Abstract

G protein-coupled receptors (GPCRs) are responsible for the majority of cellular responses to hormones and neurotransmitters as well as the senses of sight, olfaction and taste. They represent the largest family of membrane receptors that instigate signalling through nucleotide exchange on heterotrimeric G proteins. GPCRs are the single most important drug target in the pharmaceutical industry. Many structures recently became available, but unfortunately all of them contain artefacts that were needed to obtain good crystals. We filter out those artefacts by studying the mobility in distance space using the Random Forest (RF) method. We found several trends comparing the structures: the extracellular sides of transmembrane helices (TMs) 5, 6 and 7 move towards each other upon agonist binding and away from each other upon inverse agonist binding. Also the extracellular side of TM2, around the π-bulge, moves towards TM5, TM6 and TM7 upon agonist binding. This correlated with an upward movement of TM3. Fusion of GPCRs with T4 Lysozyme didn’t appear to have a consistent influence on the TM bundle, apart from a few disturbances of the intracellular loops. The structure stabilizing antibodies, on the other hand, lead to significant disturbances of the helices. We also found that π-bulges of differing length are present in several helices, these differ in length and location between different subfamilies of GPCRs.

The RF method detects important distances and is resistant to noise: Agonist binding (1) induces inward motions (2) of the extracellular side of TMS-7 resembling a clothes pin-like motion. This is followed by an outward movement of the cytosolic side of TMS-7 (3), allowing the G protein to bind (4) and become activated (5).
Introduction

GPCRs
Structure
Activation
The rhodopsin model
Ligand binding GPCRs
Approach
Random Forest
Variable Importance

GPCRs

G protein-coupled receptors (GPCRs) are responsible for the majority of cellular responses to external signals. They represent the largest and arguably the most diverse superfamily of membrane receptors represented in every eukaryotic cell. Over 1000 human genes encode GPCRs. The ligands that bind or otherwise activate these receptors are heterogeneous and include photons, odours, pheromones, hormones, ions, neurotransmitters and proteases. GPCRs transmit signals from outside the cell to amplification cascades controlling sight, taste, smell, slow neurotransmission, cell division, etc. 62

At first, in 1994, Attwood and Findlay categorized the superfamily into six classes (A–F) based on sequence homology and functional similarity. Later, our comprehensive phylogenetic analysis of the human repertoire provided the GRAFS classification. This system grouped the mammalian GPCR repertoire into five main families; Rhodopsin (Class A), Adhesion (Class B), Glutamate (Class C), Frizzled (Class F), and Secretin (Class B). 6 The Rhodopsin family is the largest with 683 members in humans 7, characterized by short N-termini and interactions with a broad variety of ligands. The Glutamate family is distinguished with long N-termini which act as the endogenous ligand binding region. The Adhesion receptors have long N-termini which contain a plethora of multiple domains while the Frizzled receptors have long cysteine-rich N-termini 8.

This scala of different GPCRs has a scala of different G protein they can bind. Subsequent to initial cloning efforts, cloning by homology has defined the human G-proteins to derive from 35 genes, 16 encoding alpha-subunits, five beta and 14 gamma. All function as guanine nucleotide exchange on-off switches and are mechanistically similar to other proteins that are enzymatic GTPases. 9 Heterotrimeric G proteins can be broadly categorized into four major classes based on the identity of the β subunit: G_s, G_i/o, G_q/11, and G_12/13. 11

Structure

To understand the function of GPCRs at the molecular level, it is fundamental to investigate the nature of the structural rearrangements that couple ligand binding to receptor-dependent activation of downstream signalling pathways. All GPCRs contain a bundle of seven helices spanning the membrane (7TM) connected by three intracellular (ICL) and three extracellular loops (ECL). Rhodopsin was the first GPCR whose crystal structure was determined to high resolution. 35 Since then a number of structures were elucidated by crystallography. Recently more and more class A GPCR structures besides Rhodopsin became available. All showing a highly...
GPCRs are very flexible proteins and are located in the fatty environment of the cell membrane. To allow these flexible proteins to assume a rigid structure, which is necessary in a crystal structure, they were modified to great extent. These modifications include thermostabilizing mutations, T4 Lysozyme fusion (Fig. 1, 2, Anti- or Nano-body binding or even a combination of several (Fig). Among Class A GPCR there are some residues that are highly conserved. In each transmembrane helix there is at least one residue that is highly conserved. In the figure above you can see the most conserved residues shown in red. In TM1 the Asn at position 130 (Oliveira numbering) is highly conserved. In TM2 it's L220. In TM3 there is a highly conserved motif consisting of E/D339, R340 and Y341 the so-called E/DRY motif. In TM4 it's W420. In TM5 P520 and Y528 are highly conserved. In TM6 C617, W618 and P620 forming the CWxP motif and in TM7 there's a motif called the NPxxY motif consisting of N729, P730 and Y733. There are many theories about the roles of these residues in the transduction of the signal between the ligand binding pocket and the G protein binding pocket. Modifications

