Investigation of Structure and Allosteric Modulation
of Family C GPCRs
by Sequence-, Structure- and Ligand-based Approaches

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Johann Wolfgang Goethe - Universität
in Frankfurt am Main

von

Swetlana Derksen
aus Orsk, Russland

Frankfurt 2009
(D 30)

vom Fachbereich Biowissenschaften der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan : Prof. Dr. V. Müller

Erster Gutachter : Prof. Dr. Gisbert Schneider

Zweiter Gutachter : Prof. Dr. Manfred Schubert-Zsilavecz

Datum der Disputation : 11 August 2009

Im Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe-Universität wurde diese Arbeit im Zeitraum von Januar 2006 bis November 2008 am Institut für Organische Chemie und Chemische Biologie in der Arbeitsgruppe Chemie- und Bioinformatik (Prof. Dr. Gisbert Schneider) angefertigt.

Danksagung

Als erstes möchte ich meinen Eltern danken, die mir das Leben und mein Studium in Deutschland ermöglicht haben. Sie haben mich zur Selbstständigkeit und Stärke erzogen und damit das wichtigste mitgegeben, um meinen Weg nach eigenen Vorstellungen zu gehen.

Ich danke Prof. Dr. G. Schneider für sein Vertrauen und die Freiheit eigene Ideen bei der Umsetzung meines Projekts verwirklichen zu dürfen. Er hat mir das zielgerichtete, kritische und interdisziplinäre wissenschaftliche Denken vermittelt, mich stets bei meinen Vorhaben unterstützt und zu neuen Zielen ermutigt.

Prof. Dr. T. Weil gilt mein besonderer Dank für die erfolgreiche Zusammenarbeit sowie die Intergation in ihr Team bei Merz Pharmaceuticals. An der Stelle bedanke ich mich bei Merz Pharmaceuticals für die großzügige finanzielle Unterstützung und den Zugang zu technischen Mitteln, die mir die Durchführung meiner Arbeit ermöglicht haben. Die tatkräftige fachliche Unterstützung verdanke ich bei meinem Projekt den Mitarbeitern, insbesondere Björn Krüger, Dr. Mirko Hechenberger und seiner Gruppe sowie Dr. Udo Meyer, Dr. Sibylle Müller und Dr. Lutz Franke. Für die Hilfestellung bei der Einarbeitung in das Themengebiet und für die bereits etablierten Methoden bedanke ich mich bei Dr. Steffen Renner und Dr. Tobias Noeske.

Ich bedanke mich bei Prof. D. Steinhilber und PD. B. Spänkuch für die Bereitschaft zur Teilnahme an der Prüfungskommission.

Ich danke allen Mitgliedern von modlab für ihre Unterstützung, insbesondere Eugen Proschak und Kristina Grabowski für die tolle gemeinsame Zeit und zahlreiche Diskussionen, nicht nur über Projekte sondern auch das Leben eines Doktoranden an sich.

Meine Energie und Freude in dieser Zeit verdanke ich am meisten meinen Freunden, Sina Kazemi, Björn Krüger, Elena Schmidt, Christina Wendel und Benjamin Stauch. Ihr habt alles mitgemacht und zusammen haben wir uns durch schwierige Momente gerettet.

1. Introduction

In therapeutic research, the knowledge of the molecular basis of a given pathology allows to reveal the role of one or more proteins as molecular effectors in a disease process (Wermuth 2008). The protein class of G Protein-Coupled Receptors (GPCRs) plays a major role in diverse signaling pathways. Enabling most physiological processes, GPCRs are suitable biological targets for a wide range of therapeutic approaches (Conn *et al.* 2009). Approximately 25% of marketed drugs act by modulation, activation or inhibition of GPCRs (Overington *et al.* 2006, Jacoby *et al.* 2006). Especially important in the treatment of human central nervous system (CNS) disorders is the possibility to modulate the intensity of neural signal transmission. Agonists, antagonist and allosteric modulators are therefore in the focus of pharmaceutical research. A further objective is the reduction of *off-target effects*, which makes receptor subtype selectivity a key issue in the development of GPCR-binding ligands (Wermuth 2008). The optimization of an active molecule is usually facilitated by knowledge of the structure of the target protein. Since the structures of only a few family A GPCR-ligand-complexes has been determined so far, drug design approaches for this target class rely on already known ligands and biochemical studies. Computational methods allow for additional drug discovery strategies such as ligand-based virtual screening and, by prediction of structural models, provide additional insight into receptor-ligand interactions of GPCRs.

1. Introduction

1.1 GPCRs

1.1.1 GPCR families

GPCRs are the largest family of cell-surface receptors involved in signal transmission. GPCRs recognize molecular signals, mediated by glycoproteins, photons, hormones, neurotransmitters, odorants and pheromones, thereby playing the major role in cell-cell communication (Pin *et al.* 2003, Reggio 2006, Conn *et al.* 2009). The evolution of GPCRs towards interaction with diverse activating molecules increases the variety of signals recognized by the cell and the range of signaling pathways that can be triggered. Based on sequence identity and bound ligands, six GPCR families (named A to F) can be distinguished (Bockaert and Pin 1999, Fredriksson *et al.* 2003); not all have been discovered in humans so far. GPCRs can be classified in (Reggio 2006):

- Family A: rhodopsin-like receptors,
- Family B: protein binding receptors,
- Family C: charged neurotransmitter binding receptors,
- Family D: fungal pheromone binding receptors,
- Family E: cyclic adenosine monophosphate receptors,
- Family F: "frizzled and smoothened" receptors.

The number of human GPCRs is estimated to be 720 to 800, which accounts for about 2% of the human genome, and includes ≈380 unique functional non-olfactory GPCR sequences (Jacoby *et al.* 2006). Of 367 human GPCRs, 284 belong to class A, 50 to class B, 17 to class C and 11 to the "frizzled and smoothened" family (Jacoby *et al.* 2006).

1.1.2 Family C GPCRs

The third GPCR family was named after the metabotropic glutamate receptors (mGluRs). They were identified in the mid 1980s as the first members of family C GPCRs (Sladeczek *et al.* 1985, Sugiyama *et al.* 1987). *Metabotropic* means being involved in metabolic processes and leading to the activation of a "second messenger". Family C is characterized by coupling to phospholipase C (Nicoletti *et al.* 1986) and includes the Ca^{2+}-sensing receptors (CaSR), the γ-aminobutyric acid (GABA)$_B$

2

receptors, bride of sevenless proteins (BOSS), orphan GPCR5 and GPCR6, the pheromone and the taste receptors (Pin *et al.* 2003, GPCRDB June 2006 Horn *et al.* 2003). Eight metabotropic glutamate receptors were discovered and sub-classified into three groups based on sequence similarity, common pharmacology and transduction mechanism (Pin & Acher 2002). Group-I mGluRs comprise the subtypes 1 and 5, Group-II mGluR2 and 3 and Group-III mGluR4, 6, 7 and 8. Family members of mGluR are supposed to exert their modulatory role by influencing neuronal excitability, synaptic transmission and plasticity (Pin & Acher 2002). Group-I receptors are typically localized at the postsynaptic membrane in somatodendritic domains, while Group-II and III are predominantly found presynaptically (Kew and Kemp 2005). The mGluR family has been implicated in the pathology of major neurological disorders such as Alzheimer's and Parkinson's disease, as well as depression, schizophrenia, anxiety and pain (Pin *et al.* 2003, Conn *et al.* 2009).

1.2 Structure and function of GPCRs

In spite of their functional diversity, different GPCRs are supposed to have a common membrane spanning heptahelical domain. The transmembrane domain (TM) domain of GPCRs is the functional unit that is capable of activating G proteins.

1.2.1 Structures of family A GPCRs

Until recently, the three dimensional (3D) structures of only three functionally distinct GPCRs were known (Protein Data Bank [PDB], Westbrook *et al.* 2002): the beta adrenergic receptor, the adenosine receptor and the rhodopsin subfamily, all of which are family A GPCRs. In the following, studies that contributed to the current knowledge of the structure and activation mechanism of GPCRs are reviewed.

Bovine rhodopsin

Rhodopsins are members of the largest GPCR family (A). Family A accounts for 77% of all human GPCRs (Jacoby *et al.* 2006). Each rhodopsin protein consists of approximately 350 amino-acids. Rhodopsins are activated by five protons and trigger the signaling pathway that is responsible for optical vision (Palczewski *et al.* 2000). 11-*cis*-retinal, a full inverse agonist that keeps rhodopsin in its ground inactive state can undergo photoisomerization to an all-*trans*-configuration and thereby trigger the

1. Introduction

formation of metarhodopsin. In 2000, Palczewski and co-workers published the first structure of ground state of bovine rhodopsin (BR) (Palczewski *et al.* 2000, PDB identifier: 1F88, 2.8 Å) followed by metarhodopsin I (Ruprecht *et al.* 2004) and the metarhodopsin II structure in 2006 by Salom and colleagues (Salom *et al.* 2006, PDB identifier: 2I37, 4.15 Å), which is the photoactivated, deprotonated intermediate.

β-Adrenergic receptors

In 2007, a further breakthrough in structure elucidation of GPCRs was achieved with a 2.4 Å crystal structure of the human β_2-adrenergic receptor (β_2AD) in complex with the partial inverse agonist carazolol (Cherezov *et al.* 2007, PDB identifier: 2RH1, 2.4 Å).
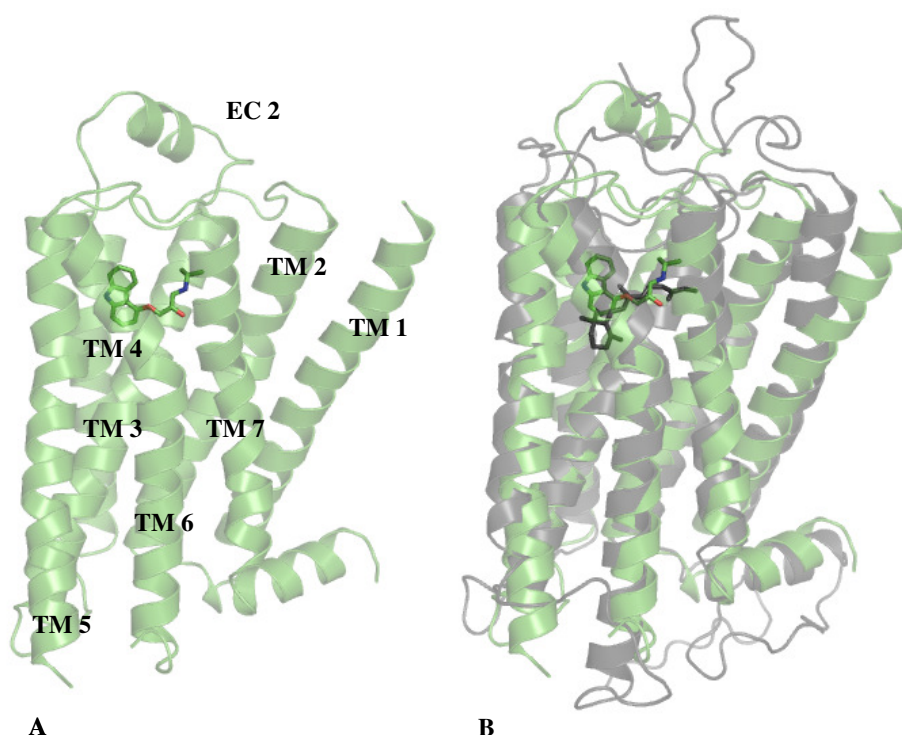
Figure 1: A) Structure of the human β_2-adrenergic receptor (green) in complex with the partial inverse agonist carazolol (Cherezov *et al.* 2007). **B)** Structure of bovine rhodopsin (grey) in complex with retinal (Okada *et al.* 2004). Receptors are shown in cartoon, ligands in stick representation.

In that study, an engineered protein was designed to reduce conformational flexibility. Furthermore, an extensive optimization of crystallization conditions to overcome drawbacks in GPCR structure determination has been performed. As carazolol is an inverse agonist, lowering its constitutive activity was crucial for the stabilization of the resolved conformation. Carazolol, being a diffusible ligand in contrast to the covalently binding retinal, provided new insights regarding receptor ligand interaction in GPCRs (Figure 1, A). Structural differences between the two receptors were striking in the

4

extracellular loop two (EC2) that folds to a β-sheet and tightly occludes the retinal binding pocket, in contrast to the more exposed α-helical part that was observed for $β_2AD$. Superposed receptor structures of BR and $β_2AD$ (Figure 1, B) show shifted or differently kinked helices that lead to adapted binding pockets for distinct ligands, whereas structurally conserved helices remain the framework for functional properties. Despite the low sequence identity of only approximately 22% in TM regions, translated helices, different secondary structures, and constraints from disulfide bonds in loops, the root-mean-square-deviation (RMSD) between $C_α$-atoms of $β_2AD$ receptor (PDB identifier: 2RH1) and BR (PDB identifier: 1GZM) reaches 2.07 Å in the TM domain (Costanzi 2008).

The structure of the β-adrenergic receptor, type one, was solved with bound cyanopindolol (PDB identifier: 2VT4, 2.7 Å, Warne *et al.* 2008) and type two with cholesterol (PDB identifier: 3D4S, 2.8 Å, Hanson *et al.* 2008).

A_{2A} adenosine receptor

In 2008, the structure of the human A_{2A} adenosine receptor in complex with an antagonist has been determined, using stabilizing techniques that were developed for $ß_2AD$. The A_{2A}-complex structure included the antagonist ZM241385 (PDB identifier: 3EML, 2.6 Å, Jaakola *et al*. 2008). In contrast to previous GPCR complexes, where the binding pocket of GPCRs is located parallel to the membrane plane, the A_{2A}-ZM241385 complex surprised by a different orientation and location of the bound ligand (Figure 2).

The extended conformation of the ligand is arranged orthogonal to the membrane plane and forms interactions to TM7 and the EC3 loop (Figure 2, B). The EC2 loop has a coiled structure that is constrained by several disulfide bridges, two of which are unique for the A_{2A} adenosine receptor. Due to the completely different interaction of the antagonist to A_{2A} adenosine receptor compared to ligand binding regions in other determined GPCR structures, the strict assumption regarding a common binding pocket of GPCRs will have to be revisited. The close proximity of the bound ligand conformations to TM3, TM5, TM6, TM7 and the extracellular loops allow for a large variety of possible interactions. The numerous side chains forming these interactions might contribute to the high diversity of recognized molecules known to interact with GPCRs (Wermuth 2008, Conn *et al.* 2009).

1. Introduction
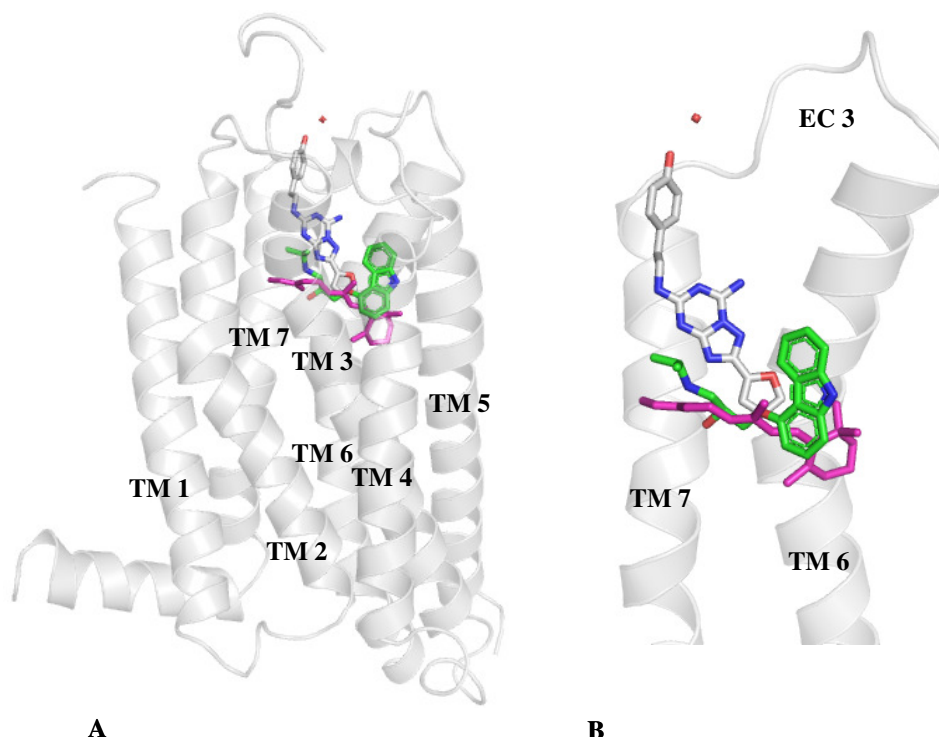
Figure 2: A) Structure of the human A_{2A} adenosine receptor in complex with ZM241385 (grey) (Jaakola *et al.* 2008) and ligand conformations of retinal (magenta, Okada *et al.* 2004) and carazolol (green, Cherezov *et al.* 2007). **B)** Enlarged section of A); all transmembrane (TM) helices besides TM7 and TM6 are omitted for clarity. EC3 means extracellular loop 2.

1.2.2 Activation of family A GPCRs

GPCRs are considered to have flexible structures that can change upon receptor activation and propagate a signal from the extracellular to the intracellular side. For rhodopsin, details on structural changes could be elucidated via site-directed spin labeling (SDSL) experiments (Farrens *et al.* 1996). These experiments involved cross-linking of cysteine residues with nitroxides, followed by electron paramagnetic resonance (EPR) spectroscopy to determine interatomic distances, which provided insight in structural changes that occur during rhodopsin photoactivation. The most pronounced movements in the transmembrane domain upon activation were observed for TM3 and TM6. A comprehensive discussion of observed movements is provided by Gouldson (Gouldson *et al.* 2004), who incorporated experimental data as distance restraints in molecular dynamics studies that aimed to obtain the active and inactive states of rhodopsin and the ß$_2$AD receptors. Solid-state nuclear magnetic resonance (NMR) (Crocker *et al.* 2006) and double electron-electron resonance (DEER) spectroscopy (Altenbach *et al.* 2008) revealed a separation of opposing residues in TM6 and TM3 at the level of the chromophore by an outward rigid-body movement of TM6

6

upon activation, starting from the proline bend region to the intracellular end of the helix.

Similar receptor activation was proposed for all GPCRs because of the conserved residues in the moving helices ("aromatic cluster") as well as residues in G-protein interaction (DRY-motif). Sheerer and colleagues were able to provide the crystal structure of the active state of rhodopsin in its ligand free form (Figure 3), bound to an 11 amino acids long C-terminal part of the G_α-subunit of the trimeric G-Protein-complex (Scheerer *et al.* 2008). In this conformation, guanosine diphosphate (GDP) is assumed to be released (Oldham *et al*. 2006).
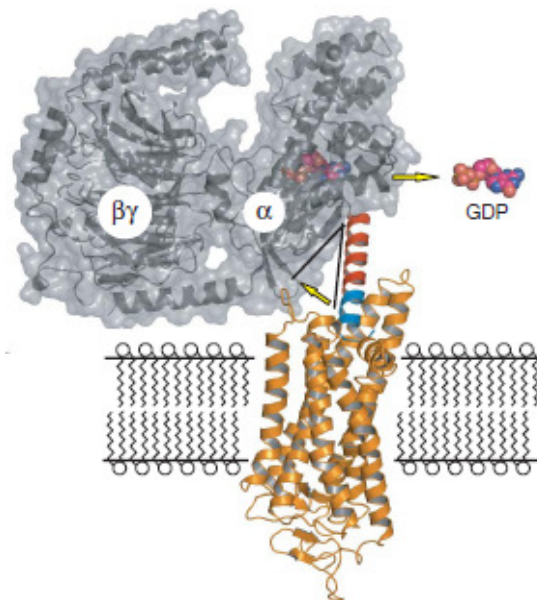
Figure 3: From Scheerer *et al.* 2008, a conceptual model of the crystal structure of the activated rhodopsin in its ligand-free conformation, bound to an 11 amino acids long C-terminal part of the G_α-subunit (synthetic construct). The trimeric G-protein was modeled by superposition (PDB identifier 1GOT; grey) with the synthetic construct (red). GDP means guanosine diphosphate.

In the activated form of rhodopsin, the "ionic lock", which is usually established between Arg135 and Glu134 from the conserved E(D)RY motif in TM3 and the side chains of Glu247 and Thr251 in TM6, is broken. Since the G-protein domain occupies the space between TM3 and TM6, it triggers the outward movement of several helices (Figure 4). The authors assume that local effects caused by G-protein binding also induce long-range stabilization effects that are propagated up to the ligand binding pocket.
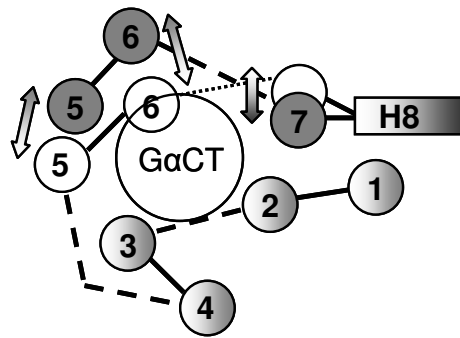
Figure 4: Schematic representation of changes in helical positions in rhodopsin (viewed from the intracellular side) between the conformation binding a G-protein to the intracellular part and the dark-state rhodopsin (adapted from Scheerer *et al.* 2008). The activated conformation is indicated in grey and the inactive in white. The area for the C-terminus of the transducin G_αt subunit (G_αCT) binding is indicated with a circle. Helix eight (H8) is a short α-helix localized orthogonal to the membrane plane and the seven TM helices.

1.2.3 Domain architecture of family C GPCRs

In addition to the TM domain that is common to GPCRs (Figure 5, A), family C receptors have a large extracellular N-terminal domain that resembles the Venus flytrap (VFT) in form and movement. Binding of an agonist triggers the closure of the bilobed structure, similar to the flower eponymous for the domain. The structure of the VFT domain has been solved for mGluR1 in presence and absence of its agonist L-glutamate and the antagonist (S)-α-methyl-4-carboxyphenylglycine, MCPG (Kunishima *et al.* 2000, Tsuchiya *et al.* 2002). Inside the VFT domain, endogenous ligands bind to the so-called orthosteric binding pocket. One of them, L-glutamate, is the major excitatory neurotransmitter in the central nervous system and mediates its actions via activation of both ionotropic and metabotropic receptor families. The VFT is connected to the TM domain by the cysteine-rich domain (CRD) (Figure 5, B).
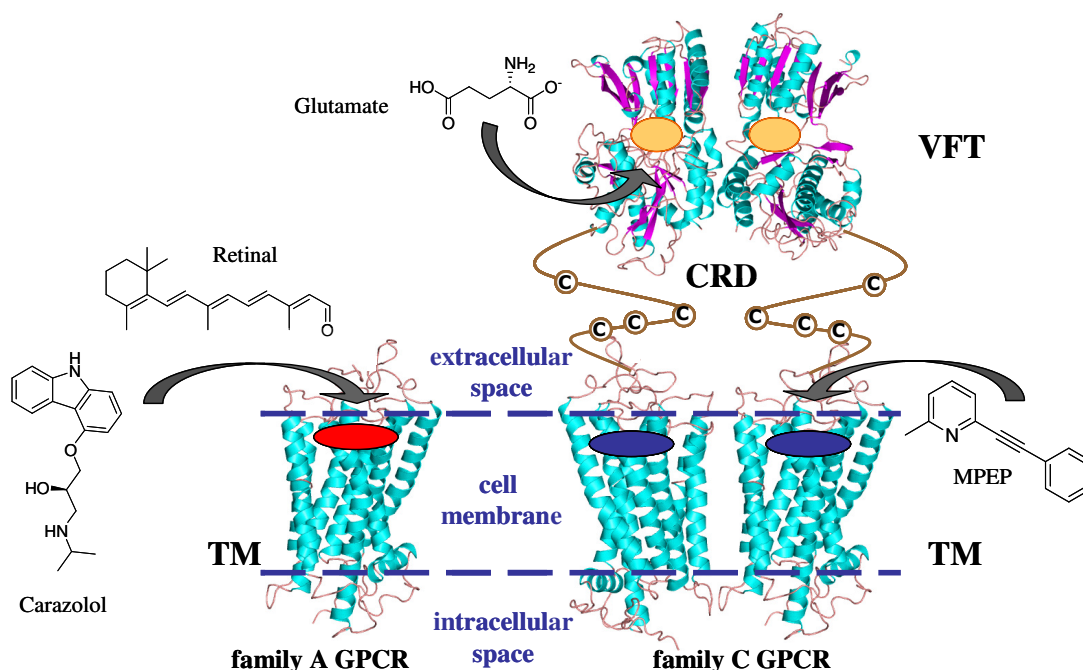
Figure 5: Schematic representation of domains and binding sites of family A and a dimer of family C GPCRs. Crystal structures of the VFT (Kunishima *et al.* 2000) and TM domain (Okada *et al.* 2004) are shown in cartoon representation. The cell membrane is indicated by dashed blue lines. The CRD structure is unknown and therefore presented schematically. Ligand binding sites are indicated by ellipses. Depicted are in red the endogenous family A binding site, in yellow the orthosteric (endogenous) and in blue the allosteric sites of family C GPCRs. Molecule structures of the family A GPCR binding ligands retinal and carazolol, and the family C GPCR binding ligands glutamate and MPEP are linked with arrows to their binding regions. Two family C GPCRs are assembled to a possible dimer (VFT dimer taken from crystal structure).

In contrast to family A GPCRs the allosteric binding site of family C GPCRs is located inside the TM domain. Allosteric modulators which bind to the allosteric site are not competing with orthosteric ligands but can change the protein conformation resulting in changes of orthosteric ligands' affinity or efficacy (Pin *et al.* 2004).

1.3 Allosteric modulation of mGluRs

Allosteric modulation of family C GPCRs is attractive for design of new drug molecules for various CNS diseases because allosteric binding sites are less conserved and thus allow for the development of subtype selective drug molecules. Although competitive agonists and antagonist that target the orthosteric binding site of mGluRs have been successfully designed (Schoepp *et al.* 1999), this remains a challenging task, because glutamate binding residues are conserved across the metabotropic family and most compounds have a poor bioavailability and CNS penetration (Kew 2004). Consequently, the allosteric binding site is considered to have a higher "druggability" than the orthosteric binding site (Kew 2004, Williams and Lindsley 2005, Malherbe *et*

al. 2006). Further advantage of allosteric modulators is that they just "tune" the activity of the receptor, because they do not activate the receptor independent from the orthosteric agonist. Until today, several potent positive and negative non-competitive or allosteric modulators have been discovered for different family C GPCRs (for reviews see Kew 2004, Kew and Kemp 2005, Noeske *et al.* 2005, Williams and Lindsley 2005, Wang and Brownell 2007; and recent studies: Vanejevs *et al.* 2008, de Paulis *et al.* 2006, Ceccarelli *et al.* 2007, Chen *et al.* 2008, Micheli *et al.* 2008).

Allosteric modulators influence the physiologic response to an endogenous agonist, exerting their effect when the latter is presented; they can be classified into negative (NAM) and positive (PAM) allosteric modulators and inverse agonists (Kew 2004, Jacoby *et al.* 2006). The allosteric binding site of family C GPCRs is localized inside the TM region, analogous to family A GPCRs (Figure 5). NAMs reduce the potency or efficacy by an orthosteric agonist, while PAMs enhance it. Under experimental conditions, the identical concentration of the agonist leads to reduced (NAM) or higher (PAM) activation of the receptor in presence of allosteric modulators. Inverse agonists reduce the constitutive activity of a GPCR. One working hypothesis is that different modulators stabilize conformationally distinct forms of the receptor (Gether and Kobilka 1998, Reggio 2006). Lu and colleagues proposed that ligand binding selectivity is determined not only by ligand binding residues, but also by the receptor conformation (Lu *et al.* 2007). They suggest that the intramolecular interaction network of GPCRs has evolved by mutations in order to support the recognition and selectivity to binding ligands.

1.4 Activation and dimerization of family C GPCRs

Dimeric structures of the VFT domains of mGluR1 bound to an antagonist, an agonist, and in the *apo*-form were resolved by X-ray crystallography (Figure 5, B) (Kunishima *et al.* 2000; Tsuchiya *et al.* 2002). A particular change in the arrangement of the domains due to the bound or non-bound ligand could be detected from these structures. The conformational change is proposed to be propagated from the VFT to the TM domain and to stabilize a particular conformation, as it was reported for the tyrosine kinase receptors, the erythropoietin receptor (Livnah *et al.* 1999) or the atrial natriuretic peptide guanylate cyclase receptor (He *et al.* 2001).

The VFT domains of mGluR1 and mGluR4 are known to be secreted as dimers (Selkirk *et al*. 2002). The participation of the C-terminus of mGluR1 splicing type *b* was demonstrated to be essential for dimerization to splicing type *a* receptor by introduction of mutations to that part (Remelli *et al.* 2008). So far, only mGluR1 and CaSR have been reported to form heterodimers. For mGluR1 and mGluR5, no heterodimers could be detected when they were co-expressed in cells (Romano *et al.* 1996, Robbins *et al.* 1999, Gama *et al.* 2001). Other family C members, $GABA_B$ and taste receptors, were thoroughly analyzed regarding heterodimerization and its implication on activation, as reviewed elsewhere (Pin *et al.* 2003).

The TM domain of mGluR5 was analyzed independently to the VFT and was found to be sufficient for functional activation by modulators (Goudet *et al.* 2004). Furthermore, it is the origin for constitutive activity for mGluRs. This finding strengthened the relation of the TM domain of family C to family A GPCRs which bind endogenous ligands in the TM domain. The TM domain is essential for interaction with ligands and participation in dimerization and G-protein coupling through protein-protein interfaces. The presence of shared structural motives like the disulfide bond between the EC2 loop and TM3, the amphipatic TM8 and conserved helix positions, shows similarities between evolutionary distinct GPCR families. These structural relationships imply evolutional neighborhood and are a preposition for structure modeling approaches of family C on the template of crystallized family A GPCR conformations, although the sequence identity is below 20%.

1.5 Protein data - from sequence to structure to function

Timely analysis of the high sequence output generated by genomics projects is a challenge for computational approaches in bioinformatics. The most recent release (v.98 from 12 January 2009) of the EMBL Nucleotide Sequence Database contains 155,114,144 sequence entries that comprise 265,307,725,081 nucleotides (www.ebi.ac.uk). Of these entries, 13,292,297 are of human origin; others belong to divisions as fungi, invertebrates, other mammals, mouse, bacteriophages, plants, prokaryotes, rodents and viruses.

