that the sampled structures do not approach the GPCR fold.



Trihelix energy	Trihelix rank	Energy (REU)	Number of contacts
-717.8	227	**-698.2**	37
-713.2	1656	-697.4	33
-716.0	515	-696.7	35
-723.1	9	-696.7	33
-718.1	198	-696.6	34
-713.0	1885	-696.3	31
-719.5	91	-696.2	37
-715.2	722	-696.1	34
-720.3	58	-696.0	35
-714.9	799	-695.7	34
		...	
		-606.4	37

(b) Trihelix sampling of HCAR1 starting from the PAR1 template finds a good candidate structure, which has a lower energy than the starting all-zero structure. The solved structure is shown in Figure 5.3. Furthermore, 37 (out of 40) class A conserved contacts are present in this structure, which indicates a high quality model. Compared to the models derived from the P2Y12 template, this time the z coordinate is 0 for all the low energy conformations.

Figure 5.2: Trihelix results for HCAR1.

5.2 Trihelix Sampling of HCAR1

We now move on to the prediction of the HCAR1 structure, which is in the Aδ subclass. The only available crystal structures in Aδ subclass are P2Y12 and PAR1, which we both use as templates. We use the transmembrane region alignment from Chapter 3, in which HCAR1 has a very low sequence identity of 24% and 29% with P2Y12 and PAR1, respectively. The TM sequence similarity to these templates is 47% and 48%. The similarity is close to the cases used to benchmark the Trihelix sampling, therefore, we apply the Trihelix sampling with the same size of steps as developed and benchmarked in Chapter 4: TM rotation around axis $\theta = -15°, 0°, +15°$, TM tilts $\alpha, \beta = -7°, 0°, +7°$, and TM translations along axis $z = -2\,\text{Å}, 0\,\text{Å}, 2\,\text{Å}$. The lowest energy structures from the sampling are listed in Figure 5.2.

The lowest energy structures based on the P2Y12 template have higher energy than the starting structure. This indicates that the suitable TM coordinates were dropped at the bi-helix and tri-helix mean-field approximations. The mean field energies are computed from structures with two and three helices only, so it is possible to get incorrect side-chain positions from the the side-chain placement algorithm. Another warning signal is that the number of conserved contacts is 32 or lower

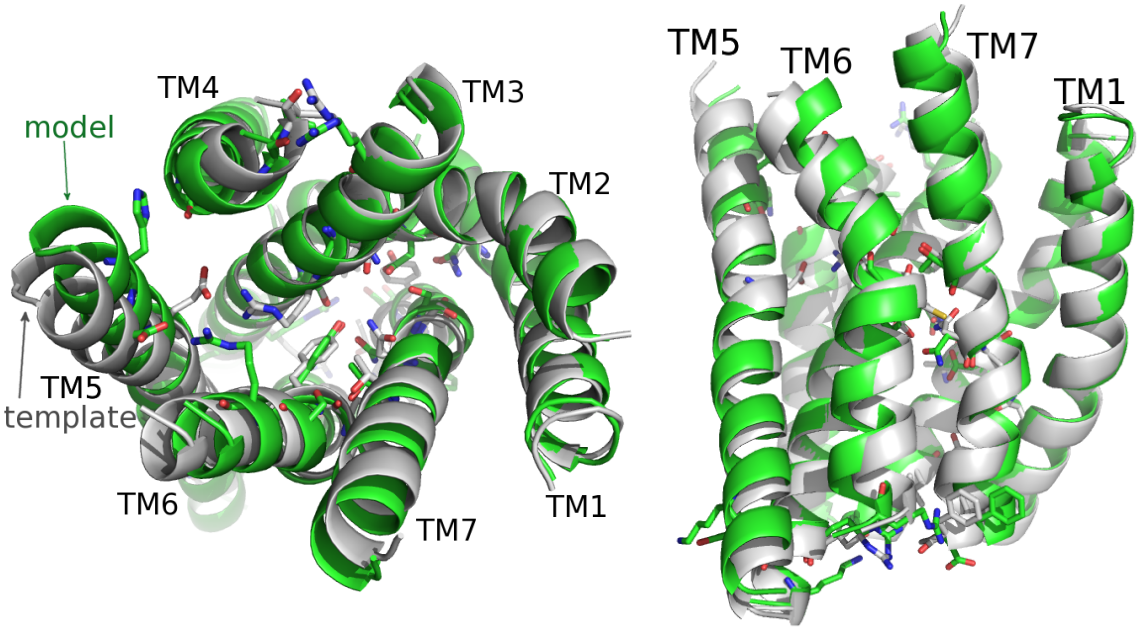


Figure 5.3: The HCAR1 model (green) is the lowest energy structure from the Trihelix run based on the PAR1 template (grey). The residues which form inter-helical hydrogen bonds are displayed in stick representation. The number of inter-helical hydrogen bonds increases from 10 in the all-zero structure to 14 (GPCR crystal structures have typically 15-20 inter-helical hydrogen bonds).

(out of 40) for all 10 lowest energy models, so the sampling is probably stuck in some local energy minimum not close to the GPCR fold. Therefore we disregard this particular run.

The results based on the PAR1 template are more reasonable. First, all sampled low energy models have lower energy than the starting structure, therefore the sampling actually decreased the energy of the template. Second, the lowest energy structure has 37 inter-helical contacts common with class A, and so this model is close to the GPCR fold. We take the lowest energy structure to be the model for the HCAR1 TM bundle. It has energy -698.2 REU (Fig. 5.2b), which is lower than energy of any models based on the P2Y12 template. The most notable difference between the two runs is that the PAR1 template kept all TM with $z = 0$.

The solved model is shown in Figure 5.3. The largest changes from the template are in the TM 5 and 6 movement, with smaller adjustments of TM 4 and 7. These changes allow for 14 inter-helical hydrogen bonds, compared to 10 in the starting structure. Sometimes the helical bundle without loops is used for ligand binding studies, but since lactate is a small molecule, we build the extra-cellular loop bundle before docking.

Loop	Residues	Length	Cysteines
N-terminus	(1) 3-14	12	6, 7
EC1	77-84	8	
EC2	153-170	18	157, 165
EC3	245-253	9	252

Table 5.2: Extra-cellular loop lengths

5.3 HCAR1 Extracellular Loops

To complete the extracellular part of the HCAR1 model, we need to build the three extra-cellular loops and the N-terminus. Their lengths are listed in Table 5.2. The EC2 loop is the longest at 18 residues, which is challenging for the common loop building approaches.

Let us briefly discuss the performance of several GPCR loop prediction methods. If the protein, nearby loops, and side-chains are at exactly the same positions as in the crystal, even long loops over 20 amino acids can often be reconstructed to RMSD <2 Å [92]. However, when the side-chains are not from the crystal structure but are predicted, the accuracy of the loop prediction decreases rapidly. Nikiforovich et al. [93] managed to solve short loops (8 -10 residues) to RMSD 3-5 Å, and long loops (>=25 residues) only to RMSD > 7 Å. GPCR loops are expected to be very flexible and it is not clear whether the low accuracy is due to inherent flexibility of the loops, or due to lack of sampling and an approximate energy model.

General loop prediction methods are benchmarked on stable, high resolution crystal structures. Stein et al. [94] implemented a robotics-inspired conformational sampling within the Rosetta framework, which solves many 12 residue loops to sub-angstrom accuracy. Even for this loop length, there are cases which cannot be solved to RMSD below 6 Å. The fraction of loops which can be solved drops down for loops 14 and 15 residues long, and no 17 residue loops are solved successfully.

Fortunately, in the case of HCAR1, the situation simplifies due to the presence of disulfide bonds, which can be used as constraints. From mutation studies, Kuei et al. [95] suggested that disulfide bonds are formed between the following cysteine pairs:

Connectivity A:	Cys 88 (TM3)	-	Cys 165 (EC2)
	Cys 6 (Nterm)	-	Cys 252 (EC3)
	Cys 7 (Nterm)	-	Cys 157 (EC2)

The cysteines C88 and C165 are highly conserved in class A and the disulfide bond between them is present in all the crystal structures in which these cysteines are present. The mutation of any of these two cysteines stops cell surface expression of the protein, so this disulfide bond is important for the folding of HCAR1 in the membrane. Mutation of the other 4 cysteines allows the protein to assemble in the membrane, but it is not sensitive to its endogenous ligand. Since C6 and C7 are

close together, we consider also an alternative connectivity:

Connectivity B:	Cys 88 (TM3)	-	Cys 165 (EC2)
	Cys 7 (Nterm)	-	Cys 252 (EC3)
	Cys 6 (Nterm)	-	Cys 157 (EC2)

We programmed a loop building algorithm, which builds random loop bundles consistent with the known disulfide constraints. This algorithm builds all loops at once, residue-by-residue, by randomly sampling backbone torsions from the selected Ramachandran distribution. From each anchor point (TM end) we grow $N = 10,000$ loop candidates, so that there are many possible combinations for loops closure. From the N^2 possibilities for each loop closure, we choose the ones which have best geometrically correct bond. As each residue is added we check if the following constraints are satisfied, otherwise we generate a new random trial:

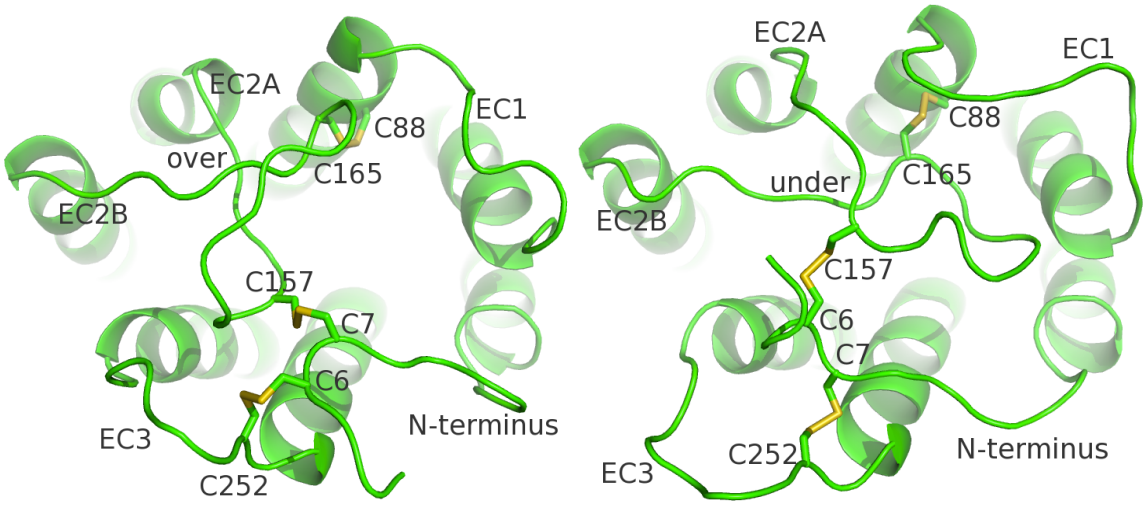
- given the number of remaining residues, each loop can close
- given the number of remaining residues, each disulfide can be formed
- each loop avoids clashes with TM and the membrane region
- each loop avoids clashes with itself

We build 1000 random loop bundles and then refine each one using Rosetta's kinematic loop builder [94]. Since the refinement is a random algorithm as well, we keep 6 candidates from each loop refinement. We use the Rosetta energy function to select the low energy models.

Interestingly, among the low energy models, the EC2 loop appears in two different topologies. In the first case (Fig. 5.4a), the EC2A part of the loop is positioned under the EC2B part. However, this is not consistent with the disulfide C88-C165 forming early during the protein assembly from the individual helices in the membrane. The second case (Fig. 5.4b), is consistent with the TM assembly as the C88-C165 forms first and C6-C252 (or C7-C252) form later. After removing the loop bundles with the wrong topology, we keep the 23 low energy bundles for the docking studies (shown in Fig. 5.5). In the current energy model, we do not see a preference for either the disulfide connectivity A or B, so we choose 10 lowest energy conformations from each and add an additional 3 conformations with interesting conformation of His 155 in EC2.

5.4 Lactate Docking to the HCAR1 Model

In this section, we study the possible binding site of lactate, which is the endogenous ligand for HCAR1. Compared to other GPCRs, a very high concentration of lactate is necessary to activate this receptor ($> 1mM$), because lactate is an intermediate of energy metabolism and is present at



(a) Rejected EC2 topology

(b) Possible EC2 topology

Figure 5.4: The highly conserved disulfide bond Cys 88 (TM3) - Cys 165 (EC2) is critical for assembly of the protein in the membrane. Likely, it forms first before the other two disulfide bonds during the protein folding process, so the EC2B segment has to be under the EC2A segment (from an extracellular view point).

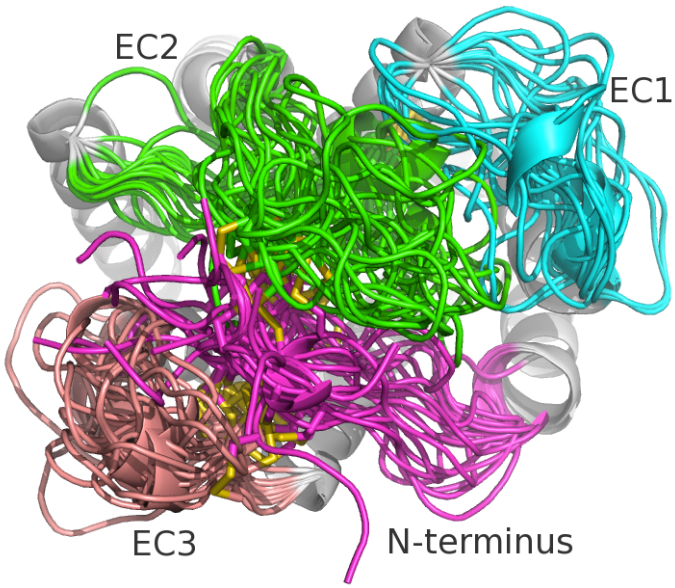


Figure 5.5: The ensemble of 20 lowest energy extracellular loop bundles. An additional 3 bundles with interesting conformation of HIS155 in EC2 were added. The disulfide bonds are partially visible (yellow). Even though the loops are constrained by three disulfide bonds, they can still fold in many different ways.

high concentrations. In addition, lactate is a very small molecule and it can fit into the binding pocket in many different orientations. This is a challenge since only a finite number of cases are considered by the docking algorithm. Therefore, we start with docking 3,5-dihydroxybenzoic acid (DHBA), which is a recently discovered agonist specific to HCAR1 [96]. Not only is DHBA a larger molecule, but only a 10× lower concentration is required to activate the receptor (150μM).

We follow the docking procedure reported in [97]. To compare the ligand poses we use the *neutral* snap binding energy, defined as the difference between the energy of the complex and the energy of the protein and the ligand evaluated in the Dreiding force field [77]. Bray et al. [98] showed that evaluating energy after *neutralization* gives better correlation of the binding energy to the experimental binding affinity than interaction energies between charged groups. Here neutralization means proton transfer between nearby charged groups, which maximizes the number of hydrogen bonds.

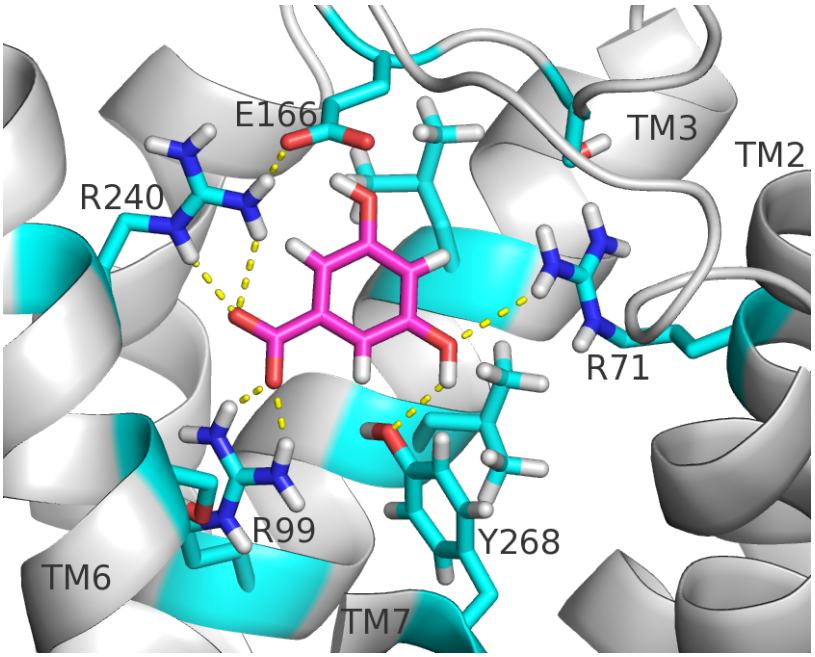
For each of the 23 candidate loop bundles, we dock DHBA into the binding site between the TMs 2, 3, 6, and 7. This produces about 2300 candidate poses. To reduce the number of poses, we use constraints coming from an analogy to the biogenic amine receptors. In biogenic amine receptors Asp 3.36 always binds with the N^+ atom in the ligand. Furthermore, this Asp or the ligand interacts with the nearby Tyr 7.43. In HCAR1, there is Arg 99 (3.36), which has no negative charge group nearby, so we expect it to bind with the negatively charged ligand. Also, Tys 268 (7.43) is conserved in this protein and in constructed models without a ligand it always interacts with Arg 99. Thus we keep only 100 poses, which minimize the sum of two distances:

- carboxyl group of the ligand to the guanidinium group of Arg 99
- OH group of Tyr 268 to the guanidinium group of Arg 99

After reducing the number of poses to 100, we annealed the ligand and side-chains within 5Å, and evaluated the neutral snap binding energies. Figure 5.6a shows the best scoring pose. The carboxyl group of the ligand interacts with Arg 99 and Arg 240, which is neutralized by Glu 166. One of the hydroxy groups on the ligand interacts with Arg 71 and Tyr 268. We see that there are a large number of charged interactions.

We repeated the same procedure for lactate. Lactate is a smaller molecule, therefore its position and orientation in the low energy poses varies more across the TM 2-3-6-7 pocket. The 8th lowest energy pose (Fig. 5.6b) shows very similar interactions as the DHBA pose. Since DHBA is a larger molecule than lactate, and a lower concentration is necessary to activate HCAR1, the docking results for DHBA are more reliable. Thus we select the 8th lactate pose. Both ligands bind in a similar area of the TM 2-3-5-6 pocket, and directly bind to residue Arg 71, Arg 99, Arg 240, and Tyr 268.

This binding mode is similar to the binding site suggested by Kuei et al. [95], except the study did not find any interaction with Tyr 268 and instead assumed direct ligand interaction with the

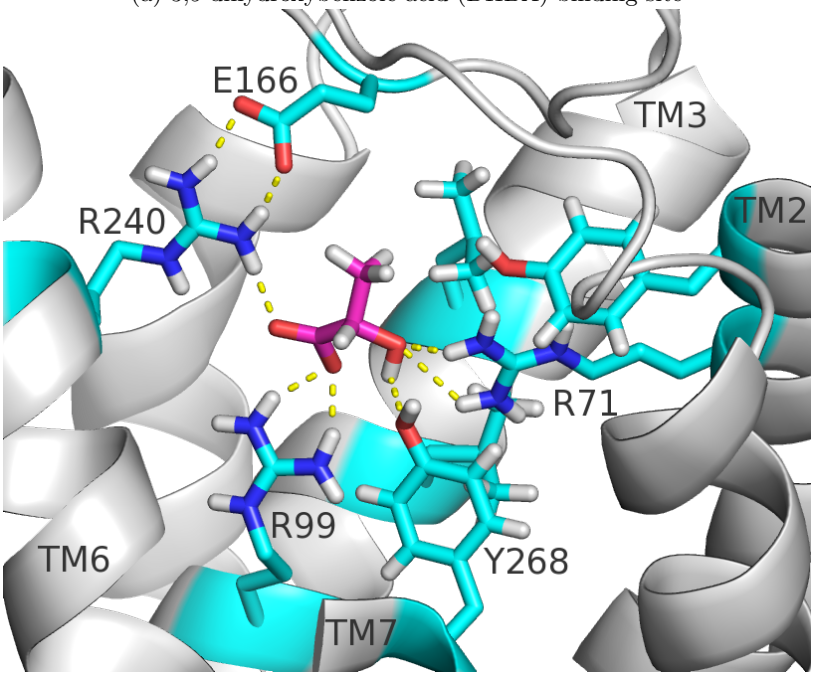


(a) 3,5-dihydroxybenzoic acid (DHBA) binding site

Neutral interaction energy between selected residues and the ligand:

	kcal/mol
ARG 99	-7.9
ARG 240	-5.8
ARG 71	-5.7
TYR 268	-4.5
GLU 166	-3.6
LEU 92	-2.9
LEU 264	-2.0

Total neutral snap binding energy: -43.6 kcal/mol



(b) Lactate binding site

Neutral interaction energy between selected residues and the ligand:

	kcal/mol
ARG 71	-7.3
ARG 99	-5.1
LEU 92	-1.7
TYR 268	-1.3
LEU 264	-1.2
GLU 166	-1.2

Total neutral snap binding energy: -30.2 kcal/mol

Figure 5.6: Predicted binding sites for the two ligands (purple). Hydrogen bonds between the ligand and nearby residues are shown in yellow.

nearby Glu 166. The mutation study shows (also [85]) that alanization of any of the following residues stops the receptor sensitivity to the lactic acid:

- Arg 71, Arg 99, Arg 240, Glu 166, Tyr 233, Thr 267

Our docking shows that the three arginines likely directly interact with the ligand. Glu 166 seems to control the charge environment near Arg 240, and Tyr 233 and Thr 267 are important for changing the receptor conformation to the active one.

5.4.1 Possible Binding in TM 3-5-6 Pocket

Since a mM concentration of lactate is required to activate the receptor and lactate is a small molecule, it is possible that two lactate molecules need to bind for the receptor activation. Here we explore this possibility. The HCAR1 binding pocket is tightly covered by the extracellular loop bundle, which includes the N-terminus. The largest opening is near TM2 and TM7, where there are also a large number of arginines. It is possible that the Arg 71 on TM2 only acts as an intermediate charge partner to a lactate molecule as it enters the binding pocket and moves deeper between TM3, 5 and 6. We repeated the docking procedure for this biding pocket.

For lactate, many possible poses were found in which the carboxyl group interacts with Arg71 and Arg240 (both or one of them). The snap binding energies were similar to docking in the 2-3-6-7- pocket. However, DHBA docking is more conclusive, as the best snap binding energy in this site (-42.1 kcal/mol) is lower than in the 2-3-6-7- pocket. It is not likely that DHBA binds between TM3-5-6 and thus it is not likely that lactate binds there either.

5.5 Conclusion

We applied the Trihelix sampling method to build a model for the HCAR1 receptor. To model its extracellular loops, we implemented a novel loop building method which uses constraints from known disulfide bonds. Since the GPCR loops are expected to be flexible, we generated many candidate loop conformations.

In order to find a binding site for lactate, which is a very small molecule, we first performed docking of 3,5-dihydroxybenzoic acid, a larger molecule with similar hydroxy-carboxylic motive. The identified binding site is in agreement with the available mutation data. The ligand binding involves many charged residues. Three arginines interact directly with the ligand, and the binding pose is stabilized by a nearby glutamic acid. The successful ligand docking validates the quality of the transmembrane bundle from the Trihelix sampling.

Chapter 6

Charge Equilibration

6.1 Force Fields

Force fields allow atomistic modeling of systems with more than 100 000 atoms, which is required for the modeling of proteins and surrounding solvent. The accuracy of force fields is key to realistic simulations. Molecular dynamics simulations most often use AMBER [78] and CHARMM [99] force fields, which parametrize the forces between atoms with several bonded and non-bonded terms. Bonded terms — bond, angle, and dihedral — capture the effect of covalent bonds. Non-bonded terms capture the electrostatic force and van der Waals interaction (both the short distance repulsion and the long distance attractive forces). In this work, we have mostly used the Rosetta force field, which includes several other empirical energy terms that were computed statistically from analysis of the database of known crystal structures. Rosetta treats the electrostatic interaction similarly to the other force fields. Accurate charge assignment is essential to determining the electrostatic energies in molecular dynamics simulations.

Although the electron density in molecules is diffuse, the short distance effects are included in the bonded terms, so the electrostatic force can be computed as interaction of point charges placed at the atomic nucleus. In this chapter we describe a method that allows one to assign the atomic charges to an arbitrary molecular system. The new method extends the successful QEQ charge equilibration method to include atomic polarization, which is a property ignored by many of the point charge models. The polarizable QEQ method has, therefore, the potential to describe qualitatively new types of phenomena.

6.2 Charge Equilibration Based On Electronegativity

Many commonly used methods derive charges from density functional calculations and fix the charges throughout molecular dynamics (MD) simulations [100, 101, 102]. Fixed charges cannot adjust to their local electrostatic environment. Reactive force fields [103, 104, 105, 106, 107], on the other

hand, change connectivity during the MD runs and thus require updates of assigned atomic charges. The charge equilibration (QEQ) scheme proposed by Rappe and Goddard [108] has been highly successful in determining geometry dependent charges that respond to changes in their environment. One advantage of this method is that it is minimally parameterized, utilizing experimental ionization potentials, electron affinities, and atomic radii. In order to model polarizable materials Zhang et al. [109, 110] proposed an extension of QEQ, which uses Gaussian charges and allows for charge polarization (PQEQ). This framework enables the assignment of atomic charges independently of atomic bonding and takes the local electrostatic environment into account. The PQEQ framework enables a description of polarized systems by allowing the position of the electron cloud to become displaced from the atomic nucleus.

In [7] we performed a constrained optimization of the QEQ atomic parameters of the elements of the first three rows of the periodic table (except for noble gases) to work well with the Gaussian charge distributions, while maintaining periodic trends and the original physical motivation. In the following sections we derive the expressions that were used for programming the PQEQ method into the LAMMPS molecular dynamics package [111]. A particularly useful feature for biological simulations is the speed at which new systems can be parametrized. For example, charges for thousands of drug candidates can be computed on the fly in virtual ligand screening. Another advantage of PQEQ is that the computation of charges is fast even for large systems such as proteins.

6.3 Polarizable QEQ with Gaussian Charges

The QEQ [108] method is a charge equilibration scheme based on minimizing the total electrostatic energy, which is expressed as a sum of internal atomic energy and pairwise Coulomb interactions:

$$E = \sum_i \left(\chi_i q_i + \frac{1}{2} J_i q_i^2\right) + \sum_{i>j} q_i q_j C_{ij}(r_{ij}). \tag{6.1}$$

Here $C_{ij}(r_{ij})$ is the screened Coulomb interaction between atom i and j. Parameters χ_i and J_i are electronegativity and idempotential of the atom i and they can be related to ionization potential (IP) and electron affinity (EA) as:

$$\chi_i = \frac{\partial E}{\partial q_i}\Big|_{q_i=0} = \frac{1}{2}(IP + EA) = \text{ electronegativity}$$
$$J_i = \frac{1}{2}\frac{\partial^2 E}{\partial q_i^2}\Big|_{q_i=0} = IP - EA = \text{ idempotential}$$

The atomic charges q_i are chosen such that they minimize the total energy, E, subject to the constraint that the sum of the charges is fixed (or that the total charge on each molecule is fixed [112]).

In the original QEQ method, the screened Coulomb interaction C_{ij} is computed as the elec-

trostatic interaction of two Slater-type orbitals. In order to be able to model atomic polarization, Zhang et al. [109, 110] developed polarizable version of QEQ (PQEQ), in which each atom has a charged core with a fixed charge q_{ic} and a variable shell charge q_{is}. The charges were modeled with a Gaussian charge distribution:

$$\rho_{ik}(\mathbf{r}) = \Big(\frac{\alpha_{ik}}{\pi}\Big)^{\frac{3}{2}} q_{ik} \exp(-\alpha_{ik}|\mathbf{r} - \mathbf{r}_{ik}|^2),$$

where index k labels core c or shell s, and the width of the distribution is described by the parameter $\alpha_{ik} = \frac{1}{2R_{ik}^2}$, where R_{ik} is the atomic radius. The electrostatic energy between two Gaussian charges at distance r can be shown to be:

$$C_{ik,jl}(r) = \frac{1}{r}\mathrm{erf}\Big(\sqrt{\frac{\alpha_{ik}\alpha_{jl}}{\alpha_{ik}+\alpha_{jl}}}r\Big).$$

Note that for Gaussian charges the energy does not diverge as $r \to 0$, the exact expression at $r = 0$ is:

$$C^0_{ik,jl} = \lim_{r\to 0} C_{ik,jl}(\mathbf{r}) = \frac{2}{\sqrt{\pi}}\sqrt{\frac{\alpha_{ik}\alpha_{jl}}{\alpha_{ik}+\alpha_{jl}}}.$$

In PQEQ, to prevent the shell from drifting away from the nucleus, the shell is attached to the core with a 2nd and/or 4th order spring force, and the total electrostatic energy is then expressed as:

$$E = \sum_i \Bigg(\Big(\chi_i + (C(r_{icis}) - C^0_{icis} + J_i)q_{ic}\Big)q_{is} + \frac{1}{2}J_i q_{is}^2 + K_2 r_{icis}^2 + K_4 r_{icis}^4 \Big) + \sum_{i>j} \Big(C(r_{icjc})q_{ic}q_{jc} + C(r_{isjs})q_{is}q_{js} + C(r_{icjs})q_{ic}q_{js} + C(r_{isjc})q_{is}q_{jc} \Big).$$

Most of the contribution to the total electrostatic energy comes from near neighbor interactions, so we do not expect that the difference between the tail of Slater and Gaussian functions to cause a large error. Nevertheless, the parameters of the original QEQ were shown to produce charges with a large mean squared error and therefore must be optimized to work with Gaussian functions [7]. After parameter optimization, the Gaussian QEQ produces the desired point charges while also enabling description new phenomena, such as polarization.

6.4 Solving for Charges Using Pseudo-Dynamics

When computing charges, minimizing the total electrostatic energy (Equation 6.1) with respect to the charges ($\frac{\partial E}{\partial q_i} = 0$) and charge conservation ($\sum q_i = Q_{total}$) leads to a system of $N + 1$ linear equations. The exact solution by matrix inversion is expensive and scales as $O(N^3)$. This approach is used by the current implementation of QEQ [113] in the MD simulation package LAMMPS [111].

Since the charges are changing slowly during the MD run, in practice only M iterations are needed every time step to obtain converged values, decreasing the scaling close to $O(MN)$. An alternative approach, pioneered by Rick et al. [112], is to evolve the charges according to an equation of motion with damping, which has linear scaling $O(N)$. We implemented this fast method.

To evolve the charge variables dynamically, we assign virtual mass m_Q to the charges q_i, and solve the corresponding equation of motion:

$$m_Q \ddot{q}_i = F_{Q,i} - \gamma_Q \dot{q}_i + \lambda \tag{6.2}$$

$$\lambda = -\frac{1}{N_Q} \sum_i (F_{Q,i} - \gamma_Q \dot{q}_i) \tag{6.3}$$

where the generalized force is $F_{Q,i} = -\frac{\partial E}{\partial q_i}$ and λ is the Lagrange multiplier enforcing charge conservation. In to the equation of motion $\frac{\partial E}{\partial q_i} = 0$, we added the attenuation force $F_{Q,\text{friction}} = -\gamma_Q \dot{q}_i$, so that the dynamics minimizes the total energy.

The value of the virtual mass m_Q must be chosen carefully. Smaller values of m_Q lead to larger changes in atomic charges each timestep. For m_Q too small the numerical integration will become unstable. For large m_Q the charges will converge slowly. For fastest convergence, m_Q and γ_Q should be set to critically damp the system. This method of solving the system of N linear equations scales as $O(N)$ and thus scales well to large systems.

The side effect of solving the charges with damped dynamics is that the charges lag behind the values which would be obtained by exact minimization of the electrostatic energy. With a proper choice of the damping constant, this effect can keep charges approximately constant during bond vibrations. Averaging the charges over bond vibrations is preferable to letting the charges fluctuate nonphysically during bond vibrations.

6.5 Charge Convergence

Instead of m_Q and γ_Q we would like to input a parameter controlling the damping time τ and the damping ratio ζ. One can express the QEQ equation of motion (Equation 6.2) in the form of a damped harmonic oscillator:

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = 0,$$

where ζ is the damping ratio and $\zeta = 1$ corresponds to critical damping. The solution of the critically damped harmonic oscillator is $(A + Bt)e^{-\omega t}$. Let's say that the initial condition was $x(0) = A$ and $\dot{x}(0) = 0$; then the solution writes $A(1 + \omega t)e^{-\omega t}$. This attenuates to $1/e$ when $\omega t \approx 2.15$, so we can express $\omega = 2.15/\tau$ in terms of the damping time τ.

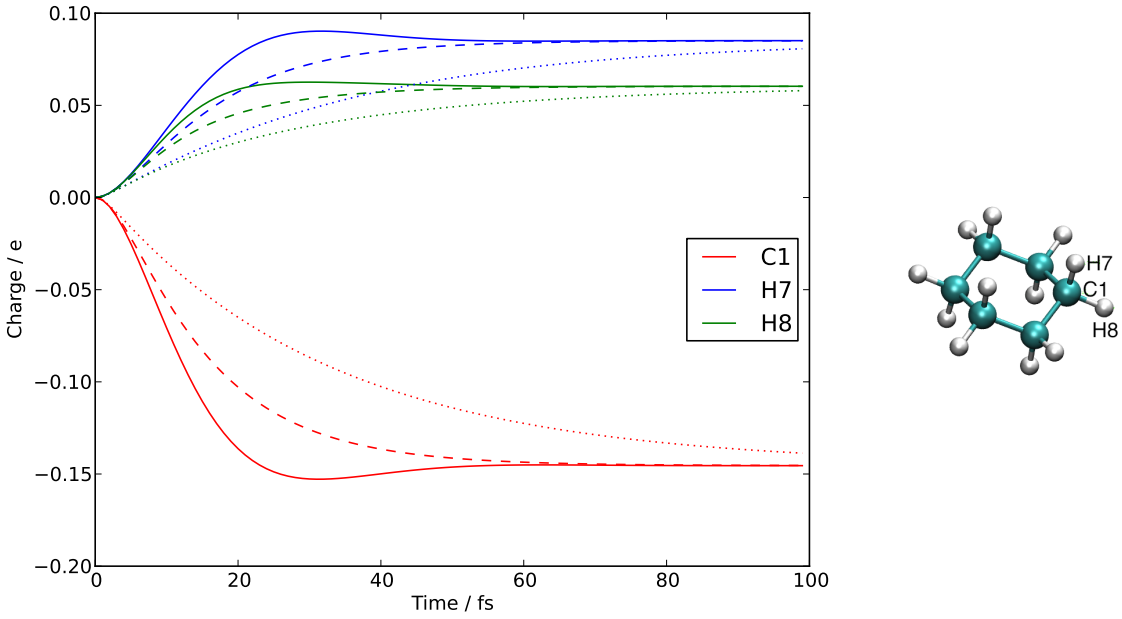


Figure 6.1: QEq pseudodynamics convergence: A cyclohexane molecule has fixed geometry and only the charges dynamics are run. Charges are plotted for a carbon C1 and its two bonded hydrogens H7 and H8. The charges for three values of ζ are plotted: solid line $\zeta = 0.6$ (underdamped), dashed line $\zeta = 1$ (critically damped), and dotted line $\zeta = 2$ (overdamped). $\tau = 15fs$.