![Image of GPCR modifications](http://www.cmbi.ru.nl/~rkant/GPCR/full_report.html)

## Activation

The repertoire of signalling activities by GPCRs is considerably more complex than envisioned in the early periods of research that led to the formulation of classical models (Fig). GPCR signalling can no longer be viewed as a single pathway consisting of a linear sequence of events. For a long time GPCRs were thought to perform a relatively straightforward role, namely coupling the binding of agonists to the activation of G proteins, which in turn leads to modulation of other downstream effector proteins. However, in recent years it has become clear that many GPCRs have much more complex signalling characteristics. Many GPCRs are constitutively active, and this allows for a fine-grained control of the amount of G-protein activation, being subject to regulation by agonist as well as...
inverse agonists. It is now clear that G proteins exhibit some promiscuity and a single GPCR can couple to and signal via G proteins from multiple classes, which results in the propagation of signals through multiple biochemical pathways to achieve different cellular responses. Also in the classical view, GPCRs signal exclusively via heterotrimeric G proteins to generate the cellular response. It is now apparent that G protein-independent signalling can occur within a cell via arrestin molecules to increase the diversity of cellular responses a single GPCR can generate. Arrestins in the classical view serve as proteins that bind phosphorylated GPCRs deactivate and, in most instances, internalize the receptor. The role arrestins play in the cell has now expanded to also include signalling functions. GPCRs can signal independently of G proteins via arrestins to mitogen-activated protein (MAP) kinases, which can regulate chemotaxis, apoptosis, cancer metastasis, and protein translation. Because of the capability of receptors to signal independently of the G protein, GPCRs are now sometimes referred to as seven-transmembrane receptors. Desensitization processes can involve multiple pathways, including phosphorylation events, arrestin-mediated receptor internalization, receptor recycling, and lysosomal degradation. The reason for this multifaceted behaviour may be the fact that, while there are only about 1000 GPCRs that can be activated by an even smaller number of endogenous agonists, these receptors need to cater for many 1000s of different messages that the whole organism needs to be able to transmit internally. They can also form homo- and heterodimers, which cooperatively modulate signaling. For more information on dimerization, click here.

The rhodopsin model

One member of the visual receptor subfamily of GPCRs, rhodopsin, is by far the best structurally defined GPCR. Rhodopsin has been studied extensively and it’s structure has been used as a template for homology modelling, but rhodopsin is actually a special case among the whole GPCR family, because it has it’s ligand covalently bound to Lys723 by a protonated Schiff base. All visual receptors from humans to squid have the 11-cis isomer of retinal bound within the 7TM bundle. In pharmacological terms, the 11-cis retinal chromophore acts as an inverse agonist, when bound it reduces the basal activity of the apoprotein opsin. Upon absorption of light, one photon is enough, the retinal isomerizes to the trans form within 200 fs. Through a series of different intermediates it decays thermally (see figure). In the transition from Meta I to the Meta II intermediate the receptor undergoes a large conformational change to generate the G-protein-binding pocket on the intracellular side of the receptor. In 1996 Farrens et al. showed that there is an outward rotation of TM6 in the transition to Meta II. This motion would create the cavity for the G protein to enter. The mechanism that connects the isomerization of retinal to the opening of the cavity remained illusive. This only began to be understood once the structures became available. Rhodopsin was the first GPCR determined at high resolution. The reason crystals of this GPCR are easiest to obtain is because they can be isolated in large amounts from the retina’s of cows. The structures of inactive rhodopsin showed the 7TM bundle structure and revealed the location of residues that are conserved across the large A GPCR family. There were several unexpected features of the structure. On the extracellular side, ECL2 was wedged between the TM helices and served as a cap on the retinal-binding site. On the intracellular side, a short amphipathic helix was found to be oriented roughly perpendicular to the seven TM helices, termed (TM8), lying parallel to the membrane. The first
clues to the structural changes occurring upon activation came from crystal structures of opsin that were determined with and without a bound undecapeptide mimic of the C-terminus of Gα. The defining feature of the opsin structure was the outward rotation of TM helix H6. There are some interesting theories of how retinal isomeration is linked to the opening of the intracellular crevice. One of them states that the photon induced retinal isomeration triggers a chain reaction of molecular switches that involve most of the highly conserved residues. The rotation of the C20 methyl group toward ECL2 and motion of the β-ionone ring toward TM5 leads to a change in orientation of TM5. Tyr528 on the intracellular side of TM5 rotates toward Arg340 on TM3. Second, rotation of the retinal C20 methyl group and motion of the β-ionone ring enable Trp618 to rotate toward the extracellular surface. The motion of Trp618 triggers an internal switch involving Asn719, Met610 and Tyr733. Together, the reorientation of Tyr518 on H5, Met610 on H6 and Tyr733 on H7 stabilize the ionic lock in an ‘open’ conformation, allowing for G protein binding and activation (illustration). Rhodopsin appears to have evolved mechanisms that stabilize the receptor not only in an inactive conformation in the dark but also in a fully active conformation on light absorption. The receptor is a robust on–off switch making use of light energy to bridge two very stable conformations.