1. Introduction

In the field of protein sequence assignment, data emerge faster than for protein structures (Figure 6 and Figure 7). In 2009, 410,518 protein sequences and 51,535 proteins structures were registered in the Swiss-Prot (v. 56.8, Bairoch *et al*. 2004) database and the PDB (v. Feb.2009, Westbrook *et al.* 2002), respectively. The number of Swiss-Prot entries with one or more pointers to the PDB is 13,713. In order to bridge this gap, computational biology is an area in great demand.
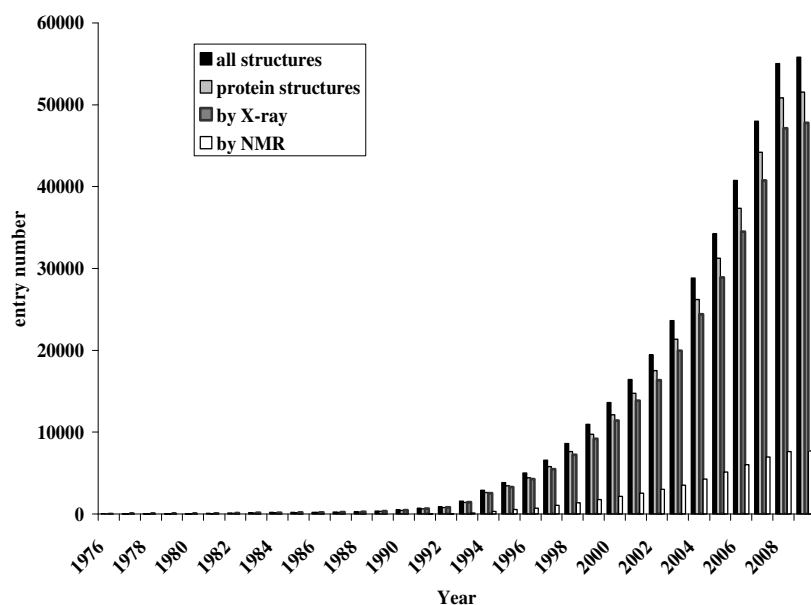
Figure 6: Number of PDB structures per year and those that were determined by X-ray or NMR, of all structures and protein structures only (v. Feb.2009, Westbrook *et al.* 2002).
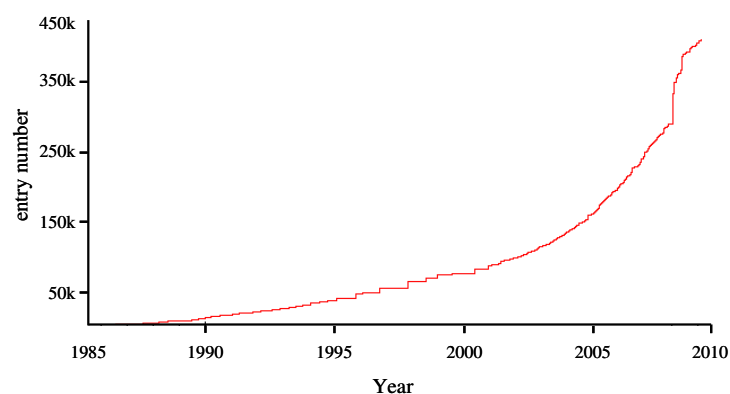
Figure 7: Number of protein sequences registered per year in the Swiss-Prot sequence database (v. 56.8, Bairoch *et al*. 2004). Numbers are shown in thousands (k).

Known protein structures have been assigned to fold- or domain-families based on their tertiary structures. Classification results are stored in public databases, *e.g.* CATH (v. 3.2.0, Cuff *et al.* 2008), which contains 114215 assigned domains, or SCOP (Andreeva *et al.* 2007) with 97178 domains. One of the objectives of the present study

was to find possibilities to interlink data on function and structure of GPCR TM domains to amino acids at particular positions, using homologous proteins and conservation analysis.

Homology between protein molecules can be detected by similarities reflected in sequence, structure, or function (Schneider and Stephens 1990, Murzin 1998, Thornton *et al*. 1999). Prediction of structure and function is routinely supported by multiple sequence alignments as well as by phylogenetic tree reconstruction of protein families (Pei and Grishin 2001). Thus, significant sequence similarity can be expected to be reflected in local structural resemblance in regions of conserved sequence motifs (Grishin 2001). These structural motifs can include domain arrangements or small molecule binding pockets, and determine the biological function of protein families. The sequence-based analysis of conserved regions was particularly focused on throughout the study.

1.6 Conservation analysis of GPCR families

In the evolutionary process, structural conservation and functional divergence are balanced. To date, 198 sequences of family C GPCRs have been determined in different species (GPCRDB, v. 10.0, Horn *et al.* 2003). In order to understand the information that is coded in amino acid sequences of GPCRs and to correlate it to structural and functional features, conservation analyses can be applied. In contrast to procedures that have been reported for family A GPCRs, and that are summarized in the following, the present study used conservation analysis for the characterization of ligand-binding and folding-determinant regions of family C GPCRs, referring to results of the family A studies for comparison.

For 111 human family A GPCRs, Bondensgaard and colleagues found a correlation between conservation patterns of residues in the ligand binding pocket and the privileged structure fragments in class A GPCR ligands (Bondensgaard *et al.* 2004). By docking of ligands with a so-called "privileged structure fragment", they analyzed the common ligand binding pockets of class A receptors. Using entropy-based methods, variable and conserved residues could be distinguished. The variable residues are considered to be responsible for selectivity in ligand recognition, while the conserved

1. Introduction

residues, typically located deeper in the binding pocket, retain a predominantly hydrophobic and aromatic character, and are recognized by different ligand types.

Universal properties of the conformation switch during activation that is present in all GPCRs were analyzed by the "Evolutionary Tracing" method (Madabushi *et al.* 2004). Based on mutagenesis data, residue clusters were defined that are important for ligand binding, G-protein recognition, and a linking core in between those regions. With the Evolutionary Tracing method, sequence position variations could be interpreted by evolutionary divergence. This led to definition of all family A GPCR common functional sites and ligand-specific functional sites, differentiating between the opsin subfamily and other family A members for the latter.

Kratochwil published the so-called ligand binding pocket vector (LPV) that captures conserved patterns across the GPCR family and that was later translated to 3D receptor pharmacophore models, in which each amino acid is represented by the according single spherical pharmacophore feature (Kratochwil *et al.* 2005). The LPV was successfully applied for clustering functional subfamilies and facilitated the creation of a receptor pharmacophore in agreement with positions from mutation studies with MPEP, DFB and EM-TBPC experiments that have been conducted for the metabotropic glutamate receptors. The LPV was calculated using 1000 family A sequences and was used for sequence alignment to other GPCR families.

Shannon entropy calculation is a widely used method to estimate the uncertainty of a variable or the order of a system (Shannon 1948). The Shannon entropy value can be interpreted as the certainty or variability of a position in multiple sequence alignments (Oliveira *et al.* 2002). The frequency of occurrence of an amino acid per position in the multiple sequence alignment follows the rule of conservation of important features (Mirzadegan *et al.* 2003). Pei and Grishin showed that by calculation of entropy-based conservation, determined from the amino acid frequency per position, the identification of conserved positions in multiple sequence alignments is feasible (Pei and Grishin 2001). These conserved positions tend to be functionally or structurally relevant. Later, it was reported that five to nine groups of amino acids cover the largest variance and suffice to find meaningful statistical patterns when applied on multiple sequence alignments of homologous proteins (Wrabl and Grishin 2005).

1.7 Prediction of GPCR 3D structures

1.7.1 Structure modeling

It has been argued that the structure of proteins with similar sequences to already resolved proteins has a higher chance to be determined using similar crystallization conditions (Jaroszewski *et al.* 2008). As a result, "difficult" and "easier" families regarding crystallization have been described by the authors. The four functionally distinct GPCRs with known structure can only represent the beginning of understanding GPCR structures, considering that up to 800 proposed GPCRs occur in the human genome. Therefore, structure modeling of GPCRs is an essential approach to generate and evaluate hypotheses on ligand binding and receptor function, and also for prospective virtual screening and flexibility studies by molecular dynamics simulations.

It has been demonstrated that the native tertiary structure of a given protein is determined solely by the protein's amino acid sequence in a given environment (Anfinsen 1973), and that proteins with similar sequences adopt similar structures (Chothia and Lesk 1986). On the other hand, it has been demonstrated for conserved structural domains that they can originate from sequences with less than 12% sequence identity (Holm and Sander 1996, Hubbard *et al*. 1997, Rost 1999). All GPCRs possess the heptahelical domain, although sequences within each family share only around 20% sequence identity within TM regions, *i.e.*, 20 to 30% for family A receptors to bovine rhodopsin (Jacoby *et al*. 2006).

In absence of experimental structures for mGluR5 or other family C GPCRs, a receptor structure can be predicted in order to provide a context for correlation of conservation profiles and prediction of ligand binding modes. The approximate tertiary structure for a protein can be predicted based on the known 3D structure of a closely related protein family member; one possible approach is homology modeling. Homology modeling is a computational approach that performs the calculation of a protein model in four steps (Schwede *et al.* 2003):

1. Structural template selection (at least one protein with determined 3D structure)
2. Alignment of template and target sequences

3. Model building (protein backbone assignment, placing and energy minimization of side-chain conformations)

4. Model evaluation of protein geometry with respect to sterical hindrance, potential energy (dihedral angles, angles, bond-length and electrostatic interactions)

In case where sequence identity is low or several templates are available, this procedure requires parameter optimization in the alignment and refinement steps.

1.7.2 Modeling of the transmembrane domain of GPCRs

Taking advantage of sequence identity between template and target structures, homology modeling methods exploit the fact that evolutionary related proteins share a similar structure. The question of whether the rhodopsin structure constitutes an acceptable template for comparative modeling of other family A GPCRs has been thoroughly investigated (Bissantz *et al.* 2003), concluding that rhodopsin is a suitable template, at least for antagonist studies. Numerous studies demonstrated the applicability of comparatively modeled structures of GPCRs for rational drug design, among them α1AD, β2AD, β1AD, 5HT2c, dopamine D3, muscarinic M1, vasopressin V1a, tachykinin binding NK1, and cannabinoid CB1 receptors (Klabunde and Hessler 2002, Evers and Klebe 2004, Bissantz *et al.* 2005, Evers and Klabunde 2005, Salo *et al.* 2004).

Recently, the accuracy of a GPCR structure that was predicted with homology modeling could be evaluated using the X-ray structure of β_2AD as target and rhodopsin (PDB identifier: 1GZM) as template (Costanzi 2008). The predicted structure reached an RMSD of 2.04 Å for backbone atoms in TM regions. Furthermore, the author reported the high impact of the EC2 loop structure and the orientation of essential bulky side chains in the binding site for the success of carazolol docking into the modeled structure.

The quality of a predicted structure depends on sequence similarity to the template structure. Close to optimal sequence alignments can only be obtained for closely related protein sequences with identities over 40% (Sanchez and Šali 1997). As sequence similarity decreases, the alignment becomes more uncertain and is likely to contain an increasingly large number of gaps and alignment errors (Rost 1999, Marti-Renom *et al.* 2000, Elofsson 2002). For distant homologues with a sequence identity below 30%, *i.e.*,

in the "twilight zone", the resulting structure models cannot be expected to have a high precision in side chain orientations. The structure model accuracy that can be achieved is comparable to a structure resolution greater than 3.5 Å but is regarded sufficient for identification of conserved patches of surface residues (Marti-Renom *et al*. 2002).

Further known problems in GPCR modeling arise from different helical structures of template and target helices defined by the composing residues, which can form π- and 3_{10}-helical regions or introduce helix kinks (Yohannan *et al*. 2004). These kinks force a helix to bend at a specific position, which changes the helix flexibility (Altenbach *et al*. 2008). When working with homology modeled structures, the introduced deviations in helical regions have to be considered (detailed discussion by Reggio 2006). Besides reasonable suspiciousness due to low accuracy and artifacts, modeled structures can be used for analysis of conserved chemical feature arrangement in a 3D context, providing a basis for comparison to known structures, as well as for understanding experimental findings. There are still a lot of open questions about GPCRs in respect of features that are essential for function, diversity and selective ligand interaction, thereby causing different modulations.

1.8 Virtual screening

High throughput screening is a common approach that allows to extract new active molecules out of a chemical library (Böhm and Schneider 2000, Bajorath 2002, Klebe 2006). In order to reduce costs, which are increasing with the size of screening libraries, virtual screening can be carried out for selection of fewer molecules for experimental screening. Essential steps for a virtual screening approach consist of:

1. Library preparation and filtering
2. Encoding of the molecules with a descriptor
3. Application of an "*in silico*" model for molecule evaluation
4. Scoring the molecules in the chemical library
5. Selection of samples with a high model score

The preparation step commonly includes the removal of molecules with undesired properties, the so-called "non-druglike" molecules (Sadowski and Kubinyi 1998) and

1. Introduction

"frequent hitters" (Schneider and Böhm 2002), as well as an application of absorption-distribution-metabolism-excretion-toxicity (ADMET) filters (van de Waterbeemd and Gifford 2003). "Drug-like" molecules are supposed to fulfill the criteria defined in Lipinski's "Rule of Five" (Lipinski *et al.* 1997). These rules are guidelines derived from the World Drug Index; they suggest that "drug-like" molecules have a molecular mass < 500 daltons, a calculated octanol/water partition coefficient $<$ five, less than five hydrogen-bond donors, and < 10 hydrogen-bond acceptors. "Frequent hitters" are molecules which bind unspecifically to different proteins or perturb the assay in another way. By ADMET prediction, the behavior of molecules in aqueous solutions, membrane permeation, metabolic clearance, reactivity and toxicity can be considered at an early step in the study (van de Waterbeemd and Gifford 2003). The more molecules can be excluded by these filters the more computationally demanding algorithms can be used for molecule evaluation afterwards.

An evaluation model can be used for the prediction of structure activity relationships by classification, probability or activity estimation, and allows for a ranking of data samples (molecules) according to their model score. Machine learning approaches can be used to develop structure-activity models for the molecule evaluation step. Among the most prominent techniques are artificial neural networks (Schneider and Wrede 1998, Noeske *et al.* 2006), support vector machines (Vapnik 1998, Byvatov and Schneider 2003) and evolutionary strategies (Rechenberg 1973, Schneider and Fechner 2005). These techniques establish a correlation of chemical data that provides a generalization or classification model. In the present study, supervised and unsupervised learning techniques, represented by Bayesian classifier (Duda *et al.* 2001) and self-organizing maps (Kohonen 1982) were applied. These techniques depend on molecules with known activity to predict previously unknown activity for molecules in a chemical library. A general measure of success is the ability of a model to enrich the number of active molecules ("hits") in a screening library.

Computationally demanding methods like docking (Alonso *et al.* 2006, Yanamala *et al.* 2008), free energy perturbation calculations (Kollmann 1993, Alonso *et al.* 2006), or application of other 3D ligand- and receptor-structure-dependent methodologies that account for conformational flexibility can only be applied on a focused set of molecules (Carlson and McCammon 2000). In virtual screening campaigns, several molecule

18

evaluation procedures can be carried out successively, decreasing the number of molecules in each run, up to a limit that is feasible for biochemical testing (Walter *et al.* 1998). The combined application of similarity search and molecule docking and scoring have been reported as feasible for virtual screening for GPCR ligands (Schneider and Böhm 2002).

1.9 Pharmacophores and molecular similarity

The concept of a *pharmacophore*, as initially formulated by Paul Ehrlich (Ehrlich 1904), defines "a composition of properties that make a molecule a drug". Its first applications in molecular modeling approaches were reported by Monty Kier in a series of papers between 1967 and 1971 (Van Drie 2007). The IUPAC (International Union of Pure and Applied Chemistry) gives a more precise definition, *i.e.*, that "a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response".

Methods that incorporate small molecules and define receptor interactions from the perspective of the ligand are referred to as *ligand-based* approaches. The missing 3D structures of macromolecules, especially GPCRs or ion channels, enforce the use of ligand-based approaches, including pharmacophores, for rational design and optimization of novel bioactive molecules. A pharmacophore model derived from known ligands can guide a strategy for discovery of active molecules. However, it should be emphasized that every pharmacophore model always defines only one specific binding mode.

The pharmacophore concept can be employed in pure ligand-based methods, as 3D quantitative structure activity relationship (3D-QSAR) (Mason *et al.* 2001), pharmacophore-based search (Sheridan *et al.* 1989) and molecular descriptors (Böhm and Schneider 2000), but also combined with structure-based design, *e.g.* pseudo-receptor models (Tanrikulu and Schneider 2008). The advantage of 2D molecule graph-based pharmacophore descriptors is that no 3D structure of the molecules is required and similarity calculations are alignment-free (pharmacophore encoding descriptors will be introduced in Chapter Molecular descriptors). The ligand-based approach can be

1. Introduction

straightforward, if a sufficient number of active analogues have already been discovered. These active molecules allow for comparison to other molecules, using their pharmacophore points. Pharmacophore types include positively or negatively charged, hydrogen bond donor or acceptor, hydrophobic, aliphatic and aromatic features. Pharmacophore points can be defined for atoms or chemical groups possibly involved in interactions with a receptor. They reference the entire set of atoms that can support the same type of interaction with the biological molecule. Since pharmacophores can be used to describe physicochemical properties of a molecule, they provide a possibility to compare molecules.

In computational approaches, various chemical structures can be evaluated in an automated manner by similarity calculation to reference molecules. Similarity search is a virtual screening approach in which molecules with unknown activity are classified according to the level of their similarity to known active molecules (Willett 1998). In that way, molecular similarity allows for the selection of probably active molecules from a pool of molecules for further biochemical testing. This strategy of discovering novel active molecules by similarity search is based on the assumption that molecules with similar structure are more likely to have similar properties (Johnson and Maggiora 1990, Brown and Martin 1997, Martin *et al.* 2002).

The combination of similarity search with pharmacophore features directs the similarity criteria towards molecules with similar interaction properties. In addition, the search is not limited to molecules with identical molecular structure, allowing to find molecules with similar pharmacophore features but a different molecule scaffold. The discovery of alternative scaffolds by *scaffold-hopping* (Schneider *et al.* 1999) provides new lead structures for further optimization (Renner *et al.* 2005).

1.10 Molecular descriptors

Molecular descriptors are used to represent physicochemical properties and biological activities. Numerical molecular descriptors allow for the generation of quantitative structure activity models. The calculation of molecular descriptors does not depend on empirically determined measurements and can therefore be performed sufficiently fast even for large molecule libraries. Examples of molecular properties that can be encoded in descriptors are the molecular mass, the polar surface area of the atoms' 3D surface, the 2D topology of the molecule, substructures or the distribution of atom types. Different molecular descriptors have been developed for virtual screening approaches (Todeschini and Consonni 2000).

One way to calculate a numerical descriptor from the molecular structure is to use the molecular graph. The molecular graph representation reduces a molecule to non-hydrogen atoms; atoms are represented by vertices and bonds are represented by edges (Balaban 1976). A molecular graph includes information about the number of bonds, bond types, molecule size and branching. These properties, as well as physicochemical properties of atoms can be incorporated as counts, binary fingerprints or combined into topological feature descriptors. Similar to graph-theoretical approaches, atoms can then be considered as starting or end points of paths in a molecular graph. Atomic features can be integrated into topological description when considered as pairs of atom types, as it was introduced in the pharmacophore concept. A binning scheme allows for subdivision of such atom combinations into discrete groups, according to the path length between the atoms.

The present study focuses on the application of topological descriptors in virtual screening, especially on the Chemically Advanced Template Search (CATS) (Schneider *et al*. 1999). The CATS descriptor is a correlation-based descriptor that encodes the frequency of atom type pairs on the molecular graph. The concept of autocorrelation for topological structure was first introduced by Broto and colleagues (Broto *et al*. 1984).

For comparison of molecules based on 3D descriptors, a 3D conformation of each molecule is required. 3D descriptors accounting for the molecular size and shape are calculated directly from the cartesian coordinates of molecule atoms and other quantities derived from the coordinates. Depending on the number of rotational bonds,

1. Introduction

several 3D conformations of one molecule are possible. While the binding conformation of a molecule has a defined 3D interaction to the receptor, which could contribute to the molecular descriptor, the correct prediction of this conformation would require the consideration of the native protein and its solution interactions, which influence the conformational freedom of the ligand.

As no single descriptor encodes all relevant information for the establishment of a predictive model, it is beneficial to use different descriptors (Sheridan and Kearsley 2002).

1.11 Ligand binding mode prediction

1.11.1 Ligand-receptor interaction

The first concept to explain drug function was a comparison to the key-and-lock mechanism introduced by Emil Fischer (Fischer 1894). He proposed that a drug and its receptor (Fischer considered enzymes) sterically fit into each other and the drug has to be bound in order to influence the function of the protein. Linus Pauling deduced from enzymatic reactions that the flexibility of the ligand allows for an amplification of the binding during the reaction by flexible adaptation (Pauling 1946).

A flexible adaptation of the ligand goes along with an optimization of ligand-receptor interaction in the binding pocket and influences the solvent. The binding site can be differently buried and is defined by amino acids and their chemical properties. Depending on the type of amino acid, it can be involved in hydrogen bonds, ionic, hydrophobic or cation-π interactions with the ligand or stabilize a metal ion. Ligands complement some of the receptor's interaction points and partially fill out the binding pocket. A binding mode is a particular set of interactions between defined atoms of ligand and receptor side chains. Binding affinity arises from an entropical and an enthalpical part, both of which can dominate receptor-ligand interactions. The Gibbs-Helmholtz-Equation defines the change in the free energy (ΔG) of a system at constant pressure (F I), depending on temperature.

$$\Delta G = \Delta H - T \cdot \Delta S , \qquad\qquad \textbf{F I}$$

where ΔH is the change in enthalpy, ΔS the change in entropy and T the temperature. The free energy is positive when the reaction is endothermic, and negative in the exothermic case. An increase in entropy lowers ΔG and increases the binding affinity. This entropical contribution results mainly from water replacement in hydrophobic regions of the binding site and flexibility constraints of the ligand, and the enthalpic contribution from building of new interactions (Connelly *et al.* 1994). Therefore the binding affinity of a ligand corresponds to the changes in free energy upon binding.

Beyond the flexible conformational adaptation of the ligand goes the induced fit theory proposed by Koshland, who states that the substrate is necessary to promote the proper orientation of catalytic groups (Koshland 1958). Most recent theory regarding the ligand binding process is influenced by increasing evidence that ligands can stabilize different active receptor conformations (Cozzini *et al.* 2008), referred to as *ligand-induced selective signaling* (LISS, by Lu *et al.* 2007) or *conformational selection*. Lu and colleagues suggested that receptor conformations play an important role in determining the binding selectivity of ligands in the human GnRH receptor (Lu *et al.* 2007).

According to *conformational selection*, a single binding site can exhibit different binding profiles to similarly affine ligands. Additionally, ligands that can bind in different binding modes were identified (McCammon 2005, Boström *et al.* 2006). The exploration of diverse protein-ligand-complexes and their ligand-free forms revealed that 75% of all intra-protein hydrogen bonds and 50 to 80% of all water-mediated intra-protein hydrogen bonds in binding pockets are preserved upon ligand binding (Arora 2005). These findings indicate that receptor flexibility and interactions with the solvent are factors that need to be accounted for in ligand binding prediction.

1.11.2 Prediction of ligand binding modes

A common computational approach to predict the binding conformation of a ligand to a receptor is the molecular docking procedure, as applied in the present study. In general, the prediction of a receptor ligand bound conformation includes the exploration of ligand flexibility and optimization of the best fit into the binding pocket, evaluated with a scoring function. A wide variety of scoring methods as well as docking algorithms were evaluated for diverse protein-ligand-complexes (Wang *et al.* 2003, Warren *et al.* 2006, Taylor *et al.* 2002, Leach *et al.* 2006). The results showed that docking methods

are able to identify the crystallographically determined conformation and recognize active molecules from a pool of decoys, but not for each of the tested receptor-ligand complexes. From the scoring performance the authors concluded that the available scoring functions cannot estimate ligands' affinities. In addition, insufficient target structure resolution and difficulties in considering water-mediated interactions that have an impact on ΔG increases the complexity of the prediction of the native binding mode (Leach *et al.* 2006).

For treatment of flexible receptors several approaches have been proposed. Common among them are molecular dynamics, flexible docking and the employment of rotamer libraries (Leach 1994) or protein ensemble grids (Knegtel *et al.* 1997). These methods allow for consideration of more than one possible conformation of the protein, thereby increasing the chance to find the binding conformation of the receptor. In addition, it avoids optimization of the flexible ligand towards a non-native binding mode for its respective class of ligands.

1.11.3 Ligand binding modes for GPCRs

Of particular interest for the present study was the ligand binding mode prediction for modeled structures of GPCRs. GPCRs were proposed to exist as a conformation ensemble that is influenced by ligands, the membrane and interacting proteins (Fanelli and De Benedetti 2005, Lu *et al.* 2007). Since interaction of ligands with GPCRs is only known for few complexes resolved so far, the definition of possible binding modes of GPCR ligands depends on molecular docking using modeled receptor structures. Costanzi used the recently resolved ß2AD-receptor-ligand-complex and reported differences in side chain orientations between the modeled and the experimental receptor conformation (Costanzi 2008). These side chain conformations have high impact on reproducibility of the ligand conformation using docking. Previously, different GPCR binding sites were explored with docking techniques for binding mode prediction for family A (Bissantz *et al.* 2003, Shacham *et al.* 2004, Evers and Klabunde 2005, Costanzi 2008) and family C GPCRs (Malherbe *et al.* 2006, Vanejevs *et al.* 2008, Yanamala *et al.* 2008). However, no structural evidence of their correctness could be provided so far. The selection of binding poses has been based on possible interaction points defined by mutation studies, as well as enrichment rates for discrimination of active molecules from non-binding ones (Bissantz *et al.* 2003, Yanamala *et al.* 2008).

The determined family A GPCR ligand complexes revealed differences between their binding sites due to extracellular loops which cover a large part of the binding site (Figure 1). In modeled family C GPCR receptor structures even higher structural uncertainties (only family A GPCRs are available as templates) can be expected than for models of family A receptors. This structure prediction problem is amplified by the possible flexibility according to a conformation ensemble of a GPCR. Ligand binding mode prediction for GPCRs has to deal with high degree of conformational freedom in the thermodynamic process of molecule interactions.

Scope of the thesis

This study focuses on structural features of a particular GPCR type, the family C GPCRs. Structure- and ligand-based approaches were adopted for prediction of novel mGluR5 binding ligand and their binding modes.

The objectives of this study were:

1. An analysis of function and structural implication of amino acids in the TM region of family C GPCRs.
2. The prediction of the TM domain structure of mGluR5.
3. The discovery of novel selective allosteric modulators of mGluR5 by virtual screening.
4. The prediction of a ligand binding mode for the allosteric binding site in mGluR5.

GPCRs are a super-family of structurally related proteins although their primary amino acid sequence can be diverse. Using sequence information a conservation analysis of family C GPCRs should be applied to reveal characteristic differences and similarities with respect function, folding and ligand binding. Using experimental data and conservation analysis the allosteric binding site of mGluR5 should be characterized regarding NAM and PAM and selective ligand binding. For further evaluation experimental knowledge about family A GPCRs as well as conservation between vertebrate rhodopsins was planned to be compared to results obtained for family C GPCRs **(Section 4.1 Conservation analysis of family C GPCRs).**

Since no receptor structure is available for any family C GPCR, discussion of conserved sequence positions between family A and C GPCRs requires the prediction of a receptor structure for mGluR5 using a family A receptor as template. In order to predict the mGluR5 structure a sequence alignment to a GPCR template protein will have to be proposed and GPCR specific features considered in structure calculation **(Section 4.1.4 Structure prediction of mGluR5)**. The obtained structure was intended to be involved in ligand binding mode prediction of newly discovered active molecules.

For discovery of novel selective mGluR modulators several ligand-based virtual screening protocols were adapted and evaluated. Prediction models were derived for selection of possibly active molecules using a diverse collection of known mGluR binding ligands. For that purpose a data collection of known mGluR binding ligands should be established and this reference collection analyzed with respect to different ligand activity classes, NAM or PAM and selective modulators. The prediction of novel NAMs and PAMs using several combinations of 2D-, 3D-, pharmacophore or molecule shape encoding methods with machine learning techniques and similarity determining methods should be tested in a prospective manner **(Section 4.2 Virtual screening for novel mGluR modulators)**. In collaboration with Merz Pharmaceuticals (Merz GmbH & Co. KGaA, Frankfurt am Main, Germany) the modulating effect of a few hundred molecules should be approved in a functional cell-based assay.

With the objective to predict a binding mode of the discovered active molecules, molecule docking should be applied using the allosteric binding site of the modeled mGluR5 structure **(Section 4.2.4 Modeling of binding modes)**. Predicted ligand binding modes are to be correlated to conservation profiles that had resulted from the sequence-based entropy analysis and information from mutation experiments, and shall be compared to known ligand binding poses from crystal structures of family A GPCRs.

2. Data

2. Data

2.1 GPCR family C protein sequences

Protein sequences of family C GPCRs were retrieved from the "information system for G protein-coupled receptors" (GPCRDB - Release 10.0, Horn *et al.* 2003). 160 sequences of different family C GPCRs were compiled: 53 metabotropic glutamate (mGluR), 24 calcium-sensing like (CaSR), 30 GABA-B (GABR), 11 orphan GPRC5, nine orphan GPCR6, four bride of sevenless proteins (BOSS), 12 taste (TR) and 17 putative pheromone receptors. Several filtering steps were carried out on the data set to remove inappropriate data; the remaining sequence numbers are given in Section 4.4.2, Table 2. Sequence filtering steps:

1. all sequences tagged as "variant", "hypothetical", "similar", "related", "probable", "splice" and "putative" were removed.
2. duplicate entries were removed

2.2 Multiple sequence alignments from GPCRDB

MSAs in the GPCRDB are automatically generated using an iterative profile alignment method (Oliveira *et al.* 1993). Two multiple sequence alignments (MSAs) were retrieved from GPCRDB (GPCRDB - Release 10.0, Horn *et al.* 2003):

1. MSA of 96 family C GPCR sequences.
2. MSA of 491 sequences of the "vertebrate rhodopsin subfamily" (family A GPCRs).

The same sequences as in Section 2.1 (these were not aligned) were retrieved as a MSA from GPCRDB. The MSA from GPCRDB should be compared to another MSA which is proposed in this study.

2.3 Family C mutation data collection

Available information on mutations in family C GPCRs from published scientific journal articles (Muehlemann *et al.* 2006, Petrel *et al.* 2003, Petrel *et al.* 2004, Schaffhauser *et al.* 2003, Xu *et al.* 2004, Hu *et al.* 2002, Knoflach *et al.* 2001, Hu *et al.*

28

2005, Litching *et al.* 1999, Winning *et al.* 2005, Malherbe *et al.* 2003, Malherbe *et al.* 2006, Jiang *et al.* 2004, Malherbe *et al.* 2003, Pagano *et al.* 2000, Miedlich *et al.* 2004) were collected and standardized to describe family C mutations in a comparable way. The accumulated data was structured as described below.

Receptor:	receptor type
Residue:	residue number
BW_position:	sequence position according to the Ballesteros-Weinstein numbering, numbering for all positions are based on the MSA, which was suggested in this study
TM/Loop:	TM or loop number
Modulator:	allosteric modulator, which was tested for affinity or effect changes using the mutated receptor
Effect:	effect on affinity or modulation caused by the mutation
Reference:	citation of the original publication

The mutation data collection is applied throughout this study in discussion of positions or residues which were sensitive to a particular mutation. It is further used for binding site definition. All mutated positions are given in Appendix, Section 10.8 (Table A 3).

2.4 Family C ligand data collection

2.4.1 Literature ligand data

A data set was compiled based on ligands published in literature, including experimentally determined activity values. 1240 mGluR binding molecules were retrieved. For each ligand the biological effect as well as the receptor subtype tested in the experiments were noted. 490 of 1240 molecules were collected from literature and patent data bases by Dr. T. Noeske before.

Properties included in the collection:

molecule structure:	2D molecule structure
molecule name:	molecule identifier, containing the authors name and ID used in the paper
ligand type:	negative (NAM) or positive (PAM) allosteric modulator

2. Data

receptor_X:	receptor subtype definition, *e.g.* mGluR5
binding_IC_{50}_X:	activity value, IC_{50}, from a binding assay given in nM, tested on the receptor subtype
functional_IC_{50}_X:	activity value, IC_{50}, from a functional assay given in nM, tested on the receptor subtype receptor_X
pIC_{50}:	negative decadic logarithm (power 10) of IC_{50}
organism:	test organism or receptor type origin (mouse, rat or human)
selectivity:	tag for indication if the ligand is subtype-selective
reference:	citation of the original publication

The placeholder X refers to the group of results associated to one of tested receptors subtypes. The collection contained an activity molecule pair for different combinations of the tested receptor, ligand and assay type. The complete ligand data collection can be found in Section 10.4 (Table A 1).

2.4.2 Ligand data

Additional ligand and activity data was provided by Merz Pharmaceuticals (Frankfurt am Main, Germany). Merz '*in house*' mGluR binding compounds were selected and values converted to the literature data set format. Selectivity was calculated from activity values determined in experimental mGluR1 and mGluR5 assays. A compound was considered as selective if the activity difference equaled factor 10. This guideline was considered to be sufficiently high to discriminate selective from nonselective molecules. Merz compounds were further used in virtual screening with shape similarity search (Section 3.11.1) and FCFP/PHRFP similarity calculations (Section 3.6.1); these applications were performed at Merz Pharmaceuticals.