Comparing the canonical damped harmonic oscillator to the QEQ equation of motion yields:

$$\gamma_Q/m_Q = 2\zeta\omega = 4.30\zeta/\tau$$
$$1/m_Q \sim \omega^2 \sim 1/\tau^2$$

where m_Q is in mass units. From a test run we approximated $m_Q \approx (\tau/11.3)^2$, where τ is in fs and m_Q in grams/mole ('real' units in LAMMPS). Now we can express the parameters controlling the dynamics in terms of the damping time constant τ and the damping ratio ζ:

$$m_Q = (\tau/11.3)^2$$
$$\gamma_Q = 4.30\zeta m_Q/\tau$$

For dynamics, we expect to use values $\tau = 30fs$ and $\zeta = 1$. When atomic coordinates are fixed and only charges are solved for, values $\tau = 15fs$ and $\zeta = 0.6$ give good convergence in about 100 steps.

By setting the time constant of the QEQ dynamics longer than a bond vibration period we can smooth the charge value assigned to each atom during a molecular dynamics run. As an example of the charge pseudynamics we choose a cyclohexane molecule and set $\tau = 15fs$. Figure 6.1 shows

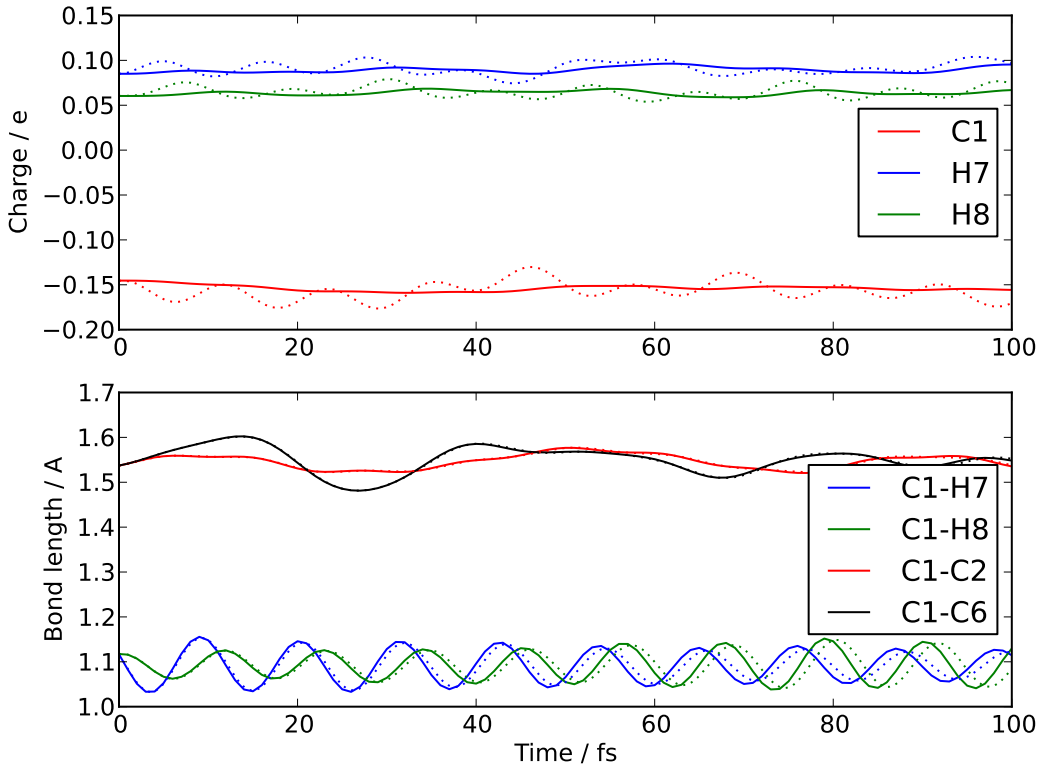


Figure 6.2: Molecular dynamics of a single cyclohexane molecule with atomic velocities corresponding to 300 K. For a large time constant $\tau = 15fs$ (solid line) the charges are averaged over the period of a bond vibration, but for a short time constant $\tau = 1.5fs$ (dotted line) the charges adjust as the bond length changes during bond vibrations. The damping ratio is set to critical damping $\zeta = 1$.

the evolution of charges during the initial run when charges are allowed to change, but the atomic geometry is fixed. Within 40 fs (40 steps), the error on all of the charges is less than 0.05 from the exact solution, and in 100 fs the error is less than 0.01.

After initialization of the charges, we assign random velocities to the atoms corresponding to 300 K and let the atomic positions evolve along with the atomic charges. Bonded and VDW parameters are from the Amber99 force field [78]. Figure 6.2 shows that the damping parameters keep the charges approximately constant during bond vibrations.

6.6 Optimization of Shell Positions Using Newton-Raphson

So far we described how to solve for the charges using dynamics in the charge degree of freedom. In PQEQ, we also need to determine the shell positions. One could split the atomic mass between the shell and core and evolve this degree of freedom as well. However, the potential confining shells near each core is spherically symmetric and approximately harmonic, and so we can find the optimal

position of shells $\mathbf{r}_{is}$ using the Newton-Raphson method. In each MD step we run one iteration of the Newton-Raphson method, and since the charges need at least 10 steps for convergence this is enough for convergence of the shell positions as well.

To get the second derivative we neglect the effect of the remote charges, so we do not need to invert the Hessian. If the force $F_{RS,i}$, position of the core $\mathbf{r}_{i,c}$, and position of the shell lie on a line, we can estimate the new position of the shell as:

$$\begin{aligned} F_{RS,i} &= -\frac{\partial E}{\partial x_{is}} \\ x_{is,new} &= x_{is,old} - \frac{E'(x_{is,old})}{E''(x_{is,old})} \\ &= x_{is,old} + \frac{F_{x,is}\Big|_{x_{is,old}}}{\frac{\partial^2 E}{\partial x_{is}^2}\Big|_{x_{is,old}}} \end{aligned}$$

Since we neglect the remote atoms for computation of the second derivative, we have:

$$\frac{\partial^2 E}{\partial x_{is}^2} = \frac{\partial^2 j(r_{is,ic})}{\partial x_{is}^2} q_{ic} q_{is} + 2K_2 + 12K_4 x_{is}^2.$$

The problem is not one-dimensional, but we can keep the above second derivative, and for each shell rotate into the direction in which the force acts.

6.7 Ewald Summation for Periodic Systems

Let us consider the total electrostatic energy of a charge-neutral periodic system:

$$E = \frac{1}{2} \sum_{i,j} \sum_{\mathbf{R}} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{R}|}, \tag{6.4}$$

where $\sum_{\mathbf{R}}$ is the sum over primitive cells of the system ($\mathbf{R} = \mathbf{0}$ is excluded when $i = j$). The $1/r$ potential is long ranged so a large number of particles have to be included in the sum. The sum 6.4 can be converted into a k-space sum [114], in which case there is a problem with convergence for small $\mathbf{K}$, caused by the Fourier transform of the point charge (delta function). If the point charges are replaced by charge distributions, the sum in the k-space converges.

Ewald [115] replaced the point charges with Gaussian distributions:

$$\rho_i(\mathbf{r}) = q_i \Big(\frac{\alpha_E}{\pi}\Big)^{\frac{3}{2}} \exp(-\alpha_E |\mathbf{r} - \mathbf{r}_i|^2)$$

and summed the total energy in k-space. The energy coming from the difference in the point charges and the Gaussian charge converges quickly in real space. In order to stay close to the current

implementation of Ewald summation in LAMMPS, we apply an analogous technique to PQEQ. We replace each Gaussian charge of width α_i with a Gaussian charge of width α_E and sum the energy in k-space; the difference of the charge distributions is summed in real space.

Note that the electrostatic energy of two Gaussian charges is $\frac{q_i q_j}{r_{ij}}\mathrm{erf}(\sqrt{\alpha_{ij}}r_{ij})$, where $\alpha_{ij} = \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}$, thus the total electrostatic energy is:

$$E = \frac{1}{2}\sum_{i,j}\sum_{\mathbf{R}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{R}|}\mathrm{erf}(\sqrt{\alpha_{ij}}|\mathbf{r}_{ij} + \mathbf{R}|),$$

where again $\mathbf{R} = \mathbf{0}$ is excluded when $i = j$. Schematically adding of the Ewald Gaussian distribution corresponds to:

$$\sum_{i,j}\mathrm{Gauss}_i\mathrm{Gauss}_j = \underbrace{\sum_{i,j}\left[\mathrm{Gauss}_i\mathrm{Gauss}_j - \mathrm{Gauss}_i\mathrm{Gauss}_j^{EW}\right]}_{\text{summed in real space}} + \underbrace{\sum_{i,j}\mathrm{Gaussian}_i\mathrm{Gaussian}_j^{EW}}_{\text{summed in k space}}$$

Thus we can split the total energy sum into two parts $E = E_{\text{real space}} + E_{\text{k space}}$. The first sum can be truncated for some short cut-off distance r_{cut}:

$$\begin{aligned} E_{\text{real space}} &= \frac{1}{2}\sum_{i,j}\sum_{\mathbf{R}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{R}|}\left[\mathrm{erf}(\sqrt{\alpha_{ij}}|\mathbf{r}_{ij} + \mathbf{R}|) - \mathrm{erf}(\sqrt{\alpha_{iE}}|\mathbf{r}_{ij} + \mathbf{R}|)\right] \\ &\approx \sum_{i<j, r_{ij}<r_{cut}} \frac{q_i q_j}{r_{ij}}\left[\mathrm{erf}(\sqrt{\alpha_{ij}}r_{ij}) - \frac{1}{2}\mathrm{erf}(\sqrt{\alpha_{iE}}r_{ij}) - \frac{1}{2}\mathrm{erf}(\sqrt{\alpha_{jE}}r_{ij})\right] \end{aligned}$$

The k-space sum requires a bit more work:

$$\begin{aligned} E_{\text{k space}} &= \frac{1}{2}\sum_{i,j}\sum_{\substack{\mathbf{R} \\ \mathbf{R}\neq 0 \text{ for } i=j}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{R}|}\mathrm{erf}(\sqrt{\alpha_{iE}}|\mathbf{r}_{ij} + \mathbf{R}|) \\ &= \frac{1}{2}\sum_{i,j}\sum_{\mathbf{R}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{R}|}\mathrm{erf}(\sqrt{\alpha_{iE}}|\mathbf{r}_{ij} + \mathbf{R}|) - \frac{1}{2}\sum_{i=j} q_i q_j 2\sqrt{\frac{\alpha_{iE}}{\pi}} \end{aligned}$$

after separating the $\mathbf{R} = 0$ term. Next, we use the following 3-dimensional Fourier transforms:

$$\begin{aligned} \frac{\mathrm{erf}(\sqrt{\alpha}r)}{r} &= \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\frac{4\pi}{k^2}e^{-\frac{k^2}{4\alpha}}e^{i\mathbf{k}\cdot\mathbf{r}} \\ \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} &= \frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta^3(\mathbf{K} - \mathbf{k}) \end{aligned}$$

and finally the k-space sum becomes:

$$E_{\text{k space}} = \frac{2\pi}{V} \sum_{i,j} \sum_{\mathbf{K}\neq 0} q_i q_j \frac{1}{K^2} e^{-\frac{K^2}{4\alpha_{iE}}} e^{-i\mathbf{K}\cdot(\mathbf{r}_i - \mathbf{r}_j)} - \sum_i q_i^2 \sqrt{\frac{\alpha_{iE}}{\pi}}$$
$$= \frac{2\pi}{V} \sum_{\mathbf{K}\neq 0} \frac{1}{K^2} \Big(\sum_i q_i e^{-\frac{K^2}{4\alpha_{iE}}} e^{-i\mathbf{K}\cdot\mathbf{r}_i} \Big) \Big(\sum_j q_j e^{i\mathbf{K}\cdot\mathbf{r}_j} \Big) - \sum_i q_i^2 \sqrt{\frac{\alpha_{iE}}{\pi}}$$

This expression converges rapidly due to K^2 in the exponential and in the denumerator, thus the summation can be terminated after a small number of terms. The derivation of Ewald summation for Gaussian charges followed the Ewald summation for point charges, and so implementation in a molecular dynamics code can follow the implementation of the usual Ewald method.

6.8 Conclusion

We extended the QEQ charge equilibration method to include the description of the atomic polarization and implemented it in Lammps. Optimization of atomic parameters for the first eighteen elements of the periodic table is presented in [7]. In the context of biological simulations it is useful for virtual ligand screening, where appropriate ligand charges can be derived on the fly for any organic molecule. Furthermore, this method includes the screening effect of the nearby atoms. For example, charges of an amino acid side-chain adjust depending on if the side-chain is exposed to the solvent or if it is buried in a hydrophobic environment. The PQEQ method is a step towards the next generation of reactive force fields.

Appendices

Appendix A

Amino Acids

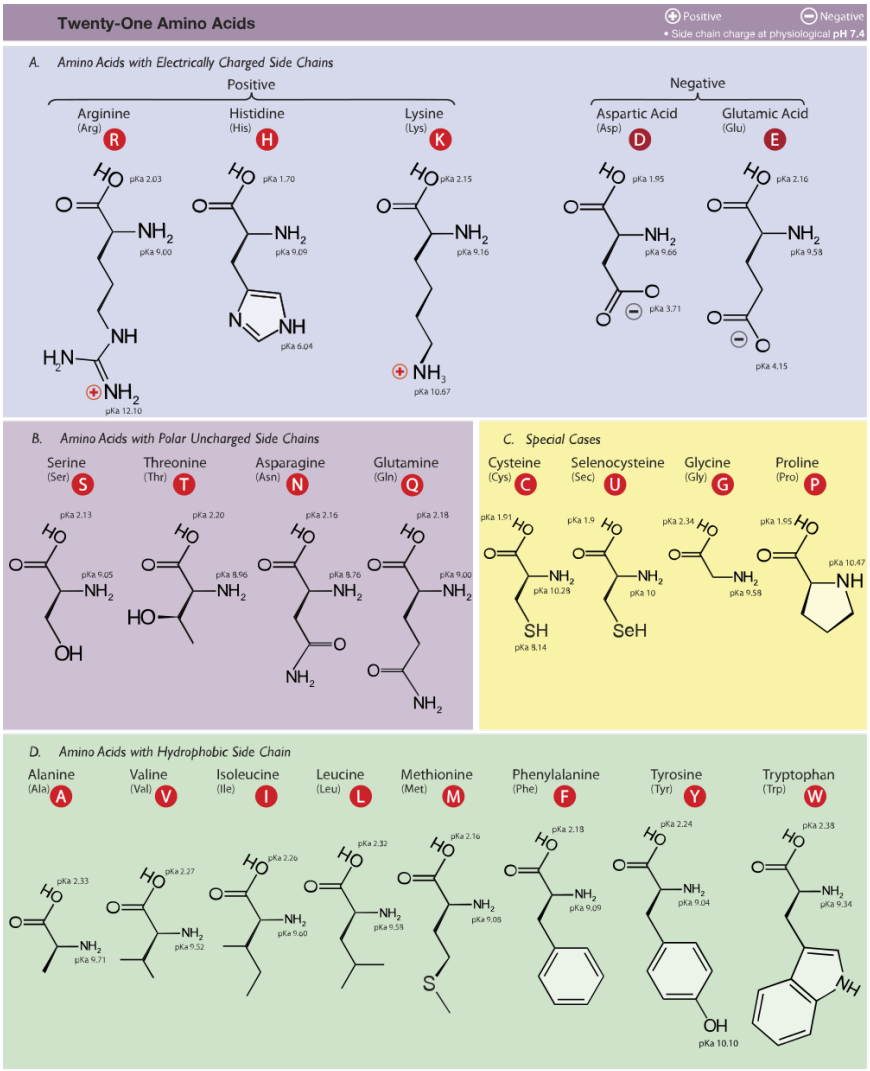


Figure A.1: Twenty one amino acids. Amino acids connect by peptide bond -CO-NH- and form long chains, proteins. Image from [116].

Appendix B

Comparison of GPCR Crystal Structures

Protein	TM1	TM2	TM3	TM4	TM5	TM6	TM7
RHO	P34-Q64 (N55)	P71-H100 (D83)	G106-C140 (R135)	E150-L172 (W161)	N200-V230 (P215)	A241-F276 (P267)	I286-M309 (P303)
RHOact	P34-Q64 (N55)	P71-H100 (D83)	G106-C140 (R135)	E150-L172 (W161)	N200-A235 (P215)	A241-T277 (P267)	I286-M308 (P303)
Beta1AR	W40-S68 (N59)	L75-R104 (D87)	S111-T144 (R139)	R155-M179 (W166)	R205-E236 (P219)	E285-F315 (P305)	D322-C344 (P340)
Beta2AR	V31-K60 (N51)	V67-M96 (D79)	N103-T136 (R131)	K147-M171 (W158)	Q197-Q229 (P211)	K267-I298 (P288)	K305-C327 (P323)
Beta2ARact	V31-K60 (N51)	V67-T96 (D79)	F104-T136 (R131)	K147-M171 (W158)	Q197-S236 (P211)	L266-I298 (P288)	K305-C327 (P323)
D3	Y32-K56 (N47)	T63-V91 (D75)	R100-V133 (R128)	C147-F170 (W158)	P186-K216 (P200)	L322-T353 (P344)	P362-F386 (P380)
H1	P29-S54 (N45)	G62-L89 (D73)	G96-Q130 (R125)	K141-L163 (W152)	T188-A216 (P202)	N408-C441 (P430)	H448-C471 (P465)
M2	V23-V50 (N41)	V57-V85 (D69)	P93-T126 (R121)	T137-V166 (W148)	A184-S213 (P198)	K384-F412 (P402)	N419-C443 (P437)
M2act	F21-V50 (N41)	V57-I86 (D69)	P93-T126 (R121)	T137-V166 (W148)	A184-S213 (P198)	R381-F412 (P402)	N419-L442 (P437)
M3	W65-V94 (N85)	V101-M130 (D113)	N137-T170 (R165)	T181-V210 (W192)	P228-E256 (P242)	L492-F515 (P505)	K522-C546 (P540)
5HT1B	L46-R76 (N67)	P83-T112 (D95)	Q119-T152 (R147)	P163-F185 (W174)	I206-R238 (P220)	K311-P338 (P329)	L348-S372 (P366)
5HT2B	L54-L81 (N72)	T89-M116 (D100)	V126-K158 (R153)	Q165-K193 (W180)	K211-K247 (P229)	T315-L349 (P339)	Q355-L382 (P377)
A2A	S7-L33 (N24)	T41-S67 (D52)	C74-I106 (R102)	G118-L141 (W129)	M174-A204 (P189)	S223-F258 (P248)	L267-R291 (P285)
A2Aact	G5-L33 (N24)	V40-S67 (D52)	C74-R107 (R102)	G118-L141 (W129)	M174-A204 (P189)	S223-F258 (P248)	L267-Y290 (P285)
S1P1	E42-K72 (N63)	P79-L104 (D91)	P114-M146 (R142)	F158-I179 (W168)	K200-T230 (L213)	S249-V280 (P271)	A293-T314 (P308)
NTS1act	I61-L89 (N82)	Q99-I129 (D113)	D139-C172 (R167)	T186-T207 (W194)	T231-T265 (P249)	A302-C332 (P323)	T341-S373 (P366)
CXCR4	N37-Y65 (N56)	M72-V99 (D84)	N106-V139 (R134)	Q145-F174 (W161)	D193-K225 (P211)	G231-L266 (P254)	C274-Y302 (P299)
CCR5	K26-N57 (N48)	M64-A91 (D76)	N98-V131 (R126)	V142-I165 (W153)	Y187-L222 (P206)	R230-T259 (P250)	C269-V300 (P294)
KappaOR	P56-R86 (N77)	A93-M121 (D105)	G127-A159 (R156)	R170-L196 (W183)	Y219-L259 (P238)	R267-L299 (P289)	L309-L333 (P327)
MuOR	V66-R95 (N86)	A102-M130 (D114)	I138-C170 (R165)	P181-M205 (W192)	T225-R258 (P244)	R273-L305 (P295)	T312-L339 (P333)
NOP	L48-L77 (N69)	A85-L113 (D97)	G119-C153 (R148)	S164-M188 (W175)	Q208-L242 (P227)	R252-G287 (P278)	E295-L322 (P316)
DeltaOR	S42-R76 (N67)	A83-M111 (D95)	E118-C151 (R146)	P162-M186 (W173)	S206-L238 (P225)	R258-L286 (P276)	P294-L321 (P315)
PAR1	W100-M131 (N120)	P136-F163 (D148)	S172-V205 (R200)	L216-L239 (W227)	G265-L297 (P282)	R305-S338 (P328)	E347-A374 (P368)
P2Y12	Y21-F51 (N43)	N58-A85 (D70)	R93-T127 (R122)	L138-I161 (W149)	E181-R222 (N201)	N235-T260 (P251)	C270-L301 (P295)
CRF1	Y116-R143 (L134)	L150-T175 (F162)	G186-V218 (L213)	R225-Y252 (W236)	D269-T296 (V279)	S304-A330 (L329)	V343-N367 (S360)
GLR	E127-L163 (L156)	T172-R199 (F184)	S217-L255 (L249)	R261-F289 (W272)	M301-R334 (A314)	Y343-F367 (V364)	G375-L403 (A397)
MGLU1	I591-L616 (T607)	R628-I649 (I638)	T654-I685 (I682)	A703-M727 (I714)	N750-T774 (L763)	E783-G807 (A800)	K811-I839 (L827)
MGLU5	P578-I603 (T594)	R615-I636 (I625)	Q641-K677 (A669)	A690-M714 (I701)	N737-T761 (L750)	E770-G794 (A787)	K798-A827 (L814)
SMO	E224-A254 (T245)	P263-F285 (F274)	L312-K344 (W339)	T357-V378 (W365)	Y397-N432 (V411)	N446-F475 (I465)	L515-V536 (S533)

Table B.1: The start and end of each transmembrane helix for known crystal structrures. The residue in parenthesis is Ballesteros-Weinstein .50 residue. Section 2.4 describes how the helix start and end were determined.

	RHO	RHOact	Beta1AR	Beta2AR	Beta2ARact	D3	H1	M2	M2act	M3	5HT1B	5HT2B	A2A	A2Aact	S1P1	NTS1act	CXCR4	CCR5	KappaOR	MuOR	NOP	DeltaOR	PAR1	P2Y12	CRF1	GLR	MGLU1	MGLU5	SMO
RHO	208	143	127	132	120	136	123	115	111	108	126	117	119	106	106	115	124	138	128	129	136	124	117	107	91	80	87	79	100
RHOact	143	204	111	121	130	124	111	113	130	107	131	112	107	112	100	134	104	123	116	114	119	116	120	116	90	79	83	81	95
Beta1AR	127	111	192	177	143	139	125	143	141	136	139	128	120	106	111	115	108	137	123	122	128	123	118	104	85	82	79	72	98
Beta2AR	132	121	177	210	159	148	136	146	142	140	149	132	128	111	118	124	116	141	132	128	136	131	123	114	90	86	86	82	106
Beta2ARact	120	130	143	159	198	133	125	132	151	126	149	126	109	110	105	133	102	122	114	115	119	117	116	120	89	75	83	81	92
D3	136	124	139	148	133	196	138	130	130	118	148	125	135	119	119	129	124	141	136	132	140	134	126	114	97	87	83	81	107
H1	123	111	125	136	125	138	189	132	129	128	140	120	125	105	112	118	106	126	123	115	120	120	114	109	92	81	82	83	96
M2	115	113	143	146	132	130	132	184	150	155	133	114	116	102	110	121	102	127	117	118	116	120	112	108	81	76	78	76	98
M2act	111	130	141	142	151	130	129	150	201	147	145	121	115	112	109	139	103	128	115	110	115	115	123	121	92	79	79	82	100
M3	108	107	136	140	126	118	128	155	147	179	132	111	110	103	103	110	92	124	114	110	114	112	113	108	78	76	76	74	91
5HT1B	126	131	139	149	149	148	140	133	145	132	210	138	117	118	114	129	106	131	120	119	125	117	121	116	103	86	82	83	101
5HT2B	117	112	128	132	126	125	120	114	121	111	138	198	106	113	91	112	104	116	103	108	108	105	102	91	82	75	79	74	95
A2A	119	107	120	128	109	135	125	116	115	110	117	106	194	145	108	114	99	128	124	117	118	123	119	101	85	82	82	79	92
A2Aact	106	112	106	111	110	119	105	102	112	103	118	113	145	199	93	116	89	111	108	105	108	111	117	96	85	74	74	71	76
S1P1	106	100	111	118	105	119	112	110	109	103	114	91	108	93	172	103	97	124	119	112	120	118	104	112	100	91	85	85	98
NTS1act	115	134	115	124	133	129	118	121	139	110	129	112	114	116	103	213	117	136	123	117	120	128	132	120	95	79	86	86	97
CXCR4	124	104	108	116	102	124	106	102	103	92	106	104	99	89	97	117	215	149	131	134	146	141	127	119	90	79	84	82	105
CCR5	138	123	137	141	122	141	126	127	128	124	131	116	128	111	124	136	149	232	151	155	166	161	143	137	106	98	96	90	108
KappaOR	128	116	123	132	114	136	123	117	115	114	120	103	124	108	119	123	131	151	197	156	159	163	131	118	95	83	89	86	101
MuOR	129	114	122	128	115	132	115	118	110	110	119	108	117	105	112	117	134	155	156	188	158	162	126	115	90	87	83	82	96
NOP	136	119	128	136	119	140	120	116	115	114	125	108	118	108	120	120	146	166	159	158	201	166	142	128	97	86	94	89	103
DeltaOR	124	116	123	131	117	134	120	120	115	112	117	105	123	111	118	128	141	161	163	162	166	199	137	124	93	84	88	85	98
PAR1	117	120	118	123	116	126	114	112	123	113	121	102	119	117	104	132	127	143	131	126	142	137	232	153	103	80	87	81	100
P2Y12	107	116	104	114	120	114	109	108	121	108	116	91	101	96	112	120	119	137	118	115	128	124	153	237	103	87	92	90	105
CRF1	91	90	85	90	89	97	92	81	92	78	103	82	85	85	100	95	90	106	95	90	97	93	103	103	195	107	86	89	106
GLR	80	79	82	86	75	87	81	76	79	76	86	75	82	74	91	79	79	98	83	87	86	84	80	87	107	178	86	89	92
MGLU1	87	83	79	86	83	83	82	78	79	76	82	79	82	74	85	86	84	96	89	83	94	88	87	92	86	86	213	178	94
MGLU5	79	81	72	82	81	81	83	76	82	74	83	74	79	71	85	86	82	90	86	82	89	85	81	90	89	89	178	202	93
SMO	100	95	98	106	92	107	96	98	100	91	101	95	92	76	98	97	105	108	101	96	103	98	100	105	106	92	94	93	212

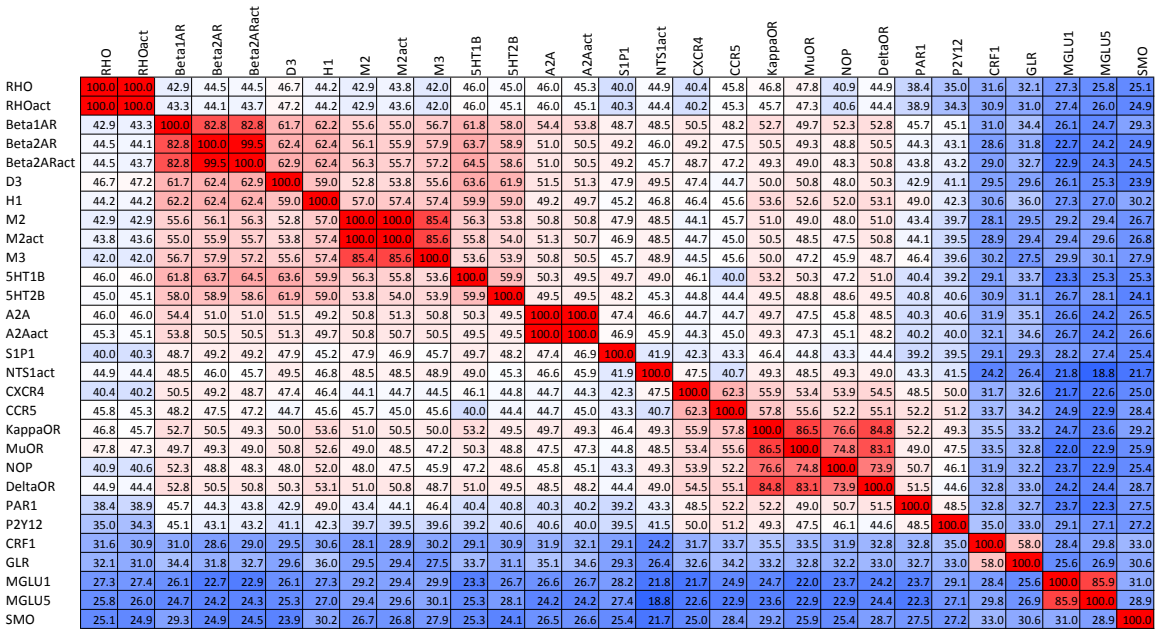
Figure B.1: Number of inter-helical contacts common to each pair of protein structures for the final sequence alignment. A contact is defined as any two heavy atoms closer than the sum of their Van der Waals radii plus 0.6 A.

	RHO	RHOact	Beta1AR	Beta2AR	Beta2ARact	D3	H1	M2	M2act	M3	5HT1B	5HT2B	A2A	A2Aact	S1P1	NTS1act	CXCR4	CCR5	KappaOR	MuOR	NOP	DeltaOR	PAR1	P2Y12	CRF1	GLR	MGLU1	MGLU5	SMO
RHO	0.00	2.22	2.12	2.07	3.21	1.68	1.96	2.25	2.66	2.40	2.15	2.00	2.36	2.72	2.26	2.91	2.16	2.08	2.50	1.85	1.86	1.99	2.62	2.91	3.21	2.88	3.14	3.25	3.03
RHOact	2.22	0.00	2.49	2.35	2.05	2.53	2.28	2.66	1.84	2.31	1.99	2.09	2.69	2.40	2.74	2.41	2.80	2.52	3.02	2.68	2.84	2.50	2.47	2.90	3.45	3.01	3.09	3.23	3.15
Beta1AR	2.12	2.49	0.00	0.60	2.54	1.38	1.53	1.34	2.09	1.53	1.51	1.91	1.89	2.58	1.77	2.70	2.75	2.24	2.32	2.32	2.20	2.21	2.96	3.14	3.14	2.85	2.95	3.11	3.09
Beta2AR	2.07	2.35	0.60	0.00	2.45	1.36	1.38	1.40	2.01	1.42	1.35	1.83	1.86	2.48	1.86	2.52	2.64	2.26	2.16	2.26	2.10	2.09	2.86	3.08	3.06	2.70	2.89	3.06	3.09
Beta2ARact	3.21	2.05	2.54	2.45	0.00	3.11	2.90	2.63	1.38	1.99	1.66	2.49	3.09	2.78	3.46	2.18	3.48	3.15	3.29	3.43	3.34	2.65	3.04	2.59	3.81	3.38	3.14	3.36	3.57
D3	1.68	2.53	1.38	1.36	3.11	0.00	1.36	1.66	2.36	1.80	1.44	1.71	1.58	2.12	1.61	2.36	2.01	1.80	1.84	1.64	1.58	1.70	2.75	2.79	2.63	2.54	2.79	2.91	2.44
H1	1.96	2.28	1.53	1.38	2.90	1.36	0.00	1.48	2.19	1.35	1.44	1.75	1.94	2.35	1.87	2.43	2.12	1.79	1.83	1.79	1.71	1.69	2.65	2.76	2.81	2.67	2.89	2.98	2.56
M2	2.25	2.66	1.34	1.40	2.63	1.66	1.48	0.00	2.23	1.10	1.96	2.12	2.06	2.72	1.95	2.96	2.52	2.13	2.19	2.12	2.06	2.08	3.26	3.16	3.16	3.06	3.04	3.19	2.97
M2act	2.66	1.84	2.09	2.01	1.38	2.36	2.19	2.23	0.00	1.72	1.31	1.89	2.44	2.14	2.68	2.02	3.12	2.88	2.82	2.89	2.72	2.55	2.83	2.87	3.15	3.19	2.89	3.09	3.40
M3	2.40	2.31	1.53	1.42	1.99	1.80	1.35	1.10	1.72	0.00	1.77	1.99	2.03	2.48	1.97	2.62	2.70	2.29	2.29	2.40	2.23	2.22	3.13	3.14	3.12	3.01	2.86	3.04	3.05
5HT1B	2.15	1.99	1.51	1.35	1.66	1.44	1.44	1.96	1.31	1.77	0.00	1.64	1.92	2.01	2.15	1.99	2.59	2.37	2.33	2.35	2.21	2.24	2.68	2.97	2.83	2.76	2.75	2.90	3.03
5HT2B	2.00	2.09	1.91	1.83	2.49	1.71	1.75	2.12	1.89	1.99	1.64	0.00	2.07	2.07	2.25	2.40	2.55	2.13	2.26	2.29	2.38	2.31	2.84	3.14	3.02	2.85	2.85	2.99	2.86
A2A	2.36	2.69	1.89	1.86	3.09	1.58	1.94	2.06	2.44	2.03	1.92	2.07	0.00	1.63	1.70	2.51	2.65	2.22	2.53	2.33	2.38	2.27	3.16	3.18	3.02	3.03	2.96	3.06	3.00
A2Aact	2.72	2.40	2.58	2.48	2.78	2.12	2.35	2.72	2.14	2.48	2.01	2.07	1.63	0.00	2.34	2.08	2.98	2.79	3.07	2.83	2.85	2.75	3.06	3.10	2.81	3.07	3.00	3.15	3.11
S1P1	2.26	2.74	1.77	1.86	3.46	1.61	1.87	1.95	2.68	1.97	2.15	2.25	1.70	2.34	0.00	3.02	2.60	2.12	2.74	2.24	2.15	2.28	3.07	3.19	2.76	2.90	2.74	2.83	2.81
NTS1act	2.91	2.41	2.70	2.52	2.18	2.36	2.43	2.96	2.02	2.62	1.99	2.40	2.51	2.08	3.02	0.00	2.89	3.06	2.63	2.82	2.67	2.52	2.43	2.83	2.99	2.90	3.00	3.19	3.31
CXCR4	2.16	2.80	2.75	2.64	3.48	2.01	2.12	2.52	3.12	2.70	2.59	2.55	2.65	2.98	2.60	2.89	0.00	1.79	2.32	1.78	1.97	1.87	2.89	2.81	3.13	2.93	3.14	3.25	2.52
CCR5	2.08	2.52	2.24	2.26	3.15	1.80	1.79	2.13	2.88	2.29	2.37	2.13	2.22	2.79	2.12	3.06	1.79	0.00	2.41	1.74	1.83	1.79	2.75	3.16	3.24	3.04	3.12	3.20	2.80
KappaOR	2.50	3.02	2.32	2.16	3.29	1.84	1.83	2.19	2.82	2.29	2.33	2.26	2.53	3.07	2.74	2.63	2.32	2.41	0.00	1.41	1.59	1.45	2.80	3.46	3.25	2.88	3.16	3.38	3.25
MuOR	1.85	2.68	2.32	2.26	3.43	1.64	1.79	2.12	2.89	2.40	2.35	2.29	2.33	2.83	2.24	2.82	1.78	1.74	1.41	0.00	1.02	0.84	2.77	2.91	2.96	2.64	3.11	3.27	2.66
NOP	1.86	2.84	2.20	2.10	3.34	1.58	1.71	2.06	2.72	2.23	2.21	2.38	2.38	2.85	2.15	2.67	1.97	1.83	1.59	1.02	0.00	1.09	2.59	3.24	2.99	2.73	3.06	3.20	2.82
DeltaOR	1.99	2.50	2.21	2.09	2.65	1.70	1.69	2.08	2.55	2.22	2.24	2.31	2.27	2.75	2.28	2.52	1.87	1.79	1.45	0.84	1.09	0.00	2.52	2.67	3.09	2.75	3.12	3.32	2.70
PAR1	2.62	2.47	2.96	2.86	3.04	2.75	2.65	3.26	2.83	3.13	2.68	2.84	3.16	3.06	3.07	2.43	2.89	2.75	2.80	2.77	2.59	2.52	0.00	2.23	3.59	3.35	3.43	3.66	3.52
P2Y12	2.91	2.90	3.14	3.08	2.59	2.79	2.76	3.16	2.87	3.14	2.97	3.14	3.18	3.10	3.19	2.83	2.81	3.16	3.46	2.91	3.24	2.67	2.23	0.00	3.43	3.24	3.27	3.52	3.31
CRF1	3.21	3.45	3.14	3.06	3.81	2.63	2.81	3.16	3.15	3.12	2.83	3.02	3.02	2.81	2.76	2.99	3.13	3.24	3.25	2.96	2.99	3.09	3.59	3.43	0.00	2.42	3.17	3.17	2.87
GLR	2.88	3.01	2.85	2.70	3.38	2.54	2.67	3.06	3.19	3.01	2.76	2.85	3.03	3.07	2.90	2.90	2.93	3.04	2.88	2.64	2.73	2.75	3.35	3.24	2.42	0.00	2.85	2.94	2.52
MGLU1	3.14	3.09	2.95	2.89	3.14	2.79	2.89	3.04	2.89	2.86	2.75	2.85	2.96	3.00	2.74	3.00	3.14	3.12	3.16	3.11	3.06	3.12	3.43	3.27	3.17	2.85	0.00	0.73	3.10
MGLU5	3.25	3.23	3.11	3.06	3.36	2.91	2.98	3.19	3.09	3.04	2.90	2.99	3.06	3.15	2.83	3.19	3.25	3.20	3.38	3.27	3.20	3.32	3.66	3.52	3.17	2.94	0.73	0.00	3.13
SMO	3.03	3.15	3.09	3.09	3.57	2.44	2.56	2.97	3.40	3.05	3.03	2.86	3.00	3.11	2.81	3.31	2.52	2.80	3.25	2.66	2.82	2.70	3.52	3.31	2.87	2.52	3.10	3.13	0.00

Figure B.2: Backbone RMSD of the TM bundle for the final sequence alignment.