Ligand binding GPCRs

In contrast, the ligand-activated GPCRs have much smaller barriers to activation. Multiple receptor conformations can be populated, which leads to high basal activity, but also provides versatility in signalling and regulation. After the rhodopsin structure another seven years of extensive research and technology developments were needed to obtain the high-resolution structure of the human β2-adrenergic receptor (β2AR) – the first example of a GPCR with a diffusible ligand. This structure was followed by other class A (rhodopsin-like) GPCRs, including β1AR, A2A adenosine (A2A AR)51, chemokine CXCR4 (CXCR4)52, dopamine D3 (D3R)53, histamine H4 (H4R)54, sphingosine 1-phosphate (S1P1)55, M2 and M3 muscarinic acetylcholine (mACHR M2 and M3)56,77, k-opioid(k-OR)58, µ-opioid(µ-OR)59, δ-opioid receptor(δ-OR)60 and the nociceptin/orphanin FQ (N/OFQ) peptide (NOP) receptors. The structures of these ligand binding GPCRs were either bound to an agonist, a partial agonist, an inverse agonist, a partial inverse agonist or an antagonist. One of them, a β2AR(PDBID: 3sn6.pdb), even has an entire G-protein bound.

Pharmacological definitions

Ligands or drugs that interact with GPCRs are defined according to their activity when added to cells that contains the specific GPCR of interest. Different ligands can stabilize different structural conformations. Definitions of ligands according to their biological activity are listed below.

Agonist: a ligand that binds to and activates a receptor and elicits a physiological response. The endogenous agonist for the β1AR, noradrenaline, is a full agonist (Figure I, red line) that elicits the maximal response for the receptor in activating a G protein.  
Basal or constitutive activity: a physiological response that occurs in the absence of an agonist or inverse agonist due to a proportion of the receptor being in the activated state.  
Inverse agonist: a ligand that binds to a receptor and inhibits or eliminates, in the case of a full inverse agonist, the basal or constitutive activity of a receptor (Figure I, green line).  
Neutral antagonist: a ligand that binds to the receptor, making it impossible for an agonist or an
inverse agonist to bind, while maintaining the basal activity (Figure I, black line).

**Partial agonist** or **weak partial agonist**: a ligand that elicits only a partial response when compared to a full agonist (Figure I, blue and yellow lines).

**Biased agonist**: many ligands can differentially activate signaling pathways mediated via a single G protein-coupled receptor. Subsequent mechanisms that may play a role include diversity of G proteins, scaffolding and signaling partners, and receptor oligomers.88