2.5 WOMBAT ligand data collection

WOMBAT is a drug data collection distributed by Sunset Molecular Discovery (LLC Santa Fe, USA). WOMBAT 2007.2 contains 203,924 entries (178,210 unique SMILES), totaling 416,405 biological activities on 1,820 unique targets (Olah *et al.* 2004). WOMBAT 2007.2 was compiled from 9,227 published papers from fourteen journals in the medicinal chemistry field between 1975 and 2007. 32% of WOMBAT ligands are GPCR binding molecules. In the present study the WOMBAT drug data

30

collection (licensed by Merz, Frankfurt am Main, Germany) was used for selection of molecules binding to targets other than mGluR.

2.6 Screening molecule libraries

2.6.1 Data sets

52 different vendor molecule data sets were used for virtual screening with machine learning and similarity search techniques. Table A 2 includes websites and versions for all applied vendor molecule collections, the collection in SD-format was provided by Merz (Merz Pharmaceuticals, Frankfurt am Main, Germany). For 3D similarity calculations the molecule conformation data bases were prepared with Phase (v.2.5, Schrödinger, LLC, New York, 2008) by Björn Krüger. The data set preparation steps for different virtual screening applications are described in detail in Section 2.6.2. The number of unique commercially available molecules was 5,124,879.

2.6.2 Ligand data preparation

Ligand data collections of various origins were used in virtual screening and molecular docking procedures. Two software packages, Pipeline Pilot (SciTegic, San Diego, USA) and Schrödinger (Schrödinger, LLC, New York, 2008), provide modules for molecule data preparation. Schrödinger software and Pipeline Pilot were licensed by and exclusively used at Merz Pharmaceuticals (Frankfurt am Main, Germany). In order to standardize the molecule atom types, all molecules involved in same procedure were prepared the same protocol. All vendor molecules were processed through a substructure filter which removed all molecules containing any of the chemical groups proposed by Hann (Hann *et al.* 1999).

Ligand preparation for Pipeline Pilot

Molecule standardization was employed using Pipeline Pilot procedures. The process comprised standardizing the stereo chemistry markings and formal charges and removal of all additional molecules from each entry besides the largest. All stereo centers were set to "unknown" if the marking was absent or checked for validity and retained. Standard formal charges were applied to common functional groups (Figure 8).

2. Data
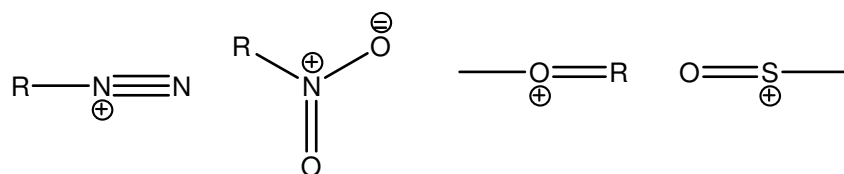
Figure 8: Functional groups with respective formal charges. R denotes the residual molecule part.

Subsequently, bases were deprotonated and acids protonated, setting charges of the functional non-hydrogen atoms involved to zero.

Ligprep module

The ligand preparation procedure Ligprep (v2.0 distributed by Schrödiger, Mannheim, Germany) was used to generate 3D molecule structures, including stereo- and protomers. The preparation includes the following actions:

1. Converting structures from 2D to 3D
2. Removing of counter ions and water molecules
3. molecule protonation at pH=5-9
4. Generation of stereoisomers
5. Performing of an energy minimization

The energy minimization is performed using OPLS force-field (Jorgensen *et al.* 1996) with default settings: "rapid search", "distance-dependent dielectric solvation model" and "no post-minimization iterations".

In the present study identical parameters were applied for all Ligprep calculations as follows:

```
para_ligprep -epik -W e,-ph,7.0,-pht,2.0 -s 32 -r 1 -bff 14 -isd
input.sdf -omae output.mae
```

Phase module

In order to perform similarity calculations for 3D structures, different conformers of each molecule were computed and stored in a data base (Phase v2.5, Schrödinger, LLC, New York, 2008). The preparation includes the following actions:

1. Generation of 3D conformations with Ligprep
2. Data base creation in Phase format
3. Multiple conformer calculation

32

In step 3. ligand conformations were generated with torsional search. Therefore the molecule is divided into core and periphery. The peripheral groups are defined to have only one rotatable bond between the terminal groups and the rest of the molecule. All rotatable bonds besides the peripheral are assigned to the core. The conformational search procedure generates all core configurations and then varies the peripheral configurations one-by-one.

Parameter for data base creation:

```
phasedb_manage -db db_name -new -mae input.mae -confs false -JOB
conf_generate
```

Parameter for calculation of multiple conformers per ligand:

```
phasedb_confsites -confs all -JOB auto_confs -db db_name
```

3. Methods

3.1 Multiple sequence alignment

ClustalW (version as described in Fukami-Kobayashi and Saito 2002) was applied for the construction of MSA (Thompson *et al*. 1994). The algorithms works as follows: First, the complete distance matrix of all pair-wise sequence distances is calculated. Based on these distances a phylogenetic tree is generated by Neighbor-Joining (Saitou and Nei 1987). The branch lengths include information about the assumed "evolutionary" distance of two sequences. The progressive alignment technique expands the MSA gradually by performing pair-wise alignment of groups of sequences according to the branching order of the phylogenetic tree. Two existing MSAs are aligned using profile alignment. ClustalW was applied with Blosum62 (Henikoff and Henikoff 1992) scoring matrix, *gap open penalty* = 7 and *gap extension penalty* = 1. Application details of ClustalW for a MSA of family C GPCRs is described in the next section.

3.2 Multiple sequence alignment of family C GPCR protein sequences

A MSA of 96 family C GPCR protein sequences was performed in several steps. First, all subfamilies (with the exception of putative pheromone receptors) of family C receptors were aligned independently from each other; this was accomplished using ClustalW. Then manual changes in the alignment were performed using MOE Sequence Editor (MOE, 2006.08 release, Chemical Computing Group, Montreal, Canada). Some of the sequences could not be aligned in the subfamily alignment without gaps in transmembrane regions and were therefore omitted. The remaining 96 sequences were included in the family C multiple sequence alignment. Sequence numbers and subfamilies are given in Section 4.1.2, Table 2. GPCR family C protein sequences and resulting alignment in Section 10.3, Figure A 1.

MSAs of subfamily sequences were aligned to each other keeping the original alignments fixed. Therefore the "partition" mode of *MOE Align* and *Blosum45* (*gap open*=7, *gap extension*=1, Kelly 1996) were used. Published alignments were also consulted in order to consider other possibilities (Jiang *et al*. 2005, Pagano *et al*. 2000, Pin *et al*. 2003, Malherbe *et al*. 2003, Kew *et al*. 2004, Xu *et al*. 2004, Surgand *et al*.

34

2006 and Petrel *et al*. 2003). The definition of TM boundaries is ambiguous, therefore they were chosen close to TM regions of BR, also considering structural properties of the residues (in TM, ionizable or charged amino acids are favored by the charged ends of phospholipids). On account of this the excised blocks start and end with amino acids, such as Asp, Glu, His, Lys, Asn, Gln and Arg, ensuring a capture of complete structural domains for further alignment. Each subfamily alignment was aligned to rhodopsin, in order to have a reference to excise the transmembrane regions at identical positions (TM1: 38-67, TM2: 72-101, TM3: 110-139, TM4: 141-171, TM5: 198-227, TM6: 249-275, TM7: 285-312, BR position numbering). After the seven transmembrane helix regions had been excised from subfamily alignments to rhodopsin, they were joined to form the family C MSA. The resulting alignment contained only TM helices, no loops, no extracellular domains and no non-TM spanning helix H8 or C-terminal domain. The entire family C GPCR sequence alignment is given in Figure A 1.

3.4 Sequence identity

Sequence identity was calculated for subfamily alignments of TM helices and for identity of sequences to rhodopsin, respectively. Given a MSA two ways of sequence identity calculations were applied:

1. the number of positions in a MSA which are identical (F II).
2. average identity of the subfamily MSA to BR (F III).

The two methods differ in a way that in the latter, the MSA is compared to an additional sequence, which is not part of the MSA. Then identity means the average identity of a subfamily to the given sequence. Since a position can be only identical (1) or not identical (0), each sequence of the MSA has to be compared separately to the given target sequence and the identities of all sequences in the MSA averaged afterwards. For the first method, all sequences at a given position need to be identical, in order to consider this position to be identical (1).

To 1.: The identity per position p is summed up for the entire MSA. A position is regarded as "identical", if the same element (depending on the alphabet) is present in each of the sequences which are included in the MSA, $seq \in MSA$.

3. Methods

$$Identity(MSA) = \sum_p \begin{cases} 1 & ,if\ \forall \{seq_A, seq_B \in MSA\ |\ seq_A^p = seq_B^p\} \\ 0 & ,if\ \exists \{seq_A, seq_B \in MSA\ |\ seq_A^p \neq seq_B^p\} \end{cases} \qquad \text{F II}$$

To 2.: The average identity of a MSA to BR is calculated for each $seq \in MSA$ independently and averaged after summing up over all N sequences in the MSA (FIII).

$$Identity(MSA, BR) = \frac{1}{N} \sum_i^N \sum_p \begin{cases} 1, & if\ seq_p^i = BR_p \\ 0, & if\ seq_p^i \neq BR_p \end{cases} \qquad \text{FIII}$$

3.5 Entropy conservation analysis

The conservation of amino acids of different GPCR families was based on a multiple alignment of their corresponding sequences and evaluated in an entropy calculation approach. For each position of the MSA, the conservation was measured applying Shannon Entropy (H, Section 3.5.3) on the frequency of occurrence for amino acids encoded by a particular scheme (Section 3.5.2). This procedure was performed to provide a description of the conservation for amino acids in TM regions of GPCRs. Each step of the procedure will be introduced in following.

3.5.1 Sequence weighting

Sequential data used for the MSA originated from different species and receptor types, leading to a biased representation in the data set. To account for this biased data bias a weighting scheme was applied (Sander and Schneider 1991). The weights are related to the density in sequence space covered by the different sequences. Therefore sequences from regions with higher local density were assigned lower weights and *vice versa*.

The weight w_a for sequence a reflects the similarity of a to other sequences (F IV), where d_{ab} is the distance (dissimilarity) between the sequences a und b. The distance between two sequences was calculated by counting the number of mismatches between sequences a and b:

$$w_a = \frac{1}{N} \sum_b^N d_{ab}, \qquad \text{F IV}$$

with $d_{ab} = d_{ba}$ and $d_{aa} = 0$.

During calculation of frequencies for each amino acid type per position, each sequence contributing to the MSA has only an impact as strong as its weight, instead of equal contribution.

3.5.2 Mapping amino acids to chemical property groups

Instead of considering amino acids independently, they were treated as chemical property groups (Wrabl and Grishin 2005), such that matching property groups were regarded as matches in the MSA. This grouping allowed for monitoring of conservation of special chemical groups rather than on the level of individual amino acids. Residue types with similar interaction or volume properties can be substituted for each other. Some amino acids can have a particular impact on folding of a protein. Glycine and proline introduce kinks in helical structures, cysteine can participate in disulfide-bonds. The amino acid grouping (Table 1) was adapted from the structural grouping of MOE| Sequence Editor (2006.08 release, Montreal, Canada).

Table 1: The definition of nine functional groups and amino acids belonging to these groups. Amino acids are given in standard three-letter code.

functional groups	amino acids
Aromatic	Trp, Phe, Tyr
Aliphatic	Met, Leu, Ile
Small	Ala, Thr, Ser
Acidic	Asp, Glu
Basic	His, Arg, Lys
Neutral	Gln, Asn
Cysteine	Cys
Glycine	Gly
Proline	Pro

3.5.3 Entropy based methods

In the present study, Shannon entropy (*H*) was evaluated for its ability to point out special sequence positions of different receptor families in combination with amino acid grouping. Based on a multiple sequence alignment the conservation level was analyzed for each position. For each position in the alignment the frequency of each letter *x* from the alphabet *X* was counted. The alphabet was composed of the nine amino acid types, as introduced in the previous section (Table 1).

Shannon entropy

Shannon entropy is often referred to as a measure of the "uncertainty" of a variable (Shannon 1948). Here, uncertainty is interpreted biologically and reflects the conservation of a sequence position, or in structural context, a 3D feature involved in receptor function or tertiary structure formation. The calculation of *H* for a position *Y* in the MSA can be performed according (FV).

$$H(Y) = -\sum_{x \in X} p(x) \log_2 p(x),$$ \hfill F V

where *x* is one amino acid type from the alphabet *X* of all considered amino acid types and *p(x)*, the frequency. The lower the value, the more certain the random variable and the more conserved the position *Y* in the MSA. When the frequency of occurrence for all amino acid groups is equal, the entropy value is largest and can be interpreted as "not conserved". Shannon entropy calculations were performed with a custom application implemented in programming language Java (Sun Microsystems, Inc. Santa Clara, CA 95054 USA).

3.5.4 Visualizing values on receptor structure

In order to analyze the calculated *H* values in a structural context, a visualization method was implemented. The *H* values for each position in the MSA were projected onto the modeled structure of one of the proteins used in the alignment. Therefore only positions present in this particular receptor structure could be displayed. In the respective PDB file *B-factor* values were replaced by *H* values. The PyMOL (v.1.0, DeLano W.L. 2002) molecule viewer was used to create the graphical representation. Two different color schemes were used:

1. colors for representation of the most conserved amino acid type (courtesy of Benjamin Stauch)

2. a colors gradient, representing the "conservation strength".

The color gradient was calculated with the *color_b.py* script (color_b.py v6.0 Copyright (c) 2004 Robert L. Campbell).

3.6 Molecular descriptors

3.6.1 Functional Class Extended-Connectivity Fingerprint

The functional class extended-connectivity fingerprint (FCFP) is a circular, substructural, vectored, value-based descriptor (Rogers and Hahn 2005). It describes the 2D position and frequency of atomic features of a molecule derived from the topological neighborhood in the molecular graph.

The generation of the fingerprint was performed in several steps starting with the generation of initial atom codes. The FCFP includes a special atom typing for all the heavy (non-hydrogen) atoms as a feature for functional coding: First, atoms are assigned abstracted interaction functions, *i.e.* hydrogen-bond acceptor, hydrogen-bond donor, positively ionized or positively ionizable, negatively ionized or negatively ionizable, aromatic and halogen. This abstraction reduces the number of possible atom types. In the second step, a representation of each atom in larger structural environment is developed in an iterative way, similar to the Morgan algorithm (Morgan 1965). The maximal distance, defining the largest diameter of the generated features in number of bonds, is an additional fingerprint property and defines the considered substructure size, for example a distance of 6 is given in FCFP_6. Enlarging the neighborhood in each step of the iteration, atom codes of an atom are updated with the atom codes of the other atoms in range. The new code is generated using a hashing scheme and is always derived from the last iteration not the initial atom codes. The type of the bond connecting to the neighboring atoms as well as their atom codes are hashed to a new number, which is added to the present array containing the initial atom codes. This procedure is repeated until the target diameter is reached. The resulting FCFP contains a list of features present in the molecule, with duplicates removed.

3. Methods

Each feature corresponds to the presence of a structural unit. Structural units are not predefined so that virtually a high number of different features is possible depending on the molecules, while only a small set is present in one molecule. Here a difference was made between a substructure of a molecule extended in any possible way and a substructure included entirely without other extension points than the defined ones. Figure 9 illustrates the difference to common substructure definitions.
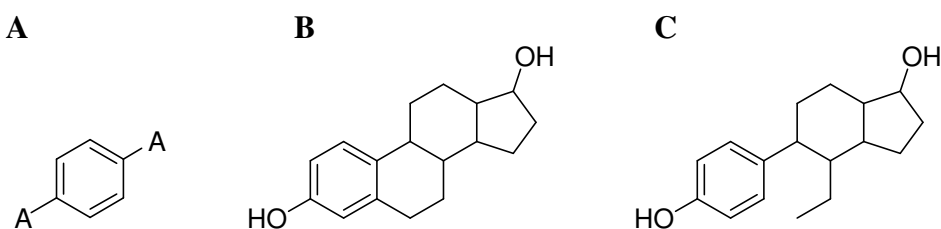
Figure 9: A benzene ring with two attachment points (marked with an **A**) as presented in **A** is part of the structure **C** but not structure **B** according to the extended connectivity definition implemented in FCFP (SciTegic, San Diego, USA)**.**

Pipeline Pilot was used to calculate the FCFP_4 and FCFP_6 descriptors (SciTegic, Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA).

3.6.2 Chemically Advanced Template Search

For calculation of topological descriptors the molecule can be represented by a molecular graph. The molecular graph describes the molecular structure in 2D, where only bonds and atomic features define the molecule.

In 1999, Schneider introduced a topological atom-pair descriptor (Chemically Advanced Template Search, CATS) which considers atom type pairs distributed over the molecule graph (Schneider *et al*. 1999). These atom types are defined as pharmacophoric features: hydrogen-bond donor, hydrogen-bond acceptor, positively charged, negatively charged, lipophilic. Using this pharmacophoric abstraction molecular interaction patterns are favored over chemical atom types by grouping many atoms similar in their interaction type to one pharmacophore type. For CATS descriptor generation all atom-pairs and the shortest paths connecting them are defined. In the next step for each distance in range from zero to nine bonds the number of different atom-pair occurrences is calculated. The resulting value vector contains 150 values resulting from 10 distance and 15 atom-pair variation per distance. The values are further scaled by the number of all non-

hydrogen atoms in the molecule. The software speedcats.com by Dr. U. Fechner was used for descriptor calculation (Fechner *et al.* 2003). No hydrogens were added to molecules beforehand.

A pharmacophore descriptor calculation variant as implemented in Pipeline Pilot, called pharmacophore fingerprints (PHRFP), has a different atom typing scheme than the CATS descriptor. The PHRFP discriminate between hydrogen bond acceptors and donors, positively and negatively charged and ionizable atoms as well as atoms in hydrophobic or aromatic groups. Therefore out of 8 atom types 28 combinations can be achieved. In the present study atom pairs in distance of 2 to 15 bonds were considered. The PH**R**FP_2 contains number of bonds and number of rotatable bonds between the features additionally.

The concept of pharmacophoric descriptors allows for scaffold hopping in virtual screening applications (Schneider *et al.* 1999, Fechner *et al.* 2003). As scaffold hopping is major goal in finding a new lead molecule and crucial for the present study, this molecule description was applied to similarity searching (Section 3.12) and molecule clustering with self-organizing maps (Section 3.13).

The two different implementations of the pharmacophoric fingerprint concept were applied due to licensing reasons, CATS (speedcats.com) in combination with in MOLMAP (by Prof. G. Schneider, Schneider and Wrede, 1998) at Johann Wolfgang Goethe-University and PHRFP in Pipeline Pilot (SciTegic, San Diego, USA) at Merz Pharmaceuticals (Merz, Frankfurt am Main, Germany).

3.6.3 Shape descriptors

Molecular surfaces define shapes of molecules as 3D objects. Based on 3D atom coordinates of a molecule, the surface can be generated near to the van der Waals surface or the solvent accessible surface (Connolly 1983). The Gaussian surface calculation method as implemented in MOE (Grant *et al.* 2001) is a smooth approximation of the Connolly Surface (Connolly 1983) and is constructed from a sum-of-Gaussians density derived from the atomic coordinates of the molecule. The surface is described by a density function $v(x)$, with x_i being the coordinate of the *i*-th atom and R_i its contact radius (FVI).

3. Methods

$$v(x) = \exp\left(-11.8\sum_i w_i \exp\left\{-K_i^2|x-x_i|^2\right\}\right),$$

F VI

$$\text{with } K_i = \frac{\pi^{1/6}(4/3)^{-1/3}}{R_i} \text{ and } w_i = \left(\frac{2^{1/2}\pi}{K_i}\right)^{3/2}.$$

The Gaussian contact surface resembles the water accessible surface, when smoothing over cavities not large enough to contain a water molecule.

Shape similarity of molecules can be calculated using shape descriptors. The spherical harmonics approach allows for rotation-invariant representations of geometric shapes (Zhan *et al.* 2006). The key feature of the spherical harmonics descriptor (SHD) is the alignment-free comparison of 3D molecular shapes. SHD is considered to be a global feature-based descriptor that is composed from spherical harmonics coefficients (Wang 2008).

To calculate the SHD of the 3D structure of a molecule, a shell model is applied. 3D coordinates can be given as Cartesian or spherical coordinates. The 3D surface point coordinates can be transformed into spherical coordinates. A 3D object can be described using points at concentric spheres around the centre of mass of the object. In order to express the point coordinates invariant to the radius, all coordinates are projected onto the unit sphere or - in other words - for each sphere with a different radius the coordinates are normalized resulting in the unit radius. Spherical coordinates define a point by two angles θ and φ, the radius is normalized by the unit sphere.

The 3D shape of an object can be decomposed into a set of orthogonal basic functions, referred to as spherical harmonics (FVII).

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\, p_l^m(\cos\theta)e^{im\phi}$$

F VII

The spherical harmonics function Y is defined by the two angles θ and φ for order l and degree m. The degree m is defined based on the l value, $m \in \{-l, -(l-1), ..., 0, ..., l-1, l\}$. By choosing the order value, several spherical harmonics functions can be solved in the range of 1 to l. This set of orthogonal spherical harmonics functions defines a complete

42

3D description, and it is similar to unit basis vector descriptions. In this manner, any surface function *r(θ,φ)* can be decomposed into its spherical harmonics (FVIII).

$$r(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m (\theta,\phi)$$ F VIII

.

The coefficients for the spherical harmonics decomposition are calculated for each combination of the order and the degree value defining unique properties of a 3D shape. The SHD used in the present work was defined as the norms of the decomposition coefficients of each degree component in every spherical harmonics order.

For each molecule, the SHD is a vector of identical length. The length is defined by the order of spherical harmonics, according to $L^2 + 2L + 1$. Here, a 100 value descriptor was used, with *L=9*. In general, the shape becomes more detailed as the order parameter increases (Figure 10).
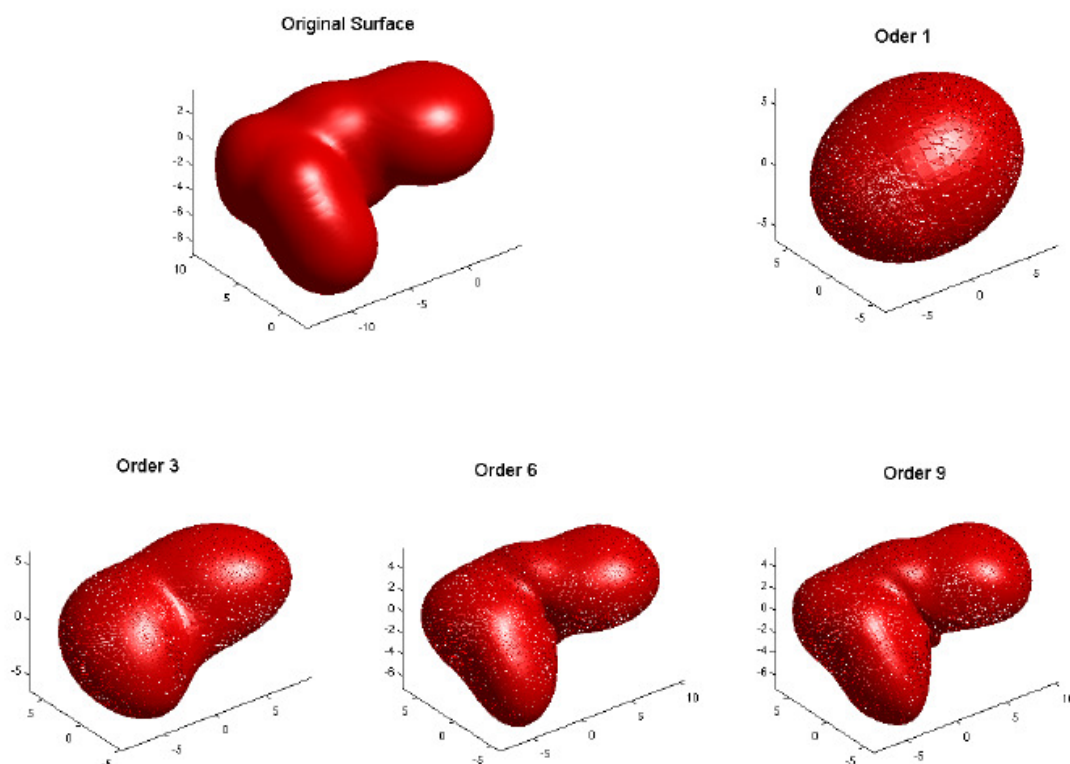
Figure 10: Influence of the order parameter on accuracy of shape description. The original surface is represented as sets of spherical harmonics functions of order one, three, five and nine. Axes represent Cartesian coordinates. Adapted from (Wang 2008), with kind permission.

3. Methods

Gaussian contact surfaces were calculated using MOE|Compute|Surfaces and the Maps|Gaussian contact function (Grant *et al*. 2001). This function was applied to a MOE database containing the 3D structures of molecules that were extracted from the literature. The 3D molecule structures were calculated with Ligprep (v2.0, Schrödinger, LLC, New York, 2008), as described in Section 2.6.2. For the Gaussian surface calculation, all heavy atoms were considered and the clipping proximity was set to 5Å. The maximum memory parameter was set to 1MB, defining the accuracy of the generated lattice.

The SHD calculation was performed with MATLAB (MATLAB, Version 2006b, The Math-Works, http://www.mathworks.com) scripts prepared by Quan Wang (Wang 2008).

3.7 Similarity measure

Similarity between two molecules can be calculated using distance measures applied on a numerical molecule description. Common measures are for example the Tanimoto coefficient (Tanimoto 1957, Johnson and Maggioga 1990, Willett 1998) and the Euclidian distance for vectorial molecular descriptors and the RMSD for atom coordinates. A detailed discussion of similarity measures for chemical similarity, including Tanimoto and Euclidian measures, can be found elsewhere (Willett 1998).

3.7.1 Tanimoto coefficient

The Tanimoto coefficient (Tanimoto 1957) is defined by the sum of products of each descriptor position i for the reference R and the target T descriptor values, x_{Ri} and x_{Ti} (F IX).

$$DA = \sum_i x_{Ti} x_{Ri} ,$$

$$DB = \sum_i x_{Ti}^2 - \sum_i x_{Ti} x_{Ri} ,$$

$$DC = \sum_i x_{Ri}^2 - \sum_i x_{Ti} x_{Ri} ,$$

$$Tanimoto = \frac{DA}{DA + DB + DC} . \qquad\qquad \text{F IX}$$

44

The coefficient's value range lies between 0 and 1. The higher the similarity, the closer is the coefficient value to 1.

3.7.2 Euclidian distance

The Euclidian distance between two vectors x_R and x_T was calculated according to F X.

$$D(x_T, x_R) = \sqrt{\sum_i (x_{Ti} - x_{Ri})^2} \qquad \text{F X}$$

with x_{Ri} and x_{Ti} the i-th vector element.

3.7.3 Root mean square deviation

The RMSD allows for the determination of structural similarity between two different three-dimensional structures of the same molecule. The RMSD between two structures is the square root of the average squared distances between equivalent atoms (F XI). One molecule conformation is regarded as the target conformation T and the other as the reference conformation R. The number of atoms is N.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (x_{Ti} - x_{Ri})^2}{N}} \qquad \text{F XI}$$

3.8 Receiver-operating characteristic analysis

The Receiver-Operating Characteristic (ROC) Analysis can be applied in virtual screening to calculate the prediction accuracy of a binary decision model (Zou *et al.* 2007, Fawcett 2006). The area under the curve (AUC) value describes the ability of a model to rank active molecules above decoys. The sensitivity (F XII) of the model is plotted against $1 -$ specificity (F XIII) at different threshold values for the binary classification (Figure 11). The specificity and sensitivity are calculated using the numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

$$sensitivity = \frac{TP}{TP + FN} \qquad \text{F XII}$$

3. Methods

$$specificity = \frac{TN}{TN + FP}$$.   F XIII

A ROC curve corresponding to random chance would connect the points (0,0) and (1,1). The AUC value summarizes the average diagnostic accuracy across the spectrum of threshold values.
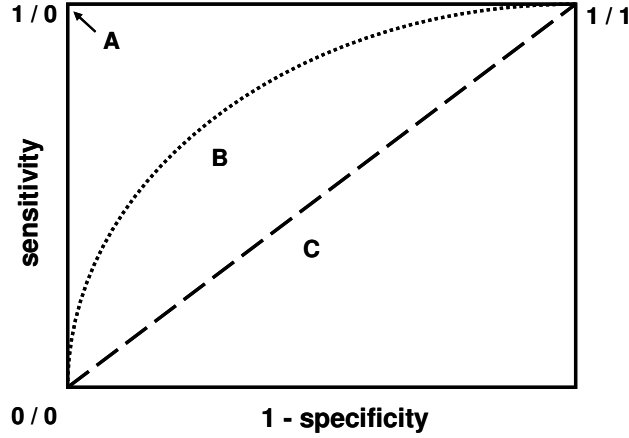
Figure 11: Three hypothetical ROC curves representing the diagnostic accuracy (adapted from Zou *et al.* 2007). Curve A (AUC equal to one) lies along the y-axis and corresponds to perfect accuracy, curve B lies between the cases A and C, the latter representing a random chance with AUC=0.5. The diagnostic accuracy improves the more a ROC curve moves towards curve A.

In case of a Bayesian classifier (Section 3.11) the ROC curve represents the prediction rate at all possible threshold values for the probability value of the binary classification in "*good*" and "*bad*" samples. Here, the ROC plots and AUC values were generated with Pipeline Pilot (SciTegic, San Diego, USA)

.

3.9 Pareto-ranking

In virtual screening as well as other applications where the result is a selection of data samples with several optimal properties, a ranking is required that favors more than one feature for optimization, especially in cases when optimization of some properties decreases the quality of others. In 1896, Pareto defined the Pareto-optimum concept (Pareto 1896). The goal of the optimization is to find a solution vector:

$x^* = \left[x_1^*, \mathsf{K}, x_n^*\right]^T$ that optimizes the function $f(x) = \left[f_1(x), \mathsf{K}, f_I(x)\right]^T$.

A solution is Pareto-optimal if there is no other solution that can have a function value closer to optimal without degrading another functional value. Given F as the set of all possible solution x, a solution is Pareto-optimal if the following holds for $\forall x \in F$:

$$\bigcap_{i \in I}(f_i(x) = f_i(x^*)) \quad \text{or} \quad \exists i \in I : f_i(x) > f_i(x^*).$$

An exemplary 2D situation is shown in Figure 12.
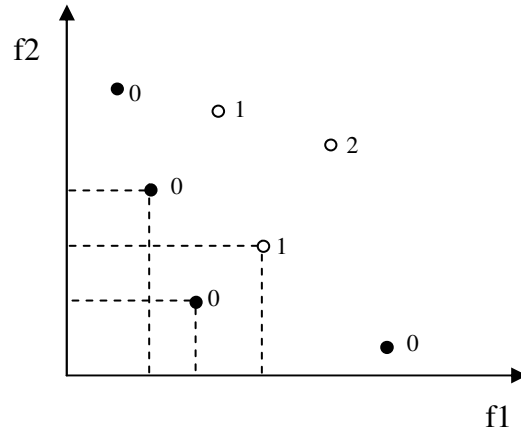
Figure 12: The Pareto-ranking principle applied to eight solutions in a 2D space where both values *f1* and *f2* should be minimized. Filled points represent the Pareto front. The dashed lines going out from a point define the area that contains other points dominating the concerned point. If the area is not populated, this solution is not dominated by any other solution and belongs to the Pareto-front. Each rank *X* (given as number) includes all solutions that are dominated by *X* many other solution ranks, as demonstrated for solutions numbered one and two.

In the present study, the Pareto-ranking was applied for a 2D optimization task, where the prediction score of two Bayesian classifiers focusing of different features were to be joined for hit list ranking (Section 3.11.2). A Pareto subset optimizer (by SciTegic, San Diego, USA) was applied. Parameters were chosen as follows: subset size: 100, number of subsets: 10, number of optimization iterations: 10000, first property: Bayesian "mGluR"-model score, second property: Bayesian "selectivity"-model, optimized property: a subset according to standard deviation, goal: score maximization.