	RHO	RHOact	Beta1AR	Beta2AR	Beta2ARact	D3	H1	M2	M2act	M3	5HT1B	5HT2B	A2A	A2Aact	S1P1	NTS1act	CXCR4	CCR5	KappaOR	MuOR	NOP	DeltaOR	PAR1	P2Y12	CRF1	GLR	MGLU1	MGLU5	SMO
RHO	100.0	100.0	20.7	21.5	21.5	27.4	20.3	21.9	22.4	20.2	21.5	21.5	24.0	23.6	21.5	22.7	22.2	23.4	19.2	23.4	21.2	22.2	18.7	18.8	12.3	11.9	10.2	10.1	14.4
RHOact	100.0	100.0	21.4	21.6	21.4	27.9	20.3	21.9	22.3	20.2	21.3	21.1	24.0	23.5	21.4	22.7	22.1	23.4	19.2	23.4	21.3	22.2	18.7	18.4	12.2	11.7	10.3	10.2	14.0
Beta1AR	20.7	21.4	100.0	68.6	69.0	42.3	37.8	30.3	30.2	34.5	45.7	40.5	37.4	36.7	29.3	22.4	24.2	26.1	28.4	26.1	27.1	28.4	21.1	23.6	14.1	12.0	9.7	11.2	8.9
Beta2AR	21.5	21.6	68.6	100.0	99.5	39.1	36.6	30.3	29.9	36.4	41.8	38.6	35.7	35.5	28.0	19.7	24.1	23.5	27.0	26.4	26.4	27.8	22.4	21.8	11.4	9.2	9.1	10.1	8.8
Beta2ARact	21.5	21.4	69.0	99.5	100.0	39.6	36.6	30.5	30.0	35.6	42.5	38.4	35.7	35.5	28.0	19.8	23.6	23.1	26.6	26.2	25.9	27.9	22.4	21.6	11.3	9.7	9.1	10.2	8.7
D3	27.4	27.9	42.3	39.1	39.6	100.0	36.4	32.3	33.5	32.3	43.6	37.6	32.7	32.5	26.8	27.8	27.8	23.9	25.8	28.9	29.3	29.5	23.7	24.5	12.6	13.8	8.5	7.3	8.7
H1	20.3	20.3	37.8	36.6	36.6	36.4	100.0	37.8	37.9	39.4	39.6	33.8	34.4	34.4	25.5	22.1	25.8	23.3	30.6	29.6	26.5	29.2	23.0	19.6	12.8	13.4	9.1	9.0	12.6
M2	21.9	21.9	30.3	30.3	30.5	32.3	37.8	100.0	100.0	73.4	35.5	31.0	26.7	26.9	28.7	24.7	23.6	22.3	29.3	28.8	27.3	28.3	24.2	22.7	11.9	8.3	11.2	8.3	10.2
M2act	22.4	22.3	30.2	29.9	30.0	33.5	37.9	100.0	100.0	72.1	35.2	31.0	26.9	26.9	28.1	24.2	23.1	22.0	28.7	28.2	26.7	28.1	23.8	22.6	11.8	8.1	11.3	8.4	10.5
M3	20.2	20.2	34.5	36.4	35.6	32.3	39.4	73.4	72.1	100.0	38.1	33.0	30.7	30.7	28.3	24.7	24.6	23.3	27.3	27.7	24.7	27.7	22.7	23.4	13.4	9.5	9.8	9.7	13.1
5HT1B	21.5	21.3	45.7	41.8	42.5	43.6	39.6	35.5	35.2	38.1	100.0	36.5	33.2	32.7	31.6	25.5	25.4	21.0	28.9	27.4	25.6	27.8	19.2	22.1	13.2	13.0	8.5	11.8	11.6
5HT2B	21.5	21.1	40.5	38.6	38.4	37.6	33.8	31.0	31.0	33.0	36.5	100.0	32.5	32.2	25.7	23.2	23.3	21.5	28.4	28.5	27.9	28.2	22.3	21.8	13.3	12.8	11.4	11.2	9.4
A2A	24.0	24.0	37.4	35.7	35.7	32.7	34.4	26.7	26.9	30.7	33.2	32.5	100.0	100.0	30.0	22.8	21.8	21.8	27.1	26.8	23.4	27.8	20.4	18.8	14.6	13.3	11.9	10.7	10.3
A2Aact	23.6	23.5	36.7	35.5	35.5	32.5	34.4	26.9	26.9	30.7	32.7	32.2	100.0	100.0	29.7	22.4	21.4	21.5	26.9	26.4	23.0	27.4	20.1	18.5	14.4	13.1	11.9	10.7	10.1
S1P1	21.5	21.4	29.3	28.0	28.0	26.8	25.5	28.7	28.1	28.3	31.6	25.7	30.0	29.7	100.0	23.6	22.8	22.2	21.4	21.4	21.6	22.2	17.0	20.0	11.0	12.8	12.6	9.1	10.8
NTS1act	22.7	22.7	22.4	19.7	19.8	27.8	22.1	24.7	24.2	24.7	25.5	23.2	22.8	22.4	23.6	100.0	25.5	22.0	29.1	23.0	28.0	24.8	20.7	20.0	12.1	11.7	7.5	5.1	10.1
CXCR4	22.2	22.1	24.2	24.1	23.6	27.8	25.8	23.6	23.1	24.6	25.4	23.3	21.8	21.4	22.8	25.5	100.0	33.8	28.4	30.1	32.0	27.7	25.7	27.2	14.8	11.9	10.3	10.2	10.1
CCR5	23.4	23.4	26.1	23.5	23.1	23.9	23.3	22.3	22.0	23.3	21.0	21.5	21.8	21.5	22.2	22.0	33.8	100.0	28.2	26.6	27.8	26.6	23.2	30.3	14.7	12.1	10.7	8.9	7.9
KappaOR	19.2	19.2	28.4	27.0	26.6	25.8	30.6	29.3	28.7	27.3	28.9	28.4	27.1	26.9	21.4	29.1	28.4	28.2	100.0	71.0	61.2	71.1	26.3	28.5	16.1	13.6	5.6	7.3	7.2
MuOR	23.4	23.4	26.1	26.4	26.2	28.9	29.6	28.8	28.2	27.7	27.4	28.5	26.8	26.4	21.4	23.0	30.1	26.6	71.0	100.0	61.9	72.0	26.4	28.7	15.7	14.4	6.8	8.9	5.8
NOP	21.2	21.3	27.1	26.4	25.9	29.3	26.5	27.3	26.7	24.7	25.6	27.9	23.4	23.0	21.6	28.0	32.0	27.8	61.2	61.9	100.0	61.4	29.2	27.2	17.6	15.6	6.8	8.9	8.5
DeltaOR	22.2	22.2	28.4	27.8	27.9	29.5	29.2	28.3	28.1	27.7	27.8	28.2	27.8	27.4	22.2	24.8	27.7	26.6	71.1	72.0	61.4	100.0	27.9	26.0	14.4	16.8	8.4	9.4	9.6
PAR1	18.7	18.7	21.1	22.4	22.4	23.7	23.0	24.2	23.8	22.7	19.2	22.3	20.4	20.1	17.0	20.7	25.7	23.2	26.3	26.4	29.2	27.9	100.0	25.2	15.6	14.8	7.3	9.5	10.6
P2Y12	18.8	18.4	23.6	21.8	21.6	24.5	19.6	22.7	22.6	23.4	22.1	21.8	18.8	18.5	20.0	20.0	27.2	30.3	28.5	28.7	27.2	26.0	25.2	100.0	14.4	17.0	10.9	11.9	8.4
CRF1	12.3	12.2	14.1	11.4	11.3	12.6	12.8	11.9	11.8	13.4	13.2	13.3	14.6	14.4	11.0	12.1	14.8	14.7	16.1	15.7	17.6	14.4	15.6	14.4	100.0	37.2	13.0	16.4	17.0
GLR	11.9	11.7	12.0	9.2	9.7	13.8	13.4	8.3	8.1	9.5	13.0	12.8	13.3	13.1	12.8	11.7	11.9	12.1	13.6	14.4	15.6	16.8	14.8	17.0	37.2	100.0	11.0	13.1	15.1
MGLU1	10.2	10.3	9.7	9.1	9.1	8.5	9.1	11.2	11.3	9.8	8.5	11.4	11.9	11.9	12.6	7.5	10.3	10.7	5.6	6.8	6.8	8.4	7.3	10.9	13.0	11.0	100.0	74.5	15.8
MGLU5	10.1	10.2	11.2	10.1	10.2	7.3	9.0	8.3	8.4	9.7	11.8	11.2	10.7	10.7	9.1	5.1	10.2	8.9	7.3	8.9	8.9	9.4	9.5	11.9	16.4	13.1	74.5	100.0	11.6
SMO	14.4	14.0	8.9	8.8	8.7	8.7	12.6	10.2	10.5	13.1	11.6	9.4	10.3	10.1	10.8	10.1	10.1	7.9	7.2	5.8	8.5	9.6	10.6	8.4	17.0	15.1	15.8	11.6	100.0

Figure B.3: Sequence identity (%) for the final sequence alignment.



	RHO	RHOact	Beta1AR	Beta2AR	Beta2ARact	D3	H1	M2	M2act	M3	5HT1B	5HT2B	A2A	A2Aact	S1P1	NTS1act	CXCR4	CCR5	KappaOR	MuOR	NOP	DeltaOR	PAR1	P2Y12	CRF1	GLR	MGLU1	MGLU5	SMO
RHO	100.0	100.0	42.9	44.5	44.5	46.7	44.2	42.9	43.8	42.0	46.0	45.0	46.0	45.3	40.0	44.9	40.4	45.8	46.8	47.8	40.9	44.9	38.4	35.0	31.6	32.1	27.3	25.8	25.1
RHOact	100.0	100.0	43.3	44.1	43.7	47.2	44.2	42.9	43.6	42.0	46.0	45.1	46.0	45.1	40.3	44.4	40.2	45.3	45.7	47.3	40.6	44.4	38.9	34.3	30.9	31.0	27.4	26.0	24.9
Beta1AR	42.9	43.3	100.0	82.8	82.8	61.7	62.2	55.6	55.0	56.7	61.8	58.0	54.4	53.8	48.7	48.5	50.5	48.2	52.7	49.7	52.3	52.8	45.7	45.1	31.0	34.4	26.1	24.7	29.3
Beta2AR	44.5	44.1	82.8	100.0	99.5	62.4	62.4	56.1	55.9	57.9	63.7	58.9	51.0	50.5	49.2	46.0	49.2	47.5	50.5	49.3	48.8	50.5	44.3	43.1	28.6	31.8	22.7	24.2	24.9
Beta2ARact	44.5	43.7	82.8	99.5	100.0	62.9	62.4	56.3	55.7	57.2	64.5	58.6	51.0	50.5	49.2	45.7	48.7	47.2	49.3	49.0	48.3	50.8	43.8	43.2	29.0	32.7	22.9	24.3	24.5
D3	46.7	47.2	61.7	62.4	62.9	100.0	59.0	52.8	53.8	55.6	63.6	61.9	51.5	51.3	47.9	49.5	47.4	44.7	50.0	50.8	48.0	50.3	42.9	41.1	29.5	29.6	26.1	25.3	23.9
H1	44.2	44.2	62.2	62.4	62.4	59.0	100.0	57.0	57.4	57.4	59.9	59.0	49.2	49.7	45.2	46.8	46.4	45.6	53.6	52.6	52.0	53.1	49.0	42.3	30.6	36.0	27.3	27.0	30.2
M2	42.9	42.9	55.6	56.1	56.3	52.8	57.0	100.0	100.0	85.4	56.3	53.8	50.8	50.8	47.9	48.5	44.1	45.7	51.0	49.0	48.0	51.0	43.4	39.7	28.1	29.5	29.2	29.4	26.7
M2act	43.8	43.6	55.0	55.9	55.7	53.8	57.4	100.0	100.0	85.6	55.8	54.0	51.3	50.7	46.9	48.5	44.7	45.0	50.5	48.5	47.5	50.8	44.1	39.5	28.9	29.4	29.4	29.6	26.8
M3	42.0	42.0	56.7	57.9	57.2	55.6	57.4	85.4	85.6	100.0	53.6	53.9	50.8	50.5	45.7	48.9	44.5	45.6	50.0	47.2	45.9	48.7	46.4	39.6	30.2	27.5	29.9	30.1	27.9
5HT1B	46.0	46.0	61.8	63.7	64.5	63.6	59.9	56.3	55.8	53.6	100.0	59.9	50.3	49.5	49.7	49.0	46.1	40.0	53.2	50.3	47.2	51.0	40.4	39.2	29.1	33.7	23.3	25.3	25.3
5HT2B	45.0	45.1	58.0	58.9	58.6	61.9	59.0	53.8	54.0	53.9	59.9	100.0	49.5	49.5	48.2	45.3	44.8	44.4	49.5	48.8	48.6	49.5	40.8	40.6	30.9	31.1	26.7	28.1	24.1
A2A	46.0	46.0	54.4	51.0	51.0	51.5	49.2	50.8	51.3	50.8	50.3	49.5	100.0	100.0	47.4	46.6	44.7	44.7	49.7	47.5	45.8	48.5	40.3	40.6	31.9	35.1	26.6	24.2	26.5
A2Aact	45.3	45.1	53.8	50.5	50.5	51.3	49.7	50.8	50.7	50.5	49.5	49.5	100.0	100.0	46.9	45.9	44.3	45.0	49.3	47.3	45.1	48.2	40.2	40.0	32.1	34.6	26.7	24.2	26.6
S1P1	40.0	40.3	48.7	49.2	49.2	47.9	45.2	47.9	46.9	45.7	49.7	48.2	47.4	46.9	100.0	41.9	42.3	43.3	46.4	44.8	43.3	44.4	39.2	39.5	29.1	29.3	28.2	27.4	25.4
NTS1act	44.9	44.4	48.5	46.0	45.7	49.5	46.8	48.5	48.5	48.9	49.0	45.3	46.6	45.9	41.9	100.0	47.5	40.7	49.3	48.5	49.3	49.0	43.3	41.5	24.2	26.4	21.8	18.8	21.7
CXCR4	40.4	40.2	50.5	49.2	48.7	47.4	46.4	44.1	44.7	44.5	46.1	44.8	44.7	44.3	42.3	47.5	100.0	62.3	55.9	53.4	53.9	54.5	48.5	50.0	31.7	32.6	21.7	22.6	25.0
CCR5	45.8	45.3	48.2	47.5	47.2	44.7	45.6	45.7	45.0	45.6	40.0	44.4	44.7	45.0	43.3	40.7	62.3	100.0	57.8	55.6	52.2	55.1	52.2	51.2	33.7	34.2	24.9	22.9	28.4
KappaOR	46.8	45.7	52.7	50.5	49.3	50.0	53.6	51.0	50.5	50.0	53.2	49.5	49.7	49.3	46.4	49.3	55.9	57.8	100.0	86.5	76.6	84.8	52.2	49.3	35.5	33.2	24.7	23.6	29.2
MuOR	47.8	47.3	49.7	49.3	49.0	50.8	52.6	49.0	48.5	47.2	50.3	48.8	47.5	47.3	44.8	48.5	53.4	55.6	86.5	100.0	74.8	83.1	49.0	47.5	33.5	32.8	22.0	22.9	25.9
NOP	40.9	40.6	52.3	48.8	48.3	48.0	52.0	48.0	47.5	45.9	47.2	48.6	45.8	45.1	43.3	49.3	53.9	52.2	76.6	74.8	100.0	73.9	50.7	46.1	31.9	32.2	23.7	22.9	25.4
DeltaOR	44.9	44.4	52.8	50.5	50.8	50.3	53.1	51.0	50.8	48.7	51.0	49.5	48.5	48.2	44.4	49.0	54.5	55.1	84.8	83.1	73.9	100.0	51.5	44.6	32.8	33.0	24.2	24.4	28.7
PAR1	38.4	38.9	45.7	44.3	43.8	42.9	49.0	43.4	44.1	46.4	40.4	40.8	40.3	40.2	39.2	43.3	48.5	52.2	52.2	49.0	50.7	51.5	100.0	48.5	32.8	32.7	23.7	22.3	27.5
P2Y12	35.0	34.3	45.1	43.1	43.2	41.1	42.3	39.7	39.5	39.6	39.2	40.6	40.6	40.0	39.5	41.5	50.0	51.2	49.3	47.5	46.1	44.6	48.5	100.0	35.0	33.0	29.1	27.1	27.2
CRF1	31.6	30.9	31.0	28.6	29.0	29.5	30.6	28.1	28.9	30.2	29.1	30.9	31.9	32.1	29.1	24.2	31.7	33.7	35.5	33.5	31.9	32.8	32.8	35.0	100.0	58.0	28.4	29.8	33.0
GLR	32.1	31.0	34.4	31.8	32.7	29.6	36.0	29.5	29.4	27.5	33.7	31.1	35.1	34.6	29.3	26.4	32.6	34.2	33.2	32.8	32.2	33.0	32.7	33.0	58.0	100.0	25.6	26.9	30.6
MGLU1	27.3	27.4	26.1	22.7	22.9	26.1	27.3	29.2	29.4	29.9	23.3	26.7	26.6	26.7	28.2	21.8	21.7	24.9	24.7	22.0	23.7	24.2	23.7	29.1	28.4	25.6	100.0	85.9	31.0
MGLU5	25.8	26.0	24.7	24.2	24.3	25.3	27.0	29.4	29.6	30.1	25.3	28.1	24.2	24.2	27.4	18.8	22.6	22.9	23.6	22.9	22.9	24.4	22.3	27.1	29.8	26.9	85.9	100.0	28.9
SMO	25.1	24.9	29.3	24.9	24.5	23.9	30.2	26.7	26.8	27.9	25.3	24.1	26.5	26.6	25.4	21.7	25.0	28.4	29.2	25.9	25.4	28.7	27.5	27.2	33.0	30.6	31.0	28.9	100.0

Figure B.4: Sequence similarity (%). Two residues are similar if their BLOSUM62 [117] entry is positive.

Appendix C

Alignment of GPCR Crystal Structures

In Figures figs. C.1 to C.7 we show the sequence alignment for the TM regions of the crystal structures. The central blue column labels the n.50 residues in BW numbering and the green color denotes extent of the TM regions for each crystal structure individually. Active structures (labels with extension act) are listed separately, as the TM length differ slightly. Prolines are highlighted in red, since they often cause helix kinks, and so are structurally important. The consensus sequence in the lower part of each figure mostly agrees with class A residues, because most of the crystal structures are from class A.

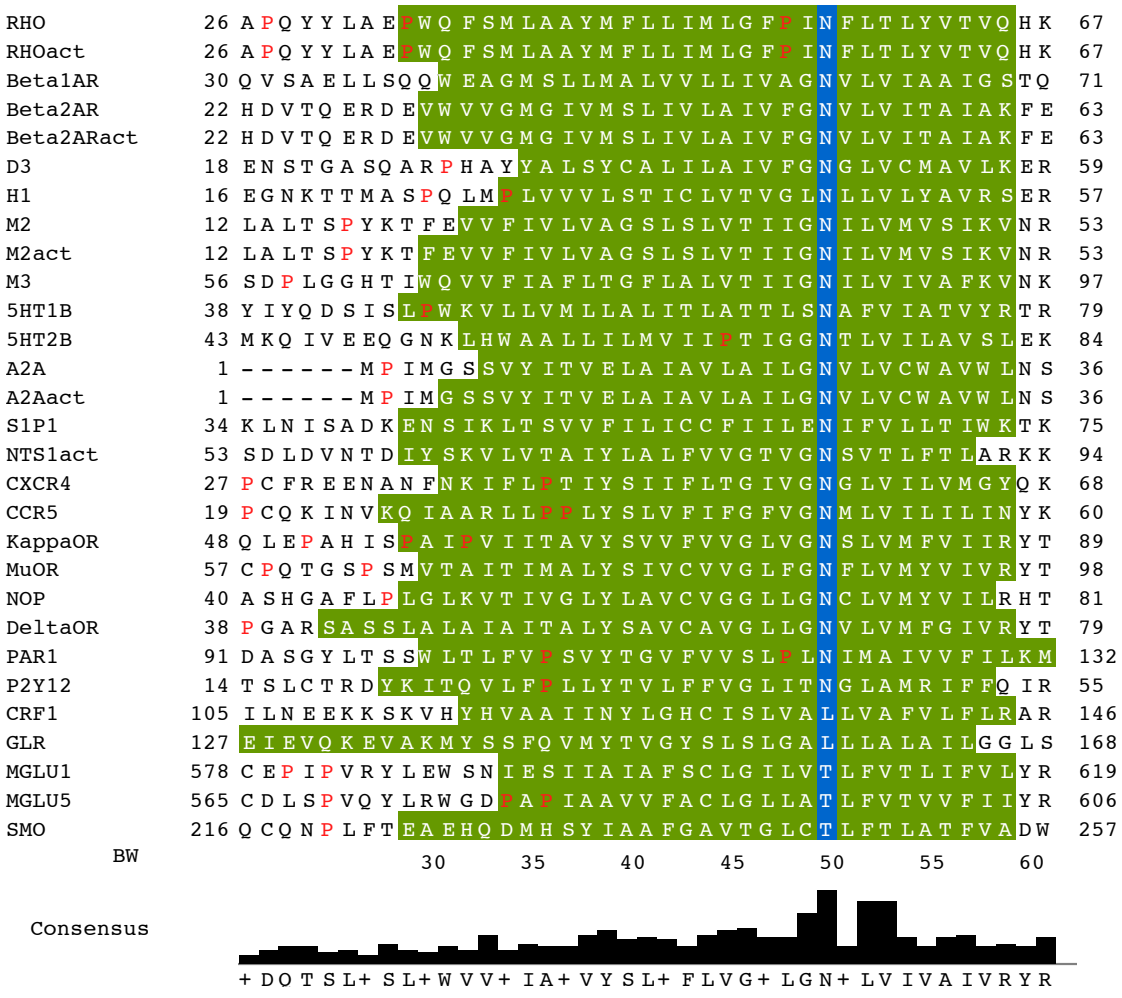


Figure C.1: TM 1 alignment for the crystal structures.

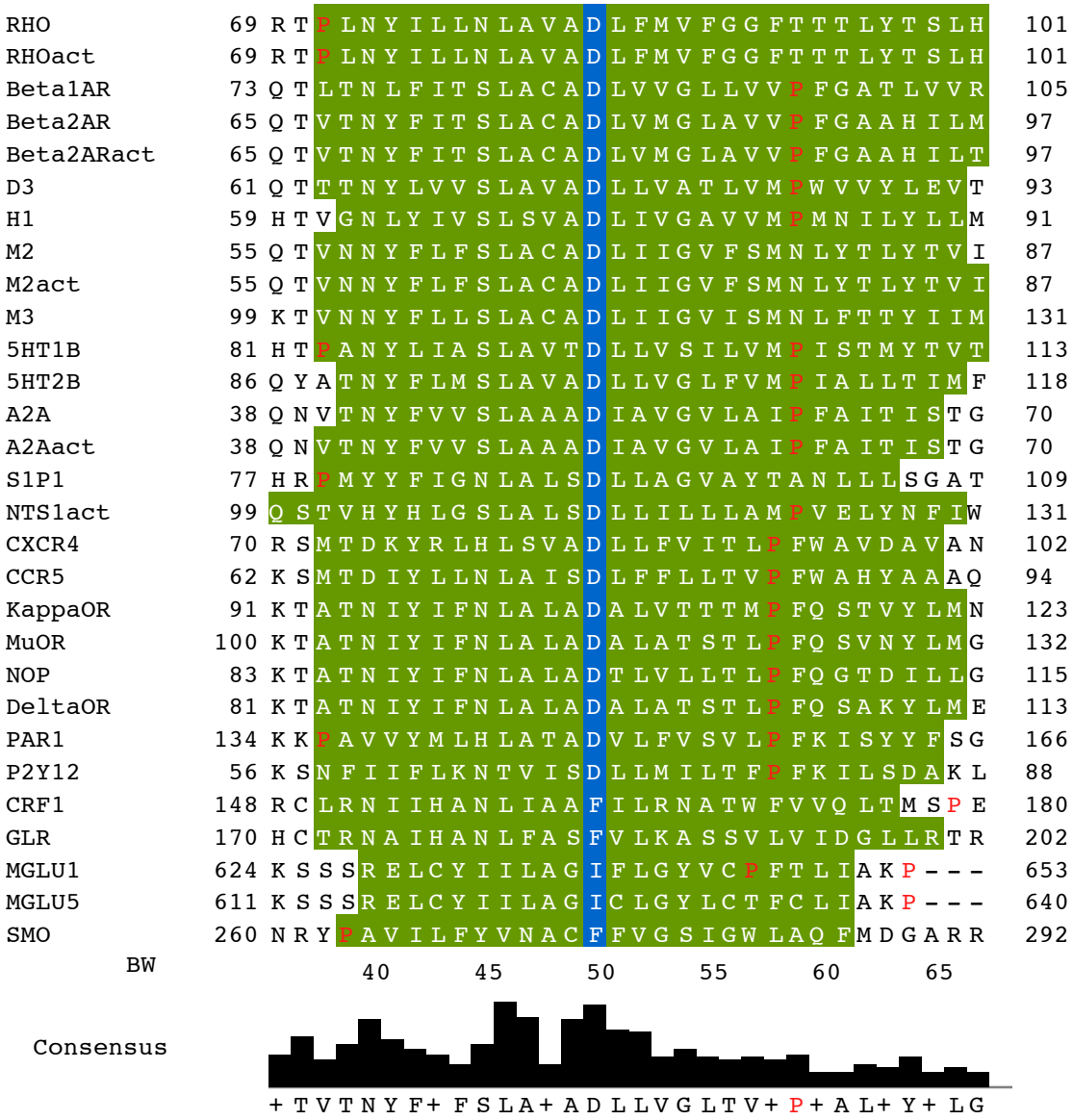


Figure C.2: TM 2 alignment for the crystal structures.

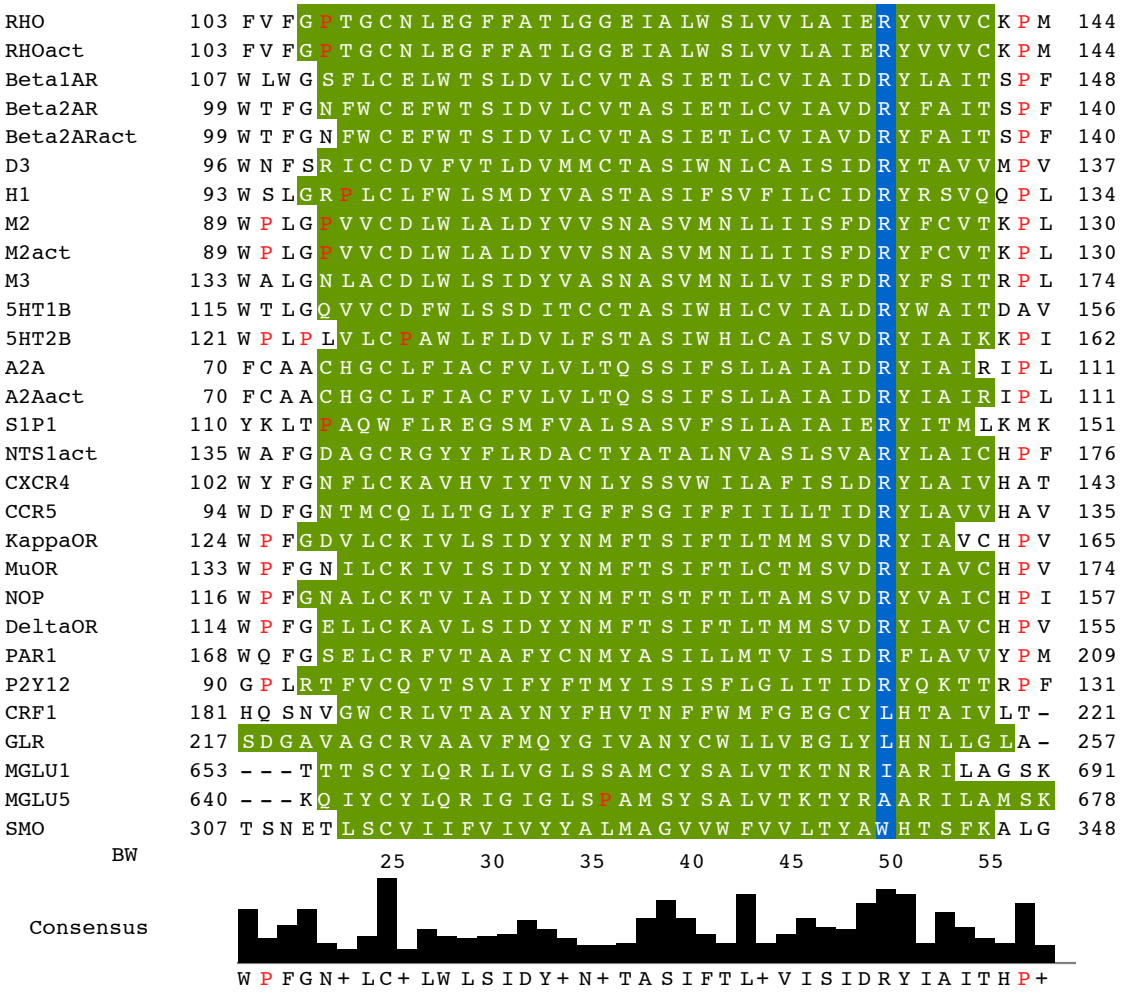


Figure C.3: TM 3 alignment for the crystal structures.

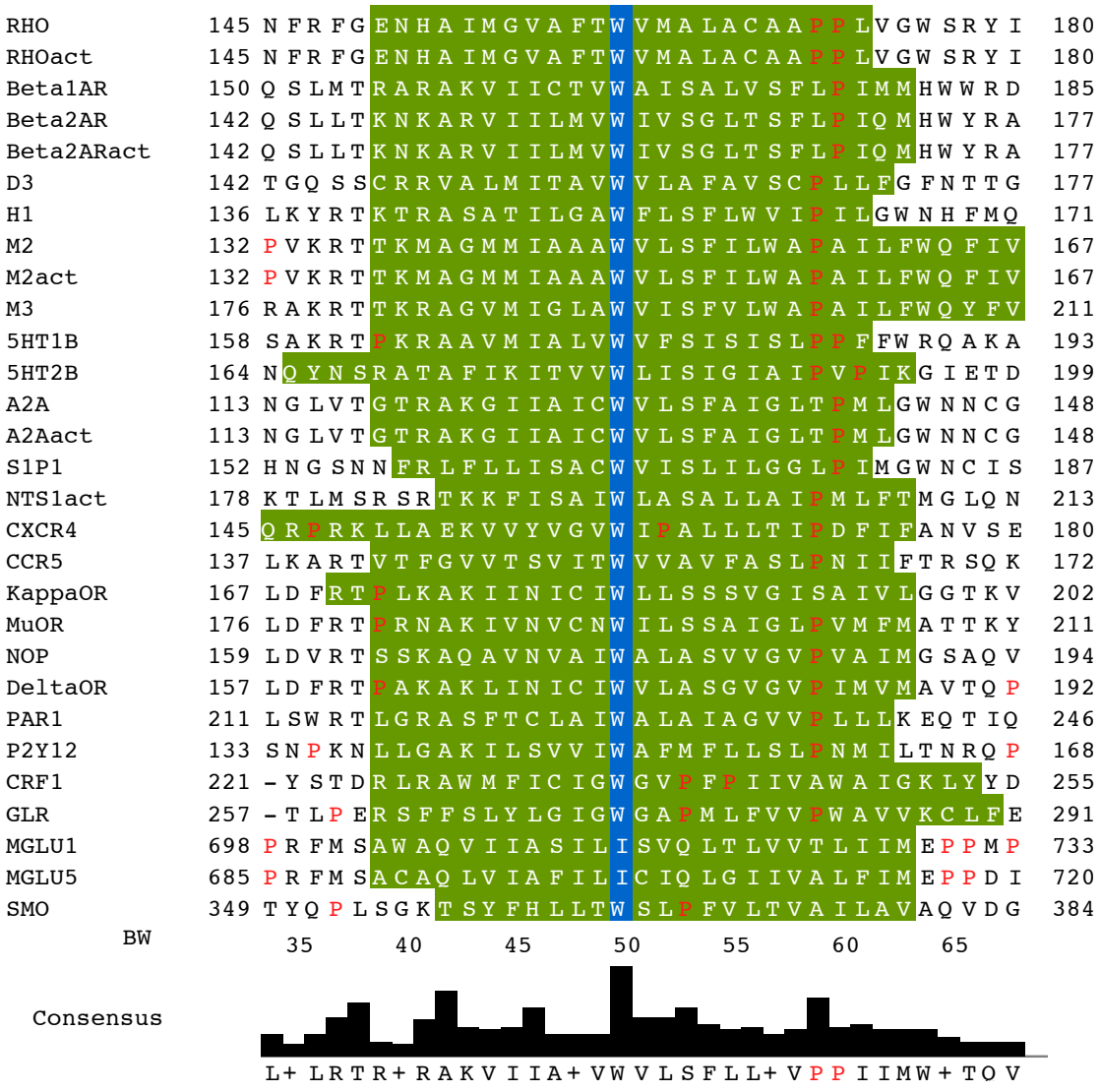


Figure C.4: TM 4 alignment for the crystal structures.