These structures of different GPCRs in different stages of activation allowed researchers to look at the structural changes that occurred upon activation and compare the different GPCRs that were in the same state of activation. The general structure of the 7TM bundle appeared to be highly similar between the different GPCRs, which was predicted. The largest differences between the different GPCRs was of course the extracellular part of the receptor, because they all have to bind to different ligands and be specific. When more and more of these structures were solved, a scala of hypotheses arose and the search for the main mechanism of activation had begun. In 2006 Schwartz et al.69 published a paper proposing a global toggle switch model for the activation mechanism to reconcile the accumulated biophysical data supporting an outward rigid-body movement of the intracellular segments, as well as an recent data derived from activating metal ion sites and tethered ligands, which suggest an opposite, inward movement of the extracellular segments of the TMs. According to this model a vertical see-saw movement of TM6, and to some degree TM7, around a pivot corresponding to the highly conserved prolines will occur during receptor activation, which may involve the outer segment of TM5. Agonists can stabilize such a proposed active conformation, where the extracellular segments of TM6 and TM7 are bent inward toward TM3, by acting as molecular glue deep in the main ligand-binding pocket between the helices whereas larger agonists, peptides and proteins can stabilize a similar active conformation by acting as Velcro at the intracellular ends of the helices and the connecting loops. They proposed that the helices move in a see-saw-like manner. Recently Mason et al.70 used a series of computational methods of the X-ray structures elucidated so far and found that upon receptor activation, the volume of the ligand binding site decreases by ~40 Å³ for the aminergic β₁ and β₂AR and the ~90 Å³ for the purinergic A₂A receptor; rhodopsin is the exception with an increase of ~100 Å³. Indicating that agonists stabilize a receptor conformation in which the extracellular sides of certain helices are closer together. This finding, combined with the finding of the see-saw-like motions of helices, results in model that could be compared to the mechanism that a clothespin uses.

**Approach**

We want to know what moves systematically upon activation of a GPCR, to elucidate the mechanism of GPCR activation. At the moment (July 2012) there are 72 structures of GPCRs available from the Protein Data Bank (PDB). As mentioned a lot of these structures were modified to great extent. One way to compare all these GPCR structures is by superposing them based on their structure and look at the differences.(sce) One disadvantage of this approach is that it's impossible to draw conclusions from 72 superposed structures, like you can see in the figure.

Another disadvantage is that superposition is relative like you can see in the picture below. Both

http://www.cmbi.ru.nl/~rkant/GPCR/full_report.html
alignments are representations of TM6, red are the supposed active structures and in cyan are the supposed inactive structures. The left one is extracted from an alignment of the entire structures, which would lead to the conclusion that there is a bending of the transmembrane helix upon activation. The right one is a superposition of TM6 alone, which would lead to the conclusion that there is no significant difference between TM6 in the active and inactive state respectively. We decided to go into distance space, because distances are variables that are independent of 3D space. We chose to use the Random Forest method to calculate which distances describe the difference between inactive and active structures the best.

Random Forest

Random forest (RF) models are non-parametric and non-linear models, attractive due to their interpretability. They are based on averaging over a large collection of decision trees, each trained on a separate bootstrap sample of the input set. The aggregate model has lower variance and is less susceptible to overfitting than a single decision tree. Gini Importance (GI) and Variable Importance (VI) are two measures of feature relevance that can be computed based on the RF model. We use the R package randomForest for training RF models. There are two parameters that influence the performance of RF: the number ntree of trees in the collection and the number mtry of variables considered for each tree split. In our experiments, we use the recommended value of mtry (square root of number of features).

In 2001 Breiman et al. proposed random forests, which add an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. Random forests are becoming increasingly popular in many scientific fields because they can cope with "small n large p" problems, complex interaction and even highly correlated predictor variables.

Each tree is built as follows:

1) If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample will be training set for growing the tree.
2) If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split of these m is used to split the node. The value of m is held constant during the forest growing.
3) Each tree is grown to the largest extent possible. There is no pruning.

Some important features, that the RF method has, are that it runs efficiently on large data bases, it can handle thousands of input variables without variable deletion and it gives estimates of what variables are important in the classification. The latter will be extensively discussed below.

Variable Importance

This is a difficult concept to define in general, because the importance of a variable may be due to its (possibly complex interaction with other variables. The random forest algorithm calculates the importance of a variable using the out-of-bag individuals according to the following logic: If randomly permuting values of a variable does not affect the predictive ability of trees on out-of-bag samples, that variable is deemed as unimportant. If the variable drastically impairs the ability
of trees to correctly predict the class of out-of-bag samples after permutation, that variable is given a high importance score. This measure of variable importance is called “permutation importance”. In addition to this measure the algorithm calculates another measure of variable importance called “Gini importance”, which is calculated as follows: Every time a split of a node is made on variable m the gini impurity criterion for the two descendent nodes is less that the parent node. Adding up the gini decreases for each individual variable over all trees in the forest gives a fast variable importance that is often very consistent with the permutation importance measure.