3.10 Diversity sampling

Diversity sampling is a procedure for diverse subset selection of data samples. Essential for the sampling is the representation of data as a valued feature vector that allows for similarity calculations between data points. The goal of diversity sampling is the

selection of samples that cover different areas in multidimensional space. The Maximum-Dissimilarity algorithm as discussed by Snarey is a common method for subset generation for model development applications (Snarey *et al.* 1997).

The Maximum-Dissimilarity algorithm for selection of n out of m samples, with $n \leq m$:

1. Select randomly a data sample
2. Calculate dissimilarity to remaining data samples
3. Add the most dissimilar sample compared to the actual data sample to the result set
4. Return to step 2 with the last selected sample as actual sample, if the number of samples is less than n.

In the present work, diversity sampling according to this algorithm was performed in order to split data into training and validation data for the Bayesian model development (Section 3.11). Molecules were compared and selected based on the FCFP_6 descriptor (Section 3.6.1) and their similarity was measured by the Tanimoto coefficient (Section 3.7.1).

3.11 Bayesian classifier

3.11.1 Theory

The Bayesian Classifier can be applied as method for binary categorization of data (Xia *et al.* 2004). A prediction model can be trained on classified patterns and then be applied for probability estimation of unclassified data. The binary classification is performed with a Bayesian Estimator using a Laplacian correction, which can be used to calculate the likelihood of a pattern based on important features learned from training data. In the case of molecules, these features are presented as molecular descriptors for both training data of "*good*" (active) and "*bad*" (inactive) molecule classes. A trained Bayesian Classifier can give the likelihood of an unknown molecule to belong to the "*good*" class, based on presence and frequency of features coded with the descriptor. The simple naïve Bayesian classifier is supposed to perform best on data with independent attributes. However, empirical results showed that the performance is well even in cases where dependency is given (Bender *et al.* 2005) and that this is not the

only requirement for optimal prediction (Domingos and Pazzani 1997). Other prediction methods are not superior in equal situations.

Given *N* samples available for training and *M* good samples being part of them, the probability of a *"good"* sample is *P(good)=M/N*. A further assumption is that feature *F* is present in *A* of *B* samples. It can be estimated that most features F_i are not important for discrimination, and therefore their probability equals the base probability *P(good)*. If a feature is sampled *K* times, its probability is corrected to *P(good)*K*, (F XIV). This correction is necessary in cases when *B* tends to become very low and without the correction *P(good|F)=A/B* would give a value close to 1. The Bayesian estimator using a Laplacian correction is given in FXV.

$$P(good \mid F) = (A + P(good) * K) / (B + K) \qquad \text{F XIV}$$

with $K = 1 / P(good)$. The relative estimation is further possible with

$$P_{relative}(good \mid F) = P(good \mid F) / P(good) \qquad \text{F XV}$$

Most features that are not important for classification will give log $P_{relative} \sim 0$. For features more frequently present in *"good"* samples $P_{relative} > 0$ and less frequent $P_{relative} < 0$. The probability function (F XV) is calculated for each feature F_i. The respective feature weights are summed up to provide a probability estimate for a new sample.

3.11.2 Model training

Two different Bayesian models were trained in order to be applied for virtual screening of selective mGluR5 binding ligands, the "mGluR"-model and the "selectivity"-model. Predictions from both models were joined by the Pareto-ranking method.

mGluR-model

An "mGluR"-model was trained using the FCFP_6 descriptor (Section 3.6.1). Three data sets were compiled from the literature data collection of family C ligands (Table A

3. Methods

1), the Merz "*in house*" screening collection (Section 2.4.2) and the WOMBAT data base (Section 2.5). Inactive molecules from the literature and Merz data sets were added to the "bad" data, since they are not presented in WOMBAT.

"*good*"

mGluR_1_5:   All mGluR1 and/or mGluR5 binding ligands from the Merz and the literature data collection with activity lower than 1000nM. Number of literature collection molecules: 870.

"*bad*"

not_mGluR: Ligands (binding to other targets than mGluR) selected from WOMBAT using the diversity sampling (Section 3.10) method considering diversity based on the FCFP_6 descriptor. Number of molecules: 152269.

bad_mGluR: All mGluR1 and/or mGluR5 binding ligands from the Merz and the literature data collections with activity higher than 1000nM. Number of literature collection molecules: 117.

Selectivity-model

An mGluR5 versus mGluR1 "selectivity"-model was trained using the FCFP_6 descriptor. Three data sets were compiled from the literature data collection of family C ligands and the Merz "*in house*" screening collection. Ligands were assumed selective if their experimental activity data for both mGluR1 and mGluR5 differed about a factor of ten or higher.

"*good*"

selective_mGluR5:   All mGluR5 binding ligands that are selective for subtype five according to experimental tests. Number of literature collection molecules: 111.

"*bad*"

selective_mGluR1:   All mGluR1 binding ligands that are selective for subtype one according to experimental tests. Number of literature collection molecules: 57

not_selective: All mGluR5 ligand that are comparably active at mGluR5 and
                             mGluR1. Number of molecules: 115

The "selectivity"-model focused on features discriminating selective mGluR5 molecules
from non-selective molecules was established. The number of molecules was less than
in the case of the mGluR-model, since information about activity values on different
receptor subtypes was not available for all ligands.

3.11.3 Retrospective validation

In order to test the prediction accuracy of the method, the complete data set was split
into training data for model development and validation data for quality assignment of
the trained model. The data split was accomplished using a FCFP_6-defined diversity
sampling (Section 3.10) method from Pipeline Pilot that chose 60% of the molecules
from "*good*" and "*bad*" samples and 40% to be used as an external test set. In the
retrospective validation, the predicted classification of molecules was compared to the
known classification. The model score provided by the model is the sum of $P_{relative}$
values of all features and is different to the normalized probability. Validation samples
with probability to belong to the "*good*" molecules higher than 0.5 (threshold was not
applied in prospective screening), were considered as "*good*" ones. For quality
assignment, the number of true positive, true negative as well as false positive and false
negative predictions was calculated.

Prediction accuracy

The accuracy of the model in predicting the likelihood of the validation data to be
"*good*" samples was evaluated with the ROC plot and AUC value (Section 3.8).

Classes' separation

The ability to use the score for binary classification was determined as follows.
Prediction score of the Bayesian model were plotted against the frequency the score was
assigned to molecules of the training and validation sets. This distribution histogram
was analyzed in terms of the discrimination between "*good*" and "*bad*" samples by the
model's prediction score. The histogram was used for definition of the score value
which defined the cut between molecules considered to be a "*good*" sample.

3. Methods

Protocol for combination of the "mGluR"- and the "selectivity"-model:

1. "mGluR"-probability prediction using the "mGluR"-model for all vendor molecules
2. Filtering of all molecules which have a model score below 24 (motivated by class separation)
3. Scoring the molecules with the "selectivity"-model
4. Definition of the Pareto front (Section 3.9)

3.12 Similarity search

3.12.1 Molecule shape similarity

A 3D molecule structure defines the volume and interaction properties exhibited by chemical groups of the molecule. These features can be used for comparison and ranking of molecules by their similarity to a reference molecule. Conformers from a test molecule are aligned in various ways to the reference molecule and the similarity is computed based on overlapping hard-sphere volumes. Atom typing allows for the introduction of additional information into the similarity evaluation.

A shape similarity search was performed using Phase (Phase v2.5, Schrödinger, LLC, New York, 2008). The molecule vendor libraries ASINEX Platinum Collection (vNov.2007, www.asinex.com) and SPECS (v2008.1, www.specs.net) were searched for molecules similar to the shape of MPEP. MacroModel atom types (MacroModel v9.6 Schrödinger, LLC, New York, 2008) were applied as additional molecule description. For shape similarity, the calculation of the 3D molecule structure was performed. The vendor libraries were prepared by Björn Krüger and used as a multi-conformer database.

Shape search with Phase (Phase v2.5, Schrödinger, LLC, New York, 2008):
```
phase_shape -screen database -shape mpep.sdf -JOB job_name -CHECKPOINT
directory -atomTypes mmod -sort
```

According to shape similarity to the reference molecule vendor molecules could be ranked. From all molecules with the `r_phase_Schape_sim`-score above 0.7 molecules with diverse scaffolds were manually selected for testing.

3.12.2 Descriptor-based similarity

Molecular descriptors allow for a quantitative comparison between molecules. The molecule structure is no longer the key description but the properties encoded by the coding scheme. Different descriptors can span different multidimensional spaces even for the same set of molecules and therefore define different neighborhood relations.

In the present study, two different descriptors were applied for similarity search, FCFP_4 (Section 3.6.1) and PHRFP (Section 3.6.2). The similarity search was performed for both descriptor types using Pipeline Pilot (SciTegic, San Diego, USA) and the Tanimoto coefficient (Section 3.7.1) for similarity detection. The same search procedure and data sets were applied in both searches. Both searches were designed to select the most similar molecules to known selective NAMs binding to mGluR5. A total of 619 mGluR5 and 261 mGluR1 ligands were selected as reference molecules from the literature (Table A 1) and the Merz Pharmaceuticals *'in house'* data collection (Section 2.4.2).

Search and ranking procedure:

1. Similarity calculation between screening molecules and known a) mGluR5 and b) mGluR1 binding NAMs with activity lower than 1000nM and filtering of those with similarity higher than a threshold value of 0.95.
2. Removal of all molecules that resulted from both steps 1b and 1a.
3. Ranking of remaining molecules according to Tanimoto similarity as calculated in step 1.

Molecules with diverse scaffolds were manually selected from the ranked molecules list for experimental testing.

3. Methods

3.13 Self-organizing maps

SOMs (or Kohonen maps) are an unsupervised machine learning technique that can be applied on various data presented as a numerical vector (Kohonen 1982). SOMs enable the representation of multi-dimensional data and preserve distance and proximity relationships. These relationships between data clusters can be projected onto a 2D map. This map depicts neighborhood in multi-dimensional space. Not the entire network can be presented when projecting down to 2D. However, the projection helps to analyze the captured data relationships and data clusters.

The number of neurons of a SOM has to be decided before the training; it should be close to the number of expected clusters. Data clustering is achieved by training of a neural net on data vectors. The aim of the training is the adaptation of the net to the data distribution in multi-dimensional space. During the training of a SOM, training data are "presented" to the net, a winner neuron is defined and the weight vectors of the winner neuron and its neighborhood are adapted according to the learning rate.

Training algorithm:

C = set of neurons, ξ = input pattern, w = weight vector, s = winner neuron, c = neuron, N_S = neighborhood of neuron s, ε = learning rate, t = number of learning patterns.

Step 1: Initialization of map A with N neurons $c_i : A = \{c_1, c_2, ..., c_N\}$; neuron weights $w_{c_i} \in R^n$ are generated by random sampling according to the distribution of the training data $p(\xi)$. Set the time parameter $t = 0$.

Step 2: Choose a training pattern ξ according to $p(\xi)$.

Step 3: Define the winner neuron with $\min_{\forall c \in C} |\xi - w_c|$.

Step 4 : Adapt neuron weights for neuron r with $\Delta w_r = \varepsilon(t) h_{rs} (\xi - w_r)$. The inter-neuron distance is calculated using the Hamming-distance d_H, further a Gaussian neighborhood function is applied for the winner neuron s.

$$h_{rs} = \exp\left(\frac{-d_H(r,s)^2}{2\sigma^2}\right)$$

using the standard deviation according to

$$\sigma(t) = \sigma_{initial} \left(\frac{\sigma_{initial}}{\sigma_{final}} \right)^{\frac{t}{t_{max}}}$$

and a learning rate

$$\varepsilon(t) = \varepsilon_{initial} \left(\frac{\varepsilon_{initial}}{\varepsilon_{final}} \right)^{\frac{t}{t_{max}}} .$$

Both parameters σ and ε undergo a time dependent adaptation based on an *initial* and a *final* value.

Step 5: $t = t + 1$.

Step 6: If $t < t_{max}$, then go to Step 2, else terminate.

The mean quantization error (*mqe*) can be calculated from equation FXVI and is a measure for the dissimilarity of neuron members to the neuron.

R_c = Receptive field of neuron *c*.

w = Weight vector centered on data of a neuron.

$$mqe = \frac{1}{N} \sum_C \sum_{\xi \in R_c} |\xi - w_c|^2 \qquad\qquad \text{F XVI}$$

In the present study, the Euclidian distance metric was used for similarity calculations (Section 2.7.2) using the implementation of the SOM-algorithm in MOLMAP (by Prof. G. Schneider, Schneider and Wrede 1998). Parameters such as the number of neurons, neuron radius and the number of training cycles were defined individually and are presented in detail in the following section.

3.13.1 Self-organizing map based virtual screening

The reference and screening data sets were encoded with CATS descriptors (Section 3.6.2). The number of neurons was defined as 20×15, the neuron radius was set to 8 and the number of training cycles to 200,000. Different vendor molecule collections (Asinex Platinum Collection, Enamine, Specs, Ubichem and Maybridge, details on vendor data bases can be found in Section 10.5, Table A 2) were clustered separately with reference molecules and joined before step five of the protocol described below.

3. Methods

Protocol:

1. All molecules from the vendor data collection were encoded with CATS.
2. Reference molecules from the literature data collection (Section 2.4.1) were encoded similarly.
3. A SOM was trained using one vendor data base and reference molecules.
4. All vendor molecules from neurons containing selective reference ligands were selected.
5. For molecules selected in step 4, a prediction of mGluR probability with the "mGluR"-model Bayesian classifier (Section 3.11) was performed.
6. Molecules were ranked according to their "mGluR"-model score; those with scores below 20 were filtered out.
7. 63 molecules were tested experimentally (13 from the Asinex, 27 from Specs, 3 from Maybridge and 20 from Enamine pick collections).

Since a single active molecule was discovered from the Asinex molecule collection, only neuron numbers that were selected in step four for that SOM are given here and will be discussed in Section 4.2.3 (0/1, 0/7, 0/14, 1/8, 12/9, 13/8, 13/10, 13/12, 14/10, 14/11, 15/9, 15/10, 16/1, 16/9, 16/10, 17/0, 17/10, 17/11, 17/12, 17/14, 18/0, 18/11, 18/13, 18/14, 19/7 19/9).

3.13.2 Self-organizing map based clustering of the ligand data collection

The literature ligand data collection (Section 2.4.1) of 1270 mGluR binding ligands was clustered based on two different molecule descriptors, the SHD (Section 3.6.3) and CATS (Section 3.6.2). Diversity and neighborhood were analyzed using the projection of a trained SOM onto a 2D map (MOLMAP application by Prof. G. Schneider). The same generic molecule names were used in order to track the distribution of particular features of molecules, like size, target or functional effect.

Ligand clustering using SHD and SOMs

SHD descriptors with 100 dimensions were used to describe the molecules. SOM training parameters were: the number of neurons 14×12, neuron radius 6 and number of training cycles 200,000.

1. Calculation of the 3D structure for each molecule (Section 2.6.2).
2. Calculation of the molecular surface based on the 3D structure of the molecule (MOE Gaussian contact surface, v.2006.08, Montreal, Canada).
3. Calculation of the SHD (MATLAB v.2006b, scripts by Quan Wang).
4. Addition of a generic molecule names to enable class visualization.
5. SOM training based on SHD.
6. Projection of ligand classes onto the SOM using generic molecule names.

Ligand clustering using CATS and SOMs

SOM training parameters were identical to the SHD based SOM training. The CATS descriptor contained 150 dimensions in contrast to SHD with 100 dimensions.

1. All molecules from the vendor data collection were encoded with the CATS descriptor.
2. Addition of a generic name molecule name to enable class visualization.
3. SOM training.
4. Projection of ligand classes onto the SOM using generic molecule names.

3.14 Experimental activity assay

Virtual screening hits were tested for their modulatory effects in a functional cell-based assay. The mGluR is activated by an endogenous transmitter acting at the orthosteric binding site of the receptor. Because allosteric modulators do not compete with endogenous agonist on the binding site, their effect can be measured by determination of increased or decreased response to the agonist. In the present study, ligand selectivity was the key issue. Therefore, computationally predicted active molecules were tested for real biological activity in the two most closely related metabotropic glutamate receptors, mGluR1 and mGluR5. Both allosteric modulator types were determined, the positive (PAM) and the negative (NAM) ones.

NAMs are defined as molecules decreasing the activation effect of the orthosteric agonist. If the inhibition rate is measured at different concentrations, it can give a full *dose response curve* (DRC) for the ligand (Figure 13, A). The DRC allows to define the

3. Methods

IC_{50} value, which is the concentration of the NAM that reduces the activation by the agonist down to 50% of the maximal activation.
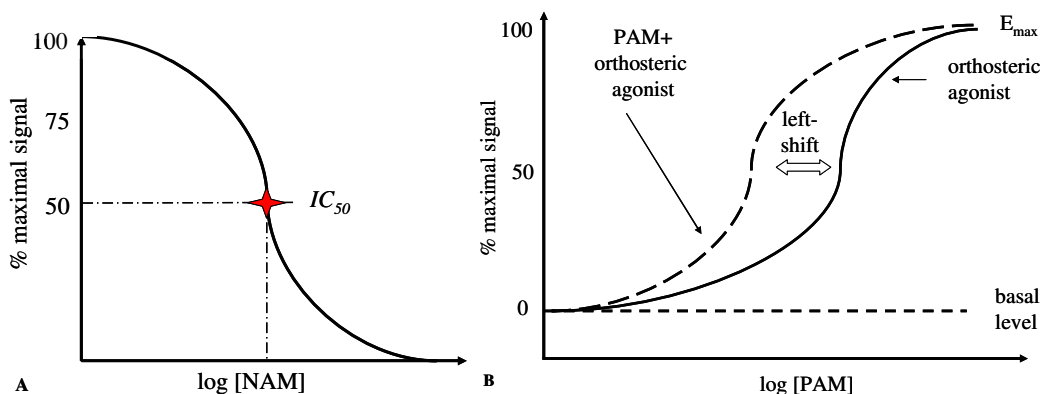
Figure 13: Schematic dose response curves of negative (NAM, **A**) and positive (PAM, **B**) allosteric modulators in a functional assay. The maximal signal is plotted versus the concentration of the allosteric modulator. The concentration to signal dependency is given as bold curve line. **A**) The IC_{50} indicates the concentration of a NAM which leads to 50% signal reduction. **B**) The endogenous agonist and by the dashed curve indicates the activation by the agonist after application of a PAM. The basal line indicates the regular receptor activity level, when no agonist is present. The potency of a PAM is expressed as left-shift of the concentration response curve.

PAMs enhance the activation of a mGluR by the orthosteric ligand. This can be detected as the maximal activation of the receptor at lower concentrations of the agonist than without the PAM (Figure 13, B). In the presence of a PAM the curve is shifted to left.

Experimental conditions for mGluR5 PAM and NAM (Vanejevs *et al.* 2008) and mGluR1 NAM (Renner *et al.* 2007) tests were published elsewhere and performed in the *in vitro* screening department at Merz Pharmaceuticals. For these experiments, the modulators were added to the test system prior to the agonist. The functional assay for mGluR5 depends on the calcium concentration increase, which is a downstream effect on mGluR5 activation. The increase of intracellular calcium after stimulation with the mGluR5 agonist was measured using a fluorometric imaging plate reader (FLIPR™) and the Ca-Kit (both Molecular Devices, CA, USA). Tested molecules were pre-incubated prior to the addition of the agonist. In case of the mGluR1 functional assay, the measured signal is the accumulation of [^{3}H]-Inositol Phosphates. All IC_{50}-value of molecules discovered in this study were determined by measurement of the particular cell signal at five different concentrations of the allosteric modulator.

3.15 Docking

3.15.1 Induced Fit Docking

The prediction of protein-ligand complexes was performed with molecular docking. Using the "Induced Fit Docking" procedure as reported by Sherman (Sherman *et al.* 2006) the receptor binding site was treated partially flexible. This approach combines in an iterative way the docking of ligands in rigid proteins implemented in Glide (Schrödinger, LLC, New York, 2008) and the modeling of receptor conformational changes with Prime (Schrödinger, LLC, New York, 2008). The complete procedure is supposed to allow for the adaptation of the receptor structure to different ligands that otherwise would not necessarily fit into the original receptor binding pocket.

IFD protocol:

1. Rigid receptor docking using softened-potential scoring.
2. Sampling of the protein for each ligand pose generated in the first step.
3. Re-docking of the ligand into low energy induced-fit structures from the previous step.
4. Scoring by accounting for the docking energy (*GlideScore*), and receptor strain and solvation terms (*PrimeEnergy*).

The softened-potential was obtained by scaling the van der Waals radii of ligand and receptor atoms by 50%. This parameter enables to find ligand poses that slightly penetrate the surface of the receptor. To remove larger steric hindrances, particular residues can be allowed to mutate to alanine, thus enlarging the binding pocket. After the first docking, 20 poses were retained for further refinement. Having the mutated residues restored, the complex structure was minimized by sampling the residue conformations within 5Å of the ligand. The adapted complexes were scored with *PrimeEnergy* terms and all solutions within an energy threshold above the lowest energy structure were retained. With the default docking, a re-docking was performed into the adapted receptors. The final scoring was performed by a combination of the Prime energy and the *GlideScore*. *GlideScore* is focused on those quantities and is softer regarding steric clashes. The combined *IFDScore* (F XVIII) is composed as follows:

$$IFDScore = GlideScore + 0.05 \text{ x } PrimeEnergy \qquad \text{F XVII}$$

3. Methods

If the top ranked structures differed less than 0.2 in their *IDFScore*, the IFD protocol was repeated for the top ranked receptor structures using the results from the first round of IFD as a starting point. The only difference was made for the default docking parameters applied in that second round (1.0 and 0.8 for the van der Waals scaling for receptor and ligand atoms, respectively, and 0.0 for both the Coulomb-van der Waals and hydrogen bond energy cut-offs).

3.15.2 Induced fit docking in mGluR5

For IFD, the receptor structure was prepared using the *Protein Preparation Wizard* protocol (Schrödinger, LLC, New York, 2008). The preparation included the addition of hydrogens, the rotation of Asn, Gln and His residues by 180° when needed to maximize hydrogen bonding. A brief relaxation was performed on each starting structure with the "Refinement Only" option using the OPLS2001 force-field (Jorgensen *et al.* 1996). The minimization was terminated when an RMSD of 2.2 Å was reached.

The first softened-potential docking was defined to generate 20 initial poses. The hydrogen bond filter was set to 0 because some ligands possessed any hydrogen-bond donors or acceptors and the hydrogen bond filter would eliminate these. All docking calculations were run in the "Standard Precision" (SP) mode of Glide.

The *trim side* and *binding side* definitions lacking known receptor-ligand complexes were defined manually. The *trim side* residues were selected to be mutated to alanine. The *trim side* included all EC2 loop residues facing the TM binding region (722, 726, 732, 734, and 735). The *binding side* was the region where the back-bone atoms were treated flexibly. Therefore, all residues facing these regions were selected, as well as their next neighboring residues (mGluR5 residue numbers 624, 628, 629, 631, 632, 641, 643, 644, 645, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 662, 710, 713, 714, 716, 717, 718, 720, 722, 724, 730, 731, 732, 733, 734, 735, 736, 737, 739, 740, 743, 744, 747, 748, 752, 777, 778, 781, 782, 784, 785, 787, 788, 789, 791, 792, 808, 809, 811, 812, 813, 815, 816).

All ligands were prepared with Ligprep (v.2.0, Section 2.6.2). The complete setup file for running the IFD is given in Section 10.1.

3.16 Structure modeling

Homology modeling is a methodology that allows to predict a 3D structure model of a protein using a related protein with known structure. Structure prediction for a given sequence via homology modeling consists of the following steps:

1. Identification of a homologue structure from the Protein Data Bank (Westbrook *et al.* 2002).
2. Sequence alignment of the template structure to the given sequence.
3. Generation of a model based on the sequence alignment.
4. Model refinement.

In general, model quality follows the rule that models built based on close homologues with sequence identity higher than 40% will possess main-chain-atoms with an RMSD error of about 1Å (Sanchez and Šali 1997).

3.16.1 MODELLER

In the present study, the software Modeller (9v1, Eswar *et al.* 2007) was used for structure prediction. The modeling method as implemented in Modeller considers a set of restraints derived from sequence alignment (Šali and Blundell 1993). During the process of structure calculation, the violation of these restraints is minimized, a procedure called "modeling by satisfaction of spatial restraints". The definition of an objective function, which is optimized during the search for the most probable model, is directed by:

- The calculation of distances and dihedral angle restraints derived empirically from a data base of protein structure alignments and expressed as conditional probability density functions.
- The extension by CHARMM (MacKerell *et al.* 1998) force-field terms enforcing proper stereochemistry.

3. Methods

3.16.2 Objective function

Šali and Blundell performed a systematic and quantitative calculation of protein structure features and expressed them as probability density functions (pdfs) that can be applied as restraints for homology modeling (Sali and Blundell 1993). The basis for the calculation was provided by the database of known protein structures from the PDB and their alignments to related proteins. Those pdfs took in account the main-chain conformation class (defined by six of the Ramachandran plot regions [Ramachandran *et al.* 1963]) of an equivalent residue, as well as the type of the modeled residue and the sequence similarity of the two equivalent local environments. The association of these residue properties of two related proteins is a major issue in homology modeling. Instead of using a value range distribution of distances between alpha-carbon atoms, residue solvent accessibilities or side-chain torsion angles were included. Thus, the back-bone conformation of a particular residue may be restrained according to the residue type, the conformation of an equivalent residue in a related protein and the local sequence similarity between the two proteins.

The objective function F is composed of dynamic and static restraint terms (F XVIII). The dynamic terms are updated when atoms are moved in order to minimize the objective function violation. F is derived from different constraints including several features:

$$F = F(R) = F_{symm} + \sum_i c_i(f_i, k_i)$$,
F XVIII

where F_{symm} is an optional symmetry term, R are Cartesian coordinates of all atoms, c is a restraint (F XIX), f is a geometric feature of a molecule and k are parameters. The molecular pdf p can be defined by other pdfs that constrain individual distances and angles; the latter can also be defined by sums of pdfs obtained from the individual homologous proteins.

$$c = -\ln p = -\ln \sum_{i=1}^{n} w_i p_i$$,
F XIX

where w_i is a weight constant.

3.16.3 Optimization of the objective function by MODELLER

Modeller implements a Beale restart conjugate gradients algorithm (Shanno and Phua 1980 and 1982) and the variable target function method (VTFM) (Braun and Gõ 1985) to the position of all non-hydrogen atoms. The restraints are sequentially introduced for optimization, resulting in a model that maximally fulfills all restraints. In a further step, the model can be energy-minimized by simulated annealing with molecular dynamics. The CHARMM function includes bond length, bond angle, dihedral angle, Lennard-Jones potential and electrostatic terms. The calculation of the restraints from the template target alignment and the optimization were performed automatically as follows.

Flowchart of comparative modeling by MODELLER

1. Read and check the alignment.
2. Generate atom coordinates, by transfer of equivalent atom coordinates from template and unknown coordinates using CHARMM topology library.
3. Calculate stereochemical, homology-derived and special restraints for the target from its alignment with the template.
4. Calculate a model that satisfies the restraints as well as possible.
   a. Start with the initial model.
   b. Partially optimize the model by applying the VTFM modifying the model by conjugate gradients.
   c. Refine the model by simulated annealing with molecular dynamics.
5. Output of the refined structure and the calculated violation terms.

3.16.4 Ballesteros and Weinstein numbering system for GPCRs

Ballesteros and Weinstein used conserved residues across family A GPCRs in order to define a sequence independent numbering system (BW-numbering), which is applicable for all GPCRs (Ballesteros and Weinstein 1995). This system defines for each helix the highest conserved position as .50, by adding the helix number, for example 1 the BW-numbering results in 1.5. Downstream sequence positions are then numbered increasingly from .50 and upstream positions decreasingly. These conserved positions in family A GPCRs are N1.5 in TM1, D2.5 in TM2, from the (D/E)RY-motif R3.5 in

3. Methods

TM3, P4.5 in TM4, P5.5 in TM5, W6.5 in TM6 and from the NPXXY-motif P7.5 in TM7.

3.16.5 Homology modeling of mGluR

Alignment

The amino acid sequence of mGluR5 was retrieved from the SwissProt database (Bairoch *et al*. 2004), accession numbers P41594 and of bovine rhodopsin from the crystallographic structure with PDB ID 1U19 (Okada *et al*. 2004). Only the structure of the TM part of mGluR5 without the extracellular domain from sequence position 576 to 865 was predicted. The TM helices were aligned in the same way as in the MSA of the mGluR subfamily to BR. For loops, the alignment was performed manually, because of high differences in length and sequence, optimizing a reasonable start conformation for further optimization. The resulting alignment can be found in Figure 16.

Structure modeling

Modeller can deal with several template structures. The structure of BR from PDB ID 1U19 (resolution 2.2 Å, Okada *et al*. 2004) was used as template for homology modeling. In two cases, the secondary structure of helices was enforced to exhibit α-helical structure, because of missing indication of non-standard α-helical structure in mGluR5, which is assumed to be caused by prolines in BR. These prolines are associated with the π-helix in TM5 and the 3^{10}-helix in TM7 in BR. Instead, modeled α-helices, composed of amino acids of mGluR5 (prepared with MOE, Montreal, Canada), were used as templates. The α-helical TM5 and TM7 templates were superposed onto the same TM helices of BR. Special constraints (special_patches) were used for modeling of the disulfide bridge between TM3 and the EC2 (Cys644 and Cys733).

Initial models were constructed with Modeller 9v1 (Šali and Blundell 1993) automodel method and parameter *slow* for optimization with molecule dynamics. Modeller scripts for calculation of the homology models are available in Section 10.2.

4. Results and Discussion

In the present study, aimed at characterization of structure and function of family C GPCRs, different strategies were applied and their results joined in a combined discussion. These approaches included ligand, sequence and structure based methods which results will be introduced in the following.

Family C GPCRs are a protein family of several receptors with different function. They can be activated by different molecules as Ca^{2+}, glutamate, γ-aminobutyric acid, sweeteners and pheromones, which are binding to the extracellular domain. The TM domain possesses a similar seven helical fold compared to family A GPCRs. Allosteric modulators bind in the TM of family C GPCRs analogous to agonist and antagonist of family A GPCRs

Beginning with the conservation analysis of family C GPCRs, structural and functional properties of the TM domain will be discussed relative to their conservation (Section 4.1). A literature survey of published mutagenesis experiments, conducted in the TM region of different receptors, will provide a detailed overview of experimental findings related to structure or function of this domain. Conservation values will be incorporated in structural context using a predicted protein structure of mGluR5 as representative for family C GPCRs. The structural discussion will be expanded by correlation to properties of family A GPCRs and known 3D structures belonging to this family.

In the second part, (Section 4.2) binding properties of allosteric modulators will be analyzed in a prospective virtual screening for novel mGluR5 ligands. The assembled ligand data collection enabled to study properties of ligand binding in mGluR with ligand-based approaches. Active molecules, discovered with computational models, will be discussed in structural context using predicted ligand binding modes. The modeled receptor structure and ligand conformations will bridge to findings from the first part, providing insight into function and structure of family C GPCRs.

4. Results and Discussion

4.1 Conservation analysis of family C GPCRs

The conservation analysis was applied in order to reveal conserved positions in the TM domain of family C GPCRs. As data basis for the analysis a multiple sequence alignment of family C GPCR sequences was generated. Shannon entropy was calculated for each position in the TM domain using the MSA. The function associated with conserved positions was defined using experimental data from studies performed on family A and family C GPCRs. The analysis was aimed at characterization of functional and structural features of family C GPCRs and the mGluR subfamily in particular. For mGluR the binding site was defined using mutation data focusing on NAM/PAM binding regions and not conserved positions important for selectivity. In the following qualitative and quantitative results will be introduced starting with data.