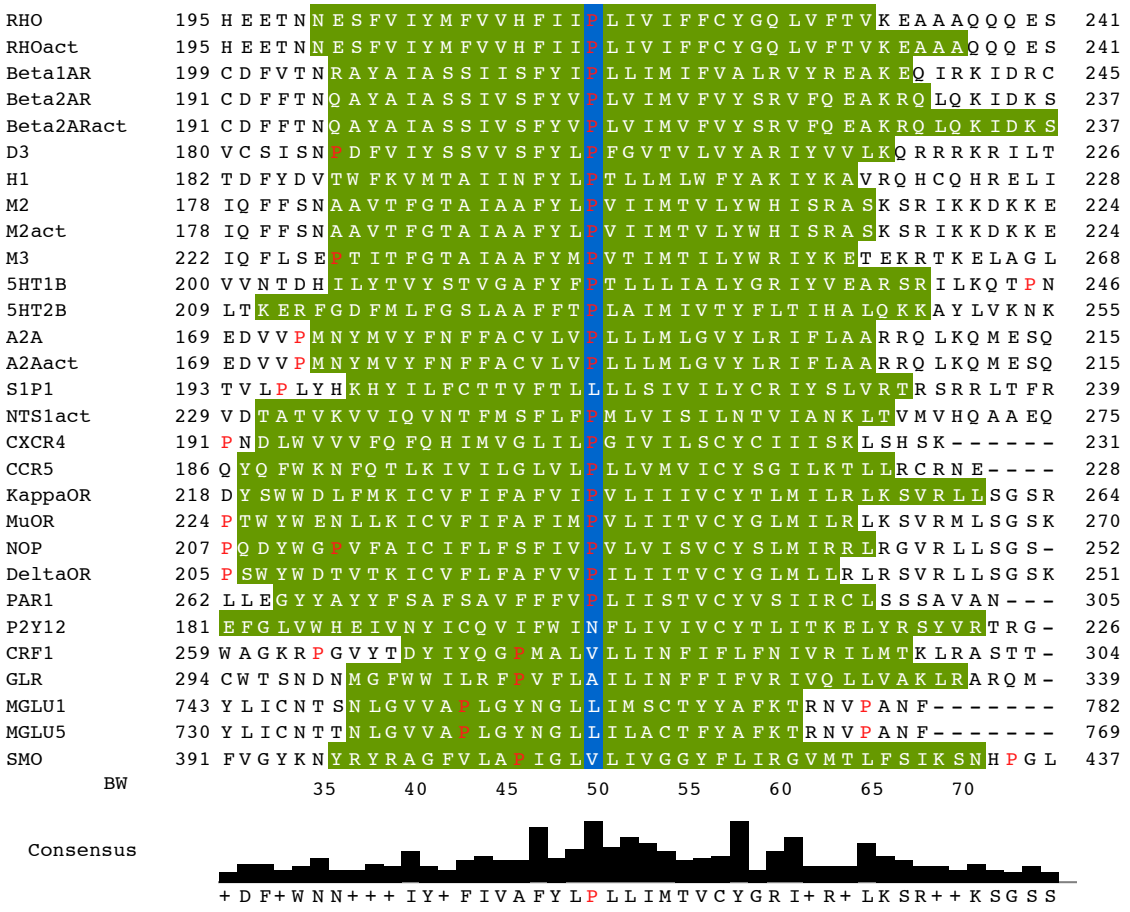


Figure C.5: TM 5 alignment for the crystal structures.

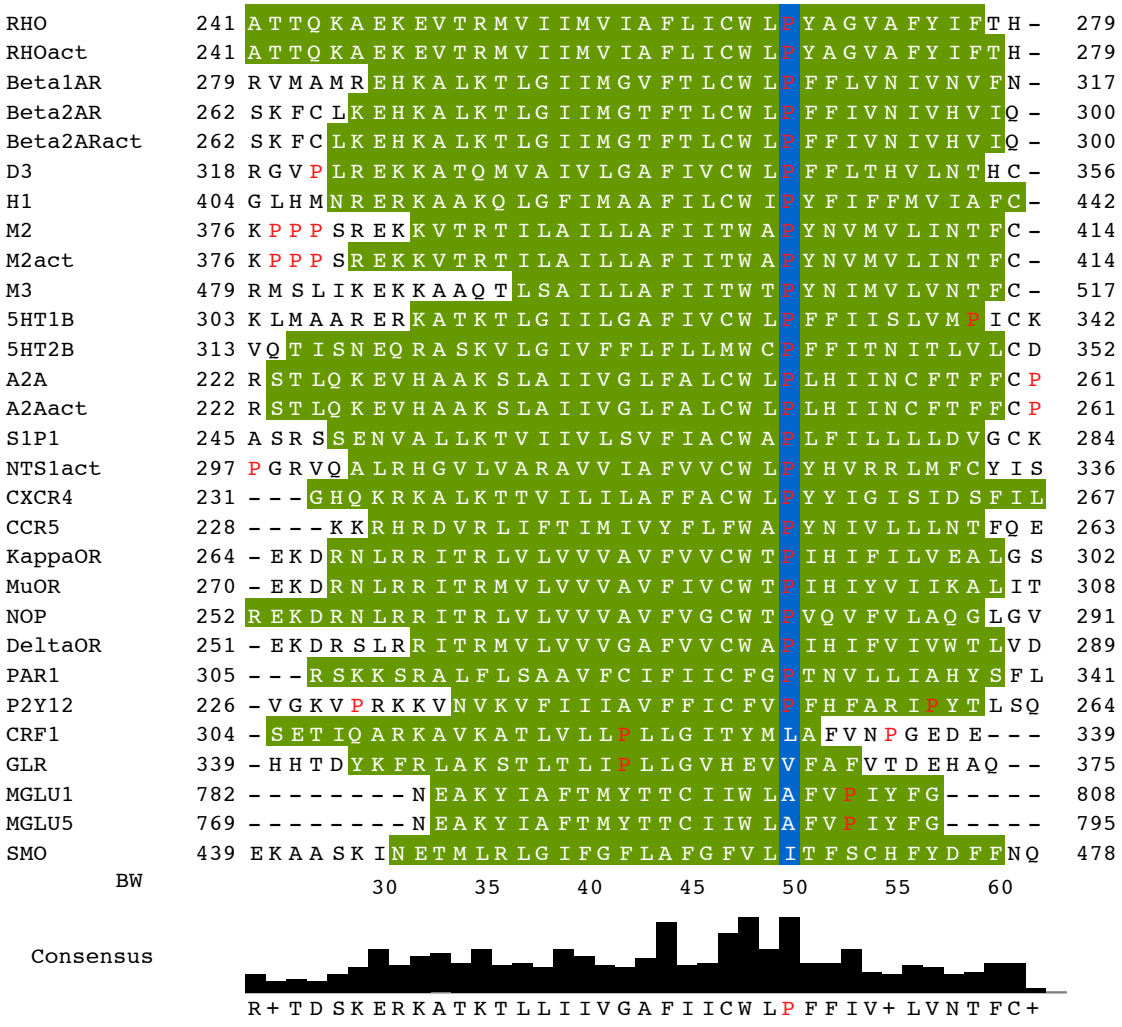


Figure C.6: TM 6 alignment for the crystal structures.

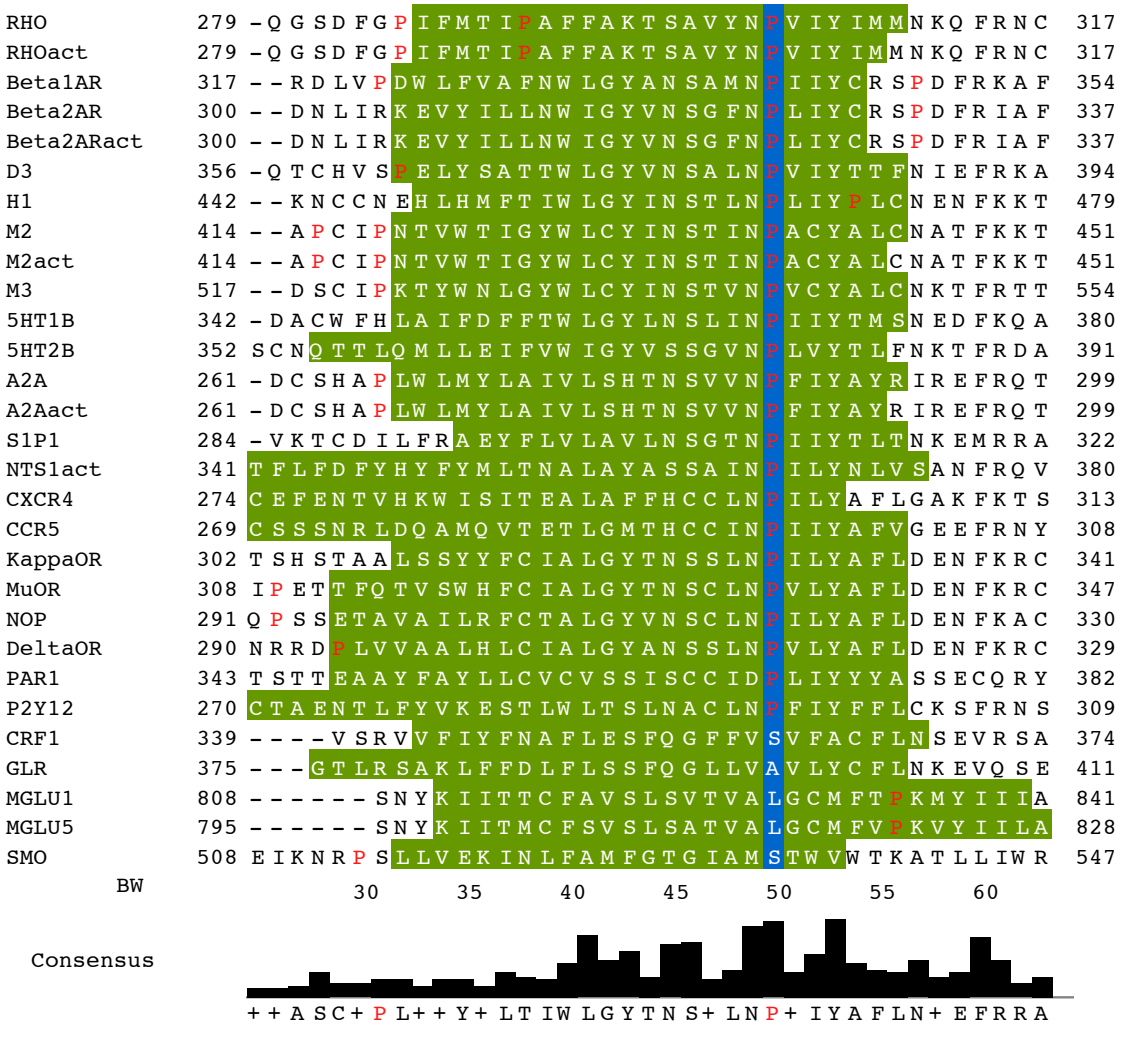


Figure C.7: TM 7 alignment for the crystal structures.

Appendix D

Sequence Alignment for All Known GPCRs

We present the alignment of all 817 human GPCR proteins, as worked out in Section 2.4 and Chapter 3. For each TM 1 though 7, the listed residues are residues with Ballesteros-Weinstein number 50. The numbers in parenthesis denote the expected range of the helical TM regions, which is estimated as the average TM region in the known crystal structures from the same class.

The following 11 proteins are likely GPCR proteins, but are not present in our alignment because their sequence similarity to known structures is too low for an unambiguous alignment (listed as Uniprot IDs):

P51810, Q5T9L3, Q5VW38, O60478, Q86V85, Q86W33, Q8N3F9, Q8NBN3, Q96K49, Q96N19, Q9NPR9.

The following 8 sequences sometimes get classifies as GPCR proteins, but they are most likely pseudogenes, because they miss one or more transmembrane regions (listed as Uniprot UDs):

A6NFC9, Q32VQ0, Q8NGA4, Q8NGU1, Q8NGY7, Q8TDU5, Q96P88, Q99463.

In section 3.4 we identified two possible alignments of TM6, but only the first one is presented in the following table. The second choice is to decrease the start, end, and BW50 residue of TM6 by 4.