Methods

Renumbering GPCR Structures

To measure distances within GPCRs the crystal structures were renumbered using Yasara according to the Oliveira numbering. All known crystal structures of GPCRs in the PDB were renumbered manually and the numbers were checked using WhatIf in Yasara. The structures were aligned with Yasara using Mustang. After alignment a java program was used to generate WhatIf-Scripts. This script measured the distances of the C-alphas of key residues between all structures. Results were analysed using a Java-application that reads the distances from the log-file generated by WhatIf. These distances were written to a table. Distances were checked manually, distances were supposed to be under 2.8A.

Measuring distances between key residues

After renumbering all GPCR structures distances were measured between key highly conserved residues. A Java-application was used that was able to generate a WhatIf-Script. This Script could be used in WhatIf within Yasara to measure distances between 7 residues within the GPCR-structures: N130, L220, R340, W420, P520, P620 and P730 (according to Oliveira numbering). This resulted in 21 distances and these values were stored in a log-file. The log-file was analysed using a Java-application which generated a text-file with distances.

Defining active GPCRs with internal distances

To see which distances could represent the active state of a GPCR, statistical analysis was necessary. Using the Random Forest method we analysed the distance-data of all GPCR-structures, looking for variables that significantly differ between active and inactive states of GPCR structures. The Random Forests scored variables according to importance with a Mean Decrease Gini score. Variables were sorted by this score and the most important distances were stored in text files. A python script was used to read these text files and create Yasara-Scenes that showed two structures, each class represented by one structure, and the most important distances were represented with arrows. These arrows were used to indicate which areas of the structures should be further investigated.

Aligning helices separately

After investigating the movement between the helices we investigated the differences within the helices. We used WhatIf to look at different collections of TMs as it calculated on which part of the helix it would be best to superpose upon. This way we investigated the intrahelical movements upon activation.
Indexing modifications PDB-structures

To see what how the GPCRs were modified to allow crystallization a special GPCR-pdf-reader was used\(^\text{46}\). This allowed for identification of thermostabilizing mutations within the GPCR-structures. These mutations were organized in a .xls file. Here a few screenshots of the GPCR-pdf-reader: 1, 2.

Results

To get an indication of what moves where upon activation, distances were measured between C-alpha's of residues shared by all GPCR structures. Distances were measured using WhatIf, these distances were analysed using the Random Forest package in R. Due to known correlation bias within the Random Forest method we chose to run the R scripts datasets with several correlation thresholds. We ran every script with the entire dataset, and datasets filtered for correlation thresholds of 0.99, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50 respectively. For the classification procedure we analysed literature to classify all structures for several classes: (1) active/inactive, (2) agonist/inverse agonist/antagonist bound, (3) lysozyme fusion. Random Forest then used these predefined hypotheses of the different GPCRs to find distances that correlate best with the different classes.

Alignment

Aligning all structures using Mustang revealed some peculiarities of the structures. The structures that had their N-terminus truncated had the extracellular part of TM1 pointing outward. The structures that still had a large part of their N-terminus, that packed on top of the 7TM bundle, had that part pointing inward. This may be due to an increased flexibility of the shortened N-terminal part of the receptor, allowing this part of the receptor to participate in crystal packing.

Another striking difference between the structures was that all CXCR4 receptors lacked a nicely packed helix 8. They all had it truncated or pointing away from the rest of the structure while having to clear structure. **Striking differences between structures.** Left: The truncation of the N-terminus clearly results in the outward pointing of the extracellular side of TM1. Right: The CXCR4 structures (Red, PDBID: 3odu, 3oe0, 3oe6, 3oe8, 3oe9) have a truncated or an unstructured TM8. They also have an extension of TM7 of 2 turns on the extracellular side compared to other structures.

Antibodies

Random Forest analysis of the large FAb5 antibodies that were co-crystallized with the β\(_2\)-AR structures (PDBID: 2r4r.pdb, 2r4s.pdb and 3kj6.pdb) compared to structures that weren't co-crystallized with this large antibody clearly showed a difference in helical packing(.sce). This is also clearly visible in the structures. **The effects of the large FAb5 co-crystallization with the β\(_2\)-AR structures.** Left: the
the FAb5 antibody in red and the β2-adrenergic receptor (Blue, PDBID: 2r4s). Middle: a close-up of the interaction between β2-AR with disturbed helices. Right: Distances that are important for classification between β2-AR structures with antibody (Red, PDBID: 2r4s) and without antibody (Cyan, PDBID: 3ny9).