4.1.1 Mutation data collection

Structural features exposed with mutation studies were collected from literature studies of family C GPCRs (references on 17 publications in Section 2.3). Mutation experiments elucidate the influence of particular amino acids on the activation of the receptor. In these studies functional implications of mutations were in each case tested in combination with allosteric modulators of different types (NAM/PAM/agonist), which allows for the definition and characterization of the binding pocket for different ligand types. Furthermore, the ligands can be compared due to different effects on their function, such as affinity changes, loss of modulation effect and activation levels of the receptor, which were caused by the respective mutation. These changes, which occur upon a mutation, can be detected by comparison to functional activity of the wild type receptor or using radioactively labeled ligands in order to track altered binding affinity.

The number of mutations in the compiled mutation data collection summed up to 157 and included four receptor types (mGluR1, mGluR5, CaSR and T1R3) of two different organisms, rat and human. Mutation experiments considering eight NAMs and five PAMs were included in the data collection, where in all cases ligand binding or efficacy has been altered in mutated receptors (Figure 14). The mutation data collection is introduced here, since it is the basis for structural projections and discussions in all following sections. A PML-script visualizing the position of the mutations, which can be applied on a mGluR5 structure using PyMOL, is given in Appendix (Section 10.6).
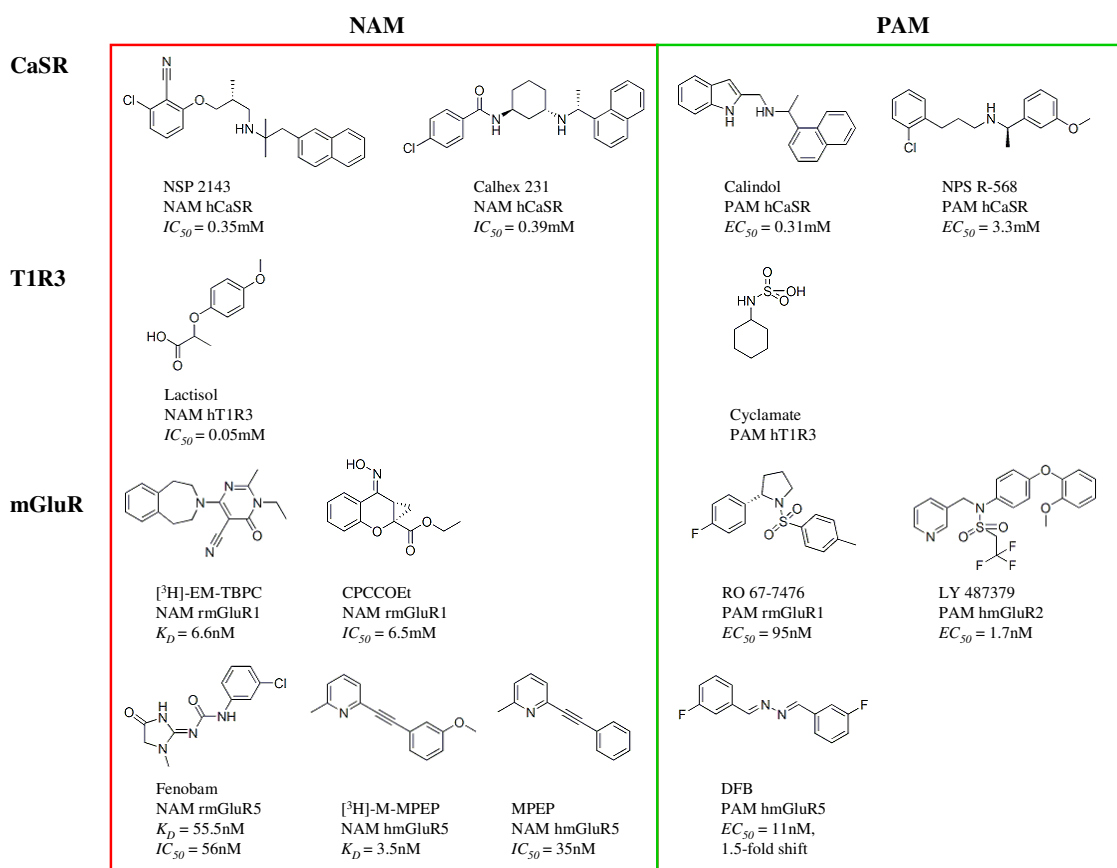
Figure 14: Ligands which were tested in mutation experiments with receptors of family C G Protein-coupled receptors (Muehlemann *et al.* 2006, Petrel *et al.* 2003, Petrel *et al.* 2004, Schaffhauser *et al.* 2003, Hu *et al.* 2002, Knoflach *et al.* 2001, Litching *et al.* 1999, Winning *et al.* 2005, Malherbe *et al.* 2003, Malherbe *et al.* 2006, Jiang *et al.* 2004, Malherbe *et al.* 2003, Pagano *et al.* 2000, Miedlich *et al.* 2004). Modulators of calcium-sensing receptors (CaSR), taste (T1R3) and metabotropic glutamate (mGluR) receptors are classified as negative (NAM) and positive (PAM) allosteric modulators. Test organisms are defined with leading letters h (human) and r (rat).

All mutated positions were aligned according to the sequence alignment of family C GPCRs, which is discussed in the next Section, and could therefore be assigned position numbers according to Ballesteros-Weinstein numbering scheme useful for projection on any GPCR.

4.1.2 Sequence alignment of family C GPCR

This section will start with a description of numerical properties and preparation details of family C sequence data. In order to motivate the final preparation procedure of the MSA of family C GPCRs difficulties experienced in handling and automated processing and the adapted solutions will be discussed. It will include a comparison of the proposed family C MSA to an automatically calculated one available at GPCRDB (v.10.0, Horn *et al.* 2003). A complete MSA of 96 family C sequences is given in Appendix (Figure A

4. Results and Discussion

1). This MSA was essential for further conservation calculations of the TM region and the template target alignment used for homology modeling of the human mGluR5 on BR as template.

A multiple sequence alignment (MSA) was performed based on 96 protein sequences selected from 160 family C sequences accessible from GPCRDB. The MSA consisted of TM helices, while all loop regions were omitted. In the following, a functional group of receptors with a similar functional profile such as mGluRs are referred to as a *receptor subfamily* and different receptors of such a subfamily as *receptor subtypes*, *e.g.* mGluR1, mGluR2. The collected sequences originated from different species and seven subfamilies defined by receptor function (Table 2).

The approach for an alignment of family C sequences to BR was guided by matching of conserved family A sequence positions in TM regions to similar amino acids in the family C sequence (Kratochwil et al. 2005). In the present study a MSA of family C sequences facilitated the alignment to BR sequence. The procedure for construction of a family C-BR TM region alignment consisted of the following steps:

1. separate MSA of each of the family C subfamilies
2. alignment of each subfamily MSA to BR separately
3. extraction of the TM region according to BR helix boundaries for each subfamily-BR-MSA
4. successive joining of subfamily-BR-MSAs by pair wise alignment according to BR sequence, which is included in each of the subfamily-BR-MSAs
   a. if joining successful, meaning the subfamily is properly aligned to the family C MSA assembled so far, then proceed with 4 if any not-aligned subfamilies are left else finish.
   b. if the new subfamily is shifted to the other subfamily MSA, than proceed with step 2 for alignment revision.

MSA of family C subfamilies
The mean identity value of family C subfamilies to BR was calculated to be 14%. The different subfamilies were unequally conserved within the subfamilies ranging from 88% (BOSS) to 11% (TASTE); therefore even functionally similar receptors are diverse in their protein sequences (Table 2). The high dissimilarities might have originated from

68

unequal distribution between species or receptors subtypes included due to data base composition. From evolutional point of view, the diversity between receptors with conserved function makes sense especially, if they are distributed to different tissues or participate in dimerization. Different receptor variants allow then a diverse but adopted controlled mechanism of response.

Table 2: Sequence data statistics retrieved from GPCRDB (Horn *et al.* 2003). Sequence numbers are given for any of the different filtering steps. In all identity calculations only the transmembrane (TM) region was considered. Family C subfamilies are metabotropic glutamate (mGluR), calcium-sensing receptors (CaSR), γ-aminobutyric acid type B (GABA-B), taste (T1R3) and "bride of sevenless proteins" receptors (BOSS). The identity bovine rhodopsin (BR) is the mean identity of all subfamily sequences to BR.

receptor subfamily	sequence number	without variant, hypothetical, similar, related, probable, splice, putative	unique	successfully aligned	identity in subfamily, TM only (%)	identity to BR, TM only (%)
mGluR	53	37	32	31	28	13
CaSR	24	18	18	17	35	14
GABA-B	30	24	23	20	12	11
Orphan GPCR5	11	9	9	9	24	10
Orphan GPCR6	9	8	4	4	49	13
BOSS	4	4	4	4	88	11
TR	12	12	11	11	11	13
Putative pheromone receptors	17	4	4	-	-	-
Human	22	22	20	20		
Total	**160**	**116**	**105**	**96**	-	**14**

Successive MSA of family C sequences to BR

Each family C subfamily MSA was aligned to BR and TM-MSAs were cut out close to rhodopsin helix boundaries according to their alignment to rhodopsin; BR positions - TM1: 38-67, TM2: 72-101, TM3: 110-139, TM4: 141-171, TM5: 198-227, TM6: 249-275, TM7: 285-312. During the process of alignment of the excised MSA blocks, the prepared alignments to rhodopsin could be revisited, so that each subfamily was aligned to both - other subfamilies and to rhodopsin. Due to low sequence identity to BR (Table 2), a particular amino acid in a single family C sequence often did not match a position in BR, but the most common amino acid or feature in the family C MSA more often did. This observation indicated that a single or few sequences can be aligned less reliably to a distantly related protein than a prealigned MSA of more closely related sequences, as

4. Results and Discussion

the family C subfamilies are. The resulting alignment contained TM helices exclusively, without any loops, no extracellular domain and no helix eight (H8) or C-terminal domains (Figure A 1).

The decision for the extraction of individual TM sequence blocks from subfamily alignments according to rhodopsin TM borders, instead of finishing the family C alignment and then align it to BR, was not only practically driven. The homology of each subfamily to BR could be tested when joining separate subfamily MSA aligned to BR, and additionally the whole family C MSA rechecked. The final alignment provided the basis for comparison of corresponding position of family C receptors to BR and through that to family A GPCRs in general.

Trials of automated alignment of family C GPCRs

In the beginning, the application of TMHMM2, a tool for TM region prediction (Krogh *et al*. 2001), failed for several sequences in finding of seven TM regions, therefore it could not be used for automated prediction and assignment of TM sequence parts for alignment. When applying TMHMM2 for the prediction of TM regions of mGluR5 only six TM regions were obtained: positions 582-604, 617-636, 692-714, 737-759, 774-796, 803-825, while TM3 was missing. However, the TM assignments were helpful in many cases as a visual guiding tool for the first whole sequence alignments and the manual extraction of the TM regions. Another failed approach included the initial alignment of subfamilies using ClustalW (Thompson *et al*. 1994) followed by an alignment of the resulted subfamily MSAs to each other. Due to considerable differences in sequence length of receptors lacking the extracellular ligand binding domains and different loop lengths the alignment within subfamilies was always displaced. For this reason TM regions had to be excised, automatically aligned and then manually corrected.

Evaluation of the proposed family C multiple sequence alignment

All comparisons between family C alignments from GPCRDB and the MSA, proposed in this study, were performed considering the same sequences. Only regions present in both MSAs were compared, therefore both MSA versions were aligned to hmGluR5. The conservation calculations with Shannon entropy (*H*) values were applied for evaluation of the quality of the proposed and the GPCRDB supported family C MSA. The comparison showed, that the proposed alignment reached a considerably higher

70

conservation, by mapping more amino acids of the same type (sum of entropy values), than the automatically generated one (Figure 15). In particular TM3 and TM4 contained positions of higher conservation, while all other helices revealed different conserved positions compared to the GPCRDB alignment.
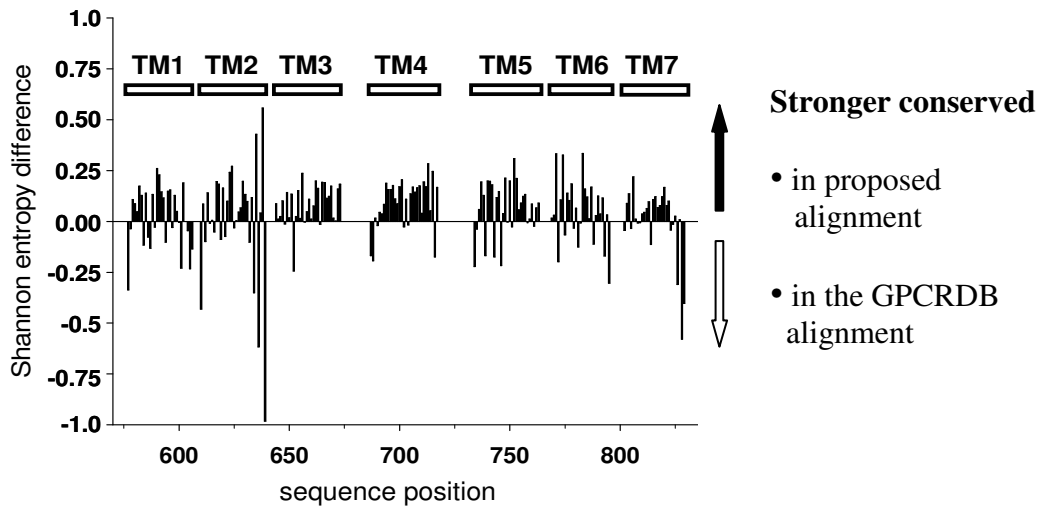
Figure 15: Differences in Shannon entropy (*H)* values calculated for two family C multiple sequence alignments (MSAs) (1. from GPCRDB (v.10.0, Horn *et al.* 2003), 2. proposed in this study) using the nine functional groups. *H* values of both MSAs were subtracted from each other. Negative values indicate stronger conservation in the GPCRDB alignment and positive in the proposed alignment.

The here proposed alignment of family C GPCRs was used for the conservation analysis instead of taking the automatically generated from GPCRDB, since in the new alignment more conserved features matched, indicating a higher reliability (Figure 15). Compared to other published alignments of several family C sequences the new alignment was constructed based on a higher number of sequences and therefore it should be more suitable for a conservation evaluation of the whole family C, since more sequence variability was captured. Besides conserved amino acids in family A GPCRs, as applied for the Ballesteros-Weinstein numbering which is supposed to facilitate the building of an alignment to family A sequences, several different alignments of family C sequences to BR were reported (Jiang *et al.* 2005, Pagano *et al.* 2000, Pin *et al.* 2003, Malherbe *et al.* 2003, Kew 2004, Xu *et al.* 2004, Surgand *et al.* 2006 and Petrel *et al.* 2003). Prior to the new alignment, a comparison of the published alignments revealed major differences in TM4, TM5 and TM7, even for identical sequence pairs thus indicating that despite conserved position different alignments could be proposed.

4. Results and Discussion

4.1.3 Entropy of conserved amino acid features

The application of entropy calculations was aimed to define conserved positions of amino acids within the TM domain and most conserved features at these locations. The nine function amino acid groups allowed for fuzzy sequence descriptions, where chemical or sterical features were used instead of single amino acids. A single amino acid is only described by one feature type, which leads to a considerable reduction of the applied chemical alphabet. Using these groups, the entropy calculation captures the conservation of the defined amino acid types. In case of H, low values indicate high conservation with respect to the amino acid typing scheme.

The calculation of H revealed positions with low, medium and high conservation in family C GPCR TM regions (Table 3). Conserved positions indicated by low H were represented by amino acids: L, I, T, G, C, A, M, V, K and P.

Table 3: Overview over the Shannon entropy (H) values calculated using the nine functional groups applied on the multiple sequence alignment of family C G protein-coupled receptor sequences. For the conserved positions the amino acids of metabotropic glutamate receptor five (mGluR5) are given in parentheses using "single letter code".

Shannon Entropy	Sequence positions and amino acids in mGluR5
$H \leq 0.6$	(G)590, (T)594, (I)620, (L)622, (C)644, (L)662, (K)665, (T)666, (I)669, (I)703, (L)705, (L)750, (A)758, (I)774, (T)777, (M)778, (L)786, (V)806, (P)820, (I)825
$H > 0.6$ & $H \leq 0.9$	587, 591, 602, 610, 613, 617, 647, 657, 660, 672, 696, 707, 711, 749, 770, 783, 784, 787, 791, 814, 821
$H > 2.1$	606, 616, 645, 689, 690, 691, 692, 734, 735, 737, 739, 747, 763, 773, 794

The vertebrate rhodopsin (VR) subfamily (family A) MSA contained closely related sequences, therefore a higher number of absolutely conserved positions was identified with the H calculation (Table 4). The high conservation of VR subfamily members originated from similar receptor function, which was not the case considering a collection of diverse family C GPCRs. The number of conserved positions was also influenced by loops, which could not be aligned in the family C MSA. Loop regions of the VR displayed lower H values, besides few positions like the disulfide bridge formed between TM3 (3.25) and EC2 (45.5) common for most GPCRs.

Table 4: Overview over the Shannon entropy (*H*) calculated using the nine functional groups using the multiple sequence alignment of the vertebrate rhodopsin family, given as sequence positions of bovine rhodopsin (BR).

Shannon Entropy	Sequence positions for BR
$H = 0$	23, 28, 31, 48, 55, 59, 63, 68, 72, 73, 75, 76, 79, 81, 105, 106, 112, 113, 121, 126, 128, 130, 131, 133, 134, 137, 138, 139, 140, 146, 148, 161, 171, 174, 176, 177, 179, 180, 182, 183, 187, 188, 189, 207, 222, 223, 226, 240, 242, 243, 246, 249, 251, 252, 253, 255, 256, 257, 267, 268, 290, 294, 296, 300, 302, 303, 305, 306, 309, 310, 345
low, $H \leq 0.3$	9, 17, 23, 24, 28, 31, 43, 44, 47, 48, 55, 57, 59, 62, 63, 66, 67, 68, 68, 69, 72, 72, 73, 73, 74, 75, 76, 76, 77, 78, 79, 80, 81, 105, 106, 110, 112, 113, 118, 121, 126, 127, 128, 130, 131, 133, 134, 135, 136, 137, 137, 138, 139, 140, 141, 142, 146, 148, 148, 160, 161, 170, 171, 171, 174, 175, 176, 177, 178, 179, 180, 180, 182, 182, 183, 183, 187, 188, 188, 189, 190, 193, 207, 215, 219, 222, 223, 226, 230, 231, 237, 239, 240, 240, 242, 243, 244, 245, 246, 246, 247, 248, 249, 250, 251, 252, 253, 255, 255, 256, 256, 257, 258, 267, 268, 269, 272, 274, 290, 290, 291, 292, 294, 294, 295, 296, 297, 298, 300, 300, 301, 302, 303, 303, 304, 305, 306, 309, 310, 311, 312, 313, 316, 317, 338, 342, 345
medium, $H > 0.3$ & $H \leq 0.5$	82, 103, 114, 115, 119, 130, 202, 216, 229, 234, 254, 263, 309, 314, 318, 340, 343, 347, 348
high, $H > 2$	1, 5, 7, 8, 16, 38, 111, 149, 194, 228, 273, 281, 282, 325, 326, 328

Nevertheless, since all GPCRs possess the heptahelical fold, it is very likely that conserved structural motifs exist, *i.e.* a special residue arrangement necessary for the GPCR activation process. The conserved positions in the VR and the family C MSAs will be discussed in following using BR and the homology modeled structure of mGluR5.

4.1.4 Structure prediction of mGluR5

Based on sequence similarities it has been hypothesized that family C receptors may originate from a fusion of a heptahelical domain of a family A receptor with a periplasmatic binding protein (Pin *et al.* 2003). Therefore, the variety of signaling pathways and orthosteric ligands managed by GPCRs could be enlarged by the extracellular ligand binding site. In the present study the focus lies on the TM domain being the binding pocket for endogenous family A ligands and allosteric family C ligands (Figure 5). In family C GPCRs the TM domain when expressed alone can be activated by allosteric modulators in a similar way as family A GPCRs (Goudet *et al.* 2004).

4. Results and Discussion

This study particularly focuses on structural arrangements and sequential relationships between family A and C. Here, the BR X-ray structure (PDB identifier: 1U19, Okada *et al*. 2004) was applied for homology modeling of mGluR5 and projection of structural and functional features of family C onto the structure of mGluR5 and those of the VR family on BR. From the sequence alignment of mGluR5 to BR and to ß2AD, sequence identity values of 14% and 13% were calculated for both determined structures respectively using only on the TM region (Figure 16). Therefore, BR was used as modeling template and β2AD, bound to carazolol, in the binding site analysis only.

The loop regions were not modeled explicitly since all available templates differ in length and secondary structure, especially for EC2, the extracellular loop connecting TM4 and TM5 and covering the binding pocket of known GPCR structures. In BR the EC2 is folded into β-sheets, it has an α-helical structure in βAD receptors and a coiled structure stabilized by three disulfide bonds in A_{2A} adenosine receptor. In the human A_{2A} adenosine receptor one additional disulfide-bond constraints EC3 exposing the binding site to the solvent (Jaakola *et al*. 2008).

```
                              1.4          1.5
                               |            |
2RH1.A       029  DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQT  066
MGR5_HUMAN   572  YLRWGDPEPIAAVVFACLGLLATLFVTVVFIIYRD--T  607
MGR1_HUMAN   585  YLEWSNIESIIAIAFSCLGILVTLFVTLIFVLYRD--T  620
1U19.A       033  EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRT  070

                          2.5          2.6
                           |            |
2RH1.A       067  -VTNYFITSLACADLVMGLAVVPFGAAHILMKMWT     100
MGR5_HUMAN   608  PVVKSSSRELCYIILAGICLGYLCTFCLIAK----    638
MGR1_HUMAN   621  PVVKSSSRELCYIILAGIFLGYVCPFTLIAK----    651
1U19.A       071  -PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFV    104

                       3.3        3.4        3.5
                        |          |          |
2RH1.A       101  FGNFWCEFWTSIDVLCVTASIETLCVIAVDRYFAIT    136
MGR5_HUMAN   639  PKQIYCYLQRIGIGLSPAMSYSALVTKTNRIARILA    674
MGR1_HUMAN   652  PTTTSCYLQRLLVGLSSAMCYSALVTKTNRIARILA    687
1U19.A       105  FGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVV-    139

                              4.4         4.5
                               |           |
2RH1.A       137  ---------SP-FK-YQSLLTKNKARVIILMVWIVSGLTSFLPIQMH  174
MGR5_HUMAN   675  GSKKKICTKKPRFMSACAQLVIAFILICIQLGIIVALFIMEPPDIMH  721
MGR1_HUMAN   688  GSKKKICTRKPRFMSAWAQVIIASILISVQLTLVVTLIIMEPPMPIL  734
1U19.A       140  ------C--KP--MS-NFRFGENHAIMGVAFTWVMALACAAPPLVG-  174
```

```
                                       45.5
                                         |
2RH1.A      175 WYRATHQEAINCYAEE--TCC--DFF-- 194
MGR5_HUMAN  722 --DYP---SI----REVYLIC--N—-T- 735
MGR1_HUMAN  735 --SYP---SI----KEVYLIC--N—-T- 748
1U19.A      175 WSRY-----I-PEGMQ—-CSCGIDYYTP 194


                     5.4        5.5        5.6
                      |          |          |
2RH1.A      195 ---T-NQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKRQL 230
MGR5_HUMAN  736 ---T-N-LGVVTPLGYNGLLILSCTFYAFKTRNVP----- 765
MGR1_HUMAN  749 ---S-N-LGVVAPLGYNGLLIMSCTYYAFKTRNVP----- 778
1U19.A      195 HEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEA- 233


                              6.4        6.5
                               |          |
2RH1.A      263 --------KFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQD- 300
MGR5_HUMAN  766 -----------ANFNEAKYIAFTMYTTCIIWLAFVPIYFGSNY--- 797
MGR1_HUMAN  779 -----------ANFNEAKYIAFTMYTTCIIWLAFVPIYFGSNY--- 810
1U19.A      234 AAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQG 281


                   7.3        7.4        7.5
                    |          |          |
2RH1.A      301 NLIRKEVYILLNWIGYVNSGFNPLIYCRSPDFRIAFQELLCL-- 342
MGR5_HUMAN  798 KIITMCFSVSLSATVALGCMFVPKVYIILAKPERNVRSAFTTST 841
MGR1_HUMAN  811 KIITTCFAVSLSVTVALGCMFTPKMYIIIAKPERNVRSAFTTSD 854
1U19        282 SDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCG 324
```

Figure 16: Sequence alignment of ß₂AD (chain A, PDB ID: 2RH1), the human mGluR5 (SwissProt-ID: P41594) and BR (chain A, PDB ID: 1U19). Identical residues for at least two sequences are highlighted in bold letters. Alpha-helical regions are given with grey background. Additional to sequence positions the Ballesteros-Weinstein notation is given above the alignment. The conserved disulfide-bond connecting EC2 and TM3 is highlighted in blue. Positions tested in mutagenesis studies on family C GPCRs affecting the binding of positive allosteric modulators are coloured green, of negative allosteric modulators in red and both types in yellow. Red letters indicate different amino acids between hmGluR1 and hmGluR5.

4.1.5 Projection of Shannon entropy values on a GPCR structure

The modeled mGluR5 structure provided a basis for conservation analysis in 3D context, therefore conserved features (Figure 18) and values (Figure 17) were projected onto receptor structure. Based on the sequence alignment used for homology modeling of mGluR5 on BR as template conservation of sequence positions between family A and family C was compared. The mutation data collection (Table A 3), which was compiled for amino acids referenced in literature, served as information source for binding site analysis, as discussed in Section 4.1.7.

4. Results and Discussion
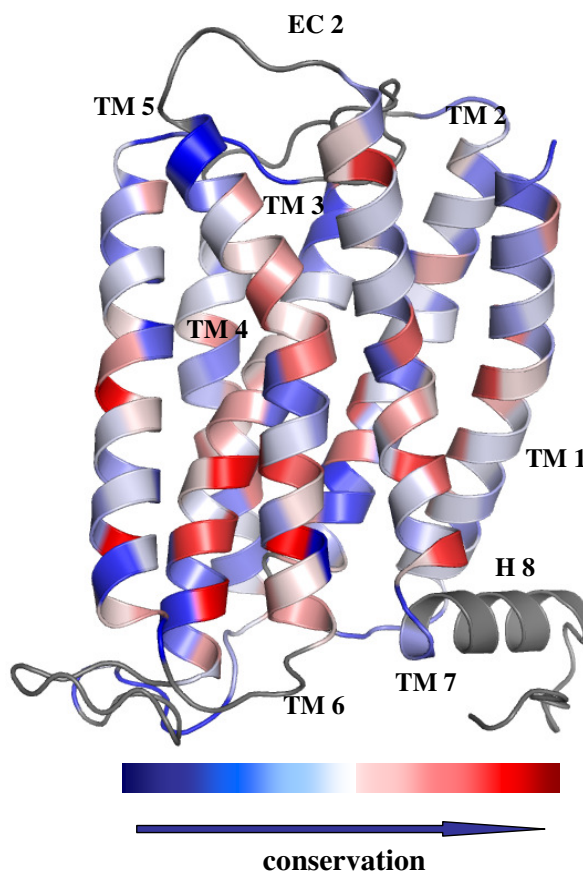
Figure 17: Projection of Shannon entropy values calculated using the nine functional groups. The colors define the level of conservation according to the color bar, from low (blue) to high (red) conservation. Grey regions were considered for calculation. The calculation was based on a family C G protein-coupled receptor alignment. The structure of the metabotropic glutamate receptor five was prepared by homology modeling using bovine rhodopsin as template.

The projection of Shannon entropy values as a color scheme onto the structure of a GPCR allowed to detect the location of conserved position in three dimensional context (Figure 17). The comparison of family C to vertebrate rhodopsins (VR) revealed several similarly conserved and positioned types as prolines (TM4, TM6 and TM7) involved in helical kinks, as well as a basic cluster at the ends of TM3 and TM6.

The most conserved features of the family C MSA were projected onto mGluR5 (Figure **18**, A) and the of the vertebrate rhodopsin family onto BR structure (Figure **18**, B). Aromatic, aliphatic, neutral and small amino acids, which are more common in membrane environment (Jones *et al.* 1994), were detected with Shannon entropy and nine functional groups as most frequently conserved types in the TM.
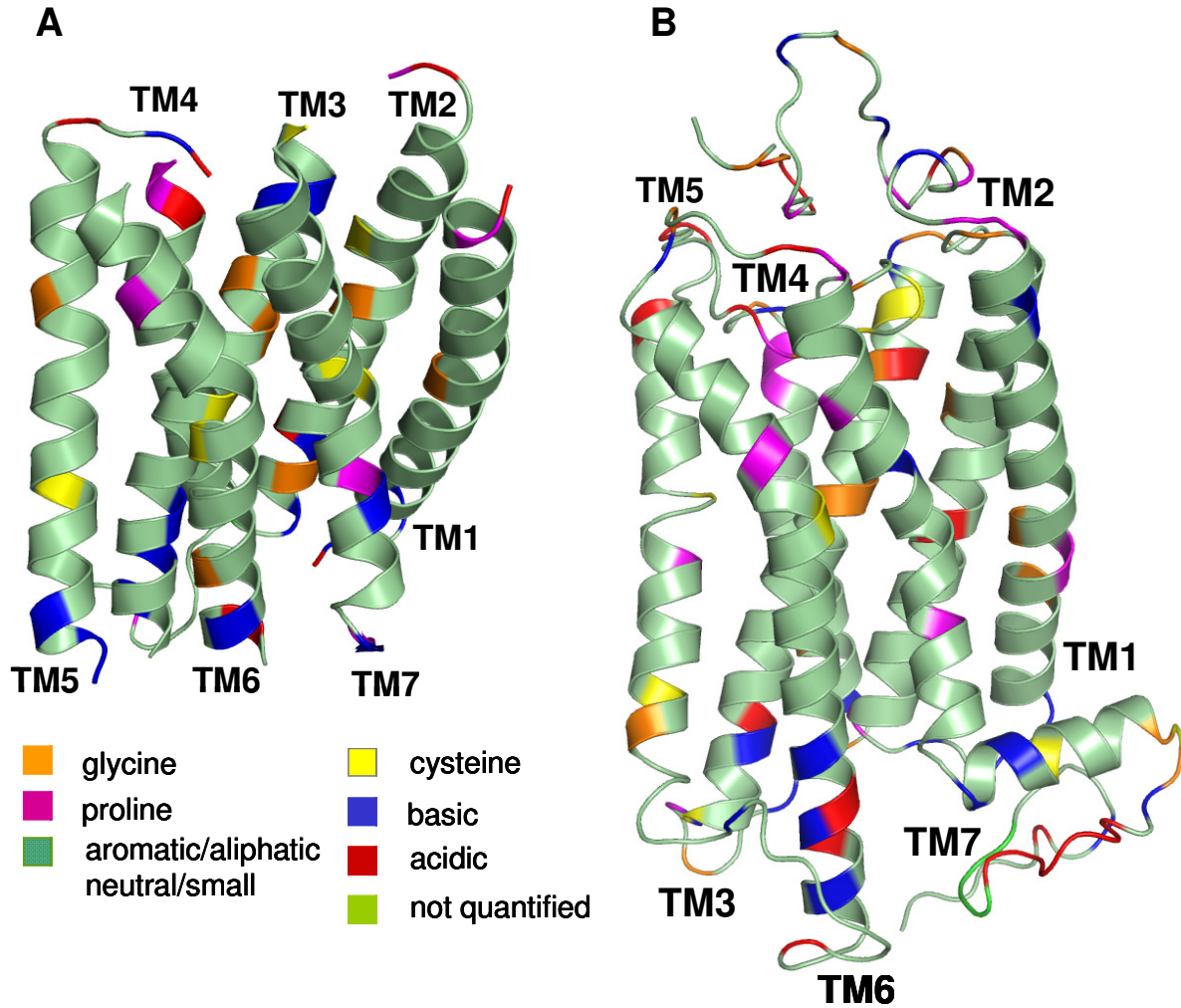
Figure 18: Most conserved features of A) family C MSA projected onto the modeled structure mGluR5 and B) family A/(Rhod)opsin/Vertebrate rhodopsin MSA projected onto the structure of BR. Nine functional amino acid groups were used for sequence encoding. These groups are condensed to six color representations as described in the legend, the seventh color represents not quantified positions.