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
P02699	RHO	A-alpha	N55 (34-64)	D83 (71-100)	R135 (106-140)	W161 (150-172)	P215 (200-230)	P267 (241-276)	P303 (286-309)
P02699	RHOact	A-alpha	N55 (34-64)	D83 (71-100)	R135 (106-140)	W161 (150-172)	P215 (200-235)	P267 (241-277)	P303 (286-308)
P07700	Beta1AR	A-alpha	N59 (40-68)	D87 (75-104)	R139 (111-144)	W166 (155-179)	P219 (205-236)	P305 (285-315)	P340 (322-344)
P07550	Beta2AR	A-alpha	N51 (31-60)	D79 (67-96)	R131 (103-136)	W158 (147-171)	P211 (197-229)	P288 (267-298)	P323 (305-327)
P07550	Beta2ARact	A-alpha	N51 (31-60)	D79 (67-96)	R131 (104-136)	W158 (147-171)	P211 (197-236)	P288 (266-298)	P323 (305-327)
P35462	D3	A-alpha	N47 (32-56)	D75 (63-91)	R128 (100-133)	W158 (147-170)	P200 (186-216)	P344 (322-353)	P380 (362-386)
P35367	H1	A-alpha	N45 (29-54)	D73 (62-89)	R125 (96-130)	W152 (141-163)	P202 (188-216)	P430 (408-441)	P465 (448-471)
P08172	M2	A-alpha	N41 (23-50)	D69 (57-85)	R121 (93-126)	W148 (137-166)	P198 (184-213)	P402 (384-412)	P437 (419-443)
P08172	M2act	A-alpha	N41 (21-50)	D69 (57-86)	R121 (93-126)	W148 (137-166)	P198 (184-213)	P402 (381-412)	P437 (419-442)
P08483	M3	A-alpha	N85 (65-94)	D113 (101-130)	R165 (137-170)	W192 (181-210)	P242 (228-256)	P505 (492-515)	P540 (522-546)
P28222	5HT1B	A-alpha	N67 (46-76)	D95 (83-112)	R147 (119-152)	W174 (163-185)	P220 (206-238)	P329 (311-338)	P366 (348-372)
P41595	5HT2B	A-alpha	N72 (54-81)	D100 (89-116)	R153 (126-158)	W180 (165-193)	P229 (211-247)	P339 (315-349)	P377 (355-382)
P29274	A2A	A-alpha	N24 (7-33)	D52 (41-67)	R102 (74-106)	W129 (118-141)	P189 (174-204)	P248 (223-258)	P285 (267-291)
P29274	A2Aact	A-alpha	N24 (5-33)	D52 (40-67)	R102 (74-107)	W129 (118-141)	P189 (174-204)	P248 (223-258)	P285 (267-290)
P21453	S1P1	A-alpha	N63 (42-72)	D91 (79-104)	R142 (114-146)	W168 (158-179)	L213 (200-230)	P271 (249-280)	P308 (293-314)
P30989	NTS1act	A-beta	N82 (61-89)	D113 (99-129)	R167 (139-172)	W194 (186-207)	P249 (231-265)	P323 (302-332)	P366 (341-373)
P61073	CXCR4	A-gamma	N56 (37-65)	D84 (72-99)	R134 (106-139)	W161 (145-174)	P211 (193-225)	P254 (231-266)	P299 (274-302)
P51681	CCR5	A-gamma	N48 (26-57)	D76 (64-91)	R126 (98-131)	W153 (142-165)	P206 (187-222)	P250 (230-259)	P294 (269-300)
P41145	KappaOR	A-gamma	N77 (56-86)	D105 (93-121)	R156 (127-159)	W183 (170-196)	P238 (219-259)	P289 (267-299)	P327 (309-333)
P41146	MuOR	A-gamma	N86 (66-95)	D114 (102-130)	R165 (138-170)	W192 (181-205)	P244 (225-258)	P295 (273-305)	P333 (312-339)
P35372	NOP	A-gamma	N69 (48-77)	D97 (85-113)	R148 (119-153)	W175 (164-188)	P227 (208-242)	P278 (252-287)	P316 (295-322)
P41143	DeltaOR	A-gamma	N67 (42-76)	D95 (83-111)	R146 (118-151)	W173 (162-186)	P225 (206-238)	P276 (258-286)	P315 (294-321)
P25116	PAR1	A-delta	N120 (100-131)	D148 (136-163)	R200 (172-205)	W227 (216-239)	P282 (265-297)	P328 (305-338)	P368 (347-374)
Q9H244	P2Y12	A-delta	N43 (21-51)	D70 (58-85)	R122 (93-127)	W149 (138-161)	N201 (181-222)	P251 (235-260)	P295 (270-301)
mod.P34998	CRF1	B	L134 (116-143)	F162 (150-175)	L213 (186-218)	W236 (225-252)	V279 (269-296)	L329 (304-330)	S360 (343-367)
P47871	GLR	B	L156 (127-163)	F184 (172-199)	L249 (217-255)	W272 (261-289)	A314 (301-334)	V364 (343-367)	A397 (375-403)
Q13255	MGLU1	C	T607 (591-616)	I638 (628-649)	I682 (654-685)	I714 (703-727)	L763 (750-774)	A800 (783-807)	L827 (811-839)
P41594	MGLU5	C	T594 (578-603)	I625 (615-636)	A669 (641-677)	I701 (690-714)	L750 (737-761)	A787 (770-794)	L814 (798-827)
Q99835	SMO	F	T245 (224-254)	F274 (263-285)	W339 (312-344)	W365 (357-378)	V411 (397-432)	I465 (446-475)	S533 (515-536)
A3KFT3	OR2M5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
A4D2G3	O2A25_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	A209 (193-226)	T253 (231-263)	P286 (266-292)
A6NCV1	O6C74_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
A6ND48	O14I1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	T252 (230-262)	P285 (265-291)
A6NDH6	O5H15_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	P254 (232-264)	P287 (267-293)
A6NDL8	O6C68_HUMAN	Olfactory	K40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	S285 (265-291)
A6NET4	OR5K3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	Y149 (138-162)	T210 (194-227)	S252 (230-262)	P285 (265-291)
A6NF89	OR6C6_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
A6NGY5	O51F1_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	I158 (147-171)	D219 (203-236)	H263 (241-273)	P298 (278-304)
A6NH00	OR2T8_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	A252 (230-262)	P285 (265-291)
A6NHA9	O4C46_HUMAN	Olfactory	Y40 (20-49)	D68 (56-84)	H120 (92-125)	W147 (136-160)	N208 (192-225)	P251 (229-261)	P282 (262-288)
A6NHG9	O5H14_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	P254 (232-264)	P287 (267-293)
A6NIJ9	O6C70_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
A6NJZ3	O6C65_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
A6NKK0	OR5H1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	P254 (232-264)	P287 (267-293)
A6NL08	O6C75_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
A6NL26	OR5BL_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	T207 (191-224)	T251 (229-261)	P284 (264-290)
A6NM03	O2AG2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
A6NM76	O6C76_HUMAN	Olfactory	N40 (20-49)	E67 (55-83)	C119 (91-124)	W146 (135-159)	T207 (191-224)	S251 (229-261)	P284 (264-290)
A6NMS3	OR5K4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	T210 (194-227)	C254 (232-263)	P284 (264-290)
A6NMU1	O52A4_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	T153 (142-166)	D215 (199-232)	L259 (237-269)	P294 (274-300)
A6NMZ5	O4C45_HUMAN	Olfactory	N37 (17-46)	D65 (53-81)	H123 (95-128)	G150 (139-163)	I211 (195-228)	P254 (232-263)	P284 (264-290)
A6NND4	O2AT4_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	W154 (143-167)	P215 (199-232)	S259 (237-269)	P292 (272-298)
B2RN74	O11HC_HUMAN	Olfactory	N57 (37-66)	E85 (73-101)	Q137 (109-142)	W164 (153-177)	N225 (209-242)	S269 (247-279)	P302 (282-308)
O00144	FZD9_HUMAN	F	T246 (225-255)	Y274 (263-285)	W338 (311-343)	W363 (355-376)	Y409 (395-430)	C461 (442-471)	G525 (507-528)
O00155	GPR25_HUMAN	A-gamma	N57 (37-66)	D85 (73-101)	R137 (109-142)	W164 (153-177)	P214 (198-231)	P259 (237-269)	P304 (284-310)
O00222	GRM8_HUMAN	C	T598 (582-607)	I629 (619-640)	I673 (645-679)	I703 (692-716)	L759 (746-770)	A796 (779-803)	L830 (814-843)
O00254	PAR3_HUMAN	A-delta	N112 (92-121)	D139 (127-155)	R191 (163-196)	W218 (207-231)	P274 (258-291)	P316 (294-326)	P355 (335-361)
O00270	GPR31_HUMAN	A-other	N32 (12-41)	D60 (48-76)	R112 (84-117)	W139 (128-152)	P189 (173-206)	P237 (215-247)	P279 (259-285)
O00398	P2Y10_HUMAN	A-delta	N52 (32-61)	D80 (68-96)	R131 (103-136)	W157 (146-170)	P210 (194-227)	P258 (236-268)	P302 (282-308)
O00421	CCRL2_HUMAN	A-gamma	N56 (36-65)	N84 (72-99)	R128 (100-133)	W156 (145-169)	P212 (196-229)	P253 (231-263)	P297 (277-303)
O00574	CXCR6_HUMAN	A-gamma	N49 (29-58)	D77 (65-93)	R127 (99-132)	W156 (145-169)	P203 (187-220)	P246 (224-256)	P285 (265-291)
O00590	ACKR2_HUMAN	A-gamma	N64 (44-73)	N92 (80-108)	K142 (114-147)	W169 (158-182)	P222 (206-239)	P265 (243-275)	P309 (289-315)
O14514	BAI1_HUMAN	Adhesion	L962 (938-970)	I991 (979-1005)	S1036 (1006-1042)	W1058 (1047-1075)	V1102 (1090-1121)	S1152 (1129-1154)	V1181 (1161-1188)
O14581	OR7AH_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
O14626	GP171_HUMAN	A-delta	S32 (12-41)	D60 (48-76)	R112 (84-117)	W139 (128-152)	F191 (175-208)	P238 (216-248)	P282 (262-288)
O14718	OPSX_HUMAN	A-alpha	N43 (23-52)	D71 (59-87)	R123 (95-128)	W149 (138-162)	P201 (185-218)	P255 (233-265)	P291 (271-297)
O14804	TAAR5_HUMAN	A-alpha	N52 (32-61)	D80 (68-96)	R132 (104-137)	W159 (148-172)	P211 (195-228)	P267 (245-277)	P302 (282-308)
O14842	FFAR1_HUMAN	A-other	N23 (3-32)	D52 (40-68)	R104 (76-109)	W131 (120-144)	P194 (178-211)	P239 (217-249)	P273 (253-279)
O14843	FFAR3_HUMAN	A-other	N32 (12-41)	D61 (49-77)	R113 (85-118)	W140 (129-153)	P196 (180-213)	P240 (218-250)	P273 (253-279)
O15218	GP182_HUMAN	A-gamma	N71 (51-80)	D99 (87-115)	R151 (123-156)	W178 (167-191)	P230 (214-247)	P273 (251-283)	P317 (297-323)
O15303	GRM6_HUMAN	C	T600 (584-609)	I631 (621-642)	I675 (647-681)	T705 (694-718)	L761 (748-772)	A798 (781-805)	L832 (816-845)
O15354	GPR37_HUMAN	A-beta	G278 (258-287)	W306 (294-322)	D358 (330-363)	I386 (375-399)	L453 (437-470)	I506 (484-516)	T545 (525-551)
O15529	GPR42_HUMAN	A-other	N32 (12-41)	D61 (49-77)	R113 (85-118)	W140 (129-153)	P196 (180-213)	P240 (218-250)	P273 (253-279)
O15552	FFAR2_HUMAN	A-other	N25 (5-34)	D55 (43-71)	R107 (79-112)	W134 (123-147)	P191 (175-208)	P237 (215-247)	P270 (250-276)
O43193	MTLR_HUMAN	A-beta	N56 (36-65)	D84 (72-100)	R136 (108-141)	W163 (152-176)	P259 (243-276)	P313 (291-323)	P352 (332-358)
O43194	GPR39_HUMAN	A-other	N48 (28-57)	D79 (67-95)	R133 (105-138)	W160 (149-173)	V234 (218-251)	P298 (276-308)	P341 (321-347)
O43603	GALR2_HUMAN	A-gamma	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P199 (183-216)	P251 (229-261)	P289 (269-295)
O43613	OX1R_HUMAN	A-beta	N64 (44-73)	D92 (80-108)	R144 (116-149)	W169 (158-182)	P227 (211-244)	P313 (291-323)	P355 (335-361)
O43614	OX2R_HUMAN	A-beta	N72 (52-81)	D100 (88-116)	R152 (124-157)	W177 (166-190)	P235 (219-252)	P319 (297-329)	P361 (341-367)
O43749	OR1F1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	H122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
O43869	OR2T1_HUMAN	Olfactory	A91 (71-100)	I119 (107-135)	D171 (143-176)	S198 (187-211)	I259 (243-276)	G303 (281-313)	N336 (316-342)
O60241	BAI2_HUMAN	Adhesion	L947 (923-955)	I976 (964-990)	S1021 (991-1027)	W1043 (1032-1060)	I1087 (1075-1106)	S1170 (1147-1172)	T1199 (1179-1206)
O60242	BAI3_HUMAN	Adhesion	L892 (868-900)	I921 (909-935)	S966 (936-972)	W988 (977-1005)	V1032 (1020-1051)	S1115 (1092-1117)	V1144 (1124-1151)
O60353	FZD6_HUMAN	F	T213 (192-222)	Y241 (230-252)	W307 (280-312)	W332 (324-345)	C378 (364-399)	T430 (411-440)	V491 (473-494)
O60403	O10H2_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	C211 (195-228)	F255 (233-265)	P288 (268-294)
O60404	O10H3_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	H123 (95-128)	W150 (139-163)	C212 (196-229)	F256 (234-266)	P289 (269-295)
O60412	OR7C2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T253 (231-263)	P286 (266-292)
O60431	OR1I1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
O60755	GALR3_HUMAN	A-gamma	N35 (15-44)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P196 (180-213)	P250 (228-260)	P288 (268-294)
O60883	ETBR2_HUMAN	A-beta	N148 (128-157)	D176 (164-192)	R228 (200-233)	W256 (245-269)	P321 (305-338)	P374 (352-384)	P413 (393-419)
O75084	FZD7_HUMAN	F	T268 (247-277)	Y296 (285-307)	W361 (334-366)	W386 (378-399)	Y432 (418-453)	I484 (465-494)	G545 (527-548)
O75388	GPR32_HUMAN	A-gamma	N61 (41-70)	D88 (76-104)	R139 (111-144)	W166 (155-179)	P228 (212-245)	P271 (249-281)	P313 (293-319)
O75473	LGR5_HUMAN	A-delta	N575 (555-584)	N603 (591-619)	R662 (634-667)	A689 (678-702)	C736 (720-753)	P781 (759-791)	P817 (797-823)
O75899	GABR2_HUMAN	C	A496 (480-505)	G527 (517-538)	V578 (550-584)	L605 (594-618)	L667 (654-678)	I705 (688-712)	L736 (720-749)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
O76000	OR2B3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	S287 (267-293)
O76001	OR2J3_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P213 (197-230)	P257 (235-267)	P290 (270-296)
O76002	OR2J2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	P255 (233-265)	P288 (268-294)
O76099	OR7C1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
O76100	OR7AA_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
O94910	LPHN1_HUMAN	Adhesion	L875 (851-883)	L903 (891-917)	L948 (918-954)	Y971 (960-988)	V1014 (1002-1033)	F1066 (1043-1068)	F1094 (1074-1101)
O95006	OR2F2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
O95007	OR6B1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A252 (230-262)	P285 (265-291)
O95013	O4F21_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	P253 (231-263)	P284 (264-290)
O95047	OR2A4_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	L121 (93-126)	W148 (137-161)	P209 (193-226)	T253 (231-263)	P286 (266-292)
O95136	S1PR2_HUMAN	A-alpha	N51 (31-60)	D79 (67-95)	R130 (102-135)	W156 (145-169)	I201 (185-218)	P248 (226-258)	P285 (265-291)
O95221	OR5F1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
O95222	OR6A2_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R127 (99-132)	W154 (143-167)	P215 (199-232)	A259 (237-269)	P292 (272-298)
O95371	OR2C1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	C149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
O95490	LPHN2_HUMAN	Adhesion	L862 (838-870)	L890 (878-904)	L935 (905-941)	Y958 (947-975)	I1001 (989-1020)	F1053 (1030-1055)	F1081 (1061-1088)
O95665	NTR2_HUMAN	A-beta	N50 (30-59)	G79 (67-95)	R133 (105-138)	W160 (149-173)	P218 (202-235)	P312 (290-322)	P355 (335-361)
O95800	GPR75_HUMAN	A-other	N60 (40-69)	D90 (78-106)	R143 (115-148)	W169 (158-182)	C214 (198-231)	P336 (314-346)	P373 (353-379)
O95838	GLP2R_HUMAN	B	S192 (168-200)	S220 (208-234)	Y284 (254-290)	G307 (296-324)	L349 (337-368)	I399 (376-401)	V432 (412-439)
O95918	OR2H2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	S252 (230-262)	P285 (265-291)
O95977	S1PR4_HUMAN	A-alpha	N64 (44-73)	D92 (80-108)	R143 (115-148)	W170 (159-183)	V215 (199-232)	P266 (244-276)	P304 (284-310)
P03999	OPSB_HUMAN	A-alpha	N52 (32-61)	G80 (68-96)	R132 (104-137)	W158 (147-171)	P212 (196-229)	P264 (242-274)	P300 (280-306)
P04000	OPSR_HUMAN	A-alpha	N71 (51-80)	D99 (87-115)	R151 (123-156)	W177 (166-190)	P231 (215-248)	P283 (261-293)	P319 (299-325)
P04001	OPSG_HUMAN	A-alpha	N71 (51-80)	D99 (87-115)	R151 (123-156)	W177 (166-190)	P231 (215-248)	P283 (261-293)	P319 (299-325)
P04201	MAS_HUMAN	A-delta	N50 (30-59)	D77 (65-93)	R130 (102-135)	W157 (146-170)	T200 (184-217)	P243 (221-253)	P277 (257-283)
P07550	ADRB2_HUMAN	A-alpha	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P211 (195-228)	P288 (266-298)	P323 (303-329)
P08100	OPSD_HUMAN	A-alpha	N55 (35-64)	D83 (71-99)	R135 (107-140)	W161 (150-174)	P215 (199-232)	P267 (245-277)	P303 (283-309)
P08172	ACM2_HUMAN	A-alpha	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P198 (182-215)	P402 (380-412)	P437 (417-443)
P08173	ACM4_HUMAN	A-alpha	N50 (30-59)	D78 (66-94)	R130 (102-135)	W157 (146-170)	P207 (191-224)	P415 (393-425)	P450 (430-456)
P08588	ADRB1_HUMAN	A-alpha	N76 (56-85)	D104 (92-120)	R156 (128-161)	W183 (172-196)	P236 (220-253)	P339 (317-349)	P374 (354-380)
P08908	5HT1A_HUMAN	A-alpha	N54 (34-63)	D82 (70-98)	R134 (106-139)	W161 (150-174)	P207 (191-224)	P360 (338-370)	P397 (377-403)
P08912	ACM5_HUMAN	A-alpha	N48 (28-57)	D76 (64-92)	R128 (100-133)	W155 (144-168)	P205 (189-222)	P457 (435-467)	P492 (472-498)
P08913	ADA2A_HUMAN	A-alpha	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P208 (192-225)	P389 (367-399)	P423 (403-429)
P0C604	OR4A8_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	V147 (136-160)	I208 (192-225)	P251 (229-261)	P282 (262-288)
P0C617	O5AL1_HUMAN	Olfactory	N58 (38-67)	D86 (74-102)	R138 (110-143)	Y164 (153-177)	S225 (209-242)	T269 (247-279)	P302 (282-308)
P0C623	OR4Q2_HUMAN	Olfactory	N42 (22-51)	D71 (59-87)	R123 (95-128)	W150 (139-163)	S211 (195-228)	P252 (230-262)	P283 (263-289)
P0C626	OR5G3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
P0C628	O5AC1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
P0C629	O10J4_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	H123 (95-128)	W150 (139-163)	P211 (195-228)	R255 (233-265)	P287 (267-293)
P0C645	OR4E1_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	C215 (199-232)	H258 (236-268)	P289 (269-295)
P0C646	O52Z1_HUMAN	Olfactory	N28 (8-37)	D56 (44-72)	R108 (80-113)	V134 (123-147)	D193 (177-210)	P237 (215-247)	P273 (253-279)
P0C7N1	OR8U8_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
P0C7N5	OR8U9_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	H149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
P0C7N8	OR9G9_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	Y148 (137-161)	P209 (193-226)	S253 (231-263)	P286 (266-292)
P0C7T2	OR2T7_HUMAN	Olfactory	N33 (13-42)	D61 (49-77)	R113 (85-118)	W140 (129-153)	P201 (185-218)	A245 (223-255)	P278 (258-284)
P0C7T3	O56A5_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	V152 (141-165)	D213 (197-230)	V257 (235-267)	P292 (272-298)
P11229	ACM1_HUMAN	A-alpha	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P200 (184-217)	P380 (358-390)	P415 (395-421)
P13945	ADRB3_HUMAN	A-alpha	N55 (35-64)	D83 (71-99)	R135 (107-140)	W162 (151-175)	P216 (200-233)	P307 (285-317)	P343 (323-349)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
P14416	DRD2_HUMAN	A-alpha	N52 (32-61)	D80 (68-96)	R132 (104-137)	W160 (149-173)	P201 (185-218)	P388 (366-398)	P423 (403-429)
P16473	TSHR_HUMAN	A-delta	N432 (412-441)	D460 (448-476)	R519 (491-524)	W546 (535-559)	A593 (577-610)	P639 (617-649)	P675 (655-681)
P18089	ADA2B_HUMAN	A-alpha	N30 (10-39)	D58 (46-74)	R110 (82-115)	W137 (126-150)	P184 (168-201)	P383 (361-393)	P420 (400-426)
P18825	ADA2C_HUMAN	A-alpha	N69 (49-78)	D97 (85-113)	R149 (121-154)	W176 (165-189)	P222 (206-239)	P397 (375-407)	P434 (414-440)
P20309	ACM3_HUMAN	A-alpha	N86 (66-95)	D114 (102-130)	R166 (138-171)	W193 (182-206)	P243 (227-260)	P506 (484-516)	P541 (521-547)
P21452	NK2R_HUMAN	A-beta	N51 (31-60)	D79 (67-95)	R131 (103-136)	W156 (145-169)	P209 (193-226)	P265 (243-275)	P304 (284-310)
P21453	S1PR1_HUMAN	A-alpha	N63 (43-72)	D91 (79-107)	R142 (114-147)	W168 (157-181)	L213 (197-230)	P271 (249-281)	P308 (288-314)
P21462	FPR1_HUMAN	A-gamma	N44 (24-53)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P213 (197-230)	P256 (234-266)	P298 (278-304)
P21554	CNR1_HUMAN	A-alpha	E133 (113-142)	A162 (150-178)	D213 (185-218)	M240 (229-253)	V285 (269-302)	G357 (335-367)	N393 (373-399)
P21728	DRD1_HUMAN	A-alpha	N41 (21-50)	D70 (58-86)	R121 (93-126)	W148 (137-161)	P206 (190-223)	P287 (265-297)	P328 (308-334)
P21730	C5AR1_HUMAN	A-gamma	N55 (35-64)	D82 (70-98)	R134 (106-139)	W161 (150-174)	P214 (198-231)	P257 (235-267)	P297 (277-303)
P21731	TA2R_HUMAN	A-alpha	N42 (22-51)	D74 (62-90)	R130 (102-135)	W157 (146-170)	G205 (189-222)	P260 (238-270)	P305 (285-311)
P21917	DRD4_HUMAN	A-alpha	N52 (32-61)	D80 (68-96)	R133 (105-138)	W160 (149-173)	P204 (188-221)	P409 (387-419)	P445 (425-451)
P21918	DRD5_HUMAN	A-alpha	N58 (38-67)	D87 (75-103)	R138 (110-143)	W165 (154-178)	P237 (221-254)	P311 (289-321)	P356 (336-362)
P22888	LSHR_HUMAN	A-delta	N377 (357-386)	D405 (393-421)	R464 (436-469)	W491 (480-504)	A538 (522-555)	P584 (562-594)	P620 (600-626)
P23945	FSHR_HUMAN	A-delta	N380 (360-389)	D408 (396-424)	R467 (439-472)	W494 (483-507)	A541 (525-558)	P587 (565-597)	P623 (603-629)
P24530	EDNRB_HUMAN	A-beta	N119 (99-128)	D147 (135-163)	R199 (171-204)	W226 (215-239)	P285 (269-302)	P338 (316-348)	P383 (363-389)
P25021	HRH2_HUMAN	A-alpha	N36 (16-45)	D64 (52-80)	R116 (88-121)	W143 (132-156)	P194 (178-211)	P249 (227-259)	P285 (265-291)
P25024	CXCR1_HUMAN	A-gamma	N57 (37-66)	D85 (73-101)	R135 (107-140)	W161 (150-174)	P214 (198-231)	P257 (235-267)	P302 (282-308)
P25025	CXCR2_HUMAN	A-gamma	N66 (46-75)	D94 (82-110)	R144 (116-149)	W170 (159-183)	P223 (207-240)	P266 (244-276)	P311 (291-317)
P25089	FPR3_HUMAN	A-gamma	N44 (24-53)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P213 (197-230)	P256 (234-266)	P299 (279-305)
P25090	FPR2_HUMAN	A-gamma	N44 (24-53)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P213 (197-230)	P256 (234-266)	P299 (279-305)
P25100	ADA1D_HUMAN	A-alpha	N114 (94-123)	D142 (130-158)	R194 (166-199)	W221 (210-234)	P266 (250-283)	P363 (341-373)	P399 (379-405)
P25101	EDNRA_HUMAN	A-beta	N98 (78-107)	D126 (114-142)	R183 (155-188)	W210 (199-223)	P267 (251-284)	P321 (299-331)	P366 (346-372)
P25103	NK1R_HUMAN	A-beta	N50 (30-59)	E78 (66-94)	R130 (102-135)	W155 (144-168)	P208 (192-225)	P263 (241-273)	P302 (282-308)
P25105	PTAFR_HUMAN	A-delta	N33 (13-42)	D63 (51-79)	R115 (87-120)	W142 (131-155)	V198 (182-215)	P247 (225-257)	P290 (270-296)
P25106	ACKR3_HUMAN	A-gamma	N62 (42-71)	D90 (78-106)	R142 (114-147)	W169 (158-182)	P224 (208-241)	P267 (245-277)	P312 (292-318)
P25116	PAR1_HUMAN	A-delta	N120 (100-129)	D148 (136-164)	R200 (172-205)	W227 (216-240)	P282 (266-299)	P328 (306-338)	P368 (348-374)
P25929	NPY1R_HUMAN	A-beta	N58 (38-67)	D86 (74-102)	R138 (110-143)	W163 (152-176)	P223 (207-240)	P278 (256-288)	P317 (297-323)
P28221	5HT1D_HUMAN	A-alpha	N56 (36-65)	D84 (72-100)	R136 (108-141)	W163 (152-176)	P209 (193-226)	P316 (294-326)	P353 (333-359)
P28222	5HT1B_HUMAN	A-alpha	N67 (47-76)	D95 (83-111)	R147 (119-152)	W174 (163-187)	P220 (204-237)	P329 (307-339)	P366 (346-372)
P28223	5HT2A_HUMAN	A-alpha	N92 (72-101)	D120 (108-136)	R173 (145-178)	W200 (189-213)	P246 (230-263)	P338 (316-348)	P377 (357-383)
P28335	5HT2C_HUMAN	A-alpha	N71 (51-80)	D99 (87-115)	R152 (124-157)	W179 (168-192)	P226 (210-243)	P326 (304-336)	P365 (345-371)
P28336	NMBR_HUMAN	A-beta	N61 (41-70)	D89 (77-105)	R141 (113-146)	W168 (157-181)	P224 (208-241)	P281 (259-291)	P321 (301-327)
P28566	5HT1E_HUMAN	A-alpha	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P194 (178-211)	P306 (284-316)	P341 (321-347)
P29274	AA2AR_HUMAN	A-alpha	N24 (4-33)	D52 (40-68)	R102 (74-107)	W129 (118-142)	P189 (173-206)	P248 (226-258)	P285 (265-291)
P29275	AA2BR_HUMAN	A-alpha	N25 (5-34)	D53 (41-69)	R103 (75-108)	W130 (119-143)	P194 (178-211)	P249 (227-259)	P287 (267-293)
P29371	NK3R_HUMAN	A-beta	N103 (83-112)	D131 (119-147)	R183 (155-188)	W208 (197-221)	P259 (243-276)	P314 (292-324)	P353 (333-359)
P30411	BKRB2_HUMAN	A-gamma	N75 (55-84)	D103 (91-119)	R155 (127-160)	W182 (171-195)	P236 (220-253)	P285 (263-295)	P329 (309-335)
P30518	V2R_HUMAN	A-beta	N55 (35-64)	D85 (73-101)	R137 (109-142)	W164 (153-177)	P217 (201-234)	P286 (264-296)	P322 (302-328)
P30542	AA1R_HUMAN	A-alpha	N27 (7-36)	D55 (43-71)	R105 (77-110)	W132 (121-145)	P192 (176-209)	P249 (227-259)	P285 (265-291)
P30550	GRPR_HUMAN	A-beta	N58 (38-67)	D86 (74-102)	R138 (110-143)	W165 (154-178)	P222 (206-239)	P279 (257-289)	P319 (299-325)
P30556	AGTR1_HUMAN	A-gamma	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	P207 (191-224)	P255 (233-265)	P299 (279-305)
P30559	OXYR_HUMAN	A-beta	N57 (37-66)	D85 (73-101)	R137 (109-142)	W161 (150-174)	P212 (196-229)	P290 (268-300)	P326 (306-332)
P30872	SSR1_HUMAN	A-gamma	N76 (56-85)	D104 (92-120)	R155 (127-160)	W182 (171-195)	P235 (219-252)	P286 (264-296)	P320 (300-326)
P30874	SSR2_HUMAN	A-gamma	N61 (41-70)	D89 (77-105)	R140 (112-145)	W167 (156-180)	P220 (204-237)	P271 (249-281)	P309 (289-315)
P30939	5HT1F_HUMAN	A-alpha	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P193 (177-210)	P308 (286-318)	P344 (324-350)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
P30953	OR1E1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
P30954	O10J1_HUMAN	Olfactory	N53 (33-62)	E81 (69-97)	R133 (105-138)	C160 (149-173)	P220 (204-237)	C264 (242-274)	P297 (277-303)
P30968	GNRHR_HUMAN	A-beta	N53 (33-62)	N87 (75-103)	R139 (111-144)	W164 (153-177)	P223 (207-240)	P282 (260-292)	P320 (300-326)
P30988	CALCR_HUMAN	B	T180 (156-188)	T224 (212-238)	Y277 (247-283)	G300 (289-317)	A341 (329-360)	V391 (368-393)	V421 (401-428)
P30989	NTR1_HUMAN	A-beta	N81 (61-90)	D112 (100-128)	R166 (138-171)	W193 (182-206)	P248 (232-265)	P318 (296-328)	P361 (341-367)
P31391	SSR4_HUMAN	A-gamma	N65 (45-74)	D93 (81-109)	R144 (116-149)	W171 (160-184)	P223 (207-240)	P274 (252-284)	P308 (288-314)
P32238	CCKAR_HUMAN	A-beta	N59 (39-68)	D87 (75-103)	R139 (111-144)	W166 (155-179)	P221 (205-238)	P328 (306-338)	P367 (347-373)
P32239	GASR_HUMAN	A-beta	N72 (52-81)	D100 (88-116)	R152 (124-157)	W179 (168-192)	P230 (214-247)	P348 (326-358)	P387 (367-393)
P32241	VIPR1_HUMAN	B	L157 (133-165)	F185 (173-199)	L240 (210-246)	W263 (252-280)	S304 (292-323)	M356 (333-358)	A385 (365-392)
P32245	MC4R_HUMAN	A-alpha	N62 (42-71)	D90 (78-106)	R147 (119-152)	W174 (163-187)	M204 (188-221)	P260 (238-270)	P299 (279-305)
P32246	CCR1_HUMAN	A-gamma	N52 (32-61)	D80 (68-96)	R131 (103-136)	W158 (147-171)	P211 (195-228)	P254 (232-264)	P298 (278-304)
P32247	BRS3_HUMAN	A-beta	N65 (45-74)	D93 (81-109)	R145 (117-150)	W172 (161-185)	P229 (213-246)	P286 (264-296)	P327 (307-333)
P32248	CCR7_HUMAN	A-gamma	N76 (56-85)	D104 (92-120)	R154 (126-159)	W183 (172-196)	P235 (219-252)	P278 (256-288)	P323 (303-329)
P32249	GP183_HUMAN	A-delta	N49 (29-58)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P208 (192-225)	P259 (237-269)	P305 (285-311)
P32302	CXCR5_HUMAN	A-gamma	N69 (49-78)	D97 (85-113)	R147 (119-152)	W174 (163-187)	P230 (214-247)	P274 (252-284)	P319 (299-325)
P32745	SSR3_HUMAN	A-gamma	N62 (42-71)	D90 (78-106)	R141 (113-146)	W168 (157-181)	P218 (202-235)	P272 (250-282)	P310 (290-316)
P33032	MC5R_HUMAN	A-alpha	N54 (34-63)	D82 (70-98)	R140 (112-145)	W167 (156-180)	M197 (181-214)	P253 (231-263)	P292 (272-298)
P33765	AA3R_HUMAN	A-alpha	N30 (10-39)	D58 (46-74)	R108 (80-113)	W135 (124-148)	P189 (173-206)	P245 (223-255)	P279 (259-285)
P34969	5HT7R_HUMAN	A-alpha	N99 (79-108)	D127 (115-143)	R180 (152-185)	W207 (196-220)	P251 (235-268)	P342 (320-352)	P381 (361-387)
P34972	CNR2_HUMAN	A-alpha	N51 (31-60)	D80 (68-96)	R131 (103-136)	W158 (147-171)	L201 (185-218)	P260 (238-270)	P296 (276-302)
P34981	TRFR_HUMAN	A-delta	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P203 (187-220)	P281 (259-291)	P317 (297-323)
P34982	OR1D2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P286 (266-292)
P34995	PE2R1_HUMAN	A-alpha	N50 (30-59)	D84 (72-100)	R135 (107-140)	A162 (151-175)	L212 (196-229)	P312 (290-322)	P348 (328-354)
P34998	CRFR1_HUMAN	B	L134 (110-142)	F191 (179-205)	L242 (212-248)	W265 (254-282)	V308 (296-327)	L358 (335-360)	S389 (369-396)
P35346	SSR5_HUMAN	A-gamma	N58 (38-67)	D86 (74-102)	R137 (109-142)	W164 (153-177)	P213 (197-230)	P263 (241-273)	P301 (281-307)
P35348	ADA1A_HUMAN	A-alpha	N44 (24-53)	D72 (60-88)	R124 (96-129)	W151 (140-164)	P196 (180-213)	P287 (265-297)	P323 (303-329)
P35367	HRH1_HUMAN	A-alpha	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P202 (186-219)	P430 (408-440)	P465 (445-471)
P35368	ADA1B_HUMAN	A-alpha	N63 (43-72)	D91 (79-107)	R143 (115-148)	W170 (159-183)	P215 (199-232)	P309 (287-319)	P345 (325-351)
P35372	OPRM_HUMAN	A-gamma	G87 (67-96)	A115 (103-131)	D166 (138-171)	N193 (182-206)	M245 (229-262)	T296 (274-306)	N334 (314-340)
P35408	PE2R4_HUMAN	A-alpha	N35 (15-44)	D65 (53-81)	R117 (89-122)	Y144 (133-157)	S193 (177-210)	P287 (265-297)	P326 (306-332)
P35410	MAS1L_HUMAN	A-delta	N93 (73-102)	D119 (107-135)	R171 (143-176)	W198 (187-211)	S236 (220-253)	P279 (257-289)	P311 (291-317)
P35414	APJ_HUMAN	A-gamma	N46 (26-55)	D75 (63-91)	R127 (99-132)	W154 (143-167)	P213 (197-230)	P263 (241-273)	P306 (286-312)
P35462	DRD3_HUMAN	A-alpha	N47 (27-56)	D75 (63-91)	R128 (100-133)	W158 (147-171)	P200 (184-217)	P344 (322-354)	P380 (360-386)
P37288	V1AR_HUMAN	A-beta	N69 (49-78)	D97 (85-113)	R149 (121-154)	W175 (164-188)	P228 (212-245)	P306 (284-316)	P345 (325-351)
P41143	OPRD_HUMAN	A-gamma	N67 (47-76)	D95 (83-111)	R146 (118-151)	W173 (162-186)	P225 (209-242)	P276 (254-286)	P315 (295-321)
P41145	OPRK_HUMAN	A-gamma	G76 (56-85)	A104 (92-120)	D155 (127-160)	I182 (171-195)	I237 (221-254)	T288 (266-298)	N326 (306-332)
P41146	OPRX_HUMAN	A-gamma	N69 (49-78)	D97 (85-113)	R148 (120-153)	W175 (164-188)	P227 (211-244)	P278 (256-288)	P316 (296-322)
P41180	CASR_HUMAN	C	T627 (611-636)	L658 (648-669)	V702 (674-708)	T732 (721-745)	L783 (770-794)	S820 (803-827)	L849 (833-862)
P41231	P2RY2_HUMAN	A-delta	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P210 (194-227)	P260 (238-270)	P303 (283-309)
P41586	PACR_HUMAN	B	L168 (144-176)	F196 (184-210)	L251 (221-257)	W274 (263-291)	S316 (304-335)	V368 (345-370)	A397 (377-404)
P41587	VIPR2_HUMAN	B	L141 (117-149)	F169 (157-183)	L227 (197-233)	W249 (238-266)	S291 (279-310)	V343 (320-345)	A372 (352-379)
P41594	GRM5_HUMAN	C	T594 (578-603)	I625 (615-636)	I669 (641-675)	I701 (690-714)	L750 (737-761)	A787 (770-794)	L814 (798-827)
P41595	5HT2B_HUMAN	A-alpha	N72 (52-81)	D100 (88-116)	R153 (125-158)	W180 (169-193)	P229 (213-246)	P339 (317-349)	P377 (357-383)
P41597	CCR2_HUMAN	A-gamma	N60 (40-69)	D88 (76-104)	R138 (110-143)	W165 (154-178)	P214 (198-231)	P258 (236-268)	P302 (282-308)
P41968	MC3R_HUMAN	A-alpha	N56 (36-65)	D84 (72-100)	R142 (114-147)	W169 (158-182)	M199 (183-216)	P257 (235-267)	P296 (276-302)
P43088	PF2R_HUMAN	A-alpha	N44 (24-53)	D77 (65-93)	R133 (105-138)	C160 (149-173)	L210 (194-227)	P264 (242-274)	P301 (281-307)
P43115	PE2R3_HUMAN	A-alpha	N66 (46-75)	D99 (87-115)	R155 (127-160)	W182 (171-195)	L239 (223-256)	P297 (275-307)	P343 (323-349)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
P43116	PE2R2_HUMAN	A-alpha	N39 (19-48)	D78 (66-94)	R134 (106-139)	Y161 (150-174)	L205 (189-222)	P279 (257-289)	P312 (292-318)
P43119	PI2R_HUMAN	A-alpha	N31 (11-40)	D60 (48-76)	R117 (89-122)	Y144 (133-157)	A193 (177-210)	P254 (232-264)	P289 (269-295)
P43220	GLP1R_HUMAN	B	L159 (135-167)	F187 (175-201)	L251 (221-257)	W274 (263-291)	A316 (304-335)	I366 (343-368)	A399 (379-406)
P43657	LPAR6_HUMAN	A-delta	N35 (15-44)	D63 (51-79)	R114 (86-119)	W141 (130-154)	P196 (180-213)	P244 (222-254)	P288 (268-294)
P46089	GPR3_HUMAN	A-alpha	N58 (38-67)	D86 (74-102)	R134 (106-139)	W161 (150-174)	V205 (189-222)	P262 (240-272)	P294 (274-300)
P46091	GPR1_HUMAN	A-gamma	N56 (36-65)	D83 (71-99)	H135 (107-140)	W162 (151-175)	P218 (202-235)	P261 (239-271)	P301 (281-307)
P46092	CCR10_HUMAN	A-gamma	N59 (39-68)	D88 (76-104)	R138 (110-143)	W166 (155-179)	P219 (203-236)	P262 (240-272)	P307 (287-313)
P46093	GPR4_HUMAN	A-delta	N35 (15-44)	D63 (51-79)	R115 (87-120)	W142 (131-155)	P193 (177-210)	P239 (217-249)	P283 (263-289)
P46094	XCR1_HUMAN	A-gamma	N49 (29-58)	D77 (65-93)	R127 (99-132)	W154 (143-167)	S197 (181-214)	P240 (218-250)	P284 (264-290)
P46095	GPR6_HUMAN	A-alpha	E89 (69-98)	A117 (105-133)	D165 (137-170)	T192 (181-205)	M236 (220-253)	L293 (271-303)	N325 (305-331)
P46663	BKRB1_HUMAN	A-gamma	N55 (35-64)	D83 (71-99)	R135 (107-140)	W162 (151-175)	P214 (198-231)	P265 (243-275)	P309 (289-315)
P47211	GALR1_HUMAN	A-gamma	N51 (31-60)	D81 (69-97)	R133 (105-138)	W160 (149-173)	P212 (196-229)	P262 (240-272)	P300 (280-306)
P47775	GPR12_HUMAN	A-alpha	N62 (42-71)	D90 (78-106)	R138 (110-143)	W165 (154-178)	M209 (193-226)	P266 (244-276)	P298 (278-304)
P47804	RGR_HUMAN	A-delta	N34 (14-43)	D62 (50-78)	R113 (85-118)	W136 (125-149)	P188 (172-204)	P226 (205-236)	A262 (242-268)
P47871	GLR_HUMAN	B	L156 (132-164)	F184 (172-198)	L249 (219-255)	W272 (261-289)	A314 (302-333)	V364 (341-366)	A397 (377-404)
P47872	SCTR_HUMAN	B	L157 (133-165)	F185 (173-199)	L240 (210-246)	W263 (252-280)	S305 (293-324)	V357 (334-359)	A385 (365-392)
P47881	OR3A1_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P213 (197-230)	S257 (235-267)	P290 (270-296)
P47883	OR3A4_HUMAN	Olfactory	T45 (25-54)	D73 (61-89)	R125 (97-130)	C152 (141-165)	P213 (197-230)	T257 (235-267)	P290 (270-296)
P47884	OR1D4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	L210 (194-227)	T254 (232-264)	P286 (266-292)
P47887	OR1E2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W158 (147-171)	P219 (203-236)	T263 (241-273)	P296 (276-302)
P47888	OR3A3_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P219 (203-236)	T263 (241-273)	P296 (276-302)
P47890	OR1G1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	W149 (138-162)	C210 (194-227)	T254 (232-264)	P287 (267-293)
P47893	OR3A2_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	L158 (147-171)	P219 (203-236)	R263 (241-273)	P296 (276-302)
P47898	5HT5A_HUMAN	A-alpha	N58 (38-67)	D86 (74-102)	R139 (111-144)	W166 (155-179)	P212 (196-229)	P300 (278-310)	P335 (315-341)
P47900	P2RY1_HUMAN	A-delta	N69 (49-78)	D97 (85-113)	R149 (121-154)	W176 (165-189)	P229 (213-246)	P275 (253-285)	P321 (301-327)
P47901	V1BR_HUMAN	A-beta	N52 (32-61)	D80 (68-96)	R132 (104-137)	W158 (147-171)	P211 (195-228)	P296 (274-306)	P335 (315-341)
P48039	MTR1A_HUMAN	A-alpha	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P199 (183-216)	P253 (231-263)	A292 (272-298)
P48145	NPBW1_HUMAN	A-gamma	N55 (35-64)	D83 (71-99)	R134 (106-139)	W163 (152-176)	P215 (199-232)	P266 (244-276)	P304 (284-310)
P48146	NPBW2_HUMAN	A-gamma	N63 (43-72)	D91 (79-107)	R142 (114-147)	W171 (160-184)	P224 (208-241)	P275 (253-285)	P313 (293-319)
P48546	GIPR_HUMAN	B	L152 (128-160)	F180 (168-194)	L241 (211-247)	W264 (253-281)	T306 (294-325)	V356 (333-358)	S389 (369-396)
P48960	CD97_HUMAN	Adhesion	L562 (538-570)	L590 (578-604)	L637 (607-643)	Y660 (649-677)	I703 (691-722)	F755 (732-757)	Y783 (763-790)
P49019	HCAR3_HUMAN	A-delta	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P200 (184-217)	P246 (224-256)	P291 (271-297)
P49146	NPY2R_HUMAN	A-beta	N68 (48-77)	D96 (84-112)	R148 (120-153)	W173 (162-186)	P231 (215-248)	P283 (261-293)	P322 (302-328)
P49190	PTH2R_HUMAN	B	L159 (135-167)	F187 (175-201)	L261 (231-267)	W284 (273-301)	A325 (313-344)	V378 (355-380)	S410 (390-417)
P49238	CX3C1_HUMAN	A-gamma	N49 (29-58)	D77 (65-93)	R127 (99-132)	W154 (143-167)	P203 (187-220)	P246 (224-256)	P290 (270-296)
P49286	MTR1B_HUMAN	A-alpha	N58 (38-67)	D86 (74-102)	R138 (110-143)	W165 (154-178)	P212 (196-229)	P266 (244-276)	A305 (285-311)
P49682	CXCR3_HUMAN	A-gamma	N71 (51-80)	D99 (87-115)	R149 (121-154)	W176 (165-189)	P227 (211-244)	P270 (248-280)	P315 (295-321)
P49683	PRLHR_HUMAN	A-beta	N78 (58-87)	D106 (94-122)	R159 (131-164)	W184 (173-197)	P237 (221-254)	P293 (271-303)	P332 (312-338)
P49685	GPR15_HUMAN	A-gamma	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P207 (191-224)	P256 (234-266)	P299 (279-305)
P50052	AGTR2_HUMAN	A-gamma	N62 (42-71)	D90 (78-106)	R142 (114-147)	W168 (157-181)	P223 (207-240)	P271 (249-281)	P315 (295-321)
P50391	NPY4R_HUMAN	A-beta	N59 (39-68)	D87 (75-103)	R139 (111-144)	W164 (153-177)	P226 (210-243)	P280 (258-290)	P319 (299-325)
P50406	5HT6R_HUMAN	A-alpha	N44 (24-53)	D72 (60-88)	R124 (96-129)	W151 (140-164)	P200 (184-217)	P283 (261-293)	P317 (297-323)
P51582	P2RY4_HUMAN	A-delta	N53 (33-62)	D81 (69-97)	R133 (105-138)	W160 (149-173)	P212 (196-229)	P260 (238-270)	P303 (283-309)
P51677	CCR3_HUMAN	A-gamma	N52 (32-61)	D80 (68-96)	R131 (103-136)	W158 (147-171)	P211 (195-228)	P254 (232-264)	P298 (278-304)
P51679	CCR4_HUMAN	A-gamma	N57 (37-66)	D85 (73-101)	R135 (107-140)	W162 (151-175)	P214 (198-231)	P257 (235-267)	P301 (281-307)
P51681	CCR5_HUMAN	A-gamma	N48 (28-57)	D76 (64-92)	R126 (98-131)	W153 (142-166)	P206 (190-223)	P250 (228-260)	P294 (274-300)
P51684	CCR6_HUMAN	A-gamma	N64 (44-73)	D92 (80-108)	R143 (115-148)	W172 (161-185)	P226 (210-243)	P269 (247-279)	P313 (293-319)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
P51685	CCR8_HUMAN	A-gamma	N53 (33-62)	D81 (69-97)	R131 (103-136)	W158 (147-171)	P210 (194-227)	P253 (231-263)	P297 (277-303)
P51686	CCR9_HUMAN	A-gamma	N66 (46-75)	D94 (82-110)	R144 (116-149)	W173 (162-186)	P226 (210-243)	P269 (247-279)	P314 (294-320)
P55085	PAR2_HUMAN	A-delta	N93 (73-102)	D121 (109-137)	R173 (145-178)	W199 (188-212)	P254 (238-271)	P302 (280-312)	P341 (321-347)
P58170	OR1D5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P286 (266-292)
P58173	OR2B6_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
P58180	OR4D2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	W210 (194-227)	P253 (231-263)	P284 (264-290)
P58181	O10A3_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
P58182	O12D2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	L120 (92-125)	W147 (136-160)	P208 (192-225)	P253 (231-263)	P286 (266-292)
P59533	T2R38_HUMAN	Taste2	N33 (13-42)	R64 (52-80)	Y117 (89-122)	L144 (133-157)	P205 (189-222)	A262 (240-272)	A294 (274-300)
P59534	T2R39_HUMAN	Taste2	N52 (32-61)	R83 (71-99)	Y135 (107-140)	L162 (151-175)	P220 (204-237)	A277 (255-287)	S308 (288-314)
P59535	T2R40_HUMAN	Taste2	S36 (16-45)	R67 (55-83)	Y119 (91-124)	L146 (135-159)	P205 (189-222)	A262 (240-272)	S293 (273-299)
P59536	T2R41_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	F107 (79-112)	L134 (123-147)	P192 (176-209)	S249 (227-259)	P280 (260-286)
P59537	T2R43_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	S244 (222-254)	P276 (256-282)
P59538	T2R31_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	S244 (222-254)	P276 (256-282)
P59539	T2R45_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	S244 (222-254)	P276 (256-282)
P59540	T2R46_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	S244 (222-254)	P276 (256-282)
P59541	T2R30_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	S244 (222-254)	P276 (256-282)
P59542	T2R19_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	C106 (78-111)	L133 (122-146)	P187 (171-204)	C244 (222-254)	S276 (256-282)
P59543	T2R20_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	V133 (122-146)	P187 (171-204)	C244 (222-254)	S276 (256-282)
P59544	T2R50_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	P187 (171-204)	F244 (222-254)	S276 (256-282)
P59551	T2R60_HUMAN	Taste2	N35 (15-44)	R66 (54-82)	Y118 (90-123)	L145 (134-158)	P205 (189-222)	S262 (240-272)	P293 (273-299)
P59922	OR2B8_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
P60893	GPR85_HUMAN	A-other	N38 (18-47)	D66 (54-82)	R119 (91-124)	W146 (135-159)	L190 (174-207)	P300 (278-310)	P336 (316-342)
P61073	CXCR4_HUMAN	A-gamma	N56 (36-65)	D84 (72-100)	R134 (106-139)	W161 (150-174)	P211 (195-228)	P254 (232-264)	P299 (279-305)
Q01718	ACTHR_HUMAN	A-alpha	N42 (22-51)	D70 (58-86)	R128 (100-133)	W155 (144-168)	M185 (169-202)	P234 (212-244)	P273 (253-279)
Q01726	MSHR_HUMAN	A-alpha	N56 (36-65)	D84 (72-100)	R142 (114-147)	W169 (158-182)	M199 (183-216)	P256 (234-266)	P295 (275-301)
Q02643	GHRHR_HUMAN	B	L144 (120-152)	F172 (160-186)	L227 (197-233)	W250 (239-267)	S292 (280-311)	I344 (321-346)	A373 (353-380)
Q03431	PTH1R_HUMAN	B	L202 (178-210)	F230 (218-244)	L306 (276-312)	W329 (318-346)	S370 (358-389)	V423 (400-425)	A456 (436-463)
Q13255	GRM1_HUMAN	C	T607 (591-616)	I638 (628-649)	I682 (654-688)	I714 (703-727)	L763 (750-774)	A800 (783-807)	L827 (811-840)
Q13258	PD2R_HUMAN	A-alpha	N34 (14-43)	D72 (60-88)	C130 (102-135)	S157 (146-170)	A208 (192-225)	P280 (258-290)	P320 (300-326)
Q13304	GPR17_HUMAN	A-delta	N77 (57-86)	D105 (93-121)	R157 (129-162)	W184 (173-197)	P232 (216-249)	P278 (256-288)	P322 (302-328)
Q13324	CRFR2_HUMAN	B	L131 (107-139)	F159 (147-173)	L209 (179-215)	W232 (221-249)	V275 (263-294)	L325 (302-327)	S356 (336-363)
Q13467	FZD5_HUMAN	F	T250 (229-259)	Y278 (267-289)	W340 (313-345)	W365 (357-378)	Y411 (397-432)	I463 (444-473)	G518 (500-521)
Q13585	MTR1L_HUMAN	A-alpha	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	P200 (184-217)	P252 (230-262)	A291 (271-297)
Q13606	OR5I1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	Y151 (140-164)	C212 (196-229)	T256 (234-266)	P289 (269-295)
Q13607	OR2F1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	V254 (232-264)	P287 (267-293)
Q13639	5HT4R_HUMAN	A-alpha	N37 (17-46)	D66 (54-82)	R118 (90-123)	W146 (135-159)	P204 (188-221)	P274 (252-284)	P309 (289-315)
Q14246	EMR1_HUMAN	Adhesion	L617 (593-625)	L645 (633-659)	L690 (660-696)	Y718 (707-735)	V761 (749-780)	L813 (790-815)	F841 (821-848)
Q14330	GPR18_HUMAN	A-delta	N40 (20-49)	D68 (56-84)	R119 (91-124)	W146 (135-159)	P199 (183-216)	P247 (225-257)	V283 (263-289)
Q14332	FZD2_HUMAN	F	T259 (238-268)	Y287 (276-298)	W352 (325-357)	W377 (369-390)	Y423 (409-444)	I475 (456-485)	G536 (518-539)
Q14416	GRM2_HUMAN	C	T582 (566-591)	V613 (603-624)	I657 (629-663)	I687 (676-700)	L738 (725-749)	A775 (758-782)	L806 (790-819)
Q14439	GP176_HUMAN	A-other	N59 (39-68)	G87 (75-103)	R142 (114-147)	W167 (156-180)	P217 (201-234)	P281 (259-291)	P317 (297-323)
Q14831	GRM7_HUMAN	C	T605 (589-614)	I636 (626-647)	I680 (652-686)	I710 (699-723)	L766 (753-777)	A803 (786-810)	L837 (821-850)
Q14832	GRM3_HUMAN	C	T591 (575-600)	V622 (612-633)	I666 (638-672)	I696 (685-709)	L747 (734-758)	A784 (767-791)	L815 (799-828)
Q14833	GRM4_HUMAN	C	T602 (586-611)	I633 (623-644)	I677 (649-683)	I707 (696-720)	L763 (750-774)	A800 (783-807)	L834 (818-847)
Q15077	P2RY6_HUMAN	A-delta	N44 (24-53)	D72 (60-88)	R124 (96-129)	W152 (141-165)	P204 (188-221)	P254 (232-264)	P298 (278-304)
Q15391	P2Y14_HUMAN	A-delta	N40 (20-49)	D67 (55-83)	R119 (91-124)	W146 (135-159)	V198 (182-215)	P248 (226-258)	P292 (272-298)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q15612	OR1Q1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	H149 (138-162)	A210 (194-227)	T254 (232-264)	P287 (267-293)
Q15614	OR8G2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	Y149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q15615	OR4D1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	W210 (194-227)	P253 (231-263)	P284 (264-290)
Q15617	OR8G1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q15619	OR1C1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q15620	OR8B8_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q15622	OR7A5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q15722	LT4R1_HUMAN	A-gamma	N36 (16-45)	D64 (52-80)	R115 (87-120)	W142 (131-155)	P193 (177-210)	P236 (214-246)	P282 (262-288)
Q15743	OGR1_HUMAN	A-delta	N39 (19-48)	D67 (55-83)	R119 (91-124)	W146 (135-159)	P197 (181-214)	P243 (221-253)	P283 (263-289)
Q15760	GPR19_HUMAN	A-other	N83 (63-92)	D111 (99-127)	R163 (135-168)	W188 (177-201)	P235 (219-252)	P291 (269-301)	P327 (307-333)
Q15761	NPY5R_HUMAN	A-beta	N59 (39-68)	D87 (75-103)	R139 (111-144)	W164 (153-177)	P223 (207-240)	P383 (361-393)	P422 (402-428)
Q16538	GP162_HUMAN	A-other	N31 (11-40)	H59 (47-75)	R115 (87-120)	W142 (131-155)	G189 (173-206)	P287 (265-297)	P321 (301-327)
Q16570	ACKR1_HUMAN	A-other	S76 (56-85)	S106 (94-122)	H155 (127-160)	W174 (163-187)	P222 (206-235)	P255 (236-265)	P301 (281-307)
Q16581	C3AR_HUMAN	A-delta	N41 (21-50)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P348 (332-365)	P392 (370-402)	P432 (412-438)
Q16602	CALRL_HUMAN	B	L156 (132-164)	F184 (172-198)	L237 (207-243)	W260 (249-277)	A301 (289-320)	L351 (328-353)	S381 (361-388)
Q49SQ1	GPR33_HUMAN	A-other	N47 (27-56)	Y74 (62-90)	R126 (98-131)	W153 (142-166)	P217 (201-234)	P260 (238-270)	P296 (276-302)
Q5JQS5	OR2BB_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	P214 (198-231)	P258 (236-268)	P291 (271-297)
Q5JRS4	O10J3_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R123 (95-128)	L150 (139-163)	P211 (195-228)	C255 (233-265)	C288 (268-294)
Q5NUL3	FFAR4_HUMAN	A-other	N58 (38-67)	D85 (73-101)	R136 (108-141)	W163 (152-176)	P219 (203-236)	P295 (273-305)	P334 (314-340)
Q5T601	GP110_HUMAN	Adhesion	L600 (576-608)	L634 (622-648)	L682 (652-688)	Y707 (696-724)	I754 (742-773)	F806 (783-808)	L835 (815-842)
Q5T6X5	GPC6A_HUMAN	C	V608 (592-617)	H640 (630-651)	I684 (656-690)	T712 (701-725)	L762 (749-773)	T797 (780-804)	L826 (810-839)
Q5T848	GP158_HUMAN	C	D431 (415-440)	S462 (452-473)	V506 (478-512)	I533 (522-546)	F589 (576-600)	I626 (609-633)	I659 (643-672)
Q5TZ20	OR2G6_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q5UAW9	GP157_HUMAN	B	S29 (5-37)	D57 (45-71)	L104 (74-110)	W129 (118-146)	E178 (166-197)	V239 (216-241)	C276 (256-283)
Q6DWJ6	GP139_HUMAN	A-other	N44 (24-53)	D73 (61-89)	R126 (98-131)	Y153 (142-166)	P195 (179-212)	P243 (221-253)	F282 (262-288)
Q6IEU7	OR5MA_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q6IEV9	OR4CB_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)
Q6IEY1	OR4F3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	P253 (231-263)	P284 (264-290)
Q6IEZ7	OR2T5_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	P214 (198-231)	A258 (236-268)	P291 (271-297)
Q6IF00	OR2T2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	A255 (233-265)	P288 (268-294)
Q6IF42	OR2A2_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	I253 (231-263)	P286 (266-292)
Q6IF63	O52W1_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	A154 (143-167)	D215 (199-232)	P259 (237-269)	P295 (275-301)
Q6IF82	O4A47_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	C120 (92-125)	W147 (136-160)	V208 (192-225)	P251 (229-261)	P282 (262-288)
Q6IF99	O10K2_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	C149 (138-162)	P210 (194-227)	C254 (232-264)	P287 (267-293)
Q6IFG1	O52E8_HUMAN	Olfactory	N48 (28-57)	D76 (64-92)	R128 (100-133)	V155 (144-168)	D215 (199-232)	P259 (237-269)	P294 (274-300)
Q6IFH4	OR6B2_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	F149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q6IFN5	O7E24_HUMAN	Olfactory	N60 (40-69)	D88 (76-104)	R140 (112-145)	F167 (156-180)	P228 (212-245)	T272 (250-282)	P305 (285-311)
Q6NV75	GP153_HUMAN	A-other	N25 (5-34)	H53 (41-69)	R108 (80-113)	W135 (124-148)	G182 (166-199)	G258 (236-268)	P294 (274-300)
Q6PRD1	GP179_HUMAN	C	I395 (379-404)	F426 (416-437)	V470 (442-476)	L497 (486-510)	L553 (540-564)	A590 (573-597)	L623 (607-636)
Q6QNK2	GP133_HUMAN	Adhesion	L581 (557-589)	V612 (600-626)	L657 (627-663)	W680 (669-697)	V723 (711-742)	F775 (752-777)	F803 (783-810)
Q6U736	OPN5_HUMAN	A-other	N51 (31-60)	D79 (67-95)	R131 (103-136)	W157 (146-170)	P211 (195-228)	P267 (245-277)	P303 (283-309)
Q6W5P4	NPSR1_HUMAN	A-other	N67 (47-76)	D94 (82-110)	R146 (118-151)	W171 (160-184)	P222 (206-239)	P289 (267-299)	P327 (307-333)
Q7RTR8	T2R42_HUMAN	Taste2	N24 (4-33)	T55 (43-71)	Y107 (79-112)	L134 (123-147)	P192 (176-209)	S249 (227-259)	S281 (261-287)
Q7RTX0	TS1R3_HUMAN	C	V583 (567-592)	L614 (604-625)	I658 (630-664)	L687 (676-700)	L740 (727-751)	S777 (760-784)	L806 (790-819)
Q7RTX1	TS1R1_HUMAN	C	L581 (565-590)	L612 (602-623)	L656 (628-662)	S687 (676-700)	L738 (725-749)	A775 (758-782)	F804 (788-817)
Q7Z3T1	OR2W3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	N254 (232-264)	P287 (267-293)
Q7Z5H4	VN1R5_HUMAN	Vomeronasal	N60 (40-69)	H91 (79-107)	Q143 (115-148)	W172 (161-185)	F222 (206-239)	D284 (262-294)	P317 (297-323)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q7Z5H5	VN1R4_HUMAN	Vomeronasal	S22 (2-31)	N53 (41-69)	Q105 (77-110)	W135 (124-148)	C191 (175-208)	L248 (226-258)	P282 (262-288)
Q7Z601	GP142_HUMAN	A-other	V173 (153-182)	S202 (190-218)	D255 (227-260)	V282 (271-295)	I324 (308-341)	A371 (349-381)	N410 (390-416)
Q7Z602	GP141_HUMAN	A-other	V35 (15-44)	H62 (50-78)	R113 (85-118)	W140 (129-153)	A191 (175-208)	P240 (218-250)	L280 (260-286)
Q7Z7M1	GP144_HUMAN	Adhesion	L673 (649-681)	L701 (689-715)	L746 (716-752)	W769 (758-786)	V812 (800-831)	A872 (849-874)	F898 (878-905)
Q86SM5	MRGRG_HUMAN	A-other	N30 (10-39)	D57 (45-73)	R103 (75-108)	W130 (119-143)	L170 (154-187)	P213 (191-223)	P247 (227-253)
Q86SM8	MRGRE_HUMAN	A-other	N42 (22-51)	D69 (57-85)	Q121 (93-126)	W148 (137-161)	L188 (172-205)	P232 (210-242)	P266 (246-272)
Q86SP6	GP149_HUMAN	A-other	S51 (31-60)	D78 (66-94)	F131 (103-136)	W158 (147-171)	L205 (189-222)	P325 (303-335)	P359 (339-365)
Q86SQ3	EMR4_HUMAN	Adhesion	L205 (181-213)	L233 (221-247)	L278 (248-284)	Y306 (295-323)	I348 (336-367)	L400 (377-402)	F432 (412-439)
Q86SQ4	GP126_HUMAN	Adhesion	S879 (855-887)	L908 (896-922)	M955 (925-961)	W979 (968-996)	M1032 (1020-1051)	F1083 (1060-1085)	F1111 (1091-1118)
Q86SQ6	GP123_HUMAN	Adhesion	L32 (8-40)	A63 (51-77)	I108 (78-114)	G140 (129-157)	I186 (174-205)	F272 (249-274)	L303 (283-310)
Q86VZ1	P2RY8_HUMAN	A-other	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P205 (189-222)	P252 (230-262)	P290 (270-296)
Q86Y34	GPR97_HUMAN	Adhesion	L283 (259-291)	L316 (304-330)	L363 (333-369)	W387 (376-404)	T442 (430-461)	L492 (469-494)	C518 (498-525)
Q8IWK6	GP125_HUMAN	Adhesion	L774 (750-782)	I805 (793-819)	I850 (820-856)	G882 (871-899)	I927 (915-946)	F1012 (989-1014)	V1043 (1023-1050)
Q8IXE1	OR4N5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8IYL9	PSYR_HUMAN	A-delta	N32 (12-41)	D60 (48-76)	R112 (84-117)	W139 (128-152)	P195 (179-212)	P241 (219-251)	P287 (267-293)
Q8IZ08	GP135_HUMAN	A-other	N121 (101-130)	D149 (137-165)	R209 (181-214)	W234 (223-247)	P289 (273-306)	P346 (324-356)	P384 (364-390)
Q8IZF2	GP116_HUMAN	Adhesion	L1027 (1003-1035)	L1061 (1049-1075)	L1111 (1081-1117)	Y1136 (1125-1153)	I1182 (1170-1201)	F1234 (1211-1236)	L1263 (1243-1270)
Q8IZF3	GP115_HUMAN	Adhesion	L416 (392-424)	L450 (438-464)	I498 (468-504)	Y523 (512-540)	I569 (557-588)	F620 (597-622)	L649 (629-656)
Q8IZF4	GP114_HUMAN	Adhesion	S263 (239-271)	V291 (279-305)	L339 (309-345)	W363 (352-380)	T423 (411-442)	L472 (449-474)	F500 (480-507)
Q8IZF5	GP113_HUMAN	Adhesion	L785 (761-793)	L819 (807-833)	L866 (836-872)	Y891 (880-908)	I936 (924-955)	L988 (965-990)	L1017 (997-1024)
Q8IZF6	GP112_HUMAN	Adhesion	L2758 (2734-2766)	L2787 (2775-2801)	M2834 (2804-2840)	W2858 (2847-2875)	I2905 (2893-2924)	F2955 (2932-2957)	F2983 (2963-2990)
Q8IZF7	GP111_HUMAN	Adhesion	L461 (437-469)	L495 (483-509)	I543 (513-549)	Y568 (557-585)	I614 (602-633)	F665 (642-667)	A694 (674-701)
Q8IZP9	GPR64_HUMAN	Adhesion	L643 (619-651)	L672 (660-686)	M719 (689-725)	W743 (732-760)	I797 (785-816)	F848 (825-850)	F876 (856-883)
Q8N0Y3	OR4N4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8N0Y5	OR8I2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S209 (193-226)	S253 (231-263)	P286 (266-292)
Q8N127	O5AS1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	A254 (232-264)	P287 (267-293)
Q8N146	OR8H3_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8N148	OR6V1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	S252 (230-262)	P285 (265-291)
Q8N162	OR8H2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8N349	OR2LD_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	P253 (231-263)	P286 (266-292)
Q8N628	OR2C3_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	S255 (233-265)	P288 (268-294)
Q8N6U8	GP161_HUMAN	A-delta	N45 (25-54)	N73 (61-89)	R125 (97-130)	W152 (141-165)	P200 (184-217)	P285 (263-295)	P321 (301-327)
Q8NDV2	GPR26_HUMAN	A-alpha	N23 (3-32)	N52 (40-68)	R104 (76-109)	W131 (120-144)	S184 (168-201)	P256 (234-266)	P291 (271-297)
Q8NFJ5	RAI3_HUMAN	C	S42 (26-51)	V75 (65-86)	L119 (91-125)	S142 (131-155)	L187 (174-198)	A226 (209-233)	L257 (241-270)
Q8NFJ6	PKR2_HUMAN	A-other	N71 (51-80)	D99 (87-115)	R153 (125-158)	W178 (167-191)	P235 (219-252)	P290 (268-300)	T330 (310-336)
Q8NFN8	GP156_HUMAN	C	I64 (48-73)	S95 (85-106)	L145 (117-151)	L173 (162-186)	L235 (222-246)	G272 (255-279)	F297 (281-310)
Q8NFZ6	VN1R2_HUMAN	Vomeronasal	N104 (84-113)	D135 (123-151)	Q187 (159-192)	W217 (206-230)	C273 (257-290)	L330 (308-340)	P364 (344-370)
Q8NG75	OR5T1_HUMAN	Olfactory	N54 (34-63)	D82 (70-98)	R134 (106-139)	Y161 (150-174)	T222 (206-239)	T266 (244-276)	P299 (279-305)
Q8NG76	O2T33_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	A252 (230-262)	P285 (265-291)
Q8NG77	O2T12_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	A252 (230-262)	P285 (265-291)
Q8NG78	OR8G5_HUMAN	Olfactory	N77 (57-86)	D105 (93-121)	G157 (129-162)	Y184 (173-197)	P245 (229-262)	S289 (267-299)	P322 (302-328)
Q8NG80	OR2L5_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	P253 (231-263)	P286 (266-292)
Q8NG81	OR2M7_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NG83	OR2M3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NG84	O2AK2_HUMAN	Olfactory	N57 (37-66)	D85 (73-101)	R137 (109-142)	W164 (153-177)	P225 (209-242)	T269 (247-279)	P302 (282-308)
Q8NG85	OR2L3_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	P253 (231-263)	P286 (266-292)
Q8NG92	O13H1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P209 (193-226)	S253 (231-263)	P286 (266-292)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NG94	O11H1_HUMAN	Olfactory	N57 (37-66)	E85 (73-101)	Q137 (109-142)	W164 (153-177)	N225 (209-242)	S269 (247-279)	P302 (282-308)
Q8NG95	OR7G3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NG97	OR2Z1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NG98	OR7D4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NG99	OR7G2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	L149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NGA0	OR7G1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	M149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGA1	OR1M1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGA2	OR7A2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	Q123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P288 (268-294)
Q8NGA5	O10H4_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R123 (95-128)	W150 (139-163)	C212 (196-229)	F256 (234-266)	P289 (269-295)
Q8NGA6	O10H5_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	C211 (195-228)	F255 (233-265)	P288 (268-294)
Q8NGA8	O4F17_HUMAN	Olfactory	N35 (15-44)	D63 (51-79)	R115 (87-120)	W142 (131-155)	S203 (187-220)	P246 (224-256)	P277 (257-283)
Q8NGB2	OR4C5_HUMAN	Olfactory	N63 (43-72)	D91 (79-107)	C143 (115-148)	W170 (159-183)	I231 (215-245)	P266 (246-276)	P297 (277-303)
Q8NGB4	OR4S1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NGB6	OR4M2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	L122 (94-127)	W149 (138-162)	C210 (194-227)	P255 (233-265)	P286 (266-292)
Q8NGB8	O4F15_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	S149 (138-162)	S210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGB9	OR4F6_HUMAN	Olfactory	N42 (22-51)	N70 (58-86)	R122 (94-127)	W149 (138-162)	S210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGC0	O5AU1_HUMAN	Olfactory	N93 (73-102)	D121 (109-137)	R173 (145-178)	Y200 (189-213)	C261 (245-278)	T305 (283-315)	P338 (318-344)
Q8NGC1	O11G2_HUMAN	Olfactory	N80 (60-89)	E108 (96-124)	R160 (132-165)	W187 (176-200)	L248 (232-265)	S292 (270-302)	P325 (305-331)
Q8NGC2	OR4E2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGC3	O10G2_HUMAN	Olfactory	N46 (26-55)	D75 (63-91)	R127 (99-132)	W154 (143-167)	C215 (199-232)	P259 (237-269)	P290 (270-296)
Q8NGC4	O10G3_HUMAN	Olfactory	N42 (22-51)	D71 (59-87)	R123 (95-128)	W150 (139-163)	C211 (195-228)	P255 (233-265)	P286 (266-292)
Q8NGC5	OR6J1_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	C209 (193-226)	I253 (231-263)	P286 (266-292)
Q8NGC6	OR4KH_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	C213 (197-230)	P256 (234-266)	P287 (267-293)
Q8NGC7	O11H6_HUMAN	Olfactory	N60 (40-69)	E88 (76-104)	R140 (112-145)	W167 (156-180)	P228 (212-245)	T272 (250-282)	P305 (285-311)
Q8NGC8	O11H7_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NGC9	O11H4_HUMAN	Olfactory	N52 (32-61)	E80 (68-96)	R132 (104-137)	W159 (148-172)	T220 (204-237)	T264 (242-274)	P297 (277-303)
Q8NGD0	OR4M1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P255 (233-265)	P286 (266-292)
Q8NGD1	OR4N2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGD2	OR4K2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGD3	OR4K5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGD4	OR4K1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGD5	OR4KE_HUMAN	Olfactory	N42 (22-51)	D71 (59-87)	R123 (95-128)	W150 (139-163)	C211 (195-228)	P254 (232-264)	P285 (265-291)
Q8NGE0	O10AD_HUMAN	Olfactory	N42 (22-51)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGE1	OR6C4_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
Q8NGE2	O2AP1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
Q8NGE3	O10P1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P288 (268-294)
Q8NGE5	O10A7_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGE7	OR9K2_HUMAN	Olfactory	N67 (47-76)	D95 (83-111)	R147 (119-152)	Y174 (163-187)	T235 (219-252)	A279 (257-289)	P310 (290-316)
Q8NGE8	OR4D9_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	V210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGE9	OR9Q2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	C210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGF0	O52B6_HUMAN	Olfactory	N62 (42-71)	D90 (78-106)	R141 (113-146)	L168 (157-181)	D229 (213-246)	P273 (251-283)	P309 (289-315)
Q8NGF1	O52R1_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	C125 (97-130)	M152 (141-165)	D213 (197-230)	P257 (235-267)	P292 (272-298)
Q8NGF3	O51D1_HUMAN	Olfactory	N55 (35-64)	D83 (71-99)	R135 (107-140)	L162 (151-175)	D223 (207-240)	P267 (245-277)	P301 (281-307)
Q8NGF4	O5AP2_HUMAN	Olfactory	N48 (28-57)	D76 (64-92)	R128 (100-133)	F155 (144-168)	C216 (200-233)	T260 (238-270)	P293 (273-299)
Q8NGF6	O10W1_HUMAN	Olfactory	N33 (13-42)	E61 (49-77)	R113 (85-118)	V140 (129-153)	P201 (185-218)	C245 (223-255)	P278 (258-284)
Q8NGF7	OR5BH_HUMAN	Olfactory	N40 (20-49)	G68 (56-84)	R120 (92-125)	Y147 (136-160)	A208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NGF8	OR4B1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	C120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NGF9	OR4X2_HUMAN	Olfactory	N34 (14-43)	E62 (50-78)	H114 (86-119)	W141 (130-154)	S202 (186-219)	P245 (223-255)	P276 (256-282)
Q8NGG0	OR8J3_HUMAN	Olfactory	N42 (22-51)	N70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGG1	OR8J2_HUMAN	Olfactory	N42 (22-51)	N70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGG2	OR5T2_HUMAN	Olfactory	N81 (61-90)	D109 (97-125)	R161 (133-166)	Y188 (177-201)	T249 (233-266)	T293 (271-303)	P326 (306-332)
Q8NGG3	OR5T3_HUMAN	Olfactory	N72 (52-81)	D100 (88-116)	H152 (124-157)	Y179 (168-192)	T240 (224-257)	T284 (262-294)	P317 (297-323)
Q8NGG4	OR8H1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R121 (93-126)	Y148 (137-161)	S209 (193-226)	T253 (231-263)	P286 (266-292)
Q8NGG5	OR8K1_HUMAN	Olfactory	N48 (28-57)	D76 (64-92)	R128 (100-133)	Y155 (144-168)	S216 (200-233)	T260 (238-270)	P293 (273-299)
Q8NGG6	OR8BC_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	Y148 (137-161)	P209 (193-226)	S253 (231-263)	P286 (266-292)
Q8NGG7	OR8A1_HUMAN	Olfactory	N59 (39-68)	D87 (75-103)	R139 (111-144)	Y166 (155-179)	T226 (210-243)	S270 (248-280)	P303 (283-309)
Q8NGG8	OR8B3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	P210 (194-227)	S254 (232-264)	P286 (266-292)
Q8NGH3	OR2D3_HUMAN	Olfactory	N58 (38-67)	D86 (74-102)	R138 (110-143)	W165 (154-178)	P226 (210-243)	S270 (248-280)	P303 (283-309)
Q8NGH5	O56A1_HUMAN	Olfactory	N49 (29-58)	D77 (65-93)	R129 (101-134)	V156 (145-169)	D217 (201-234)	I261 (239-271)	P296 (276-302)
Q8NGH6	O52L2_HUMAN	Olfactory	N60 (40-69)	D88 (76-104)	R140 (112-145)	L167 (156-180)	D228 (212-245)	P272 (250-282)	P307 (287-313)
Q8NGH7	O52L1_HUMAN	Olfactory	N60 (40-69)	D88 (76-104)	C140 (112-145)	L167 (156-180)	D228 (212-245)	P272 (250-282)	P307 (287-313)
Q8NGH8	O56A4_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	I152 (141-165)	D213 (197-230)	V257 (235-267)	P292 (272-298)
Q8NGH9	O52E4_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D211 (195-228)	P255 (233-265)	P290 (270-296)
Q8NGI0	O52N2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	F151 (140-164)	D212 (196-229)	A256 (234-266)	P292 (272-298)
Q8NGI1	O56B2_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	A158 (147-171)	D219 (203-236)	V263 (241-273)	P297 (277-303)
Q8NGI2	O52N4_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	F151 (140-164)	D212 (196-229)	P256 (234-266)	P292 (272-298)
Q8NGI3	O56B1_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	V158 (147-171)	D219 (203-236)	I263 (241-273)	P297 (277-303)
Q8NGI4	OR4DB_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	W149 (138-162)	W210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGI6	OR4DA_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	W210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGI7	O10V1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	Q123 (95-128)	L150 (139-163)	P211 (195-228)	C255 (233-265)	P288 (268-294)
Q8NGI8	O5AN1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	Y150 (139-163)	S211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGI9	OR5A2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	Y150 (139-163)	S211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGJ0	OR5A1_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	Y152 (141-165)	S213 (197-230)	T257 (235-267)	P290 (270-296)
Q8NGJ1	OR4D6_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	W210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGJ2	O52H1_HUMAN	Olfactory	N50 (30-59)	D78 (66-94)	H130 (102-135)	S157 (146-170)	D218 (202-235)	P262 (240-272)	P297 (277-303)
Q8NGJ3	O52E1_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	C121 (93-126)	F148 (137-161)	D208 (192-225)	P252 (230-262)	S287 (267-293)
Q8NGJ4	O52E2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	S124 (96-129)	F151 (140-164)	D211 (195-228)	P255 (233-265)	P290 (270-296)
Q8NGJ5	O51L1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	L151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NGJ6	O51A4_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	S151 (140-164)	D211 (195-228)	P255 (233-265)	P290 (270-296)
Q8NGJ7	O51A2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	S151 (140-164)	D211 (195-228)	P255 (233-265)	P290 (270-296)
Q8NGJ8	O51S1_HUMAN	Olfactory	N50 (30-59)	D78 (66-94)	R130 (102-135)	S157 (146-170)	D217 (201-234)	P261 (239-271)	P296 (276-302)
Q8NGJ9	O51T1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	C151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NGK0	O51G2_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	L154 (143-167)	D215 (199-232)	P259 (237-269)	P294 (274-300)
Q8NGK1	O51G1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NGK2	O52B4_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	H124 (96-129)	S151 (140-164)	D212 (196-229)	S256 (234-266)	P291 (271-297)
Q8NGK3	O52K2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D212 (196-229)	T256 (234-266)	P291 (271-297)
Q8NGK4	O52K1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NGK5	O52M1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	L151 (140-164)	D213 (197-230)	P257 (235-267)	P292 (272-298)
Q8NGK6	O52I1_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	T153 (142-166)	D214 (198-231)	P258 (236-268)	P294 (274-300)
Q8NGK9	OR5DG_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	H124 (96-129)	Y151 (140-164)	T212 (196-229)	T256 (234-266)	P289 (269-295)
Q8NGL0	OR5L2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGL1	OR5DI_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	Y150 (139-163)	T211 (195-228)	T255 (233-265)	P288 (268-294)
Q8NGL2	OR5L1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	S287 (267-293)
Q8NGL3	OR5DE_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	Y151 (140-164)	T212 (196-229)	T256 (234-266)	P289 (269-295)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NGL4	OR5DD_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	Y151 (140-164)	S212 (196-229)	T256 (234-266)	P289 (269-295)
Q8NGL6	O4A15_HUMAN	Olfactory	N70 (50-79)	D98 (86-114)	R150 (122-155)	W177 (166-190)	T238 (222-255)	P281 (259-291)	P312 (292-318)
Q8NGL7	OR4P4_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	C147 (136-160)	T208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NGL9	OR4CG_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NGM1	OR4CF_HUMAN	Olfactory	N40 (20-49)	D69 (57-85)	R121 (93-126)	W148 (137-161)	N209 (193-226)	P252 (230-262)	P283 (263-289)
Q8NGM8	OR6M1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	S207 (191-224)	S251 (229-261)	P284 (264-290)
Q8NGM9	OR8D4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGN0	OR4D5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NGN1	OR6T1_HUMAN	Olfactory	K42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	S210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGN2	O10S1_HUMAN	Olfactory	N55 (35-64)	D83 (71-99)	R136 (108-141)	W163 (152-176)	C224 (208-241)	P268 (246-278)	P299 (279-305)
Q8NGN3	O10G4_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R121 (93-126)	W148 (137-161)	C209 (193-226)	P253 (231-263)	P284 (264-290)
Q8NGN4	O10G9_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R121 (93-126)	W148 (137-161)	C209 (193-226)	P253 (231-263)	P284 (264-290)
Q8NGN5	O10G8_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R121 (93-126)	W148 (137-161)	C209 (193-226)	P253 (231-263)	P284 (264-290)
Q8NGN6	O10G7_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R121 (93-126)	W148 (137-161)	C209 (193-226)	P253 (231-263)	P284 (264-290)
Q8NGN7	O10D4_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	P252 (230-262)	P283 (263-289)
Q8NGN8	OR4A4_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	A208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NGP0	OR4CD_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	H120 (92-125)	W147 (136-160)	N208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NGP2	OR8J1_HUMAN	Olfactory	N42 (22-51)	N70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P288 (268-294)
Q8NGP3	OR5M9_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	S208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NGP4	OR5M3_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	S208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NGP6	OR5M8_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	Y148 (137-161)	S209 (193-226)	T253 (231-263)	L286 (266-292)
Q8NGP8	OR5M1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGP9	O5AR1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGQ1	OR9G4_HUMAN	Olfactory	N57 (37-66)	D85 (73-101)	R137 (109-142)	Y164 (153-177)	S225 (209-242)	S269 (247-279)	P302 (282-308)
Q8NGQ2	OR6Q1_HUMAN	Olfactory	N44 (24-53)	E72 (60-88)	R126 (98-131)	W153 (142-166)	S214 (198-231)	T258 (236-268)	P291 (271-297)
Q8NGQ3	OR1S2_HUMAN	Olfactory	N55 (35-64)	D83 (71-99)	H135 (107-140)	W162 (151-175)	P223 (207-240)	T267 (245-277)	P300 (280-306)
Q8NGQ4	O10Q1_HUMAN	Olfactory	N46 (26-55)	E74 (62-90)	R126 (98-131)	L153 (142-166)	P215 (199-232)	C259 (237-269)	P292 (272-298)
Q8NGQ5	OR9Q1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T253 (231-263)	P286 (266-292)
Q8NGQ6	OR9I1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	N210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NGR1	O13A1_HUMAN	Olfactory	N60 (40-69)	D88 (76-104)	R140 (112-145)	W167 (156-180)	N228 (212-245)	A272 (250-282)	P305 (285-311)
Q8NGR2	OR1L6_HUMAN	Olfactory	N79 (59-88)	D107 (95-123)	R159 (131-164)	C186 (175-199)	P247 (231-264)	S291 (269-301)	P324 (304-330)
Q8NGR3	OR1K1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	C123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P288 (268-294)
Q8NGR4	OR5C1_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	G153 (142-166)	T214 (198-231)	T258 (236-268)	P291 (271-297)
Q8NGR5	OR1L4_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	C150 (139-163)	P211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGR6	OR1B1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	W151 (140-164)	P216 (200-233)	T260 (238-270)	P293 (273-299)
Q8NGR8	OR1L8_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	C150 (139-163)	R211 (195-228)	S255 (233-265)	P287 (267-293)
Q8NGR9	OR1N2_HUMAN	Olfactory	N59 (39-68)	D87 (75-103)	R139 (111-144)	W166 (155-179)	P227 (211-244)	S271 (249-281)	P304 (284-310)
Q8NGS0	OR1N1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NGS1	OR1J4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGS2	OR1J2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGS3	OR1J1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGS4	O13F1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGS5	O13C4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P293 (273-299)
Q8NGS6	O13C3_HUMAN	Olfactory	N72 (52-81)	D100 (88-116)	R152 (124-157)	W179 (168-192)	P240 (224-257)	T284 (262-294)	P323 (303-329)
Q8NGS7	O13C8_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P293 (273-299)
Q8NGS8	O13C5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P293 (273-299)
Q8NGS9	O13C2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P293 (273-299)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NGT0	O13C9_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P293 (273-299)
Q8NGT1	OR2K2_HUMAN	Olfactory	N71 (51-80)	D99 (87-115)	R151 (123-156)	W178 (167-191)	P238 (222-255)	A282 (260-292)	P315 (295-321)
Q8NGT2	O13J1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGT5	OR9A2_HUMAN	Olfactory	N41 (21-50)	E69 (57-85)	R118 (90-123)	W145 (134-158)	S206 (190-223)	S250 (228-260)	P283 (263-289)
Q8NGT7	O2A12_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	S253 (231-263)	P286 (266-292)
Q8NGT9	OR2A1_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	S253 (231-263)	P286 (266-292)
Q8NGU2	OR9A4_HUMAN	Olfactory	N41 (21-50)	E69 (57-85)	R122 (94-127)	W149 (138-162)	S210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGU4	OR2I1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGU9	GP150_HUMAN	A-other	N60 (40-69)	D93 (81-109)	R147 (119-152)	W169 (158-182)	P244 (228-261)	P310 (288-320)	P349 (329-355)
Q8NGV0	OR2Y1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P209 (193-226)	S253 (231-263)	P286 (266-292)
Q8NGV5	O13D1_HUMAN	Olfactory	N74 (54-83)	D102 (90-118)	H154 (126-159)	W181 (170-194)	L242 (226-259)	S286 (264-296)	P319 (299-325)
Q8NGV6	OR5H6_HUMAN	Olfactory	N58 (38-67)	D86 (74-102)	R138 (110-143)	F165 (154-178)	T226 (210-243)	P270 (248-280)	P303 (283-309)
Q8NGV7	OR5H2_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	F154 (143-167)	T215 (199-232)	P259 (237-269)	P292 (272-298)
Q8NGW1	OR6B3_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	F149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NGW6	OR6K6_HUMAN	Olfactory	N70 (50-79)	E98 (86-114)	R150 (122-155)	C177 (166-190)	S237 (221-254)	S281 (259-291)	P314 (294-320)
Q8NGX0	O11L1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGX1	O2T34_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	W154 (143-167)	P215 (199-232)	A259 (237-269)	P292 (272-298)
Q8NGX2	O2T35_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NGX3	O10T2_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R123 (95-128)	G150 (139-163)	P211 (195-228)	C254 (232-264)	P287 (267-293)
Q8NGX5	O10K1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	C149 (138-162)	P210 (194-227)	C254 (232-264)	P287 (267-293)
Q8NGX6	O10R2_HUMAN	Olfactory	N62 (42-71)	E90 (78-106)	R142 (114-147)	A169 (158-182)	P230 (214-247)	C274 (252-284)	P307 (287-313)
Q8NGX8	OR6Y1_HUMAN	Olfactory	N47 (27-56)	E75 (63-91)	R127 (99-132)	W154 (143-167)	P215 (199-232)	M259 (237-269)	P292 (272-298)
Q8NGX9	OR6P1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGY0	O10X1_HUMAN	Olfactory	N58 (38-67)	E86 (74-102)	R138 (110-143)	C165 (154-178)	T226 (210-243)	F270 (248-279)	P300 (280-306)
Q8NGY1	O10Z1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	F149 (138-162)	S210 (194-227)	C254 (232-264)	P287 (267-293)
Q8NGY2	OR6K2_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	H121 (93-126)	C148 (137-161)	A211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NGY3	OR6K3_HUMAN	Olfactory	N58 (38-67)	E86 (74-102)	R138 (110-143)	C165 (154-178)	T225 (209-242)	S269 (247-279)	P302 (282-308)
Q8NGY5	OR6N1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGY6	OR6N2_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGY9	OR2L8_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	P253 (231-263)	P286 (266-292)
Q8NGZ0	O2AJ1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	P254 (232-264)	P287 (267-293)
Q8NGZ2	O14K1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NGZ3	O13G1_HUMAN	Olfactory	N39 (19-48)	D67 (55-83)	R119 (91-124)	M146 (135-159)	D207 (191-224)	P251 (229-261)	P284 (264-290)
Q8NGZ4	OR2G3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NGZ5	OR2G2_HUMAN	Olfactory	N45 (25-54)	Y73 (61-89)	R125 (97-130)	W152 (141-165)	P213 (197-230)	T257 (235-267)	P290 (270-296)
Q8NGZ6	OR6F1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	S210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NGZ9	O2T10_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NH00	OR2T4_HUMAN	Olfactory	N74 (54-83)	D102 (90-118)	R154 (126-159)	W181 (170-194)	P242 (226-259)	A286 (264-296)	P319 (299-325)
Q8NH01	O2T11_HUMAN	Olfactory	N39 (19-48)	D67 (55-83)	C119 (91-124)	W146 (135-159)	P207 (191-224)	A251 (229-261)	P284 (264-290)
Q8NH02	O2T29_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	W153 (142-166)	P214 (198-231)	A258 (236-268)	P291 (271-297)
Q8NH03	OR2T3_HUMAN	Olfactory	N47 (27-56)	D75 (63-91)	R127 (99-132)	W154 (143-167)	P215 (199-232)	A259 (237-269)	P292 (272-298)
Q8NH04	O2T27_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NH05	OR4Q3_HUMAN	Olfactory	N42 (22-51)	D71 (59-87)	R123 (95-128)	W150 (139-163)	C211 (195-228)	P254 (232-264)	P285 (265-291)
Q8NH06	OR1P1_HUMAN	Olfactory	N56 (36-65)	D84 (72-100)	R136 (108-141)	W163 (152-176)	P223 (207-240)	T267 (245-277)	P302 (282-308)
Q8NH07	O11H2_HUMAN	Olfactory	N57 (37-66)	E85 (73-101)	Q137 (109-142)	W164 (153-177)	N225 (209-242)	P269 (247-279)	P302 (282-308)
Q8NH08	O10AC_HUMAN	Olfactory	N43 (23-52)	E70 (58-86)	R126 (98-131)	C153 (142-166)	P214 (198-231)	C258 (236-268)	P291 (271-297)
Q8NH09	OR8S1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	N210 (194-227)	S254 (232-264)	S285 (265-291)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NH10	OR8U1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NH16	OR2L2_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	P253 (231-263)	P286 (266-292)
Q8NH18	OR5J2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	T210 (194-227)	T254 (232-264)	L287 (267-293)
Q8NH19	O10AG_HUMAN	Olfactory	N33 (13-42)	E61 (49-77)	R113 (85-118)	W140 (129-153)	P201 (185-218)	A245 (223-255)	P278 (258-284)
Q8NH21	OR4F5_HUMAN	Olfactory	N35 (15-44)	D63 (51-79)	R115 (87-120)	W142 (131-155)	S203 (187-220)	P246 (224-256)	P277 (257-283)
Q8NH37	OR4C3_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	N208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NH40	OR6S1_HUMAN	Olfactory	N44 (24-53)	E72 (60-88)	R124 (96-129)	W151 (140-164)	S213 (197-230)	S257 (235-267)	P290 (270-296)
Q8NH41	OR4KF_HUMAN	Olfactory	N66 (46-75)	D94 (82-110)	R146 (118-151)	W173 (162-186)	S234 (218-251)	P277 (255-287)	P308 (288-314)
Q8NH42	OR4KD_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NH43	OR4L1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q8NH48	OR5B3_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	A208 (192-225)	T252 (230-262)	P285 (265-291)
Q8NH49	OR4X1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NH50	OR8K5_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	S210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NH51	OR8K3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	L122 (94-127)	Y149 (138-162)	S210 (194-227)	T253 (231-263)	P286 (266-292)
Q8NH53	O52N1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	H124 (96-129)	F151 (140-164)	D212 (196-229)	P256 (234-266)	P292 (272-298)
Q8NH54	O56A3_HUMAN	Olfactory	N46 (26-55)	D74 (62-90)	R126 (98-131)	L153 (142-166)	D214 (198-231)	I258 (236-268)	P293 (273-299)
Q8NH55	O52E5_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	I151 (140-164)	D209 (193-226)	P253 (231-263)	S288 (268-294)
Q8NH56	O52N5_HUMAN	Olfactory	N50 (30-59)	D79 (67-95)	R131 (103-136)	F158 (147-171)	D219 (203-236)	P263 (241-273)	P299 (279-305)
Q8NH57	O52P1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NH59	O51Q1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	C124 (96-129)	I151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q8NH60	O52J3_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	N211 (195-228)	P255 (233-265)	P290 (270-296)
Q8NH61	O51F2_HUMAN	Olfactory	N56 (36-65)	D84 (72-100)	R136 (108-141)	L163 (152-176)	D224 (208-241)	P268 (246-278)	P303 (283-309)
Q8NH63	O51H1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	L151 (140-164)	D212 (196-229)	P256 (234-266)	P292 (272-298)
Q8NH64	O51A7_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	A149 (138-162)	D209 (193-226)	P253 (231-263)	P288 (268-294)
Q8NH67	O52I2_HUMAN	Olfactory	N72 (52-81)	D100 (88-116)	R152 (124-157)	T179 (168-192)	D240 (224-257)	P284 (262-294)	P320 (300-326)
Q8NH69	OR5W2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NH70	O4A16_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	M147 (136-160)	I208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NH72	OR4C6_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	I208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NH73	OR4S2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	P251 (229-261)	P282 (262-288)
Q8NH74	O10A6_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	L287 (267-293)
Q8NH76	O56B4_HUMAN	Olfactory	N48 (28-57)	D76 (64-92)	R128 (100-133)	M155 (144-168)	D216 (200-233)	I260 (238-270)	P293 (273-299)
Q8NH79	OR6X1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	S208 (192-225)	A252 (230-262)	P285 (265-291)
Q8NH80	O10D3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P254 (232-264)	P285 (265-291)
Q8NH81	O10G6_HUMAN	Olfactory	N63 (43-72)	D91 (79-107)	R143 (115-148)	W170 (159-183)	C231 (215-248)	P275 (253-285)	S306 (286-312)
Q8NH83	OR4A5_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	M147 (136-160)	I207 (191-224)	P250 (228-260)	P281 (261-287)
Q8NH85	OR5R1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	Y149 (138-162)	S210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NH87	OR9G1_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	Y148 (137-161)	P209 (193-226)	S253 (231-263)	L286 (266-292)
Q8NH89	O5AK3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NH90	O5AK2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	P122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8NH92	OR1S1_HUMAN	Olfactory	N55 (35-64)	D83 (71-99)	H135 (107-140)	W162 (151-175)	P223 (207-240)	T267 (245-277)	P300 (280-306)
Q8NH93	OR1L3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	I149 (138-162)	P210 (194-227)	S254 (232-264)	P286 (266-292)
Q8NH94	OR1L1_HUMAN	Olfactory	N92 (72-101)	D120 (108-136)	R172 (144-177)	F199 (188-212)	P260 (244-277)	S304 (282-314)	P336 (316-342)
Q8NH95	YI035_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R124 (96-129)	W151 (140-164)	P212 (196-229)	T256 (234-266)	P295 (275-301)
Q8NHA4	O2AE1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P209 (193-226)	A253 (231-263)	S286 (266-292)
Q8NHA6	OR2W6_HUMAN	Olfactory	N48 (28-57)	D76 (64-92)	R128 (100-133)	W155 (144-168)	P216 (200-233)	T260 (238-270)	P293 (273-299)
Q8NHA8	OR1FC_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8NHB1	OR2V1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	A255 (233-265)	P288 (268-294)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q8NHB7	OR5K1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NHB8	OR5K2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)
Q8NHC4	O10J5_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	F149 (138-162)	P209 (193-226)	C253 (231-263)	P286 (266-292)
Q8NHC5	O14AG_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	T251 (229-261)	P284 (264-290)
Q8NHC6	O14L1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	S211 (195-228)	S255 (233-265)	P288 (268-294)
Q8NHC7	O14CZ_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	L147 (136-160)	C208 (192-225)	S252 (230-262)	P285 (265-291)
Q8NHC8	OR2T6_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q8TCB6	O51E1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D212 (196-229)	P255 (233-265)	P290 (270-296)
Q8TCW9	PKR1_HUMAN	A-other	N80 (60-89)	D108 (96-124)	R162 (134-167)	W187 (176-200)	P244 (228-261)	P299 (277-309)	T339 (319-345)
Q8TDS4	HCAR2_HUMAN	A-delta	N45 (25-54)	D73 (61-89)	R125 (97-130)	W152 (141-165)	P200 (184-217)	P246 (224-256)	P291 (271-297)
Q8TDS5	OXER1_HUMAN	A-other	N110 (90-119)	D138 (126-154)	R190 (162-195)	W217 (206-230)	P264 (248-281)	P310 (288-320)	P352 (332-358)
Q8TDS7	MRGRD_HUMAN	A-delta	N45 (25-54)	D72 (60-88)	R123 (95-128)	W150 (139-163)	T191 (175-208)	P236 (214-246)	P273 (253-279)
Q8TDT2	GP152_HUMAN	A-other	N47 (27-56)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P204 (188-221)	P252 (230-262)	P292 (272-298)
Q8TDU6	GPBAR_HUMAN	A-other	N32 (12-41)	G61 (49-77)	R110 (82-115)	W132 (121-145)	P176 (160-193)	P239 (217-249)	P277 (257-283)
Q8TDU9	RL3R2_HUMAN	A-other	N57 (37-66)	D87 (75-103)	R139 (111-144)	W166 (155-179)	P216 (200-233)	P261 (239-271)	P306 (286-312)
Q8TDV0	GP151_HUMAN	A-other	N54 (34-63)	D84 (72-100)	C136 (108-141)	W162 (151-175)	P214 (198-231)	P265 (243-275)	P303 (283-309)
Q8TDV2	GP148_HUMAN	A-other	S67 (47-76)	D96 (84-112)	T146 (118-151)	W173 (162-186)	C231 (215-248)	L282 (260-292)	T317 (297-323)
Q8TDV5	GP119_HUMAN	A-other	N22 (2-31)	D51 (39-67)	R103 (75-108)	W130 (119-143)	A177 (161-194)	P240 (218-250)	P276 (256-282)
Q8TE23	TS1R2_HUMAN	C	T580 (564-589)	L611 (601-622)	I655 (627-661)	T686 (675-699)	L738 (725-749)	S775 (758-782)	S804 (788-817)
Q8WXD0	RXFP2_HUMAN	A-delta	G432 (412-441)	A460 (448-476)	E519 (491-524)	I545 (534-558)	L600 (584-617)	I652 (630-662)	N687 (667-693)
Q8WXG9	GPR98_HUMAN	Adhesion	C5921 (5897-5929)	A5947 (5935-5961)	N5994 (5964-6000)	S6017 (6006-6034)	C6071 (6059-6090)	L6120 (6097-6122)	V6148 (6128-6155)
Q8WZ84	OR8D1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	F149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q8WZ92	OR5P2_HUMAN	Olfactory	N38 (18-47)	D66 (54-82)	R118 (90-123)	Y145 (134-158)	T206 (190-223)	T250 (228-260)	P283 (263-289)
Q8WZ94	OR5P3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	T254 (232-264)	P287 (267-293)
Q8WZA6	OR1E3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	C122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q92633	LPAR1_HUMAN	A-alpha	N67 (47-76)	D95 (83-111)	R146 (118-151)	W172 (161-185)	T217 (201-234)	P273 (251-283)	P308 (288-314)
Q92847	GHSR_HUMAN	A-beta	N61 (41-70)	D89 (77-105)	R141 (113-146)	W168 (157-181)	P224 (208-241)	P278 (256-288)	P320 (300-326)
Q969F8	KISSR_HUMAN	A-gamma	N60 (40-69)	D88 (76-104)	R140 (112-145)	W167 (156-180)	P216 (200-233)	P278 (256-288)	P320 (300-326)
Q969N4	TAAR8_HUMAN	A-alpha	N49 (29-58)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P208 (192-225)	P272 (250-282)	P307 (287-313)
Q969V1	MCHR2_HUMAN	A-other	N52 (32-61)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P209 (193-226)	P266 (244-276)	P303 (283-309)
Q96AM1	MRGRF_HUMAN	A-delta	N62 (42-71)	D89 (77-105)	R141 (113-146)	W168 (157-181)	C210 (194-227)	Y254 (232-264)	P288 (268-294)
Q96CH1	GP146_HUMAN	A-other	N45 (25-54)	G73 (61-89)	H127 (99-132)	W152 (141-165)	P199 (183-216)	P248 (226-258)	P293 (273-299)
Q96G91	P2Y11_HUMAN	A-delta	N46 (26-55)	D75 (63-91)	R127 (99-132)	W154 (143-167)	P217 (201-234)	P263 (241-273)	P318 (298-324)
Q96K78	GP128_HUMAN	Adhesion	L449 (425-457)	M478 (466-492)	L551 (521-557)	W575 (564-592)	I635 (623-654)	L686 (663-688)	F717 (697-724)
Q96KK4	O10C1_HUMAN	Olfactory	N41 (21-50)	E69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	T253 (231-263)	P286 (266-292)
Q96LA9	MRGX4_HUMAN	A-delta	N45 (25-54)	D72 (60-88)	R120 (92-125)	W147 (136-160)	C188 (172-205)	P231 (209-241)	P269 (249-275)
Q96LB0	MRGX3_HUMAN	A-delta	N45 (25-54)	D72 (60-88)	R120 (92-125)	W147 (136-160)	C188 (172-205)	P231 (209-241)	P269 (249-275)
Q96LB1	MRGX2_HUMAN	A-delta	N48 (28-57)	D75 (63-91)	R127 (99-132)	W154 (143-167)	F195 (179-212)	P238 (216-248)	P276 (256-282)
Q96LB2	MRGX1_HUMAN	A-delta	N45 (25-54)	D72 (60-88)	R120 (92-125)	W147 (136-160)	C188 (172-205)	P231 (209-241)	P269 (249-275)
Q96P65	QRFPR_HUMAN	A-delta	N63 (43-72)	D91 (79-107)	R143 (115-148)	W170 (159-183)	P226 (210-243)	P288 (266-298)	P329 (309-335)
Q96P66	GP101_HUMAN	A-delta	N49 (29-58)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P204 (188-221)	P414 (392-424)	P451 (431-457)
Q96P67	GPR82_HUMAN	A-other	N33 (13-42)	N61 (49-77)	R117 (89-122)	W160 (149-173)	F212 (196-229)	P264 (242-274)	P306 (286-312)
Q96P68	OXGR1_HUMAN	A-delta	N51 (31-60)	D79 (67-95)	R131 (103-136)	W158 (147-171)	P210 (194-227)	P256 (234-266)	L299 (279-305)
Q96P69	GPR78_HUMAN	A-alpha	N23 (3-32)	H52 (40-68)	Q104 (76-109)	W131 (120-144)	P184 (168-201)	P256 (234-266)	P291 (271-297)
Q96PE1	GP124_HUMAN	Adhesion	L784 (760-792)	I815 (803-829)	L860 (830-866)	G892 (881-909)	I935 (923-954)	C1035 (1012-1037)	F1066 (1046-1073)
Q96R08	OR5BC_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	S208 (192-225)	T252 (230-262)	P285 (265-291)
Q96R09	OR5B2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	Y147 (136-160)	V208 (192-225)	T252 (230-262)	P285 (265-291)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q96R27	OR2M4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q96R28	OR2M2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	A254 (232-264)	P287 (267-293)
Q96R30	OR2V2_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	A255 (233-265)	P288 (268-294)
Q96R45	OR2A7_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	L121 (93-126)	W148 (137-161)	P209 (193-226)	T253 (231-263)	P286 (266-292)
Q96R47	O2A14_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P209 (193-226)	S253 (231-263)	P286 (266-292)
Q96R48	OR2A5_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R122 (94-127)	W149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q96R54	O14A2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	T252 (230-262)	P285 (265-291)
Q96R67	OR4CC_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	C120 (92-125)	W147 (136-160)	N208 (192-225)	P251 (229-261)	P282 (262-288)
Q96R69	OR4F4_HUMAN	Olfactory	N35 (15-44)	D63 (51-79)	R115 (87-120)	W142 (131-155)	S203 (187-220)	P246 (224-256)	P277 (257-283)
Q96R72	OR4K3_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	C210 (194-227)	P253 (231-263)	P284 (264-290)
Q96R84	OR1F2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R121 (93-126)	W148 (137-161)	P209 (193-226)	T253 (231-263)	P286 (266-292)
Q96RA2	OR7D2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q96RB7	OR5MB_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R121 (93-126)	Y148 (137-161)	S209 (193-226)	T253 (231-263)	P286 (266-292)
Q96RC9	OR8B4_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R121 (93-126)	Y148 (137-161)	S209 (193-226)	S253 (231-263)	P286 (266-292)
Q96RD0	OR8B2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	P210 (194-227)	S254 (232-264)	P286 (266-292)
Q96RD1	OR6C1_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
Q96RD2	O52B2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	I151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q96RD3	O52E6_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D211 (195-228)	P255 (233-265)	P290 (270-296)
Q96RI0	PAR4_HUMAN	A-delta	N95 (75-104)	D122 (110-138)	R174 (146-179)	W201 (190-214)	P256 (240-273)	P298 (276-308)	P337 (317-343)
Q96RI8	TAAR6_HUMAN	A-alpha	N50 (30-59)	D78 (66-94)	R130 (102-135)	W157 (146-170)	P209 (193-226)	P273 (251-283)	P308 (288-314)
Q96RI9	TAAR9_HUMAN	A-alpha	N50 (30-59)	D78 (66-94)	R130 (102-135)	W157 (146-170)	P209 (193-226)	P273 (251-283)	P308 (288-314)
Q96RJ0	TAAR1_HUMAN	A-alpha	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P202 (186-219)	P266 (244-276)	P301 (281-307)
Q99500	S1PR3_HUMAN	A-alpha	N57 (37-66)	D85 (73-101)	R136 (108-141)	W162 (151-175)	I207 (191-224)	P258 (236-268)	P295 (275-301)
Q99527	GPER1_HUMAN	A-other	N77 (57-86)	D105 (93-121)	R155 (127-160)	W182 (171-195)	P226 (210-243)	P274 (252-284)	P321 (301-327)
Q99677	LPAR4_HUMAN	A-delta	N57 (37-66)	D85 (73-101)	R136 (108-141)	W163 (152-176)	P216 (200-233)	P264 (242-274)	P308 (288-314)
Q99678	GPR20_HUMAN	A-other	N71 (51-80)	D99 (87-115)	R148 (120-153)	W175 (164-188)	P208 (192-225)	P256 (234-266)	P294 (274-300)
Q99679	GPR21_HUMAN	A-alpha	N46 (26-55)	D75 (63-91)	R127 (99-132)	W154 (143-167)	A203 (187-220)	P267 (245-277)	C301 (281-307)
Q99680	GPR22_HUMAN	A-other	N59 (39-68)	D88 (76-104)	R140 (112-145)	W165 (154-178)	T223 (207-240)	P328 (306-338)	P364 (344-370)
Q99705	MCHR1_HUMAN	A-other	N127 (107-136)	D158 (146-174)	R210 (182-215)	W237 (226-250)	P289 (273-306)	P340 (318-350)	P377 (357-383)
Q99788	CML1_HUMAN	A-gamma	N58 (38-67)	D85 (73-101)	R137 (109-142)	W164 (153-177)	P232 (216-249)	P275 (253-285)	P313 (293-319)
Q99835	SMO_HUMAN	F	T245 (224-254)	F274 (263-285)	W339 (312-344)	W365 (357-378)	V411 (397-432)	I465 (446-475)	S533 (515-536)
Q9BPV8	P2Y13_HUMAN	A-delta	N62 (42-71)	D89 (77-105)	R141 (113-146)	W168 (157-181)	V220 (204-237)	P270 (248-280)	P314 (294-320)
Q9BXA5	SUCR1_HUMAN	A-delta	N41 (21-50)	D69 (57-85)	R120 (92-125)	W147 (136-160)	P199 (183-216)	P247 (225-257)	P292 (272-298)
Q9BXB1	LGR4_HUMAN	A-delta	N556 (536-565)	N584 (572-600)	R643 (615-648)	F670 (659-683)	A717 (701-734)	P762 (740-772)	P798 (778-804)
Q9BXC0	HCAR1_HUMAN	A-delta	N33 (13-42)	D61 (49-77)	R113 (85-118)	W140 (129-153)	P188 (172-205)	P235 (213-245)	P275 (255-281)
Q9BXC1	GP174_HUMAN	A-delta	N37 (17-46)	D65 (53-81)	R116 (88-121)	W142 (131-155)	P197 (181-214)	P245 (223-255)	P289 (269-295)
Q9BXE9	VN1R3_HUMAN	Vomeronasal	N23 (3-32)	N54 (42-70)	Q106 (78-111)	W137 (126-150)	C193 (177-210)	T248 (226-258)	P282 (262-288)
Q9BY15	EMR3_HUMAN	Adhesion	L371 (347-379)	L399 (387-413)	L444 (414-450)	Y472 (461-489)	I515 (503-534)	L567 (544-569)	F595 (575-602)
Q9BY21	GPR87_HUMAN	A-delta	N60 (40-69)	D87 (75-103)	R139 (111-144)	W166 (155-179)	V218 (202-235)	P267 (245-277)	P311 (291-317)
Q9BZJ6	GPR63_HUMAN	A-other	N99 (79-108)	D127 (115-143)	R179 (151-184)	W203 (192-216)	P254 (238-271)	P329 (307-339)	P368 (348-374)
Q9BZJ7	GPR62_HUMAN	A-alpha	N32 (12-41)	D58 (46-74)	R113 (85-118)	W137 (126-150)	P184 (168-201)	P256 (234-266)	P287 (267-293)
Q9BZJ8	GPR61_HUMAN	A-alpha	N60 (40-69)	D86 (74-102)	R140 (112-145)	W167 (156-180)	P220 (204-237)	P302 (280-312)	P338 (318-344)
Q9GZK3	OR2B2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q9GZK4	OR2H1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	P208 (192-225)	S252 (230-262)	P285 (265-291)
Q9GZK6	OR2J1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	P255 (233-265)	P288 (268-294)
Q9GZK7	O11A1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P288 (268-294)
Q9GZM6	OR8D2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	Y149 (138-162)	T210 (194-227)	S254 (232-264)	P287 (267-293)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q9GZN0	GPR88_HUMAN	A-other	N49 (29-58)	D77 (65-93)	R138 (110-143)	W166 (155-179)	Q204 (188-221)	P299 (277-309)	P333 (313-339)
Q9GZP7	VN1R1_HUMAN	Vomeronasal	N63 (43-72)	N94 (82-110)	Q146 (118-151)	W176 (165-189)	S231 (215-248)	V288 (266-298)	P322 (302-328)
Q9GZQ4	NMUR2_HUMAN	A-beta	N63 (43-72)	D91 (79-107)	R144 (116-149)	W171 (160-184)	P228 (212-245)	P283 (261-293)	P324 (304-330)
Q9GZQ6	NPFF1_HUMAN	A-beta	N61 (41-70)	D89 (77-105)	R141 (113-146)	W166 (155-179)	P228 (212-245)	P285 (263-295)	P326 (306-332)
Q9H1C0	LPAR5_HUMAN	A-delta	N40 (20-49)	D68 (56-84)	R119 (91-124)	W146 (135-159)	P203 (187-220)	P249 (227-259)	P294 (274-300)
Q9H1Y3	OPN3_HUMAN	A-alpha	N59 (39-68)	D87 (75-103)	R139 (111-144)	W162 (151-175)	P214 (198-231)	P270 (248-280)	P306 (286-312)
Q9H205	O2AG1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	S210 (194-227)	A254 (232-264)	P287 (267-293)
Q9H207	O10A5_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	S255 (233-265)	P288 (268-294)
Q9H208	O10A2_HUMAN	Olfactory	N29 (9-38)	E57 (45-73)	R109 (81-114)	W136 (125-149)	P197 (181-214)	S241 (219-251)	P274 (254-280)
Q9H209	O10A4_HUMAN	Olfactory	N43 (23-52)	E71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P288 (268-294)
Q9H210	OR2D2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	S254 (232-264)	P285 (265-291)
Q9H228	S1PR5_HUMAN	A-alpha	N54 (34-63)	D82 (70-98)	R133 (105-138)	W159 (148-172)	I204 (188-221)	P266 (244-276)	P303 (283-309)
Q9H244	P2Y12_HUMAN	A-delta	N43 (23-52)	D70 (58-86)	R122 (94-127)	W149 (138-162)	N201 (185-218)	P251 (229-261)	P295 (275-301)
Q9H255	O51E2_HUMAN	Olfactory	N41 (21-50)	D69 (57-85)	R121 (93-126)	V148 (137-161)	D209 (193-226)	P253 (231-263)	P288 (268-294)
Q9H2C5	O52A5_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	T151 (140-164)	D213 (197-230)	L257 (235-267)	P292 (272-298)
Q9H2C8	O51V1_HUMAN	Olfactory	N51 (31-60)	D79 (67-95)	R131 (103-136)	I158 (147-171)	D219 (203-236)	P263 (241-273)	P298 (278-304)
Q9H339	O51B5_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	L147 (136-160)	D208 (192-225)	T252 (230-262)	P287 (267-293)
Q9H340	O51B6_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	C120 (92-125)	L147 (136-160)	D208 (192-225)	T252 (230-262)	P287 (267-293)
Q9H341	O51M1_HUMAN	Olfactory	N54 (34-63)	D82 (70-98)	R134 (106-139)	I161 (150-174)	D222 (206-239)	P266 (244-276)	P301 (281-307)
Q9H342	O51J1_HUMAN	Olfactory	N45 (25-54)	E73 (61-89)	S128 (100-133)	S165 (154-178)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q9H343	O51I1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	L151 (140-164)	D212 (196-229)	P256 (234-266)	P291 (271-297)
Q9H344	O51I2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	A149 (138-162)	D210 (194-227)	P254 (232-264)	P289 (269-295)
Q9H346	O52D1_HUMAN	Olfactory	N45 (25-54)	D73 (61-89)	R125 (97-130)	L152 (141-165)	D213 (197-230)	P257 (235-267)	P293 (273-299)
Q9H3N8	HRH4_HUMAN	A-alpha	N33 (13-42)	D61 (49-77)	R112 (84-117)	W140 (129-153)	P186 (170-203)	P318 (296-328)	P355 (335-361)
Q9H461	FZD8_HUMAN	F	T292 (271-301)	Y320 (309-331)	W421 (394-426)	W446 (438-459)	Y492 (478-513)	V546 (527-556)	G601 (583-604)
Q9HAR2	LPHN3_HUMAN	Adhesion	L879 (855-887)	L907 (895-921)	L952 (922-958)	Y975 (964-992)	I1018 (1006-1037)	F1070 (1047-1072)	F1098 (1078-1105)
Q9HB89	NMUR1_HUMAN	A-beta	N78 (58-87)	D106 (94-122)	R159 (131-164)	W186 (175-199)	P243 (227-260)	P312 (290-322)	P353 (333-359)
Q9HBW0	LPAR2_HUMAN	A-alpha	N50 (30-59)	D78 (66-94)	R129 (101-134)	W155 (144-168)	V200 (184-217)	P256 (234-266)	A292 (272-298)
Q9HBW9	ELTD1_HUMAN	Adhesion	L442 (418-450)	L470 (458-484)	L515 (485-521)	Y538 (527-555)	I581 (569-600)	F633 (610-635)	F661 (641-668)
Q9HBX8	LGR6_HUMAN	A-delta	N579 (559-588)	N608 (596-624)	C667 (639-672)	L694 (683-707)	C743 (727-760)	P788 (766-798)	P824 (804-830)
Q9HBX9	RXFP1_HUMAN	A-delta	N423 (403-432)	D451 (439-467)	K510 (482-515)	W536 (525-549)	A591 (575-608)	P643 (621-653)	P678 (658-684)
Q9HC97	GPR35_HUMAN	A-delta	N38 (18-47)	D66 (54-82)	R114 (86-119)	W141 (130-154)	P183 (167-200)	P232 (210-242)	A273 (253-279)
Q9HCU4	CELR2_HUMAN	Adhesion	L2391 (2367-2399)	L2419 (2407-2433)	L2464 (2434-2470)	W2487 (2476-2504)	A2530 (2518-2549)	L2578 (2555-2580)	F2606 (2586-2613)
Q9NPB9	ACKR4_HUMAN	A-gamma	N59 (39-68)	D87 (75-103)	R137 (109-142)	W162 (151-175)	P212 (196-229)	P255 (233-265)	P300 (280-306)
Q9NPC1	LT4R2_HUMAN	A-gamma	N69 (49-78)	D99 (87-115)	R150 (122-155)	W177 (166-190)	P224 (208-241)	P270 (248-280)	P316 (296-322)
Q9NPG1	FZD3_HUMAN	F	T217 (196-226)	Y245 (234-256)	W311 (284-316)	W336 (328-349)	Y382 (368-403)	V434 (415-444)	V495 (477-498)
Q9NQ84	GPC5C_HUMAN	C	T64 (48-73)	T97 (87-108)	L141 (113-147)	T164 (153-177)	L219 (206-230)	V258 (241-265)	V292 (276-305)
Q9NQN1	OR2S1_HUMAN	Olfactory	N43 (23-52)	D71 (59-87)	R123 (95-128)	W150 (139-163)	P211 (195-228)	T255 (233-265)	P294 (274-300)
Q9NQS5	GPR84_HUMAN	A-other	N38 (18-47)	D66 (54-82)	R118 (90-123)	W145 (134-158)	G190 (174-207)	P334 (312-344)	P367 (347-373)
Q9NS66	GP173_HUMAN	A-other	N40 (20-49)	D68 (56-84)	R121 (93-126)	W148 (137-161)	M192 (176-209)	P301 (279-311)	P337 (317-343)
Q9NS67	GPR27_HUMAN	A-other	N36 (16-45)	D64 (52-80)	R120 (92-125)	W148 (137-161)	L189 (173-206)	P299 (277-309)	P335 (315-341)
Q9NS75	CLTR2_HUMAN	A-delta	N56 (36-65)	D84 (72-100)	R136 (108-141)	W163 (152-176)	P213 (197-230)	P262 (240-272)	P302 (282-308)
Q9NSD7	RL3R1_HUMAN	A-gamma	N99 (79-108)	D128 (116-144)	R180 (152-185)	W222 (211-235)	P279 (263-296)	P341 (319-351)	P386 (366-392)
Q9NYM4	GPR83_HUMAN	A-beta	N89 (69-98)	D117 (105-133)	R169 (141-174)	W194 (183-207)	P251 (235-268)	P307 (285-317)	P342 (322-348)
Q9NYQ6	CELR1_HUMAN	Adhesion	L2483 (2459-2491)	L2511 (2499-2525)	V2556 (2526-2562)	W2579 (2568-2596)	V2622 (2610-2641)	L2670 (2647-2672)	L2698 (2678-2705)
Q9NYQ7	CELR3_HUMAN	Adhesion	L2554 (2530-2562)	L2582 (2570-2596)	L2627 (2597-2633)	W2650 (2639-2667)	V2693 (2681-2712)	F2740 (2717-2742)	L2768 (2748-2775)
Q9NYV7	T2R16_HUMAN	Taste2	S25 (5-34)	R56 (44-72)	Y103 (75-108)	L130 (119-143)	P188 (172-205)	T242 (220-252)	S273 (253-279)