Variables with the highest variable importance score were mostly distances within helices, so describing differences in vertical distances between the structures of the different classes. After visual inspection with Yasara the structures bound to a FAb5 antibody clearly showed distorted helices. These structures also had the disadvantage that large parts of the extracellular side of several transmembrane helices are missing. The large influence of this artefact on the helical packing of the transmembrane helices was a reason to leave out these three structures from further analysis. Leaving out these structures resulted in a larger amount of common C-alpha's among the remaining GPCR structures, resulting in more variables and a larger part of the 7TM bundle taken into account in further calculations. The amount of common Ca's went up from 157 to 203.

Active vs. Inactive

In this picture you see rhodopsin structure 1F88.pdb (sce). The residues are coloured by occurrence in the top ranking distances determined by Mean Decrease Gini from Random Forest. Highest occurrence (Yellow), lowest/no occurrence (Dark Blue). The most important residue in classification between active and inactive structures is Y733 located at the intracellular side of TM7. This residue is highly conserved and is hypothesised to stabilize the active receptor conformation by stabilizing the open conformation of the ‘ionic lock’ formed by R340 and E600, allowing for G protein binding and activation. Another interesting finding is that the intracellular half of TM3 is also deemed important by Random Forest. After closer inspection by comparing active structures with inactive structures using superposition of the entire receptors, we found that activated GPCRs show an upward movement of TM3, while tightening a little bit.
showing more consistent helical packing and no bending. This is best visible in the extracellular side of TM3 in the A2A-AR structures (2nd from right).

When superposing all bovine rhodopsin structures we see that the inactive rhodopsins show variance in the intracellular loops, while there is almost no distortion in the active conformations. This indicates that there is a strict on/off-switch in rhodopsin, with one active conformation and several inactive states with some basal activity. Also the upward motion and helical tightening of TM3 is clearly visible. Another clearly visible change is the breaking of the ionic lock (fig).

T4 vs. no T4

When comparing GPCR structures that were fused to T4-lysozyme at ICL3 with the ones that weren't fused to T4-lysozyme, we didn't see very significant differences between the classes. They did classify correctly, but there were no variables that really jumped out. Only some minor differences in the intracellular loops, which weren't very systematic.

Agonist vs. Inverse Agonist

Another interesting classification was comparing structures with an agonist bound to structures with an inverse agonist bound. Classification between these classes was performed well by RF. The most important variables involved distances between C-alpha's of the extracellular sides of TM5, TM6 and TM7 (fig). After evaluation of the structures it clearly showed that these halves of the helices moved toward each other close to the binding pocket.

Bulges

While renumbering the structures an interesting observation was made. Renumbering was done by superposing structures that were already renumbered with structures that still needed renumbering. We found that helices differed between species. Some GPCR structures showed short \( \pi \)-helices where other showed regular \( \alpha \)-helices. These bumps in the \( \alpha \)-helical structure are called \( \pi \)-bulges. We used the updated version of DSSP, DSSP 2.0, to analyze the structures and look at which parts of the TMs contain short spans of \( \pi \)-helices. We combined the outputs of the DSSP files and mapped this occurrences of these \( \pi \)-bulges in a table. (DSSP-files, table) For example the structures of the S1P-lipid receptor lack a \( \pi \)-bulge in TM5 near the binding pocket. All other GPCR structures show a bulge between in TM5 in the direction of TM3. This might indicate an alternative entrance for the ligand to enter, because the flexible ligand could also enter through the side of the hydrophobic 7TM bundle while it's located in the cell membrane. All rhodopsin structures have a bulge in the middle of TM2, the opioid lacks this bulge. Another interesting bulge makes a difference between squid and bovine rhodopsin. At the extracellular side of TM2 squid rhodopsin has a bulge, which is absent in bovine rhodopsin. Another quite interesting bulge is the residue after
Tyr733, it is present in all GPCRs so far, except for the β1, β2 and CXCR4 receptors. The missence of this bulge is always accompanied by a Proline at location 808, which may allow TM8 to stay at the same angle as the other GPCRs. We further investigated the occurrence of these bulges by looking at the DSSP files of the crystal structures (results). We found π-helices of differing lengths in every transmembrane helix except for TM6.

Sidechain orientation

When comparing active and inactive structures visually, while focussing on the sidechain orientation of highly conserved residues, we found some irregularities between species. If we compare the orientation of Y528, Y733 and R340, we see that the two Tyrosines hypothesised to stabilize the Arginine of the ionic lock are differently oriented between species. As you can see in the figures below Residues R340, Y528 and Y733 extracted from whole structure alignments, displayed in tube mode. From left to right:

Residues from Rhodopsins aligned, residues from A2A AR structures aligned and residues from β2 ARs aligned. The difference in orientation upon activation of Y528 differs between species. In contrast to the other species Y528 of the A2A AR structures moves outward upon activation instead of inward in the other species.