For eight of nine amino acid types an equivalent most conserved type was discovered in both MSAs, family C and VR (Figure 19). The implication of these positions is discussed in the following sections.
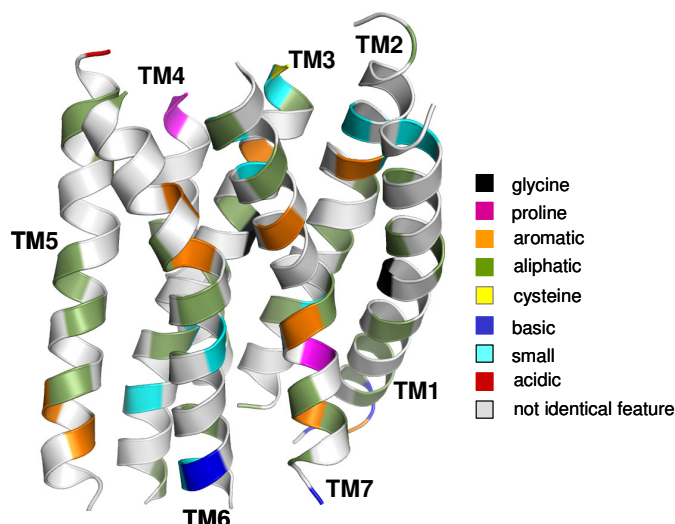
Figure 19: Projection of the most conserved amino acid groups at corresponding positions identical in the transmembrane region (TM) of family C and family A (VR) G protein-coupled receptors. Identical positions are shown in colors representing the special amino acid types (eight types as given in the legend, "neutral" is missing, since identical positions were not observed for that type) and in light grey the not identical features. Not aligned positions are omitted.

4.1.6 Conservation of structural and functional features

Based on the idea that structural motifs are more conserved than sequences (Holm and Sander 1996, Hubbard *et al*. 1997, Rost 1999), GPCR common sites were evaluated with respect to conservation in family C. Several different structural and functional receptor features, helix-helix contacts, dimerization, function, motifs and the ligand binding site, were analyzed and compared for the vertebrate rhodopsin and family C GPCRs. The mutation data collection (Table A 3) was applied for experimental details on family C, an equivalent collection published by Madabushi completed the data for family A GPCRs (Madabushi *et al.* 2004) as several studies focusing on a particular GPCR feature.

Helix-helix Contacts

In the homology model of mGluR5 (Figure 20, A) 10 out of 20 highly ($H \leq 0.6$) conserved family C positions were identified at helix-helix contact areas; two residues, 3.25 and 7.36, are close to EC2 loop. These 10 residues with Shannon entropy values below 0.6 (Table 3) were G590 (1.46) and L622 (2.51) between TM1 and TM2, I620 (2.49) from TM2 pointing towards TM3, also L662 (3.43), I669 (3.46) and K665 (3.5) in TM3 in contact to I774 (6.37) and M778 (6.41) in TM6 near the intracellular side and A758 (5.58) from TM5 facing T666 (3.47) in TM3. The disulfide-bridge to EC2 formed by residue C644 (3.25) in TM3 was also conserved.
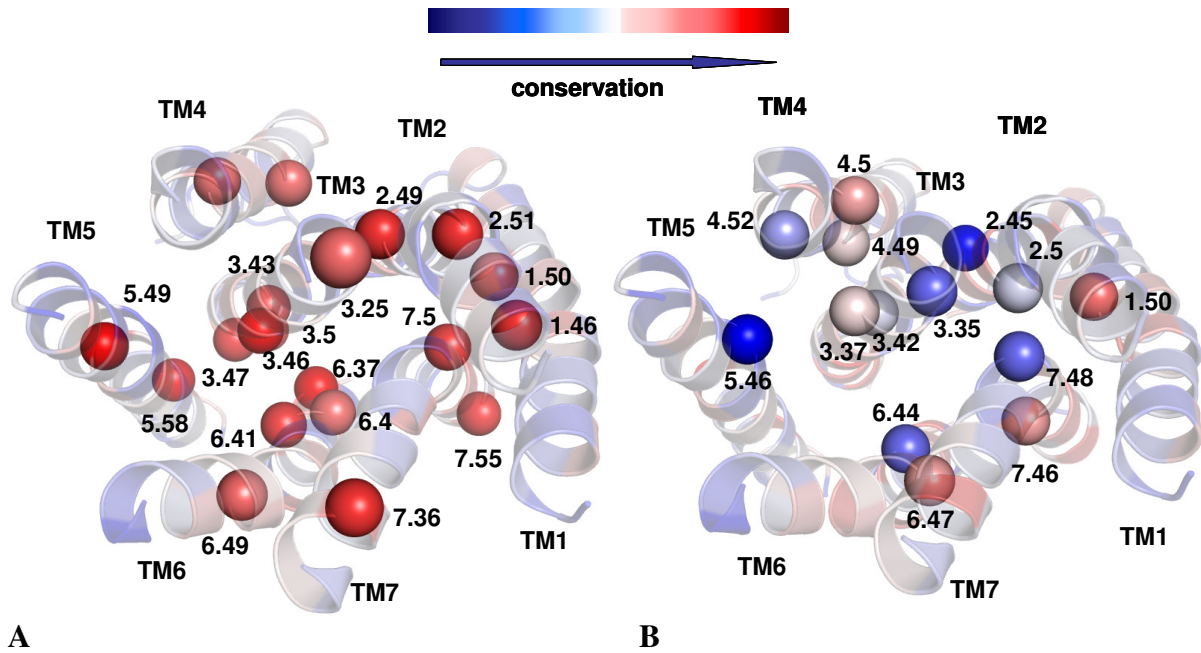
Figure 20: Projection of Shannon entropy values (*H*) calculated with nine functional groups onto the transmembrane region (TM) of mGluR5. Cα-atoms of residues are presented as spheres and colored according to *H* conservation values for the family C G protein-coupled receptors (GPCR) multiple sequence alignment. **A)** Positions with $H \leq 0.6$ (Table 3). **B)** *H* values of positions at family A GPCR helix-helix contacts (Palczewski *et al.* 2000, Teller *et al.* 2003). Positions are matched according to the bovine rhodopsin to family C GPCR alignment (Figure 16).

A different picture emerged when helix-helix contacts from rhodopsin (Palczewski *et al.* 2000, Teller *et al.* 2003) were analyzed on the mGluR5 structure regarding their conservation in family C GPCRs (Figure 20, B). Rhodopsin contact sites did not correspond to conserved family C positions. In the conformational state of the inactive rhodopsin, with retinal in the binding site, other residues form close contacts.

Helical packing plays a crucial role in the folding and oligomerization of integral membrane proteins (Russ and Engelman 2000). The conserved helix-helix contacts define a mutual arrangement of helices in TM domains (Chugunov *et al.* 2007). The location of the most conserved positions in family C GPCRs lead to the conclusion that helix-helix-contacts are conserved regions facilitating the helical packing of the TM.

Dimerization

Prediction of dimerization sites for GPCRs is still a challenging task, as it can be seen from modeling and experimental studies suggesting different possible dimerization interfaces for rhodopsin involving helices I/II/VII or IV/V/VI (Filipek *et al.* 2004, Filipek 2005, Filizola and Weinstein 2005, Salom *et al.* 2006). Using Shannon entropy

4. Results and Discussion

conservation analysis for comparison of these proposed interfaces no preference of a more common interface could be observed (Table 5).

Table 5: Shannon entropy (H) of residues proposed as important for receptor dimerization of rhodopsin, from modeling studies ([1] Filipek *et al.* 2004, [2] Filipek 2005) and from experimental results ([3] Filizola and Weinstein 2005, [4] Salom *et al.* 2006). Positions are given for the metabotropic glutamate receptor five (mGluR5) and bovine rhodopsin (BR). The most conserved feature according to the nine functional groups is given for the family C GPCR and the vertebrate rhodopsins (VR).

Position in mGluR5	H for family C	Most conserved group in family C	Position in BR	H for VR	Most conserved group in VR	Reference
584	1.67	aliphatic	F45	1.35	aromatic	D (4)
588	1.67	aliphatic	M49	1.23	aliphatic	D (4)
591	0.79	aliphatic	F52	1.61	small	D (4)
598	0.95	aliphatic	L59	0	aliphatic	D (3)
634	1.98	aliphatic	Y96	1.88	aromatic	D (4)
638	1.91	acidic	H100	1.88	aromatic	D (4)
698	2.03	aliphatic	H152	1.24	basic	D (2)
701	1.03	aliphatic	M155	1.31	aliphatic	D (2)
704	1.18	neutral	A158	1.47	small	D (3)
705	0.56	aliphatic	F159	1.13	aromatic	D (3)
708	1.78	aliphatic	V162	0.95	aliphatic	D (2)
712	1.14	aliphatic	A166	1.77	small	D (3)
715	2.02	acidic	A169	1.35	small	D (3)
717	1.68	proline	P171	0.02	proline	D (3)
735	2.42	basic	D199	1.46	neutral	D (1)
737	2.52	acidic	E201	1.32	acidic	D (2)
738	1.36	aliphatic	S202	0.38	small	D (1)
779	1.39	aromatic	I259	1.28	aliphatic	D (3)

Filipek proposed that S202, D199 and E196 of two monomers are involved in receptor dimerization in rhodopsin (Filipek *et al.* 2004, Filipek 2005). According to the mGluR5 model the end of EC2 connecting to TM5 contains three residues, N734, N737 and T742, which could be part of the proposed interface between TM5 and TM4. In case that TM5 is involved in dimerization, residues at TM4 pointing away from the receptor could be regarded as potential interacting points.

For family C receptors it is known (Pin *et al*. 2003) that dimers could be linked together by a disulfide bond through one or two cysteins located in the fifth loop of the VFT. Remelli and co-workers showed that the C-terminus of the metabotropic glutamate receptor 1b regulates the dimerization of the receptor (Remelli *et al.* 2008). Since neither the C-terminus nor the VFT were part of the conservation study, these findings could not be evaluated. With the exception of the reported heterodimerization of the mGluR1 and calcium sensing receptor, there has been no evidence for the heterodimerization of other mGluRs, in fact mGluR1 and mGluR5 do not form heterodimers when co-expressed in cells (Romano *et al.* 1996, Robbins *et al.* 1999, Gama *et al.* 2001).

Regarding family A GPCR dimerization, associations between TM1 and TM8 has been found in rhodopsin crystal structures, the physiological relevance of this contact remains unproven (Cherezov *et al*. 2007). For family A GPCRs, hypotheses involving TM6 in the $ß_2$AD (Hebert *et al*. 1996) and TM4 in the D_2 dopamine receptors (Guo *et al*. 2003) describe possible dimer interfaces, respectively.

However no support for a conserved dimerization site based on the here proposed type of conservation analysis could be identified for any of the reported sites. Taking into account that residues in N- and C-terminal domains were experimentally tracked for involvement in receptor dimerization and the fact that functionally distinct GPCRs could form homo as well as heterodimers, it might be unreasonable to expect one absolutely conserved dimerization region in the TM domain.

Receptor activation

Receptor activation data was correlated to H values calculated for VR subfamily as representatives of family A GPCRs. 16 mutations causing constitutive activity taken from the Madabushi collection (Madabushi *et al*. 2004) of family A GPCRs were compared with conservation values calculated on the vertebrate rhodopsin subfamily (Table 6). Besides two moderately conserved residues (2.53 and 3.35) in the vertebrate rhodopsin family out of 16 considered, all other positions were found to be highly conserved ($H \leq 0.6$) indicating that residues at these positions play an important role in receptor activation (Madabushi *et al.* 2004). These positions were further investigated in the MSA of family C GPCRs (Table 6).

4. Results and Discussion
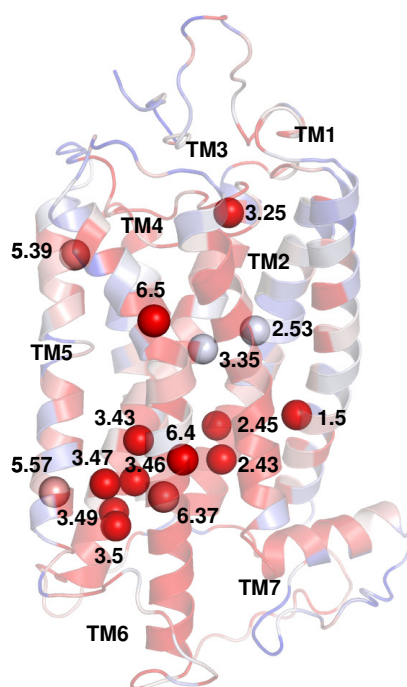
Figure 21: Residues described by Madabushi and co-workers experimentally determined to be important for receptor activation (Madabushi *et al.* 2004). If mutated, the receptor turned out to be constitutively active. Structure of BR colored according to conservation calculated with H for the vertebrate rhodopsin subfamily MSA, Cα-atoms of concerned residues are shown in sphere representation. Residues are given in Table 6.

In family C GPCRs, four residues (2.43, 2.45, 2.53, 3.35), which were not highly ($H > 0.6$) conserved, were exclusively located at contact regions of TM3 to TM2 and to TM4; G120 is also low conserved ($H = 1.35$) in family A GPCRs. Nine of the 16 considered positions were also conserved ($H \leq 0.6$) in family C GPCRs, especially seven (3.43, 3.46, 3.47, 3.50, 5.57, 6.37, 6.40) at the intracellular end of TM3, TM5 and TM6 in contact to each other, 1.5 at TM1 close to TM2 and 3.25 at TM3 in contact to EC2.

From the positions described in Table 6 two have been proven so far to be important for function of family C GPCRs (Table A 3). The human response to lactisol could be introduced into the rat T1R3 by a mutation L740F (5.39) (Winning *et al*. 2005). S787F (6.5) is a naturally occurring mutation increasing calcium sensitivity for hCaSR (Hu *et al*. 2005). From the highly conserved ($H \leq 0.6$) residues of family C GPCRs shown in Figure 20, other than these nine residues might be important for receptor function.

Table 6: Residues described by Madabushi and co-workers experimentally determined to be important for receptor activation (Madabushi *et al.* 2004). If mutated, the receptor displayed constitutive activity. *HV* stands for Shannon entropy value calculated using the nine functional groups for amino acids. For each position the most conserved feature is given. Positions highly ($H \leq 0.6$) conserved in family C and in family A GPCRs are highlighted in bold.

BW numbering	Residue in mGluR5	*HV* for family C	Conserved group in family C	Position in BR	*HV* for VR	Conserved feature in VR
1.50	**T594**	**0.6**	**small**	**N55**	**0**	**neutral**
2.43	S614	2.05	glycine	L76	0.02	aliphatic
2.45	E616	2.45	acidic	N78	0.04	neutral
2.53	G624	1.44	glycine	M86	1.42	aliphatic
3.25	**C644**	**0.54**	**cysteine**	**C110**	**0.04**	**cysteine**
3.35	S654	2	glycine	G120	1.35	glycine
3.43	**L662**	**0**	**aliphatic**	**L128**	**0**	**aliphatic**
3.46	**K665**	**0.4**	**basic**	**L131**	**0**	**aliphatic**
3.47	**T666**	**0.48**	**small**	**A132**	**0**	**small**
3.49	R668	1.87	basic	E134	0	acidic
3.50	**I669**	**0.09**	**aliphatic**	**R135**	**0.02**	**basic**
5.39	V740	0.92	aliphatic	V204	0.62	aliphatic
5.57	**A758**	**0.28**	**small**	**C222**	**0.9**	**cysteine**
6.37	**I774**	**0**	**aliphatic**	**V254**	**0.52**	**aliphatic**
6.40	**T777**	**0.59**	**small**	**M257**	**0**	**aliphatic**
6.50	A787	0.86	small	P267	0	proline

G protein-coupling

G protein-coupling is major part of the receptor activation, since it is responsible for signal propagation inside the cell. Depending on the G-proteins different pharmacological processes can be activated, therefore special G-proteins and/or groups of G-proteins interact with particular receptors or receptor oligomers (Jacoby *et al.* 2006, Kew and Kemp 2005). For selective G-proteins binding, a protein-protein interface between the receptor and the G-proteins is needed (Scheerer *et al.* 2008). Several positions which are involved in G-protein coupling were proposed from modeling studies (Madabushi *et al.* 2004, Filipek *et al.* 2004) and some revealed by experimental tests (Palczewski *et al.* 2000, Acharya *et al.* 1997, Madabushi *et al.* 2004).

According to our conservation calculations, all residues summarized in Table 7 are highly conserved ($H \leq 0.6$) in the vertebrate rhodopsin subfamily. Nevertheless, only few correspond to similar conserved types in family C, which could be due to interactions with completely different G-proteins in comparison with family A GPCRs or uncertainties in intracellular site alignment and modeling. The opposite was observed

4. Results and Discussion

for the metabotropic glutamate receptor subfamily (mGluR), there 13 of 24 reported positions are absolutely conserved, with H values equal to zero.

Table 7: Residues which were reported as important for G protein-coupling and the Shannon entropy (H) values calculated for the metabotropic glutamate receptor (mGluR) and the vertebrate rhodopsin subfamilies ((1) Palczewski *et al.* 2000, (3) Madabushi *et al.* 2004, (4) Filipek *et al.* 2004, (8) Acharya *et al.* 1997). GP stands for proposed and G for experimentally validated changes on G protein-coupling.

Ballesteros Weinstein numbering	H for mGluR	H for family C	H for vertebrate rhodopsins	Reference
1.63	0.97	2.23	0.04	GP (4)
2.39	**0**	0.8	0.02	G (1, 3)
2.40	**0**	1.04	0.02	G (3)
2.45	**0**	2.45	0.04	G (3)
3.49	0.48	1.87	0	G (3)
3.5	**0**	0.09	0.02	G (3)
3.51	1.23	2.09	0.02	G (8, 3)
3.52	**0**	1.84	0.2	G (8)
3.53	**0**	0.88	0	G (1, 3)
3.54	0.75	1.3	0	G (8,1)
IC2	0.41	2.06	0.02	GP (4)
IC2	0.75	2.53	0.82	GP (4)
4.36	**0**	1.63	1.02	GP (4)
4.37	**0**	1.63	0.02	GP (4)
5.58	0.96	1.42	0	G (3)
5.61	**0**	0.95	0	G (1)
6.32	**0**	1.04	0	G (8)
6.33	**0**	0.89	0.02	G (8, 1)
6.34	**0**	0.92	0	G (8)
6.36	1.47	2.32	0	G (1)
6.37	**0**	0	0.52	G (3)
7.53	0.66	1.59	0	G (3)
7.57	1.34	2.11	0	G (3)
7.59	0.2	1.79	0.11	GP (4)

The new structure of squid rhodopsin (2Z73, Murakami and Kouyama 2008) revealed great differences in the helix lengths compared to the bovine form, which were explained by the authors as essential for different light absorption capabilities of the to rhodopsin types. Therefore, it can be concluded that the G-protein coupling interface is optimized in distinct signal transduction pathways and might be only conserved between receptors binding to similar G-proteins. The correlation of G-protein interfaces to signaling molecules it outside the scope of present analysis.

The intracellular side of the TM3 domain which interacts with the G protein contains the "DRY"-motif (Scheerer *et al.* 2008), named after the conserved amino acids at

positions 3.49, 3.50 and 3.51. According to our alignment the DRY-motif seems to be not conserved in family C GPCRs. Instead, a BasicXXBasicXXBasic-motif was found conserved at positions from 3.46 to 3.52 with hydrophobic or neutral residues in-between. Only the first basic position was preserved in family C receptors, both other basic positions were acidic for taste receptors. The mGluR subfamily showed a conserved **KXXRXXR** amino acid cluster. The DRY-motif is part of the *ionic lock* proposed to stabilize the inactive conformation of BR (Scheerer *et al*. 2008). The ionic interaction in BR is established between Arg135 and Glu134 from the conserved E(D)RY motif in TM3, and the side chains of Glu247 and Thr251 in TM6. According to the modeled structure of mGluR5, close contacts between residues in TM3 and TM6 could be observed: Lys665, Arg668 in TM3 and Glu770, Tyr773 in TM6 (Figure 22).
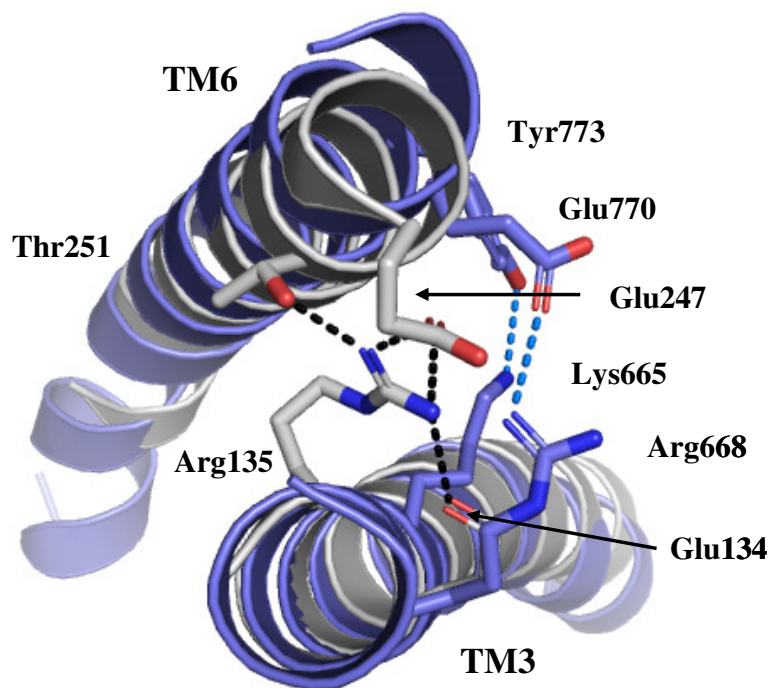
Figure 22: Superposition of bovine rhodopsin (BR, grey) with the modeled structure of mGluR5 (blue) focusing on the *ionic lock*. All helices besides TM3 and TM6 all loops are omitted for clarity reasons. Hydrogen bonds are indicated with dashed lines in black for BR and blue for mGluR5.

In GPCRs, where constitutive activity could be observed (Cherezov *et al.* 2007, Rasmussen *et al.* 2007), no *ionic lock* is proposed to stabilize the ground state. A_{2A} adenosine and ßAD receptor structures revealed that the DRY-sequence participates in interactions to the IC2 loop. The IC2 loop is helically structured in A_{2A} adenosine and $ß_1AD$ receptors and the presence of that helix correlates with constitutive activity profiles (Jaakola *et al.* 2008). Therefore, the modeled structure might be biased towards interactions insight the DRY-region which are present in the template structure.

4. Results and Discussion

Nevertheless, charged side chains are in close proximity in the mGluR5 model and might be a comparable cluster which links TM3 and TM6 together, both helices are involved in receptor activation of GPCRs (Crocker *et al*. 2006, Altenbach *et al.* 2008).

In general, TM3 revealed the most inconvenient amino acid types according to common amino acid distributions proposed by Jones for TM regions (Jones 1994). The reason for that is the localization of TM3 inside the helix bundle with limited contacts to the lipid bilayer.

4.1.7 Conservation of the ligand binding site

Ligand binding in the TM domain of family C GPCRs

For the definition of the binding pocket information from mutation studies conducted in the TM of family C GPCRs was used (Table A 3). This information included the particular mutation, the tested ligand type (NAM/PAM, Figure 14) and the impact of a mutation on ligand effect compared to the wild-type receptor. These positions were considered important for ligand binding in family C GPCRs and evaluated with Shannon entropy values. Additionally a comparison was made to known family A binding sites (Table 8). A position was considered to be important for ligand binding, if at least one ligand was affected in binding or effect by a mutation at this position. It should be noted that mutations at the same position have different effect on the tested ligands. A total of 36 different positions were considered for binding site analysis, they were composed from the mutation data collection and the binding site of carazolol. According to changes tested with mutations 25 positions are involved in allosteric modulation and 12 are essential for functional activation by the agonist.

Residues which were proposed to be involved in interaction with carazolol (Cherezov *et al.* 2008, supplementary material), a diffusible ligand, were analyzed with respect to conservation values of family C GPCRs and the mGluR subfamily and mutation data (Table 8). According to the conservation calculation for the mGluR subfamily out of 16 highly conserved ($H \leq 0.6$) positions in mGluR eight are also conserved ($H = 0$) in VR subfamily. 11 out of the same 16 positions could be proven my mutagenesis studies to be important for allosteric modulation by ligands for family C GPCRs. These results are in agreement to the fact of similar location for the allosteric binding sites of family C GPCRs and the endogenous binding pocket of family A GPCRs.

Table 8: Mutations tested on family C GPCRs with respect changes in ligand effect of different ligand types (PAM, NAM, agonist). All ligands used for these studies are presented in Figure 14. Conservation values calculated using Shannon entropy (H) using the nine functional groups (overall H ranges from 0 to 2.55) as well as the most conserved group are presented for: family C, the metabotropic glutamate receptor (mGluR) subfamily and the vertebrate rhodopsin (VR). Interacting residues of the ß$_2$AD with carazolol (Cherezov *et al.* 2007), an inverse agonist, are included for comparison.

BW numbering	Bovine rhodopsin position	H for VR	Effected ligand type in family C	Conserved group family C/ mGluR	H for family C	H for mGluR	ß$_2$AD interactions with carazolol
2.64	T97	1.2	PAM, NAM	Aromatic/ Aliphatic	1.12	0.66	
3.27	L112	0.7	NAM	Aliphatic	1.42	1.14	
3.28	E113 (retinal)	**0**	PAM	Basic	0.71	1.0	W109 hydrophobic
3.29	G114	0.55	NAM	Basic	2.11	**0**	
3.32	A117	1.41	NAM	Aliphatic	1.29	0.22	D113 h-bond
3.33	T118	0.06	NAM	Glycine	1.54	**0**	D114 hydrophobic
3.36	G121	**0**	PAM, NAM, agonist	Aromatic	1.2	1.81	V117 hydrophobic
3.37	E122	1.18	agonist	Small	1.05	0.64	T118 hydrophobic
3.39	A124	0.77	PAM, NAM	Cysteine/Small	1.82	1.54	
3.4	L125	0.78	NAM	Aromatic	0.91	**0**	
4.45	W161	**0**	PAM	Small	1.9	1.1	
4.46	V162	0.95	PAM	Aliphatic	0.5	**0.35**	
5.32	Y191	-				-	F193 hydrophobic
5.38	F203	**0**				1.14	Y199 hydrophobic
5.39	V204	0.62	agonist	Aliphatic	0.92	0.8	
5.42	M207	**0**	NAM	Aliphatic/Small	2.06	1.6	S203 h-bond
5.43	F208	0.93	PAM, NAM	Aliphatic	1.52	1.01	
5.44	V209	1.28	PAM	Glycine	1.46	1.25	
5.46	H211	1.1	PAM, agonist	Small	2.43	1.84	S207 hydrophobic
6.44	F261	**0**	PAM, NAM	Small	1.84	**0**	
6.48	W265	**0**	PAM, NAM	Aromatic	1.14	**0**	W286
6.49	L266	1.57	agonist	Aliphatic	0.53	**0**	
6.5	P267	**0**	agonist	Small	0.86	**0**	
6.51	Y268	**0**	PAM, NAM, agonist	Aromatic	1.24	**0**	F289
6.52	A269	**0.02**	agonist	Aliphatic	1.18	**0**	**F290** hydrophobic
6.53	G270	1.70	agonist	Proline	1.35	**0**	
6.54	V271	1.33	agonist	Aliphatic	0.88	**0**	
6.55	A272	0.2	NAM	Aromatic	1.66	**0**	N293 hydrophobic
7.32	P285	1.70	PAM, NAM, agonist	Aliphatic	1.66	0.6	

4. Results and Discussion

7.33	I286	1.42	PAM, NAM	Small	1.71	1.0	
7.35	M288	1.0	NAM	Small	1.28	**0**	Y308 hydrophobic
7.36	T289	**0**	NAM	Aliphatic	**0.52**	**0**	
7.37	I290	0.1	PAM, NAM	Aliphatic/Small	1.45	0.4	
7.39	A292	**0.02**				**0**	N312 h-bond
7.4	F293	0.90	NAM	Small	1.64	1.22	
7.43	K296 (retinal)	**0**	agonist	Aliphatic/small	1.56	**0**	Y316 h-bond

It is possible that different ligands bind within the TM region of family C GPCRs in different binding modes like in the case of family A GPCRs, the adenosine (Jaakola *et al.* 2008) and the ß$_2$AD receptors (Cherezov *et al.* 2007). Chen discovered for mGluR group I that CPPHA does not interact with the receptor the same way as MPEP (Chen *et al.* 2008) and emphasized even the hypothesis that different binding modes might trigger different signal transduction, since it is known that mGluR5 can couple to multiple G protein types (Pin *et al.* 2003). Depending on the mutation and the ligand, a single mutation could have different or even opposite effects on ligand binding for the same receptor, as it is the case for 3.36, 3.4, 6.51, 6.55, and 7.4. Amino acid positions in the TM crucial for interaction with allosteric modulators according to mutagenesis studies in mGluRs are highlighted in Figure 23.
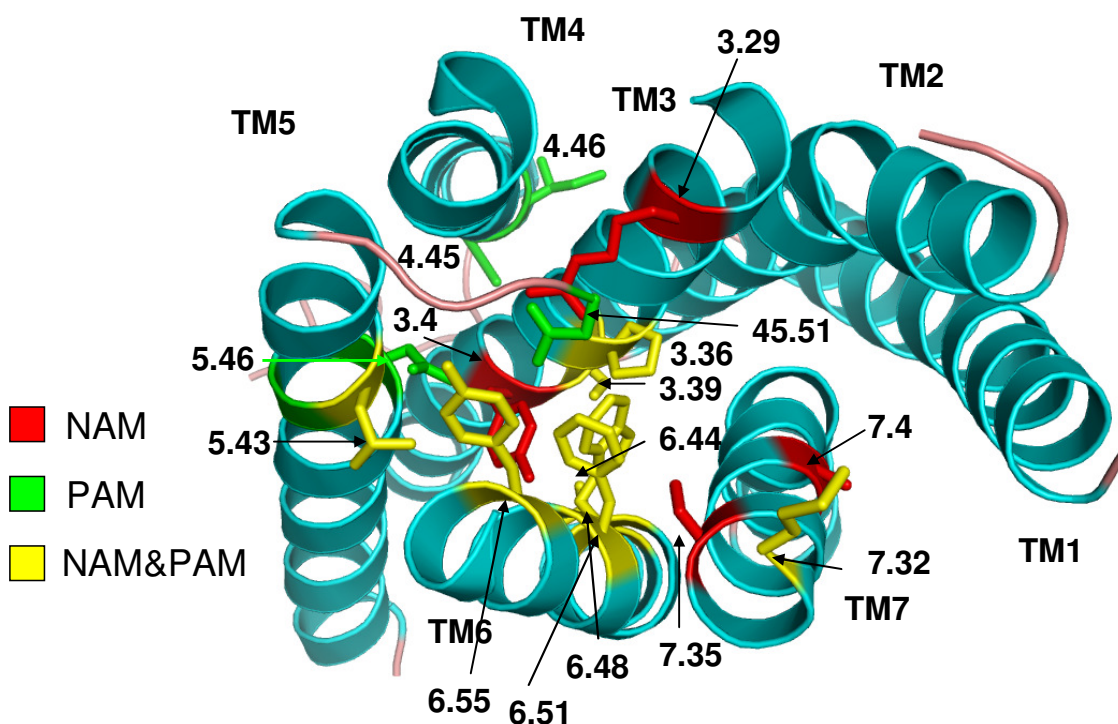
Figure 23: Mutated positions of mGluR receptors, which show impact on ligand binding or allosteric effect: NAM (12), green: PAM (12), yellow PAM and NAM (8). The homology modeled structure of mGluR5 with view from the extracellular site is shown in cartoon representation. Overview about mutated positions is given in Table 8.