Uniprot	Name	Class	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Q9NYV8	T2R14_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P190 (174-207)	S246 (224-256)	S277 (257-283)
Q9NYV9	T2R13_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P190 (174-207)	C247 (225-257)	S278 (258-284)
Q9NYW0	T2R10_HUMAN	Taste2	N24 (4-33)	R54 (42-70)	Y106 (78-111)	I133 (122-146)	F185 (169-202)	G242 (220-252)	S274 (254-280)
Q9NYW1	TA2R9_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P190 (174-207)	V247 (225-257)	S279 (259-285)
Q9NYW2	TA2R8_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P193 (177-210)	S250 (228-260)	S282 (262-288)
Q9NYW3	TA2R7_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P193 (177-210)	S250 (228-260)	S282 (262-288)
Q9NYW4	TA2R5_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y103 (75-108)	L130 (119-143)	P180 (164-197)	A237 (215-247)	S269 (249-275)
Q9NYW5	TA2R4_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y106 (78-111)	L133 (122-146)	Q188 (172-205)	A245 (223-255)	S277 (257-283)
Q9NYW6	TA2R3_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y107 (79-112)	L134 (123-147)	P192 (176-209)	A249 (227-259)	S281 (261-287)
Q9NYW7	TA2R1_HUMAN	Taste2	N24 (4-33)	R55 (43-71)	Y103 (75-108)	I130 (119-143)	P186 (170-203)	I243 (221-253)	S274 (254-280)
Q9NZD1	GPC5D_HUMAN	C	T37 (21-46)	V70 (60-81)	L114 (86-120)	S137 (126-150)	L180 (167-191)	V219 (202-226)	L254 (238-267)
Q9NZH0	GPC5B_HUMAN	C	T70 (54-79)	T103 (93-114)	V147 (119-153)	M170 (159-183)	L210 (197-221)	A249 (232-256)	V284 (268-297)
Q9NZP0	OR6C3_HUMAN	Olfactory	N39 (19-48)	E67 (55-83)	R119 (91-124)	W146 (135-159)	T207 (191-224)	S251 (229-261)	P284 (264-290)
Q9NZP2	OR6C2_HUMAN	Olfactory	N40 (20-49)	E68 (56-84)	R120 (92-125)	W147 (136-160)	T208 (192-225)	S252 (230-262)	P285 (265-291)
Q9NZP5	O5AC2_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	Y151 (140-164)	T212 (196-229)	T256 (234-266)	P289 (269-295)
Q9P1P4	TAAR3_HUMAN	A-alpha	N49 (29-58)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P209 (193-226)	P271 (249-281)	P306 (286-312)
Q9P1P5	TAAR2_HUMAN	A-alpha	N61 (41-70)	D89 (77-105)	R141 (113-146)	W168 (157-181)	P221 (205-238)	P277 (255-287)	P312 (292-318)
Q9P1Q5	OR1A1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P209 (193-226)	T253 (231-263)	P285 (265-291)
Q9P296	C5AR2_HUMAN	A-gamma	N53 (33-62)	D80 (68-96)	L132 (104-137)	W159 (148-172)	P212 (196-227)	P248 (228-258)	P288 (268-294)
Q9UBS5	GABR1_HUMAN	C	L605 (589-614)	G636 (626-647)	W687 (659-693)	L717 (706-730)	L779 (766-790)	L817 (800-824)	T848 (832-861)
Q9UBY5	LPAR3_HUMAN	A-alpha	N48 (28-57)	D76 (64-92)	R127 (99-132)	W153 (142-166)	A198 (182-215)	P254 (232-264)	P290 (270-296)
Q9UGF5	O14J1_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	C208 (192-225)	A252 (230-262)	P285 (265-291)
Q9UGF6	OR5V1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	S254 (232-264)	P287 (267-293)
Q9UGF7	O12D3_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	R120 (92-125)	W147 (136-160)	A208 (192-225)	P253 (231-263)	P286 (266-292)
Q9UHM6	OPN4_HUMAN	A-alpha	N88 (68-97)	D116 (104-132)	R168 (140-173)	W195 (184-208)	P247 (231-264)	P311 (289-321)	P347 (327-353)
Q9UHX3	EMR2_HUMAN	Adhesion	L551 (527-559)	L579 (567-593)	L624 (594-630)	Y652 (641-669)	I695 (683-714)	L747 (724-749)	F775 (755-782)
Q9UJ42	GP160_HUMAN	A-other	N39 (19-48)	D66 (54-82)	Y120 (92-125)	W147 (136-160)	S186 (170-203)	P254 (232-264)	A289 (269-295)
Q9UKL2	O52A1_HUMAN	Olfactory	N44 (24-53)	D72 (60-88)	R124 (96-129)	V151 (140-164)	D213 (197-230)	L257 (235-267)	P292 (272-298)
Q9UKP6	UR2R_HUMAN	A-other	N69 (49-78)	D97 (85-113)	R148 (120-153)	W174 (163-187)	P223 (207-240)	P273 (251-283)	P312 (292-318)
Q9ULV1	FZD4_HUMAN	F	T234 (213-243)	Y262 (251-273)	W327 (300-332)	W352 (344-365)	Y398 (384-419)	C450 (431-460)	G492 (474-495)
Q9ULW2	FZD10_HUMAN	F	S242 (221-251)	Y270 (259-281)	W334 (307-339)	W359 (351-372)	Y405 (391-426)	C457 (438-467)	G519 (501-522)
Q9UNW8	GP132_HUMAN	A-delta	N60 (40-69)	E88 (76-104)	R140 (112-145)	F167 (156-180)	P211 (195-228)	P257 (235-267)	P305 (285-311)
Q9UP38	FZD1_HUMAN	F	S333 (312-342)	C361 (350-372)	T426 (399-431)	A451 (443-464)	V497 (483-518)	T549 (530-559)	S617 (599-620)
Q9UPC5	GPR34_HUMAN	A-other	N72 (52-81)	D100 (88-116)	R152 (124-157)	W179 (168-192)	I230 (214-247)	P281 (259-291)	P324 (304-330)
Q9Y271	CLTR1_HUMAN	A-delta	N41 (21-50)	D69 (57-85)	R121 (93-126)	W148 (137-161)	P201 (185-218)	P248 (226-258)	P292 (272-298)
Q9Y2T5	GPR52_HUMAN	A-alpha	N58 (38-67)	D87 (75-103)	R139 (111-144)	W166 (155-179)	A215 (199-232)	P280 (258-290)	C314 (294-320)
Q9Y2T6	GPR55_HUMAN	A-delta	N38 (18-47)	D70 (58-86)	R119 (91-124)	W146 (135-159)	P193 (177-210)	P241 (219-251)	V285 (265-291)
Q9Y3N9	OR2W1_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P210 (194-227)	T254 (232-264)	P287 (267-293)
Q9Y4A9	O10H1_HUMAN	Olfactory	N42 (22-51)	E70 (58-86)	R122 (94-127)	W149 (138-162)	C211 (195-228)	F255 (233-265)	P288 (268-294)
Q9Y585	OR1A2_HUMAN	Olfactory	N42 (22-51)	D70 (58-86)	R122 (94-127)	W149 (138-162)	P209 (193-226)	T253 (231-263)	P285 (265-291)
Q9Y5N1	HRH3_HUMAN	A-alpha	N52 (32-61)	D80 (68-96)	R132 (104-137)	W160 (149-173)	P210 (194-227)	P373 (351-383)	P409 (389-415)
Q9Y5P0	O51B4_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	C119 (91-124)	L146 (135-159)	D207 (191-224)	T251 (229-261)	P286 (266-292)
Q9Y5P1	O51B2_HUMAN	Olfactory	N40 (20-49)	D68 (56-84)	C120 (92-125)	F147 (136-160)	D208 (192-225)	T252 (230-262)	P287 (267-293)
Q9Y5X5	NPFF2_HUMAN	A-beta	N165 (145-174)	D193 (181-209)	R245 (217-250)	W270 (259-283)	P333 (317-350)	P391 (369-401)	P432 (412-438)
Q9Y5Y3	GPR45_HUMAN	A-other	N52 (32-61)	D80 (68-96)	R132 (104-137)	W156 (145-169)	P207 (191-224)	P282 (260-292)	P321 (301-327)
Q9Y5Y4	PD2R2_HUMAN	A-gamma	N50 (30-59)	D77 (65-93)	R129 (101-134)	W156 (145-169)	P218 (202-235)	P261 (239-271)	P301 (281-307)
Q9Y653	GPR56_HUMAN	Adhesion	C418 (394-426)	V453 (441-467)	L500 (470-506)	W524 (513-541)	V581 (569-600)	L625 (602-627)	F655 (635-662)