Comparing helices seperately

After looking at the interhelical distances we looked at what moved intrahelical. WhatIf was used to superpose a large part of each TM and assessed which part of the helix was best soothed to superpose upon. We found some interesting movements in several helices. When we compared active A2A AR structures with inactive ones we saw some trends.

Left to right: TM3-7. Red: active A2A ARs. Cyan: inactive A2A ARs. TM3 seems so straighten its backbone. TM4 has an irregularity at the extracellullar side in 3vg9.pdb. TM5 seems to wind up a little bit. TM6 seems to differ nothing at all. TM7 shows an inward movement of the extracellullar side, which encompasses a larger part of the helix in 3qak.pdb.
Discussion

To elucidate which distances describe the motion of GPCRs upon activation the best we used the variable importance score of Random Forest. It is important to keep in mind that the GPCR structures have a lot of artefacts and probably are a poor representation of what actually happens in nature. There are no structures resolved that don't have artefacts that could have detrimental effects on the structures. To allow them to crystallize the GPCRs were either fused to T4 lysozyme at the intracellular or extracellular side, bound to a large FAb antibody or Nanobody, mutated to improve thermostability or they formed weird anti-parallel dimers. All these modifications were necessary to force these flexible proteins into the rigid environment of crytal packing. We wanted to see what moves upon activation and that's very hard to do when all mobility is lost. We tried to separate facts from fiction and let Random Forest indicate what might be important to look at. We found some trends that could describe the mechanism of activation of GPCRs.

Activation mechanism GPCRs

When comparing active structures to inactive structures and structures with an agonist bound to structures with an inverse agonist bound, we saw some interesting systematic movements. The most important distances upon GPCR activation described the inward motion of Y733, a highly conserved Tyrosine at the intracellular side of TM7. It has been hypothesised that this residue stabilizes the active conformation by interacting with R340, thereby stabilizing the open conformation of the "ionic lock" consisting of R340 and E600. Another important movement was the outward motion of intracellular side of TM6. This last finding can't be seen as a real result, because we used this motion to classify active structures. One surprising group of important distances described the upward motion of TM3, which will be discussed below.

Upward movement of TM3

When comparing active structures with inactive structures, quite an interesting trend was revealed. All active structures showed an upward movement of TM3, while performing a kind of screwing motion. This upward movement differed per species in the sense that at the extracellular side of TM3 they behaved differently. For example in rhodopsin in moves away from the 7TM bundle. It has been hypothesised that this might have something to do with the fact that some GPCRs occur in dimers or oligomers, and this outward movement might induce coactivating GPCRs in these dimers or oligomers to increase the intensity of a signal upon activation. This mechanism, if occurring, might differ greatly between species and that could explain the fact that only the intracellular part of TM3 makes a significant movement upon activation among all GPCRs. This upward might pull the entire (E/D)RY motif upwards, allowing the 'ionic lock' to break more easily.

Clothespin mechanism

We saw that upon agonist binding TM5-7 moved toward each other around the ligand binding site. We have seen that upon activation and G protein binding, the intracellular parts of TM6 moves outward, leaving a crevice on the intracellular side in which the G protein can dock. We also showed that when you superpose TM6 extracted from the different structures, this helix doesn’t have a hingepoint where the bending angle of the helix changes, but that the helix itself stays rigid and it actually rolls outwards. Upon activation TM3 straightens its backbone and obtains a good straight helical packing, while also showing an upward screwing motion. Our idea of how a GPCR gets activated, formed by our observations, is as follows.