The importance of several aromatic residues in the TM cleft was discovered with mutagenesis studies (Malherbe *et al.* 2003, Malherbe *et al.* 2006). There, it could be shown that the binding of fenobam, DFB and MPEP (mGluR5) as well as EM-TBPC (mGluR1) depend on these aromatic amino acids in TM6: 6.48, 6.51, 6.55 and TM3: 3.4. Position 6.48 is essential for NAM binding (Figure 23). There a W784F mutation in mGluR5 reduces and W784A abolishes fenobam binding (Malherbe *et al.* 2006). Additionally, a mutation at that position (W784A or W784A in mGluR5) can increase PAM effect of DFB (Muehlemann *et al.* 2006). It seemed that the "aromatic cluster" in TM5 and TM6, F261 (6.44), W265 (6.48) and Y268 (6.51) might contribute less to the selectivity of ligand binding than to a general activation mechanism in mGluRs, since NAMs and PAMs equally depend on these positions and the position were conserved.

The conservation analysis revealed that positions 3.4, 6.48, 6.51 and 6.55 are conserved (H=0) in the mGluR subfamily, but not conserved overall in family C. The GABA-B and the Taste subfamilies contain sequences, which have different amino acid types at these positions (Figure A 1). Here, the conservation analysis pointed out a difference between mGluR and the GABA-B and Taste subfamilies as part of family C GPCRs.

Selective ligand binding

Family C GPCRs bind their endogenous ligands in the extracellular VFT and no endogenous ligands are known for the TM region. Therefore, in contrast to family A GPCRs, one might assume that residues in the TM binding site are less conserved, as it would be in case of homologous receptors binding their endogenous ligands. Subtype selective ligand binding facilitates the reduction of side-effects in medical treatment. Not conserved positions which were discovered in mutation studies as contributing to ligand binding should be considered as promising sites for selective ligand binding. Changing amino acid type at position 7.32 abolishes ligand binding for different receptors (rmGluR5a, hCaSR, hmGluR1b, rmGluR1a and hT1R3), having different non conserved amino acids with a special interaction for each receptor (Malherbe *et al.* 2006, Muehlemann *et al.* 2006, Jiang *et al.* 2005, Hu *et al.* 2005). This position is not conserved in family C (H=1.66), which points out its importance in selective ligand recognition.

4. Results and Discussion

Conservation calculation for the metabotropic glutamate receptor subfamily (50%) revealed higher conservation than for the whole family C (8%). TM3 and TM6 undergoing conformational changes during activation (Altenbach *et al.* 2008) possess the highest number of conserved positions. All mutations from the mutation data collection which has been tested at mGluR are facing the TM binding site in the modeled receptor structure of mGluR5 (Figure 24) but are not identically conserved, what might be essential for development of selective ligands.

For a subfamily like mGluRs conservation and selectivity can behave a follows on examples of 3.39 (selective) and 3.4 (non-selective). The mutation Ser3.39 abolishes DFB, fenobam and MPEP binding and effect in mGluR5 (Mühlemann *et al.* 2006, Malherbe *et al.* 2006, Pagano *et al.* 2000), but is not conserved in family C (H=1.82) or mGluRs (H=1.54) either. Close to it, position 3.4 is conserved in mGluRs and involved in NAM binding of mGluR1 (Malherbe *et al.* 2003a) and mGluR5 (Malherbe *et al.* 2006).
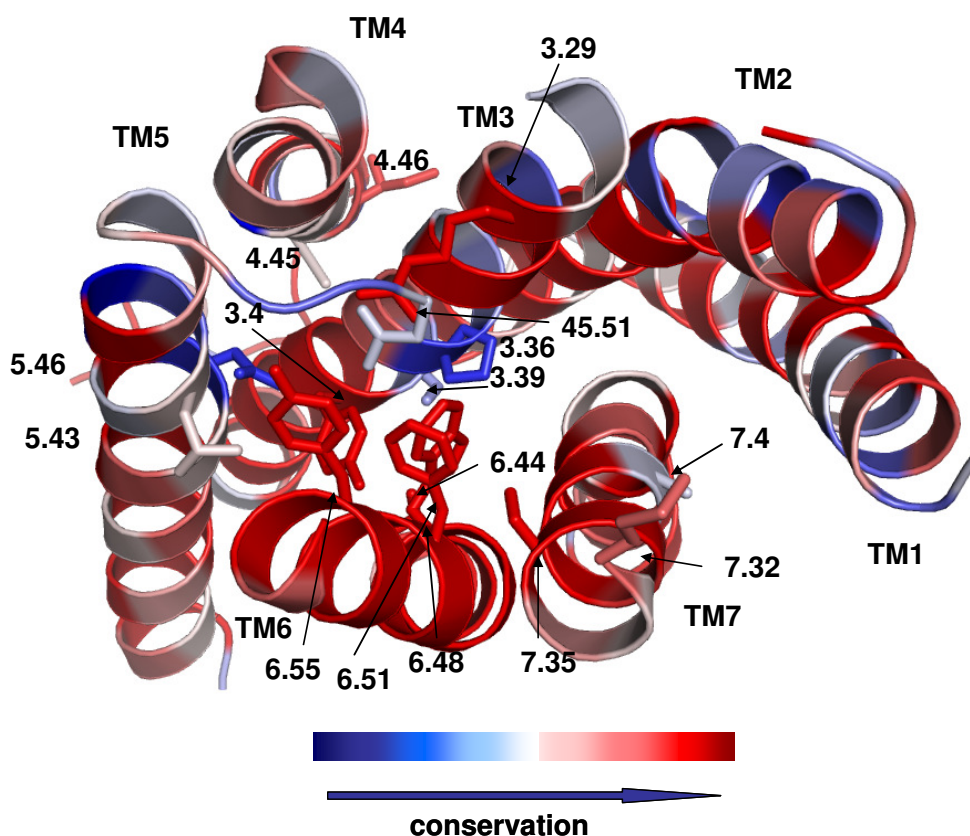
Figure 24: predicted mGluR5 receptor structure colored according to *H* values of the mGluR subfamily. Loops besides short part of extracellular loop 2 are omitted. Colors are applied as a color gradient, from red over white to blue representing high, middle and low conservation. Residues of mGluR5 at positions tested in mutation studies with mGluRs as important for ligand binding (Table A 3) are shown in stick representation.

It is not known in detail how the activation takes place and how allosteric modulators could influence the conformational changes during this event. From the correlation experiment using mutation data, conservation values und crystal structures of a covalently bound and a diffusible ligand, similarities in the binding site of family C and A GPCRs could be defined. This discussion will be continued in Section 4.2.4 using predicted mGluR5 ligand binding modes.

4.1.8 Conservation analysis - conclusions

The GPCR protein family allows for signal trafficking through cell membranes using a mechanism of activation by the transmembrane domain (Farrens *et al.* 1996, Altenbach *et al.* 2008). The goal of the conservation analysis was the analysis of function and structural implication of amino acids in the TM region of family C GPCRs and their evaluation in structural context. The performed Entropy calculations allowed the identification of features within the TM region of 96 diverse family C GPCRs with functional importance. These features were compared to published structurally important sites of family A GPCRs. As a proof of concept experimentally determined positions controlling receptor function were determined at low entropy values in either the mGluR as the vertebrate rhodopsin subfamily. All structural features and positions involved in helical packing and activation in family C receptors could be interpreted using the 3D receptor model of mGluR5.

Entropy calculations in ligand binding site of family C GPCRs allowed the discrimination of differently conserved positions in family C and the mGluR subfamily. These differences mark important positions in the **binding site** in several family C GPCRs:

- Exclusively subfamily conserved positions ($H \leq 0.3$) might define particular receptor function or the ligand binding site, but not subtype selectivity (3.29, 3.32, 3.33, 3.4, 6.44, 6.48, 6.55, 7.35, 7.36, 7.37, and 7.4).
- Not conserved position neither in family C nor mGluR ($H > 0.3$), but experimentally determined as important for allosteric ligand binding but not agonist effect alone, might define receptor subtype selectivity (2.64, 3.27, 3.28, 3.39, 5.42, 5.43, 5.44, 7.32, 7.33)

4. Results and Discussion

With a focus on structural motifs crucial for functional activity as a) helix-helix contacts, b) receptor dimerization interfaces, c) G protein interaction sites d) general structural motifs and e) ligand binding sites similarities between the two GPCR families could be evaluated. Especially in case of helix-helix contacts several conserved positions (1.46, 2.51, 2.49, 3.43, 3.46, 3.47, 3.5, 5.58 6.37 and 6.41) were identified, which were different to analogous helix-helix contacts in BR and might compose a conserved pattern coding for TM packing in family C GPCRs. A conserved sequence pattern around the DRY-motif (Scheerer *et al.* 2008), which is involved in G-protein coupling, was observed in mGluR5.

The conserved helix-helix contacts, the location of the G-protein interface and positions involved in ligand binding in the predicted structure model of mGluR5 were evidence of appropriate sequence alignment and modeling template selection. By projection of mutated positions onto the structure of mGluR5 the ligand binding site of family C could be visualized and an overlap with the one of family A GPCR detected based on known complex structures (Table 8). The location of NAM and PAM binding regions of family C GPCRs projected on mGluR5 revealed similarities to findings published by Bissantz for family A GPCRs (Figure 25, Bissantz *et al.* 2003). Bissantz proposed a smaller binding region for agonists which is completely included in the antagonists' binding region (Figure 25, B). Here, in the mutation data collection for family C GPCRs both NAM and PAM regions overlapped and were located between TM2, TM3, TM5, TM6 and TM7 according to the predicted mGluR5 structure (Figure 25, A). The broad distribution of positions involved in ligand interaction (Figure 25) of family C GPCRs visualize the diversity of different ligand classes recognized by the subfamilies.
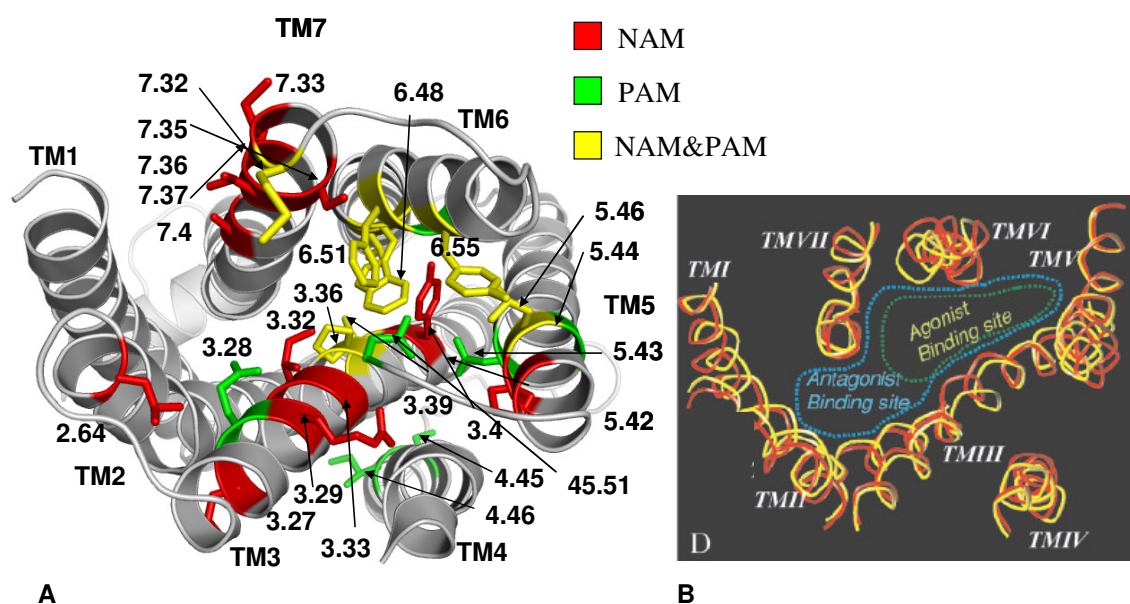
Figure 25: TM ligand binding region of GPCRs. A) mGluR5 model with residues from mutation studies, which influence the binding or effect of negative (NAM) and/or positive (PAM) allosteric modulators in family C GPCRs (Table 8). B) D3 antagonist model with agonist and antagonist binding regions for family A GPCRs (Bissantz *et al.* 2003).

4.2 Virtual screening for novel mGluR modulators

The discovery of mGluR5 selective modulators is important for antipsychotic treatment (Conn *et al.* 2009). The ability of virtual approaches to discriminate mGluR NAMs versus PAMs and selective molecules from non-selective using different molecule encoding methods will be analyzed in the beginning of this section. The subsequent prospective virtual screening approaches which were applied here will be discussed with respect to discovery of structurally distinct molecules compared to reference molecules. Which structural features of the "hit" molecules were important for modulation and can bind to similar regions in the binding pocket will be emphasized using reference molecules and binding mode prediction in the modeled structure of mGluR5.

4.2.1 Ligand data analysis

Optimization of small molecules binding to mGluRs has been successfully carried out and described in literature and patents (Williams and Lindley 2005, Wang and Brownell 2007, Conn *et al.* 2009). Published structural information about molecules and their activity values was used to compile a training data set for virtual screening purposes. 1246 unique molecules were collected and categorized in functional classes, such as

4. Results and Discussion

NAM or PAM, and assigned experimentally determined receptor selectivity for mGluR1, mGluR2 and mGluR5 and their activity range (Table 9).

Table 9: Quantitative overview over the ligand data set of mGluR binding ligands and their activity. Negative (NAM) and positive (PAM) allosteric modulators were differentiated regarding the target receptor and activity values.

NAM & PAM				
Activity ≤ 1000nM	X	inactive	X	X
Receptor		mGluR1 & mGluR5	mGR5	mGR1
Number	**977**	**117**	**698**	**279**

NAM						
Activity ≤1000nM	X	X	X	X	X	
Receptor		mGR5	mGR1	mGR5	mGR1	mGR1/5
Selective				yes	yes	no
Number	**870**	**615**	**255**	**111**	**57**	**15**

PAM				
Activity ≤ 1000nM	X	X		
Receptor		mGR5	mGR1	mGluR 2/4/7
Number	**150**	**83**	**24**	**43**

The ligand data collection was primary used to set up training sets for virtual screening purposes and to classify/cluster ligands into different classes (Figure 26).
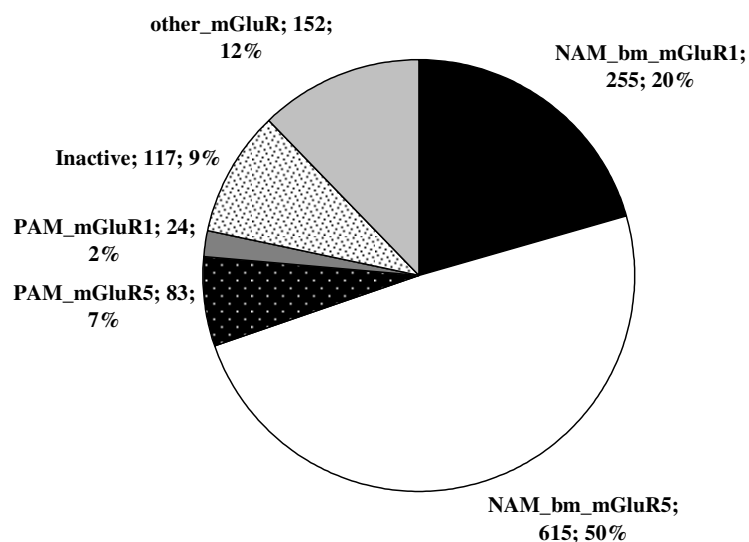
Figure 26: Composition of the data collection containing molecules, which were tested for activity on mGluR receptors. 1246 molecules are grouped according to their activity profile (NAM or PAM), activity range below 1μM (**bm**) and target molecule type. The number of ligands per class is given as number and percentage value.

In order to evaluate the diversity of the compiled data collection it was analyzed regarding the included variety of structural fragments. Therefore most frequent ring assemblies and contiguous chains present were enumerated (Pipeline Pilot, SciTegic, San Diego, USA). The frequency of occurrence of the structural fragments represents the structure and scaffold diversity of the data set (Figure 27).

4. Results and Discussion

Figure 27: Ring assemblies of known mGluR binding allosteric modulators in the literature data collection sorted by decreasing frequency (as indicated by bold numbers).

The frequency of substructures and atomic properties in the data set are relevant to determine their importance for machine learning techniques or encoding by molecular descriptor. Data set composition can bias the relevance of a feature by over- or under-representation of this feature in the data set. Frequency detection for single rings, linker constructs and frequent substituents could give an indication which substructures are dominating the data collection and would be expected to be favored in screening applications featuring substructural similarity (Figure 28).

rings

| 2607 | 1115 | 302 | 260 | 221 | 215 | 205 |

| 170 | 161 | 159 | 146 | 101 | 101 |

linkers

| 315 | 133 | 52 |

substructures

| 569 | 244 | 208 | 215 | 139 | 133 | 82 |

Figure 28: Frequencies of most frequent substructures counted in the literature data collection of known mGluR binding allosteric modulators.

After the analysis of structural diversity the metabotropic data collection was utilized as data pool for selection of particular representative molecules for model training and similarity calculations.

4.2.2 Clustering of mGluR ligands using Spherical Harmonics Descriptors

The ligand data collection consisted of 1246 ligands binding to mGluR which were published in scientific journals or patents. The collection was analyzed with a SOM, a machine learning technique, based on two different descriptor types, the CATS and the SHD descriptors. The CATS descriptor is defined on the topological (2D) representation of a molecule, the molecular graph, and captures pharmacophoric features and their distribution in a molecule. In contrast the SHD relies on the 3D conformation of a molecule and includes information about size and 3D volume. Both methods where tested regarding their ability to be used for representation of functionally diverse clusters, these are mGluR1 and mGluR5 modulation or NAM and PAM effects. Considered functional clusters included different functional effect types such as NAM and PAM, different volumes and targeted receptors. A SOM was trained for each of the

4. Results and Discussion

descriptors and clusters defined using the 2D map projection of neighbourhood and neuron population.

Both descriptor types were suitable for discrimination of NAM and PAM clusters but not without overlapping neurons (Figure 29 and Figure 30 both C/D). PAMs were clustered in regions of the map with high quantization error. Quantization errors are observed in neurons where the data samples clustered are less similar to each other. NAMs outnumbered the PAM samples and were clustered more accurately. The reason for the better representation was that during the training more neurons were adapted to NAMs and NAM samples were presented more frequently.
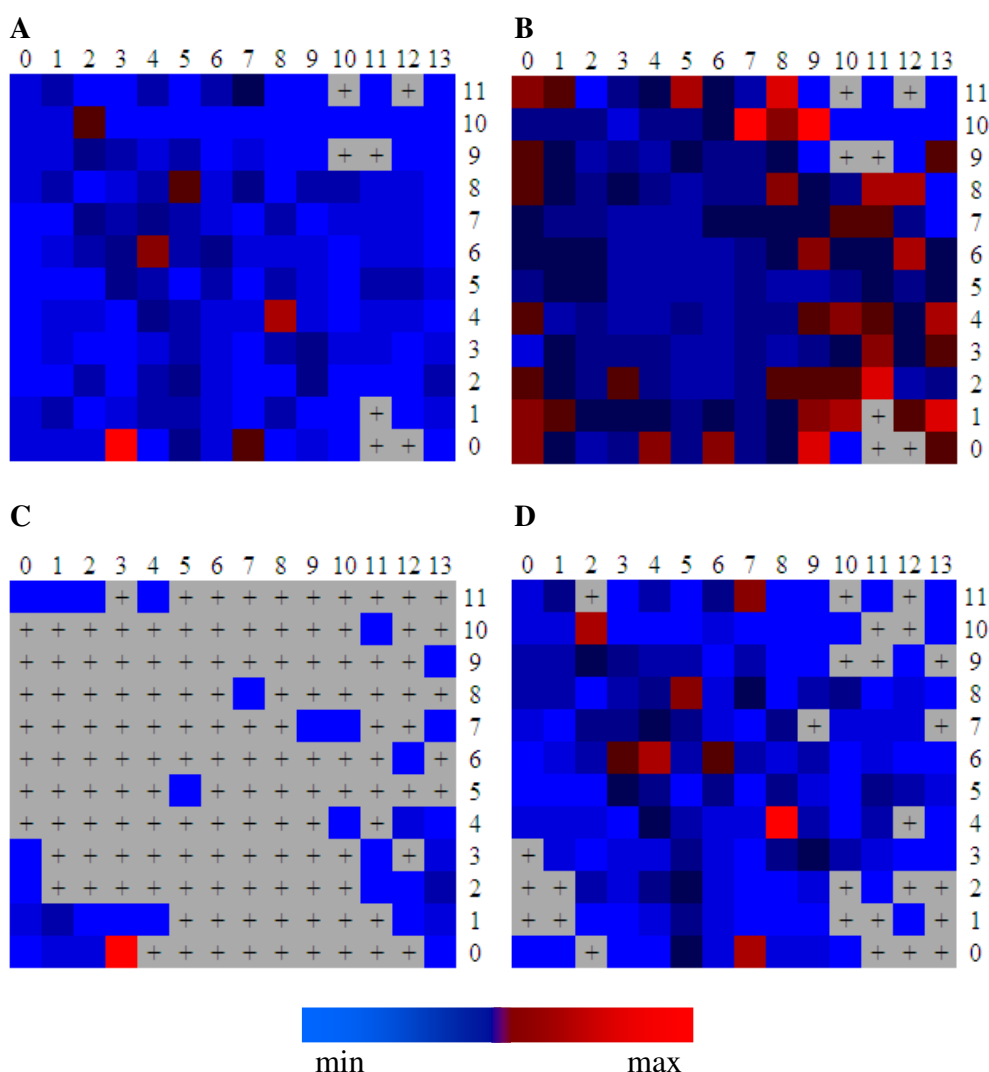
Figure 29: A self organizing map trained on 1246 mGluR ligands described with spherical harmonics descriptor. A) Overall frequency, B) quantization error. Neurons and frequencies of C) PAM and D) NAM ligands, exclusively. The color scale indicates the number of assigned patterns or the quantization error from minimal to maximal. Grey fields represent empty neurons where no samples of the projected type were assigned to.

Both the clustering with SHD and CATS could not separate the classes exactly. Evaluation criterion was the number of neurons which only NAMs or PAMs had exclusively been assigned to. Using SHD 19 out of 34 PAM (55%) neurons were also assigned at least one NAM. The CATS descriptor lead to 10 (41%) mixed neurons out of 24 PAM. The separation of PAM and NAM was more accurate with the pharmacophore description of CATS than the shape description by SHD. Still, based on these descriptor types NAMs and PAMs were not classified accurately.
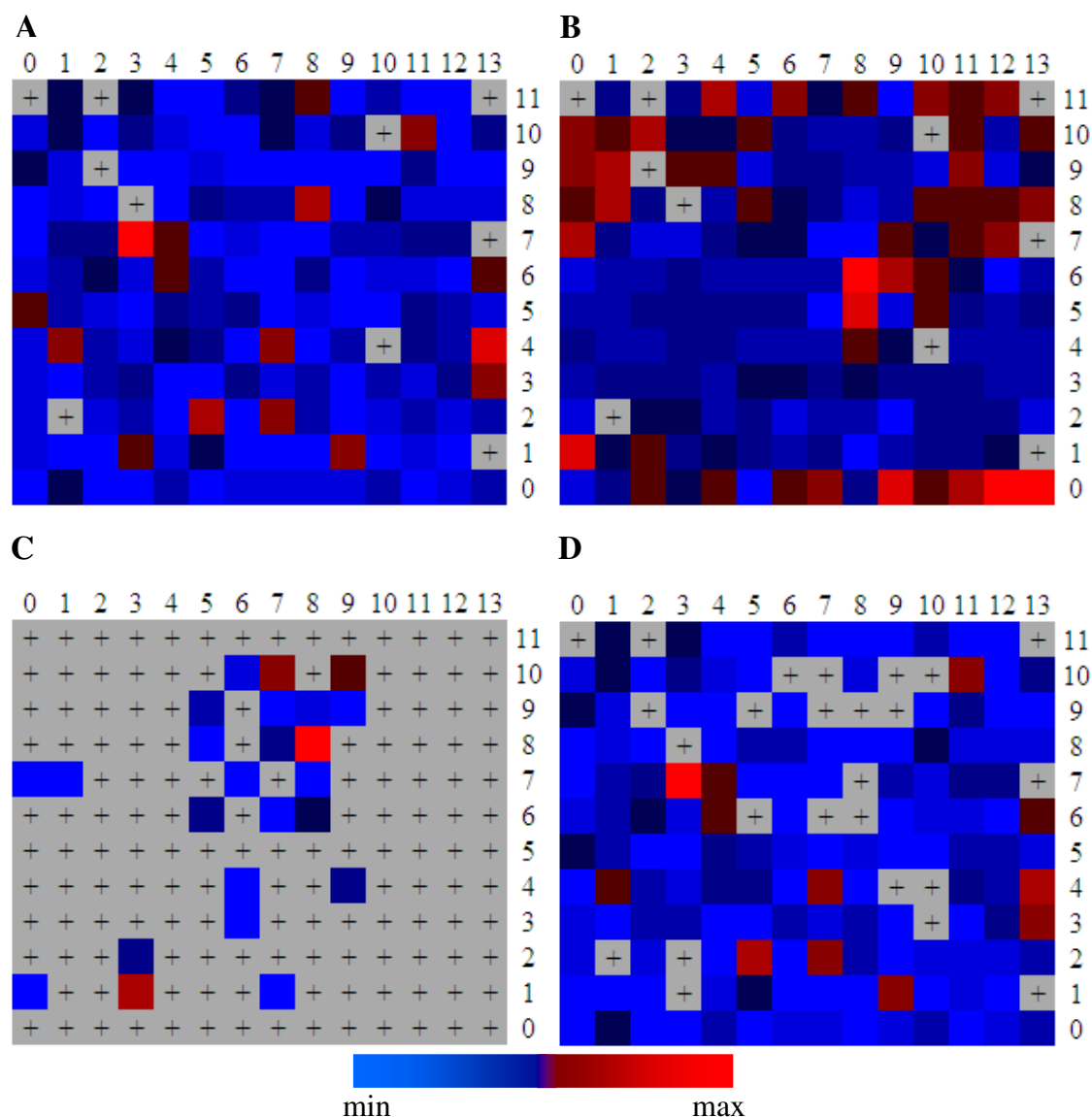
Figure 30: A self organizing map trained on 1246 mGluR ligands described with chemicals advanced template search (CATS) descriptors. Projection shows the A) Overall frequency, grey fields represent empty neurons, B) quantization error, including only C) PAM or D) NAM. The color scale indicates the number of assigned patterns or the quantization error from minimal to maximal.

Clusters for large (> 24 non-hydrogen atoms), medium sized (> 18 and ≤ 24 non-hydrogen atoms) and small (≤ 18 non-hydrogen atoms) ligand classes were observed indicating the diversity of molecules regarding that respective feature (Figure 31).

4. Results and Discussion
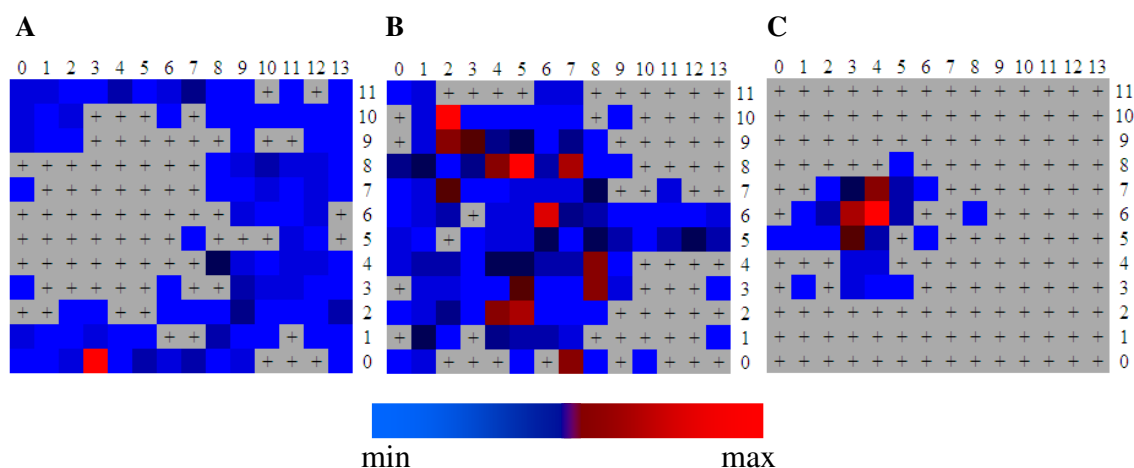
Figure 31: Self-organizing map (SOM) of the ligand data collection using spherical harmonics descriptors for molecule representation. The trained SOM is the same as in Figure 29. Here, molecules with different atom numbers are projected A) 491 large (>24), B) 633 medium (>18 and <24) and C) 146 small (<18). The color scale indicates the number of assigned patterns from minimal (blue) to maximal (red).

The SOM trained with SHD (Figure 29) and the one with the CATS (Figure 30) descriptor were analyzed with respect to separation of target receptor clusters, mGluR1 and mGluR5 (Figure 32). The distribution of clusters for receptor targets mGluR1 and mGluR5 revealed that using SHD in 61 of 85 (71%) mGluR1 neurons, mGluR5 samples were clustered as well. For the map trained with the CATS descriptor a lower overlap was detected, 34 of 71 (47%) neurons. Further, it was observed that neuron clusters frequently represented the scaffold series of the data set. Using these descriptors the neurons were adapted to dominating similarities insight scaffold clusters, which included selective and nonselective molecules. Therefore the SOMs possessed low discriminative power regarding mGluR1 and mGluR5 selectivity.
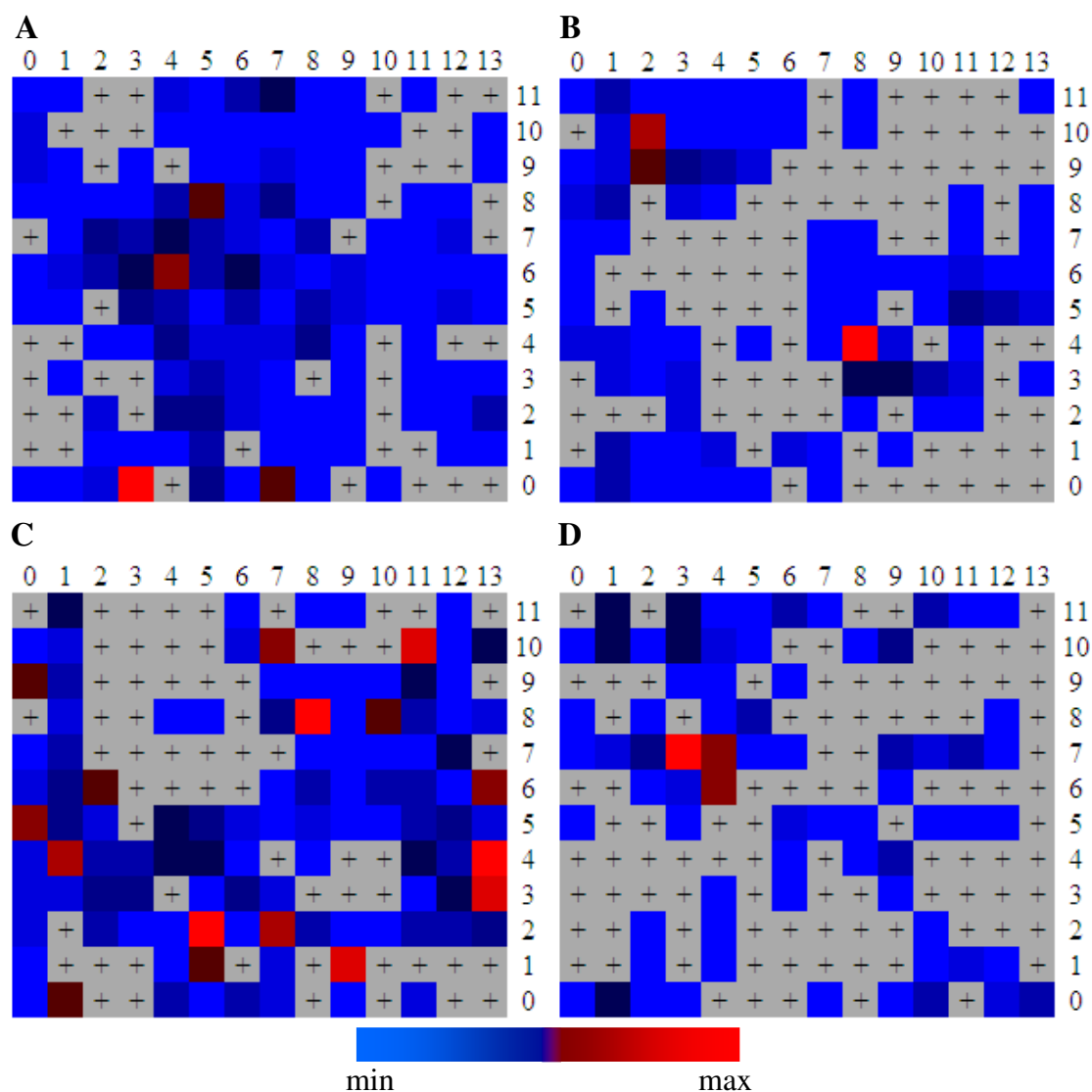
Figure 32: Projection of ligands binding to mGluR5 (left) and mGluR1 (right) resulted from SOM trained on spherical harmonics descriptor (SHD, top) and chemically advanced template search (CATS, bottom) descriptors. A) mGluR5 with SHD, B) mGluR1 with SHD, C) mGluR5 with CATS and D) mGluR1 with CATS. The trained SOMs are given in Figure 29 and Figure 30. The color scale indicates the number of assigned samples from minimal to maximal.