Appendix E

Invoking QEQ in Lammps

E.1 Installing PQEQ Package

We implemented the QEQ and PQEQ methods (described in Chapter 6) in package USER_PQEQ for the general purpose molecular dynamics code LAMMPS [111]. The current PQEQ distribution is built on LAMMPS version 4Sep13. The package is not yet fully integrated into Lammps and needs to be installed last, because it modifies the following Lammps files:

```
atom_vec_hybrid.cpp, .h
atom.cpp, h.
min.cpp, .h
pair_reax_c.cpp
reaxc_nonbonded.cpp
reaxc_types.h
respa.cpp, .h
verlet.cpp, .h
```

The following commands compile LAMMPS with PQEQ:

```
tar -xzf lammps-4Sep13.tar.gz
tar -xzf pqeq-$version.tar.gz
cp USER-PQEQ/wag/Makefile.ion lammps-4Sep13/src/MAKE/
cp -r USER-PQEQ lammps-4Sep13/src
cd lammps-4Sep13/src
# for serial installation, install fake mpi:
#cd STUBS
#make
#cd ..
make yes-rigid
```

```
make yes-user-eff
make yes-user-reaxc
make yes-user-pqeq
make ion
```

E.2 Implementing Time Integration in LAMMPS

PQEQ implementation consists of two parts: computing the virtual forces on the charges (in the file pair_coul_pqeqgauss.cpp), and numerically integrating the equation of motion (fix_pqeq.cpp). The time integration is very similar to the simple NVE time integration implemented in fix_nve.cpp. The appropriate hooks get invoked at the following points in the timestep:

- INITIAL_INTEGRATE
- FINAL_INTEGRATE
- INITIAL_INTEGRATE_RESPA
- FINAL_INTEGRATE_RESPA

We need to add damping and charge conservation. Damping adds hooks at (similar to fix_viscous.cpp but MIN_POST_FORCE is not used):

- POST_FORCE
- POST_FORCE_RESPA

Charge conservation adds hooks at (like fix_shake.cpp but PRE_NEIGHBOR is not needed):

- POST_FORCE
- POST_FORCE_RESPA

Note: Instead of relying on setting the total force on the charges to zero, one might consider explicitly enforcing the total charge to stay exactly 0 (or an integer), which would be done with a hook at INITIAL_INTEGRATE (as in fix_recenter.cpp).