A ligand binding GPCR without a ligand bound or with a neutral agonist bound has a basal activity,
It's mostly in the inactive conformation and occasionally in the active conformation, in which it can activate a G protein or another pathway. Upon agonist binding the active conformation is stabilized, this resembles the open form of a clothespin. The activating ligand keeps the extracellular sides of TM5-7 close together and straightens TM3 and making it screw towards the extracellular side. The other helices don't bend at their hinge points, but stay in the same general shape. Like in a clothespin the helices move outward on the intracellular side. Due to the upward movement of TM3, the ionic lock can be more easily broken. Whilst the Arginine flips upward, it's kept there by Y733 and Y528 by cation-π interactions. This way a crevice is created at the intracellular side, allowing for the C-terminus of the α subunit of a G protein to dock into the 7TM bundle. Upon binding to the activated GPCR the G protein exchanges its GDP to a GTP and activates downstream pathways. An inverse agonist stabilizes the inactive conformation by pushing the extracellular side of the TMs outward, not allowing for the crevice on the intracellular side to occur. **Clothespin mechanism.** Left: The clothespin mechanism: Agonist binding (1) induces inward motions (2) of the extracellular side of TM5-7 resembling a clothes pin-like motion. This is followed by an outward movement of the cytosolic side of TM5-7 (3), allowing the G protein to bind (4) and become activated (5). Right: metaphor used to describe GPCR activation.

Further research

It would be very interesting to look into the movement of TM3. Not only the upward movement but also the helical tightness. When you compare TM3 among active and inactive GPCR structures it looks as if the helices tighten up upon activation. An interesting way to compare helical tightness might be by describing the helices as rotation matrices and comparing them to see if there are differences between the direction of the helical axes, between the rotation about the helical axes and between the translation along the helical axes. We've seen that the helical packing of the transmembrane helices can differ a lot between species. Most of the helices aren't perfect 4 helices, there are π-bulges and parts with 3.10 helical packing. It's almost not scientifically sound to call the transmembrane helices, they are more like transmembrane rolls. Especially the tightening of TM3 is intriguing, in the inactive state it's more like a loose roll and in the active state a tight, perfectly straight, helix. I can imagine that a better packed helix might be more rigid, like when one tightly curls up the opening of a plastic bag. Looking at the technique we used, using distances and RF to find important distances which cooperatively describe important motions. It might be good to test this technique on proteins that have their mechanism extensively described. It would also be interesting to not only look at the Co's of the 7TM bundle but also at the distances between side-chains of residues. This increases the amount of variables enormously, but Random Forest should be able to cope with this large amount of data. A difficulty that would arise is how to define the coordinates of the sidechain, because Co's are largely shared among the GPCR family (except for the bulges). The residues on the other hand differ largely of course. It might be wise to only look at the residues that are highly conserved, and therefore are highly important, because these are shared between all structures and the atoms furthest from the backbone can be used.

It would also be an interesting challenge to find trends in the sequences that would allow these bulges to be predicted beforehand. This would be of great benefit for the improvement of homology modelling. Recently the occurrence of π-bulges in TM2 and TM5 have been mapped by Gonzalez et al. They found practically the same things as we did. They also found that both Pro in TM2 and (W/F)x(F/L)G in ECL1 form a common structural feature in the varying extracellular domain of GPCRs. They found a similar structural feature surrounding the bulge in TM5.
showed with superimposition that P520 remains at the same position in space and that a Tyrosine a few turns earlier was also structurally conserved. This Tyrosine has number 507 in the β2-AR, number 508 in the A2A-AR, and number 509 in the S1P1-R. Using this structurally conserved residue a method could be devised that predicts the occurrence of bulges at the extracellular side of TM5 and TM2.

Acknowledgements

References

Agonist-Induced Conformational Changes in Bovine Rhodopsin: Insight into Activation of G-Protein-Coupled Receptors
Supriyo Bhattacharya, Spencer E. Hall and Nagarajan Vaidehi


http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3311957/figure/F1/ (Article)

10. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3311957/figure/F1/ (Article)


24. http://www.bris.ac.uk/synaptic/receptors/mglur/ (Article)


45. G.Vriend. WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8, 52-56 (1990) (Article)


63. Snyder, B. Where are the new drugs?: The push to improve the pipeline. (Article) Picture by William Oldham. (2005)

64. The Scripps Research Institute. GPCR network: Understanding Human GPCR Biology. (Article)


http://www.cmbi.ru.nl/~rkant/GPCR/full_report.html
Thanks

I'd like to take this opportunity to thank the people that helped me to accomplish this research. First of all I'd like to thank Gert Vriend for allowing me to do this internship. I've learned more about doing research, proteins and biology during this internship than I have during the rest of my education. I think it's very cool that you're so down to earth when it come to your students, you don't see a lot of professors caring about their interns like you do. I enjoyed learning about these interesting, important and diverse receptors. It's a bit of a bummer that we got scooped on the π-bulge story, but as you said: "Internships are allowed to fail. It's all about learning." I hope I can get a nice PhD position in the near future and I hope we'll stay in touch.