E.3 Invoking PQEQ in LAMMPS

PQEQ can be invoked by adding these commands to the Lammps input file:

```
units           real
atom_style      pqeq
pair_style      coul/pqeqgauss  R_cut_inner  R_cut_outer
fix             pqeq all pqeq damping_time damping_ratio
```

A sample input file reads:

```
boundary        f f f
units           real
neighbor        2.0 nsq
atom_style      pqeq
pair_style      coul/pqeqgauss 10 12
read_data       twoatom.data
fix             1 all pqeq 20.0 1.0
thermo          1
dump            1 all custom 1 dump.out id x y z q
timestep        1
run             100
```

Note that minimization is not implemented for PQEQ.

E.3.1 Data File

The PQEQ parameters are read from the data file. The `Atom` lines have data in the following order:

```
id  type  charge  x y z  Rsx Rsy Rsz
```

If `hybrid` atom style is used, the order or data specific to PQEQ is:

```
charge  Rsx Rsy Rsz
```

The `Pair Coeffs` lines have data in the following order:

```
id  chi  idem  Rcore  polar  Qcore  Rshell  K2  K4
```

The current implementation allows both polarizable and nonpolarizable atoms in the same run. If `polar==0`, then only `id chi idem Rcore` are used; the atom property q is the total charge on the atom, and R_{core} is used as the width of the Gaussian charge. If `polar==1`, then all parameters are used and the total charge on the atom is $q + Q_{\text{core}}$.

Note that the parameters have to be converted to the units consistent with `units real`. When `units real` is used in Lammps the unit of energy is kcal/mol and so the parameters χ =`chi`, J =`idem`, K_2, and K_4 have to be in kcal/mol. The constant in the Coulomb force:

$$c_{\text{ES}} = \frac{1}{4\pi\epsilon_0}$$

is $c_{\text{ES}} = 332.06371$, when the unit of energy is kcal/mol, unit of distance is A, and unit of charge is $|e|$.

The order of parameters on the `Atoms` line depends on the selected atom style:

```
pqeq:         id type q coord(3x) rsxs(3x)
pqeq-hybrid:  q rsx(3x)
full:         id molecule type q coord(3x)
full-hybrid:  molecule q
hybrid:       id type coord(3x) substyles
```

E.g. for:

```
 atom_style      hybrid full pqeq
```

The following order of inputs should be entered in the data file:

```
id type coord(3x) molecule q q rsx(3x)
```

E.3.2 Caveat for Reading PQEQ Parameters

Values for off-diagonal coefficients are ignored since the QEQ formalism is diagonal in the atom types. However, when hybrid pair style is used, the coefficients must be explicitly specified in the input file, so that the interaction is included.

Warning: there is only one flag for a hybrid atom pair style to mark whether the parameters are diagonal or not, so the usual LAMMPS checks do not work, and one has to manually set all coefficients for all pair styles used.

AMBER and CHARMM forcefields use arithmetic mixing rule for the Van der Waals parameters, so with pair styles `lj/charmm/coul/charmm` and `lj/cut typically pair_modify` one has to use `mix arithmetic`. If one used hybrid atom style (which does not allow mixing), one has to specify the off-diagonal parameters manually. The off-diagonal parameters can be computed as:

$$\epsilon = \sqrt{\epsilon_1 \epsilon_2} \tag{E.1}$$

$$\sigma = 0.5 * (\sigma_1 + \sigma_2) \tag{E.2}$$

E.4 Usage with ReaxC Implementation

Coulomb energy included in the ReaxC code has to be turned off, and Coulomb energy from PQEQ is used instead. This is done with a keyword `coulomb_off yes` and has to be accompanied with `checkqeq no`, since the QEQ from ReaxC is not used. For example:

```
pair_style reax/c control_file coulomb_off yes checkqeq no
```

The current implementation of the `coulomb_off` option might not be the most efficient. We simply commented out the computation of Coulomb interactions in the ReaxC code.

E.5 Usage with Force Fields in Lammps

Force fields do not include Coulomb force for bonded atoms (1-2 and 1-3 interactions) and 1-4 interactions are often accounted for partially. In LAMMPS this behavior is controlled by the variable `special_bonds`. We want to keep this behavior for computing physical forces on atoms, but we need to include our modified Coulomb interaction into the computation for the auxiliary forces F_Q and F_{RS}. For QEQ there is no issue. For PQEQ, it should be checked that this asymmetrical handling of forces (action $\neq$ reaction) is fine. The quantity counted into the total energy is the one corresponding to the real forces.

Bibliography

[1] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "Atomic-level description of ubiquitin folding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, pp. 5915–5920, Apr 2013.

[2] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, and A. Tramontano, "Critical assessment of methods of protein structure prediction (CASP) - round X," *Proteins: Structure, Function, and Bioinformatics*, vol. 82, pp. 1–6, 2014.

[3] V. Yarov-Yarovoy, J. Schonbrun, and D. Baker, "Multipass membrane protein structure prediction using Rosetta," *Proteins*, vol. 62, pp. 1010–1025, Mar 2006.

[4] I. Kufareva, V. Katritch, R. Stevens, and R. Abagyan, "Advances in {GPCR} Modeling Evaluated by the {GPCR} Dock 2013 Assessment: Meeting New Challenges," *Structure*, vol. 22, no. 8, pp. 1120 – 1139, 2014.

[5] D. Latek, P. Pasznik, T. Carlomagno, and S. Filipek, "Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison," *PLoS ONE*, vol. 8, p. e56742, 02 2013.

[6] J. K. Bray, R. Abrol, W. A. Goddard, B. Trzaskowski, and C. E. Scott, "SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, pp. E72–78, Jan 2014.

[7] S. Naserifar, D. J. Brooks, V. Cvicek, Q. Zhang, and W. A. G. III, "Quantum Optimized Charge Equilibration (PQEq) Parameters for Predicting Accurate Charges in Molecules and Solids." California Institute of Technology, 2014.

[8] J. L. Sharman, H. E. Benson, A. J. Pawson, V. Lukito, C. P. Mpamhanga, V. Bombail, A. P. Davenport, J. A. Peters, M. Spedding, A. J. Harmar, and NC-IUPHAR, "IUPHAR-DB: updated database content and new features," *Nucleic Acids Research*, vol. 41, no. D1, pp. D1083–D1088, 2013.

[9] I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, and M. L. Tress, "Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes," *Human Molecular Genetics*, 2014.

[10] R. C. Stevens, V. Cherezov, V. Katritch, R. Abagyan, P. Kuhn, H. Rosen, and K. Wuthrich, "The GPCR Network: a large-scale collaboration to determine human GPCR structure and function," *Nat Rev Drug Discov*, vol. 12, pp. 25–34, Jan 2013.

[11] S. L. Garland, "Are GPCRs Still a Source of New Targets?," *Journal of Biomolecular Screening*, 2013.

[12] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, and B. K. Kobilka, "Crystal structure of the β2 adrenergic receptor-Gs protein complex," *Nature*, vol. 477, pp. 549–555, Sep 2011.

[13] P. Ghanouni, Z. Gryczynski, J. J. Steenhuis, T. W. Lee, D. L. Farrens, J. R. Lakowicz, and B. K. Kobilka, "Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor," *J. Biol. Chem.*, vol. 276, pp. 24433–24436, Jul 2001.

[14] G. Swaminath, Y. Xiang, T. W. Lee, J. Steenhuis, C. Parnot, and B. K. Kobilka, "Sequential binding of agonists to the beta2 adrenoceptor. Kinetic evidence for intermediate conformational states," *J. Biol. Chem.*, vol. 279, pp. 686–691, Jan 2004.

[15] D. M. Rosenbaum, C. Zhang, J. A. Lyons, R. Holl, D. Aragao, D. H. Arlow, S. G. Rasmussen, H. J. Choi, B. T. Devree, R. K. Sunahara, P. S. Chae, S. H. Gellman, R. O. Dror, D. E. Shaw, W. I. Weis, M. Caffrey, P. Gmeiner, and B. K. Kobilka, "Structure and function of an irreversible agonist-β_2 adrenoceptor complex," *Nature*, vol. 469, pp. 236–240, Jan 2011.

[16] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "OPM: orientations of proteins in membranes database," *Bioinformatics*, vol. 22, pp. 623–625, Mar 2006.

[17] P. Barth, B. Wallner, and D. Baker, "Prediction of membrane protein structures with complex topologies using limited constraints," *Proceedings of the National Academy of Sciences*, vol. 106, no. 5, pp. 1409–1414, 2009.

[18] R. Fredriksson, M. C. Lagerstrom, L. G. Lundin, and H. B. Schioth, "The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints," *Mol. Pharmacol.*, vol. 63, pp. 1256–1272, Jun 2003.

[19] T. K. Attwood and J. B. C. Findlay, "Fingerprinting G-protein-coupled receptors," *Protein Engineering*, vol. 7, no. 2, pp. 195–203, 1994.

[20] L. F. Kolakowski, "GCRDb: a G-protein-coupled receptor database," *Recept. Channels*, vol. 2, no. 1, pp. 1–7, 1994.

[21] P. Joost and A. Methner, "Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands," *Genome Biol.*, vol. 3, p. RESEARCH0063, Oct 2002.

[22] J. A. Ballesteros and H. Weinstein, "Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors," in *Receptor Molecular Biology* (S. C. Sealfon, ed.), vol. 25 of *Methods in Neurosciences*, pp. 366 – 428, Academic Press, 1995.

[23] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.3r1." Available at `www.pymol.org`, August 2010.

[24] A. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, and M. M. Babu, "Molecular signatures of G-protein-coupled receptors," *Nature*, vol. 494, no. 7436, pp. 185–194, 2013.

[25] J. Li, P. C. Edwards, M. Burghammer, C. Villa, and G. F. Schertler, "Structure of Bovine Rhodopsin in a Trigonal Crystal Form," *Journal of Molecular Biology*, vol. 343, no. 5, pp. 1409 – 1438, 2004.

[26] H. Choe, Y. Kim, J. Park, T. Morizumi, E. Pai, N. Krauss, K. Hofmann, P. Scheerer, and O. Ernst, "Crystal structure of metarhodopsin II," *Nature*, vol. 471, no. 7340, pp. 651–5, 2011.

[27] T. Warne, M. Serrano-Vega, J. Baker, R. Moukhametzianov, P. Edwards, R. Henderson, A. Leslie, C. Tate, and G. Schertler, "Structure of a beta1-adrenergic G-protein-coupled receptor," *Nature*, vol. 454, no. 7203, pp. 486–91, 2008.

[28] V. Cherezov, D. Rosenbaum, M. Hanson, S. Rasmussen, F. Thian, T. Kobilka, H. Choi, P. Kuhn, W. Weis, B. Kobilka, and R. Stevens, "High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor," *Science (New York, N.Y.)*, vol. 318, no. 5854, pp. 1258–65, 2007.

[29] E. Chien, W. Liu, Q. Zhao, V. Katritch, G. Han, M. Hanson, L. Shi, A. Newman, J. Javitch, V. Cherezov, and R. Stevens, "Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist," *Science (New York, N.Y.)*, vol. 330, no. 6007, pp. 1091–5, 2010.

[30] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. Han, T. Kobayashi, R. Stevens, and S. Iwata, "Structure of the

human histamine H1 receptor complex with doxepin," *Nature*, vol. 475, no. 7354, pp. 65–70, 2011.

[31] K. Haga, A. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. Weis, T. Okada, B. Kobilka, T. Haga, and T. Kobayashi, "Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist," *Nature*, vol. 482, no. 7386, pp. 547–51, 2012.

[32] A. C. Kruse, A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P. M. Sexton, A. Christopoulos, C. C. Felder, P. Gmeiner, J. Steyaert, W. I. Weis, K. C. Garcia, J. Wess, and B. K. Kobilka, "Activation and allosteric modulation of a muscarinic acetylcholine receptor," *Nature*, vol. 504, pp. 101–106, Dec 2013.

[33] A. Kruse, J. Hu, A. Pan, D. Arlow, D. Rosenbaum, E. Rosemond, H. Green, T. Liu, P. Chae, R. Dror, D. Shaw, W. Weis, J. Wess, and B. Kobilka, "Structure and dynamics of the M3 muscarinic acetylcholine receptor," *Nature*, vol. 482, no. 7386, pp. 552–6, 2012.

[34] C. Wang, Y. Jiang, J. Ma, H. Wu, D. Wacker, V. Katritch, G. Han, W. Liu, X. Huang, E. Vardy, J. McCorvy, X. Gao, X. Zhou, K. Melcher, C. Zhang, F. Bai, H. Yang, L. Yang, H. Jiang, B. Roth, V. Cherezov, R. Stevens, and H. Xu, "Structural basis for molecular recognition at serotonin receptors," *Science (New York, N.Y.)*, vol. 340, no. 6132, pp. 610–4, 2013.

[35] W. Liu, D. Wacker, C. Gati, G. Han, D. James, D. Wang, G. Nelson, U. Weierstall, V. Katritch, A. Barty, N. Zatsepin, D. Li, M. Messerschmidt, S. Boutet, G. Williams, J. Koglin, M. Seibert, C. Wang, S. Shah, S. Basu, R. Fromme, C. Kupitz, K. Rendek, I. Grotjohann, P. Fromme, R. Kirian, K. Beyerlein, T. White, H. Chapman, M. Caffrey, J. Spence, R. Stevens, and V. Cherezov, "Serial femtosecond crystallography of G protein-coupled receptors," *Science (New York, N.Y.)*, vol. 342, no. 6165, pp. 1521–4, 2013.

[36] V. Jaakola, M. Griffith, M. Hanson, V. Cherezov, E. Chien, J. Lane, A. Ijzerman, and R. Stevens, "The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist," *Science (New York, N.Y.)*, vol. 322, no. 5905, pp. 1211–7, 2008.

[37] F. Xu, H. Wu, V. Katritch, G. Han, K. Jacobson, Z. Gao, V. Cherezov, and R. Stevens, "Structure of an agonist-bound human A2A adenosine receptor," *Science (New York, N.Y.)*, vol. 332, no. 6027, pp. 322–7, 2011.

[38] M. Hanson, C. Roth, E. Jo, M. Griffith, F. Scott, G. Reinhart, H. Desale, B. Clemons, S. Cahalan, S. Schuerer, M. Sanna, G. Han, P. Kuhn, H. Rosen, and R. Stevens, "Crystal structure of a lipid G protein-coupled receptor," *Science (New York, N.Y.)*, vol. 335, no. 6070, pp. 851–5, 2012.

[39] J. White, N. Noinaj, Y. Shibata, J. Love, B. Kloss, F. Xu, J. Gvozdenovic-Jeremic, P. Shah, J. Shiloach, C. Tate, and R. Grisshammer, "Structure of the agonist-bound neurotensin receptor," *Nature*, vol. 490, no. 7421, pp. 508–13, 2012.

[40] B. Wu, E. Chien, C. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. Bi, D. Hamel, P. Kuhn, T. Handel, V. Cherezov, and R. Stevens, "Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists," *Science (New York, N.Y.)*, vol. 330, no. 6007, pp. 1066–71, 2010.

[41] Q. Tan, Y. Zhu, J. Li, Z. Chen, G. Han, I. Kufareva, T. Li, L. Ma, G. Fenalti, J. Li, W. Zhang, X. Xie, H. Yang, H. Jiang, V. Cherezov, H. Liu, R. Stevens, Q. Zhao, and B. Wu, "Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex," *Science (New York, N.Y.)*, vol. 341, no. 6152, pp. 1387–90, 2013.

[42] H. Wu, D. Wacker, M. Mileni, V. Katritch, G. Han, E. Vardy, W. Liu, A. Thompson, X. Huang, F. Carroll, S. Mascarella, R. Westkaemper, P. Mosier, B. Roth, V. Cherezov, and R. Stevens, "Structure of the human κ-opioid receptor in complex with JDTic," *Nature*, vol. 485, no. 7398, pp. 327–32, 2012.

[43] A. Manglik, A. Kruse, T. Kobilka, F. Thian, J. Mathiesen, R. Sunahara, L. Pardo, W. Weis, B. Kobilka, and S. Granier, "Crystal structure of the μ-opioid receptor bound to a morphinan antagonist," *Nature*, vol. 485, no. 7398, pp. 321–6, 2012.

[44] A. Thompson, W. Liu, E. Chun, V. Katritch, H. Wu, E. Vardy, X. Huang, C. Trapella, R. Guerrini, G. Calo, B. Roth, V. Cherezov, and R. Stevens, "Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic," *Nature*, vol. 485, no. 7398, pp. 395–9, 2012.

[45] S. Granier, A. Manglik, A. Kruse, T. Kobilka, F. Thian, W. Weis, and B. Kobilka, "Structure of the δ-opioid receptor bound to naltrindole," *Nature*, vol. 485, no. 7398, pp. 400–4, 2012.

[46] C. Zhang, Y. Srinivasan, D. Arlow, J. Fung, D. Palmer, Y. Zheng, H. Green, A. Pandey, R. Dror, D. Shaw, W. Weis, S. Coughlin, and B. Kobilka, "High-resolution crystal structure of human protease-activated receptor 1," *Nature*, vol. 492, no. 7429, pp. 387–92, 2012.

[47] K. Zhang, J. Zhang, Z. Gao, D. Zhang, L. Zhu, G. Han, S. Moss, S. Paoletta, E. Kiselev, W. Lu, G. Fenalti, W. Zhang, C. Mller, H. Yang, H. Jiang, V. Cherezov, V. Katritch, K. Jacobson, R. Stevens, B. Wu, and Q. Zhao, "Structure of the human P2Y12 receptor in complex with an antithrombotic drug," *Nature*, vol. 509, no. 7498, pp. 115–8, 2014.

[48] K. Hollenstein, J. Kean, A. Bortolato, R. Cheng, A. Dor, A. Jazayeri, R. Cooke, M. Weir, and F. Marshall, "Structure of class B GPCR corticotropin-releasing factor receptor 1," *Nature*, vol. 499, no. 7459, pp. 438–43, 2013.

[49] F. Siu, M. He, C. de Graaf, G. Han, D. Yang, Z. Zhang, C. Zhou, Q. Xu, D. Wacker, J. Joseph, W. Liu, J. Lau, V. Cherezov, V. Katritch, M. Wang, and R. Stevens, "Structure of the human glucagon class B G-protein-coupled receptor," *Nature*, vol. 499, no. 7459, pp. 444–9, 2013.

[50] H. Wu, C. Wang, K. Gregory, G. Han, H. Cho, Y. Xia, C. Niswender, V. Katritch, J. Meiler, V. Cherezov, P. Conn, and R. Stevens, "Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator," *Science (New York, N.Y.)*, vol. 344, no. 6179, pp. 58–64, 2014.

[51] A. S. Dore, K. Okrasa, J. C. Patel, M. Serrano-Vega, K. Bennett, R. M. Cooke, J. C. Errey, A. Jazayeri, S. Khan, B. Tehan, M. Weir, G. R. Wiggin, and F. H. Marshall, "Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain," *Nature*, vol. advance online publication, pp. –, 2014.

[52] C. Wang, H. Wu, V. Katritch, G. Han, X. Huang, W. Liu, F. Siu, B. Roth, V. Cherezov, and R. Stevens, "Structure of the human smoothened receptor bound to an antitumour agent," *Nature*, vol. 497, no. 7449, pp. 338–43, 2013.

[53] C. Wang, H. Wu, T. Evron, E. Vardy, G. W. Han, X. P. Huang, S. J. Hufeisen, T. J. Mangano, D. J. Urban, V. Katritch, V. Cherezov, M. G. Caron, B. L. Roth, and R. C. Stevens, "Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs," *Nat Commun*, vol. 5, p. 4355, 2014.

[54] R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander, and G. Vriend, "A series of PDB related databases for everyday needs," *Nucleic Acids Res.*, vol. 39, pp. D411–419, Jan 2011.

[55] C. E. Scott, R. Abrol, K. H. Ahn, D. A. Kendall, and W. A. Goddard, "Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations," *Protein Sci.*, vol. 22, pp. 101–113, Jan 2013.

[56] K. H. Ahn, C. E. Scott, R. Abrol, W. A. Goddard, and D. A. Kendall, "Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding," *Proteins*, vol. 81, pp. 1304–1317, Aug 2013.

[57] T. U. Consortium, "Activities at the Universal Protein Resource (UniProt)," *Nucleic Acids Research*, vol. 42, no. D1, pp. D191–D198, 2014.

[58] B. J. Feldman, S. M. Rosenthal, G. A. Vargas, R. G. Fenwick, E. A. Huang, M. Matsuda-Abedini, R. H. Lustig, R. S. Mathias, A. A. Portale, W. L. Miller, and S. E. Gitelman,

"Nephrogenic syndrome of inappropriate antidiuresis," *N. Engl. J. Med.*, vol. 352, pp. 1884–1890, May 2005.

[59] E. Schipani, K. Kruse, and H. Juppner, "A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia," *Science*, vol. 268, pp. 98–100, Apr 1995.

[60] O. Edvardsen, A. L. Reiersen, M. W. Beukers, and K. Kristiansen, "tGRAP, the G-protein coupled receptors mutant database," *Nucleic Acids Res.*, vol. 30, pp. 361–363, Jan 2002.

[61] J. Kazius, K. Wurdinger, M. van Iterson, J. Kok, T. Back, and A. P. Ijzerman, "GPCR NaVa database: natural variants in human G protein-coupled receptors," *Hum. Mutat.*, vol. 29, pp. 39–44, Jan 2008.

[62] J. R. Doyle, J. M. Lane, M. Beinborn, and A. S. Kopin, "Naturally occurring HCA1 missense mutations result in loss of function: potential impact on lipid deposition," *J. Lipid Res.*, vol. 54, pp. 823–830, Mar 2013.

[63] D. Wootten, J. Simms, L. J. Miller, A. Christopoulos, and P. M. Sexton, "Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. 5211–5216, 2013.

[64] C. Kuei, J. Yu, J. Zhu, J. Wu, L. Zhang, A. Shih, T. Mirzadegan, T. Lovenberg, and C. Liu, "Study of GPR81, the lactate receptor, from distant species identifies residues and motifs critical for GPR81 functions," *Mol. Pharmacol.*, vol. 80, pp. 848–858, Nov 2011.

[65] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *J. Mol. Biol.*, vol. 305, pp. 567–580, Jan 2001.

[66] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, no. 1, 2011.

[67] I. Letunic and P. Bork, "Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy," *Nucleic Acids Research*, vol. 39, no. suppl 2, pp. W475–W478, 2011.

[68] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, "Jalview Version 2a multiple sequence alignment editor and analysis workbench," *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.

[69] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, and P. Bradley, "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules," *Meth. Enzymol.*, vol. 487, pp. 545–574, 2011.

[70] P. Barth, J. Schonbrun, and D. Baker, "Toward high-resolution prediction and design of transmembrane helical protein structures," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, pp. 15682–15687, Oct 2007.

[71] K. J. Kohlhoff, D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman, and V. S. Pande, "Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways," *Nat Chem*, vol. 6, pp. 15–21, Jan 2014.

[72] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovi?, D. Baker, and F. Players, "Algorithm discovery by protein folding game players," *Proceedings of the National Academy of Sciences*, vol. 108, no. 47, pp. 18949–18953, 2011.

[73] R. Abrol, S.-K. Kim, J. K. Bray, A. R. Griffith, and W. A. G. III, "Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins," *Methods*, vol. 55, no. 4, pp. 405 – 414, 2011. Membrane Protein Technologies for Structural Biology.

[74] R. Abrol, J. K. Bray, and W. A. Goddard, "Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations," *Proteins: Structure, Function, and Bioinformatics*, vol. 80, no. 2, pp. 505–518, 2012.

[75] V. W. Tak Kam and W. A. Goddard, "Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements," *Journal of Chemical Theory and Computation*, vol. 4, no. 12, pp. 2160–2169, 2008.

[76] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, "Improved prediction of protein side-chain conformations with SCWRL4," *Proteins*, vol. 77, pp. 778–795, Dec 2009.

[77] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: a generic force field for molecular simulations," *The Journal of Physical Chemistry*, vol. 94, no. 26, pp. 8897–8909, 1990.

[78] J. Wang, P. Cieplak, and P. A. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?," *Journal of Computational Chemistry*, vol. 21, no. 12, pp. 1049–1074, 2000.

[79] S. Chaudhury, S. Lyskov, and J. J. Gray, "PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta," *Bioinformatics*, vol. 26, pp. 689–691, Mar 2010.

[80] E. H. Baugh, S. Lyskov, B. D. Weitzner, and J. J. Gray, "Real-time PyMOL visualization for Rosetta and PyRosetta," *PLoS ONE*, vol. 6, no. 8, p. e21931, 2011.

[81] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[82] A. Leaver-Fay, J. Snoeyink, and B. Kuhlman, "On-the-fly rotamer pair energy evaluation in protein design," in *Bioinformatics Research and Applications* (I. Mandoiu, R. Sunderraman, and A. Zelikovsky, eds.), vol. 4983 of *Lecture Notes in Computer Science*, pp. 343–354, Springer Berlin Heidelberg, 2008.

[83] C. L. Roland, T. Arumugam, D. Deng, S. H. Liu, B. Philip, S. Gomez, W. R. Burns, V. Ramachandran, H. Wang, Z. Cruz-Monserrate, and C. D. Logsdon, "Cell surface lactate receptor GPR81 is crucial for cancer cell survival," *Cancer Res.*, vol. 74, pp. 5301–5310, Sep 2014.

[84] T.-Q. Cai, N. Ren, L. Jin, K. Cheng, S. Kash, R. Chen, S. D. Wright, A. K. Taggart, and M. G. Waters, "Role of {GPR81} in lactate-mediated reduction of adipose lipolysis," *Biochemical and Biophysical Research Communications*, vol. 377, no. 3, pp. 987 – 991, 2008.

[85] C. Liu, J. Wu, J. Zhu, C. Kuei, J. Yu, J. Shelton, S. W. Sutton, X. Li, S. J. Yun, T. Mirzadegan, C. Mazur, F. Kamme, and T. W. Lovenberg, "Lactate inhibits lipolysis in fat cells through activation of an orphan G-protein-coupled receptor, GPR81," *J. Biol. Chem.*, vol. 284, pp. 2811–2822, Jan 2009.

[86] A. K. P. Taggart, J. Kero, X. Gan, T.-Q. Cai, K. Cheng, M. Ippolito, N. Ren, R. Kaplan, K. Wu, T.-J. Wu, L. Jin, C. Liaw, R. Chen, J. Richman, D. Connolly, S. Offermanns, S. D. Wright, and M. G. Waters, "(D)-beta-Hydroxybutyrate inhibits adipocyte lipolysis via the nicotinic acid receptor PUMA-G," *Journal of Biological Chemistry*, vol. 280, no. 29, pp. 26649–26652, 2005.

[87] K. Ahmed, S. Tunaru, C.-D. Langhans, J. Hanson, C. W. Michalski, S. Koelker, P. M. Jones, J. G. Okun, and S. Offermanns, "Deorphanisation of GPR109B as a receptor for the beta-Oxidation intermediate 3-OH-octanoic acid and its role in the regulation of lipolysis," *Journal of Biological Chemistry*, 2009.

[88] K. Ahmed, S. Tunaru, and S. Offermanns, "GPR109A, GPR109B and GPR81, a family of hydroxy-carboxylic acid receptors," *Trends Pharmacol. Sci.*, vol. 30, pp. 557–562, Nov 2009.

[89] K. Ahmed, S. Tunaru, C. Tang, M. Mller, A. Gille, A. Sassmann, J. Hanson, and S. Offermanns, "An Autocrine Lactate Loop Mediates Insulin-Dependent Inhibition of Lipolysis through {GPR81}," *Cell Metabolism*, vol. 11, no. 4, pp. 311 – 319, 2010.

[90] D. Langin, "Adipose Tissue Lipolysis Revisited (Again!): Lactate Involvement in Insulin Antilipolytic Action," *Cell Metabolism*, vol. 11, no. 4, pp. 242 – 243, 2010.

[91] S. Offermanns, S. L. Colletti, T. W. Lovenberg, G. Semple, A. Wise, and A. P. IJzerman, "International Union of Basic and Clinical Pharmacology. LXXXII: Nomenclature and Classification of Hydroxy-carboxylic Acid Receptors (GPR81, GPR109A, and GPR109B)," *Pharmacol. Rev.*, vol. 63, pp. 269–290, Jun 2011.

[92] D. A. Goldfeld, K. Zhu, T. Beuming, and R. A. Friesner, "Successful prediction of the intra- and extracellular loops of four G-protein-coupled receptors," *Proceedings of the National Academy of Sciences*, vol. 108, no. 20, pp. 8275–8280, 2011.

[93] G. V. Nikiforovich, C. M. Taylor, G. R. Marshall, and T. J. Baranski, "Modeling the possible conformations of the extracellular loops in G-protein-coupled receptors," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 2, pp. 271–285, 2010.

[94] A. Stein and T. Kortemme, "Improvements to Robotics-Inspired Conformational Sampling in Rosetta," *PLoS ONE*, vol. 8, p. e63090, 05 2013.

[95] C. Kuei, J. Yu, J. Zhu, J. Wu, L. Zhang, A. Shih, T. Mirzadegan, T. Lovenberg, and C. Liu, "Study of GPR81, the lactate receptor, from distant species identifies residues and motifs critical for GPR81 functions," *Mol. Pharmacol.*, vol. 80, pp. 848–858, Nov 2011.

[96] C. Liu, C. Kuei, J. Zhu, J. Yu, L. Zhang, A. Shih, T. Mirzadegan, J. Shelton, S. Sutton, M. A. Connelly, G. Lee, N. Carruthers, J. Wu, and T. W. Lovenberg, "3,5-Dihydroxybenzoic acid, a specific agonist for hydroxycarboxylic acid 1, inhibits lipolysis in adipocytes," *J. Pharmacol. Exp. Ther.*, vol. 341, pp. 794–801, Jun 2012.

[97] W. A. Goddard, S. K. Kim, Y. Li, B. Trzaskowski, A. R. Griffith, and R. Abrol, "Predicted 3D structures for adenosine receptors bound to ligands: comparison to the crystal structure," *J. Struct. Biol.*, vol. 170, pp. 10–20, Apr 2010.

[98] J. K. Bray and W. A. G. III, "The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists," *Journal of Molecular Graphics and Modelling*, vol. 27, no. 1, pp. 66 – 81, 2008.

[99] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig,

S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: the biomolecular simulation program," *J Comput Chem*, vol. 30, pp. 1545–1614, Jul 2009.

[100] F. A. Momany, "Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid," *The Journal of Physical Chemistry*, vol. 82, no. 5, pp. 592–601, 1978.

[101] R. S. Mulliken, "Electronic population analysis on LCAO–MO molecular wave functions. I," *The Journal of Chemical Physics*, vol. 23, no. 10, pp. 1833–1840, 1955.

[102] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *The Journal of Physical Chemistry*, vol. 97, no. 40, pp. 10269–10280, 1993.

[103] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, III, "ReaxFF: A reactive force field for hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, no. 41, pp. 9396–9409, 2001.

[104] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard, III, "ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation," *The Journal of Physical Chemistry A*, vol. 112, no. 5, pp. 1040–1053, 2008.

[105] S. Naserifar, L. Liu, W. A. Goddard, III, T. T. Tsotsis, and M. Sahimi, "Toward a process-based molecular model of SiC membranes. I. Development of a reactive force field," *The Journal of Physical Chemistry C*, vol. 117, no. 7, pp. 3308–3319, 2013.

[106] S. Naserifar, W. A. Goddard, III, L. Liu, T. T. Tsotsis, and M. Sahimi, "Toward a process-based molecular model of SiC membranes. II. Reactive dynamics simulation of the pyrolysis of polymer precursor to form amorphous SiC," *The Journal of Physical Chemistry C*, vol. 117, no. 7, pp. 3320–3329, 2013.

[107] A. Jaramillo-Botero, S. Naserifar, and W. A. Goddard III, "General Multiobjective Force Field Optimization Framework, with Application to Reactive Force Fields for Silicon Carbide," *Journal of Chemical Theory and Computation*, vol. 10, no. 4, pp. 1426–1439, 2014.

[108] A. K. Rappe and W. A. Goddard III, "Charge equilibration for molecular dynamics simulations," *The Journal of Physical Chemistry*, vol. 95, no. 8, pp. 3358–3363, 1991.

[109] Q. Zhang and W. A. Goddard III, "Charge and polarization distributions at the 90 domain wall in barium titanate ferroelectric," *Applied physics letters*, vol. 89, no. 18, p. 182903, 2006.

[110] Q. Zhang, T. Cagin, and W. A. Goddard, "The ferroelectric and cubic phases in $BaTiO_3$ ferroelectrics are also antiferroelectric," *Proceedings of the National Academy of Sciences*, vol. 103, no. 40, pp. 14695–14700, 2006.

[111] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Journal of Computational Physics*, vol. 117, no. 1, pp. 1 – 19, 1995.

[112] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: Application to liquid water," *The Journal of Chemical Physics*, vol. 101, no. 7, pp. 6141–6156, 1994.

[113] H. Aktulga, J. Fogarty, S. Pandit, and A. Grama, "Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques," *Parallel Computing*, vol. 38, no. 45, pp. 245 – 259, 2012.

[114] J. Thijssen, *Computational Physics*. Cambridge University Press, 1999.

[115] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[116] Wikipedia, "Amino acid — Wikipedia, The Free Encyclopedia," 2004.

[117] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, pp. 10915–10919, Nov 1992.