Taken together, SOM clustering revealed that the chosen shape and pharmacophoric molecule description is insufficient for predicting functional differences between molecules. Still, they allow to detect a trend in receptor selectivity, as observed for the CATS clustering. It was stated that the application of different descriptor types can be beneficial for capturing more relevant molecule properties (Sheridan and Kearsley 2002). Thus, the combination of molecular shape with atom typing as well as several pharmacophore encoding techniques was applied for virtual screening for novel mGluR5 binding molecules.

4.2.3 Virtual screening

Virtual screening allows for selection of molecules with a particular activity profile out of million of compounds. The discovery of novel selective allosteric modulators would be relevant for treatment of central nervous system disorders (Pin *et al.* 2003). In the present study several procedures were applied in order to detect active and selective allosteric modulators for mGluR subtypes 1 and 5 (Figure 33).
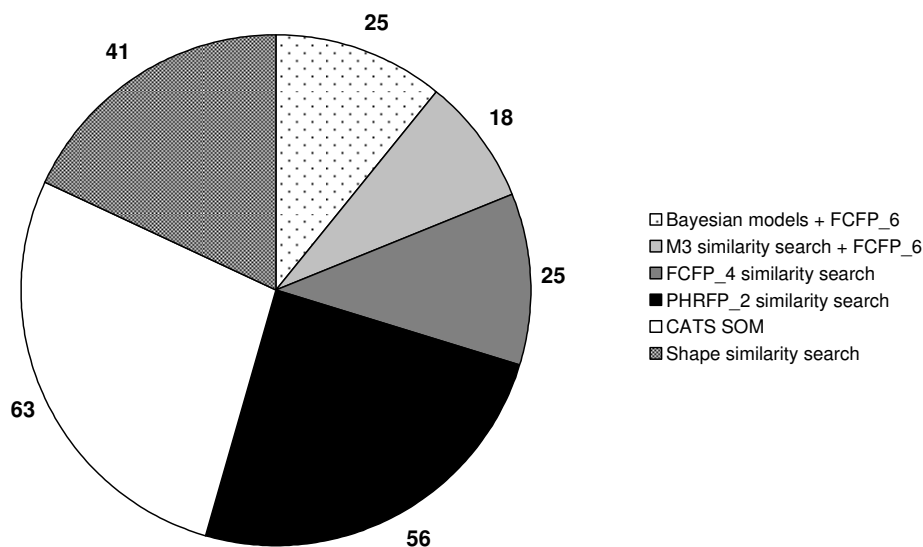
Figure 33: 228 molecules selected with different virtual screening methods (given in the legend) for experimental testing of activity on mGluR5.

Five combinations of molecule descriptors and prediction techniques were applied for prediction of mGluR binding molecules. Experimental verification was applied for 228 virtually selected molecules. From each approach a number (Figure 33) of molecules was selected for *in vitro* testing: pharmacophore based descriptors CATS with SOMs (63 molecules), PHRFP_2 (56) and FCFP_6 (25) both in similarity searches, FCFP_6 with a Bayesian model (25) and shape defined similarity (41). Each of the virtual hit molecules was tested for agonism and antagonism in a Fluorescent Imaging Plate Reader (FLIPR[TM]) functional screen (Vanejevs *et al.* 2008). The read out of the assay is Ca^{2+}-concentration and all values are IC_{50} values calculated from DRC. All molecules active on mGluR5 were tested for activity on mGluR1 in order to determine selectivity.

The variety of techniques can be anticipated to cover diverse molecular features and possibilities to evaluate their similarity. In the following, results of different similarity calculation approaches using known active molecules are described. Further on, all

molecules verified experimentally for modulating function on mGluR1 or mGluR5 are discussed regarding their novelty compared to the reference data collection and dependency on the descriptors used or virtual selection method applied.

Shape search

Since shape restricts molecules to a particular volume and can define the 3D form of the binding pocket, this feature was applied for virtual screening of novel compounds. The conformationally restricted molecule 2-methyl-6-(phenylethynyl)pyridine (MPEP, Gasparini *et al.* 1999), a selective mGluR5 NAM, which possesses an accurately predictable planar geometry was used as 3D reference molecule. Shape similarity is a scalar measure between 1.0 (maximally similar) and 0.0 (minimal similarity). The similarity is weighted using chemical atom typing. 2399 out of 382671 Specs (www.specs.net, v2008.1) molecules had a calculated shape score above 0.7 representing 0.63% of all data base molecules. 41 molecules were selected based on shape similarity and tested *in vitro*. **M1** (Figure 34) revealed functional activity (IC_{50}=2.2 μM) and selectivity for mGluR5. The `r_phase_Shape_Sim` score for **M1** was calculated to be 0.7357, maximum similarity equals to 1.0.

M1 possessed moderate activity (IC_{50}=2.2 μM) against the human mGluR5 in comparison to the applied query MPEP (IC_{50}=2 nM). Structurally similar ligands (Figure 34) were already reported in similar activity range (Roppe *et al.* 2004, Iso *et al.* 2006).

4. Results and Discussion

MPEP –
reference molecule
mGluR5 IC_{50}=2nM
Gasparini *et al.* 1999

M1
Specs AN-329/10552014
mGluR5 IC_{50}=2180nM

SIB-1893
mGluR5 IC_{50}=290nM
Varney *et al.* 1999

mGluR5
IC_{50}=3000 ± 0.9nM
Iso *et al.* 2006

mGluR5
IC_{50} >10000nM
Iso *et al.* 2006

mGluR5
IC_{50} =190 (160, 3000, 2310)nM
Roppe *et al.* 2004

mGluR5
IC_{50} =4620 (258, 8280, 5000)nM
Roppe *et al.* 2004
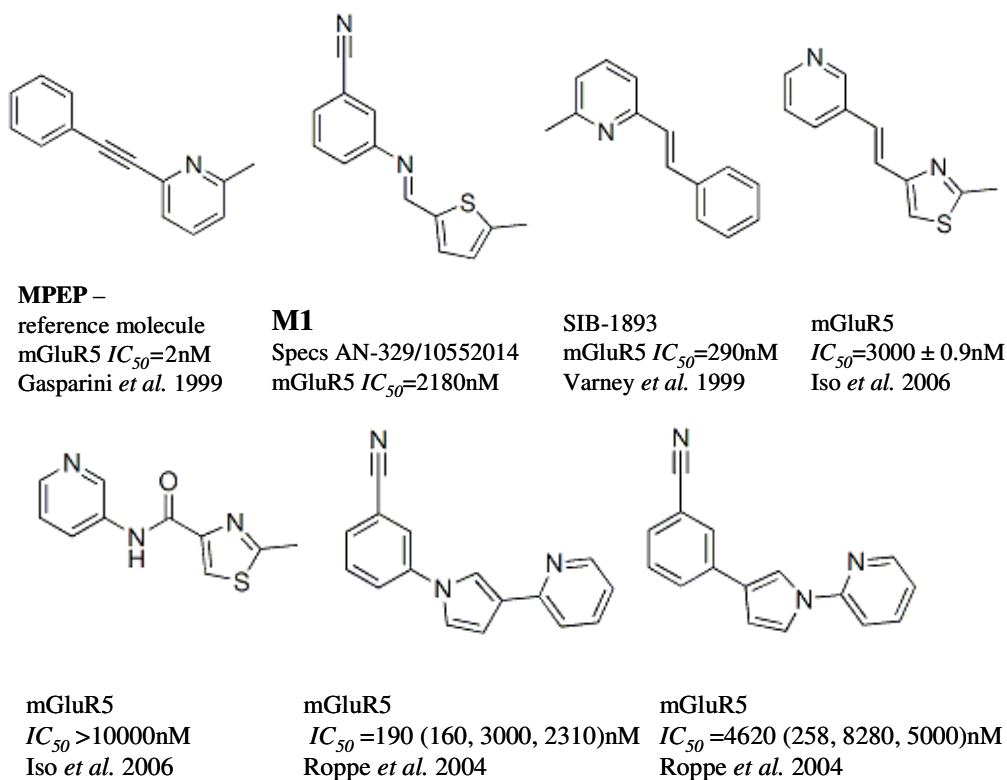
Figure 34: Reference and virtual "hit" molecules from a shape similarity search (Phase v.2.5, Schrödinger, LLC, New York, USA). Activity value for **M 1** was defined in functional NAM assay on the human mGluR5.

The discovery of **M1** proves that shape similarity was a detectable and traceable molecular feature (Figure 35). In case of mGluR binding ligands it allowed the discovery of a moderately active molecule within the scope of the ligand class of small molecules defined by MPEP or 3-[(2-methyl-4-thiazolyl)ethynyl]pyridine (MTEP) analogues, the heteroarylazole class.

A   B   C   D
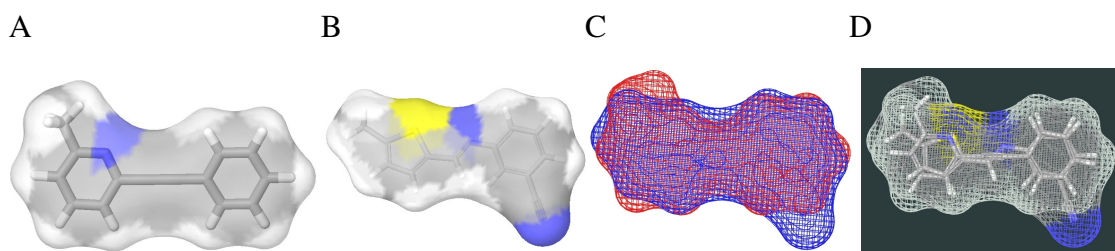
Figure 35: The query 3-[(2-methyl-4-thiazolyl)ethynyl]pyridine (MPEP, A) and the "hit" molecule (**M1**, B) of a shape similarity search. **M1** was discovered with a shape based molecule alignment (C) and additionally scored using MacroModel atom types (D). The representation of molecular shapes was calculated with Schrödinger software (Schrödinger, LLC, New York, 2008) and colored according to MacroModel atom types (A, B, C). Molecule alignment is given in wire frame representation (C, D).

Shape similarity mainly depends on the 3D structure of query and screened molecules. Therefore, it can be inferred that different query molecules would support a broader coverage of chemical space. A low enrichment of other scaffolds or molecules with

larger molecule size was observed when MPEP shape-based similarity ranking was tested on known mGluR ligands (data not shown).

Using molecular docking (IFD by Schrödinger, LLC, New York, 2008) several different docking conformations were observed for **M1**. Figure 36 shows one possible binding mode of **M1**, in which the molecule extends orthogonal to the membrane plane and might be involved in an interaction to Arg648 and Trp785.

Figure 36: Binding conformation of molecule **M1** in the modeled structure of mGluR5, predicted with molecular docking (Schrödinger, LLC, New York, 2008). A) View from extracellular site B) view from the TM between TM5 and TM6. Residues in close contact to **M1**, Trp785 and Arg648 are highlighted in stick representation and hydrogen-bonds indicated with dashed lines.

Bayesian classifier

Bayesian classifiers were applied for probability estimation of molecules' classification with respect to a classification scheme. The probability prediction was carried out with two Bayesian classifiers aiming at selection of active mGluR binding molecules ("mGluR"-model), which are selective for mGluR5 ("selectivity-model") at the same time. The two classifiers were combined for final molecule selection (procedure explained in Section 3.11). The "mGluR"-exhibited a promising prediction accuracy for the validation data set (Figure 37). The AUC value reached 0.963, which was considered accurate. Therefore the classifier applied in more than one of the virtual screening approaches (in a combination with the "selectivity"-model and as ranking methods in CATS-based clustering using a SOM) as additional scoring method.

4. Results and Discussion
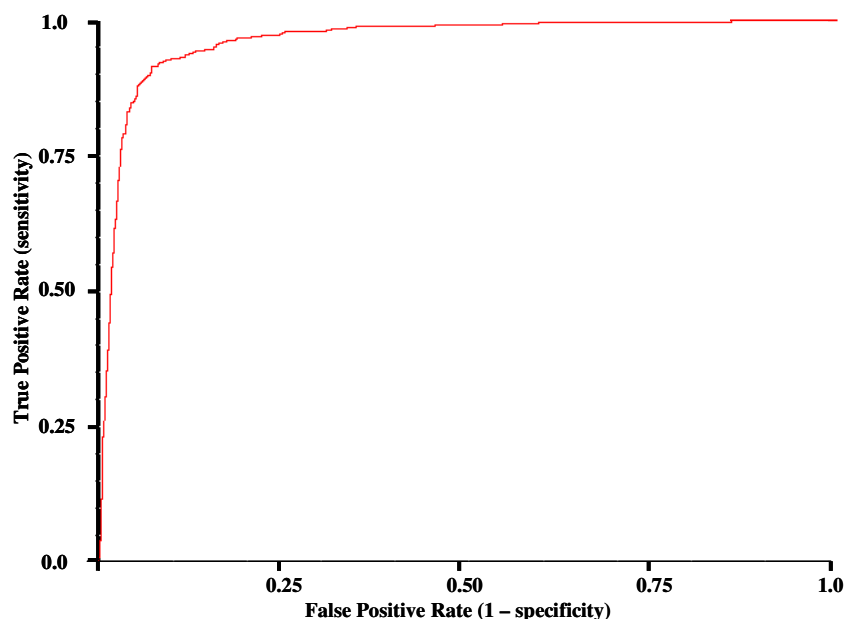
Figure 37: Receiver-Operating Characteristic (ROC) curve of the Bayesian classifier used for discrimination between mGluR active molecules and molecules binding to different target molecules ("mGluR"-model) or being inactive in tests on mGluR1 or mGluR5. The True Positive Rate is plotted against the False Positive Rate, the area under curve (AUC) equals to 0.963.

The number of 'good' and 'bad' samples was plotted versus the Bayesian classifier score for the "mGluR"- and the "selectivity"-models in a histogram, respectively (Figure 38). The histogram for the mGluR-classifier (Figure 38, A) revealed only the small overlap between active in inactive molecules regarding the "mGluR"-model score. According to prediction performance on the validation data set, the "mGluR"-model score was considered for molecule ranking.

For the selectivity classification a less predictive model ("selectivity"-model) was obtained because of fewer data samples. The score histogram showed a good discrimination between a) selective mGluR5 and selective mGluR1 and b) not selective mGluR5 molecules (Figure 38, B), but the performance was influenced by scaffold clusters (as described in Figure 27) and was therefore considered less reliable. However, the "selectivity"-model was subsequently applied after scoring with the "mGluR"-classifier. A Pareto-front definition for the 2D optimization of both model scores was calculated and 25 diverse molecules were selected for experimental testing.

106
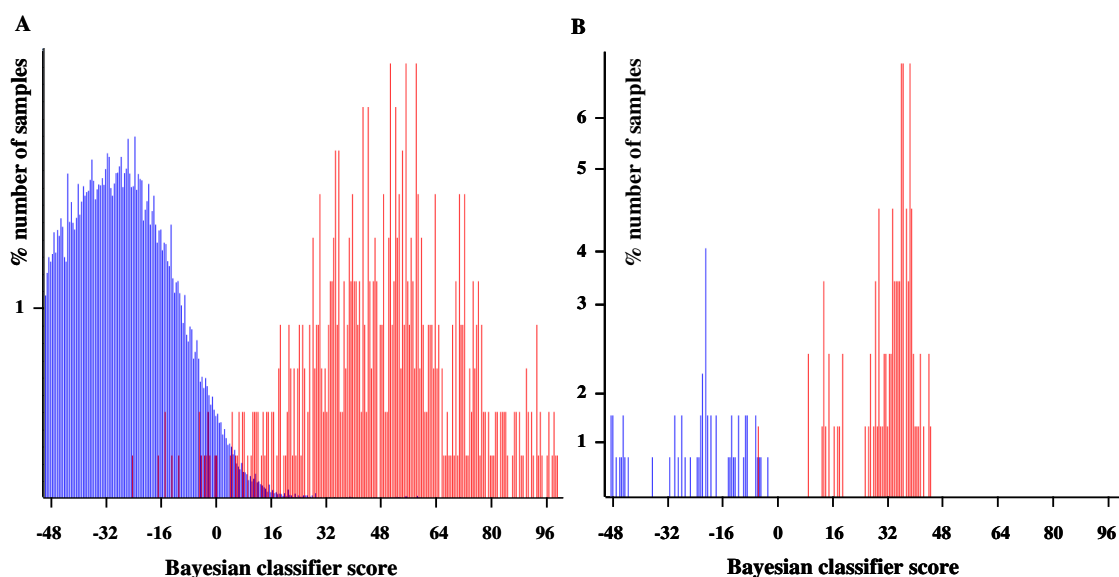
Figure 38: Histogram of Bayesian model scores (x-axis) versus frequency of samples with a particular model score (y-axis). Active (red) samples and inactive (blue) samples are indicated with colors. Training and test data are included. **A) "mGluR"-model**, active: mGluR1/5 $IC_{50} \leq 1000$nM, inactive samples: Wombat data collection and mGR1/5 $IC_{50} > 1000$nM **B) "selectivity"-model** active samples: mGR5 selective, inactive samples: mGluR1 selective and mGluR5 not selective.

In functional tests on mGluR5 three molecules **M2**, **M3** and **M4** were discovered with with IC_{50}=458nM (**M2**, Figure 39), IC_{50}=1780nM (**M3**, Figure 41) and IC_{50}=3890nM (**M4**, Figure 44), respectively. **M4** was selective for mGluR5 (IC_{50}=3890nM, IC_{50}>10µM), but possessed a lower activity than the other two molecules, **M2** and **M3**. **M3** was twice as potent as **M4** but not selective (**M3**: mGluR5 IC_{50}=1780nM, mGluR1 IC_{50}=3100nM).

M2
mGluR5 IC_{50}=458 nM mGluR1 inactive

| Bonnefous *et al.* 2005
mGluR5 NAM mol 9
IC_{50}=165 nM | Bonnefous *et al.* 2005
mGluR5 NAM mol 31
IC_{50}=11 nM | Williams & Lindsley 2005
mGluR5 PAM
IC_{50}=245nM | Iso *et al.* 2006
mGluR5 mol 38
IC_{50}=1000nM |
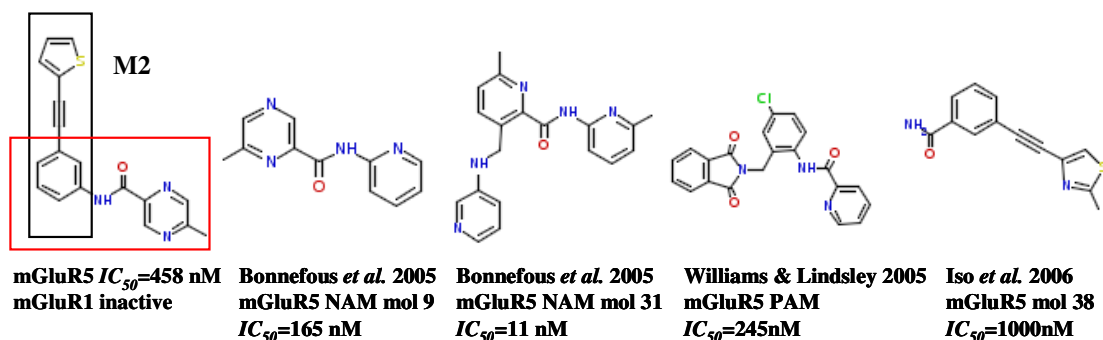
Figure 39: Molecule **M 2** discovered with Bayesian classifier prediction. The black and the red box outline parts of **M2** which are structurally similar to molecules included in the training data set such as present to the right of **M2**.

The molecule structure of **M2** contains substructures similar to MTEP analogues (outlined in black in Figure 39, Cosford *et al.* 2003) as well as molecule series reported by Bonnefous (outlined in red, Bonnefous *et al.* 2005), both ligand series were included in the reference data set (Figure 39). It should be mentioned that a PAM series (Figure

4. Results and Discussion

39, Williams and Lindsley 2005) possess a similar part to the substructure of **M2** outlined in red. The PAM is enhanced at *ortho*-position in contrast to *meta* in **M2** by the MTEP like substructure (outlined in black substructure in Figure 39).
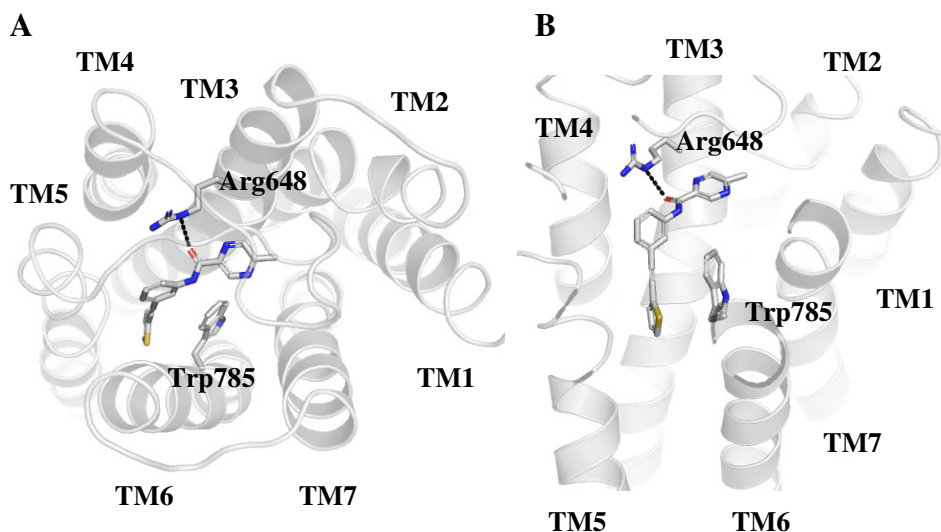
Figure 40: Binding pose of molecule **M2** in the modeled structure of mGluR5 predicted with molecular docking (IFD, Schrödinger, LLC, New York, 2008). A) View from extracellular site B) view between TM5 and TM6.

The docking conformation predicted for **M2** (Figure 40) revealed a parallel orientation of the acetylenic part (outlined in black, Figure 39) to Trp785 and the amid-linker was proposed to make a hydrogen bond interaction to Arg648. In the docked conformation (Figure 40) the acetylenic molecule part extends parallel to the helix bundle deep into the TM cleft. The non-acetylenic molecule part (outlined in red, Figure 39) shows similarities to molecule 9 (Figure 39, Bonnefous *et al*. 2005) and is oriented orthogonally to the helix bundle and closely to TM3 (Figure 40).

mGluR5 IC_{50}=1780 nM	Wang *et al*. 2007	Micheli *et al*. 2008	Micheli *et al*. 2008	Alagille *et al*. 2005
mGluR1 IC_{50}=3100 nM	mol 57	Mol 16	Mol 92	mol 9
	IC_{50}=9nM	pIC_{50}=7.8	pIC_{50}=7.9	IC_{50}=740nM
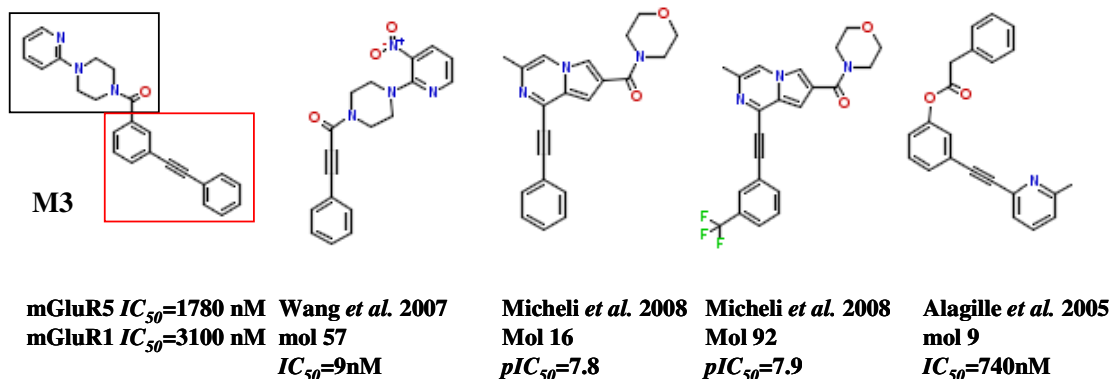
Figure 41: Molecule **M3** discovered with Bayesian classifier prediction. The black and the red box highlight parts of **M3** which are structurally similar to molecules included in the training data set such as presented to the right of **M3**.

Structurally similar molecules to **M3** (Figure 41) were reported to exhibit a middle to low nanomolar affinity to mGluR5 (Micheli *et al.* 2008, Alagille *et al.* 2005 and Wang *et al.* 2007). The acetylenic MPEP like molecule part (outlined in red, Figure 41) contains no pyridine as present in MPEP. The nitrogen in the pyridine ring might be relevant for selectivity to mGluR5, since **M3** is not selective and MPEP is. The acetylenic linker region should overlap in the binding pocket with the selective MPEP and MTEP analogues (Gasparini *et al.* 1999, Cosford *et al.* 2003, Alagille *et al.* 2005a/b, Vanejevs *et al.* 2008).

After the discovery of **M3** a similarity search with FCFP_6 descriptor was performed and 18 of the best ranked (using Tanimoto similarity Section 3.12.2) neighbors were experimentally tested. All 18 molecules were tested with IC_{50}>10μM. The reason for the activity failure of these molecules was analyzed. A flexible alignment (Phase v2.5, Schrödinger, LLC, New York, 2008) of the molecules was performed (Figure 42) and revealed that the variability introduced at the acetylenic molecule part (outlined in red, Figure 41) lead activity loss at mGluR5.
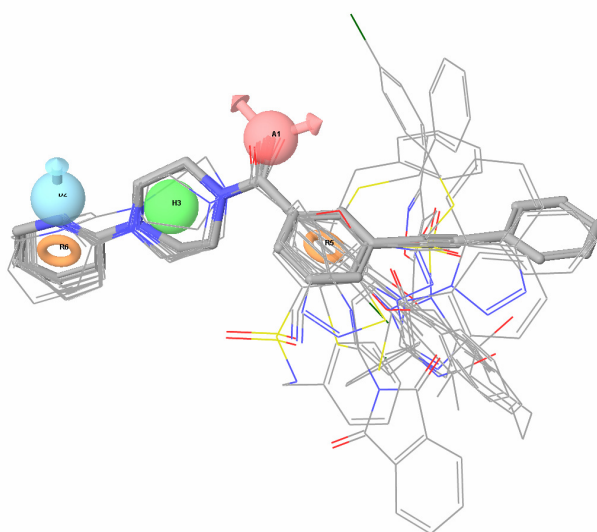
Figure 42: Flexible ligand alignment of **M3** (stick representation, Phase v2.5, Schrödinger, LLC, New York, 2008) and 18 structurally similar molecules (line representation), with pharmacophore points defining similar regions. All 18 molecules are functionally inactive on mGluR5. Pharmacophore definition: H1 (blue sphere) hydrogen bond donor, A2 (red sphere) hydrogen bond acceptor, H3 (green sphere) hydrophobic and R4-5 (yellow rings) aromatic.

The multiple ligand alignment revealed high sterical difference in the non-acetylenic part of the ligands (Figure 42). The docking suggests that this part extends deeply into the TM cleft of mGluR5 (Figure 43). Bulky substituents would be hardly tolerated in the narrow space deeper in the TM cleft. The docking poses more over revealed that the

4. Results and Discussion

pyridine nitrogen of **M3** might be involved as a donor in a hydrogen bond with Gln647, the carbonyl oxygen in another one to Arg648 (Figure 43). The distance between the non hydrogen atoms (ligand)N---O(Gln647) is 1.93Å and (ligand)=O---N(Arg648) is 2.99Å. By rotation around torsional angles (χ_3 and χ_4) of Arg648 the construction of an ideal hydrogen bond to the carbonyl atom would be feasible, but was not observed in the docking.
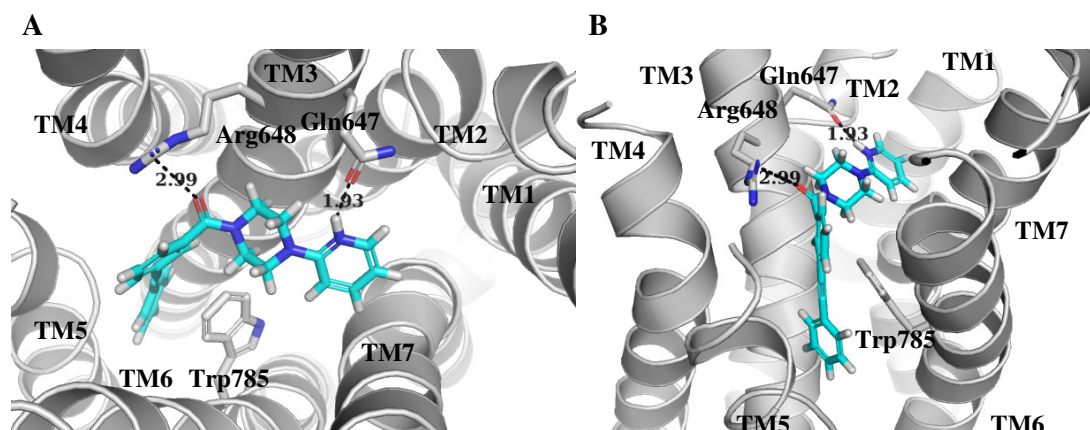
Figure 43: Binding poses of molecule **M3** in the modeled structure of mGluR5 predicted molecular docking (IFD, Schrödinger, LLC, New York, 2008). A) View from extracellular site B) view from the TM parallel site between TM5 and TM6. In both representations the EC2 loop is omitted.

M4 is a selective NAM of mGluR5 IC_{50}=3890nM. For **M4** the highest structural similarity to reference molecules was observed for a series developed by Astra Zeneca and Addex Pharmaceuticals SA (Figure 44). The oxadiazole and the triazole rings of **M4** are also present in these reference structures.

M4					
mGluR5 IC_{50}=3890 nM	AZ & NPS	AZ & NPS	AP SA	AZ & NPS	AZ & NPS
mGluR1 inactive	WO 2005/080396	WO 2005/080397	WO 2005/044797	WO 2005/080397	WO 2005/080356
	IC_{50}=199nM	IC_{50}=49nM		IC_{50}=81nM	IC_{50}=55nM
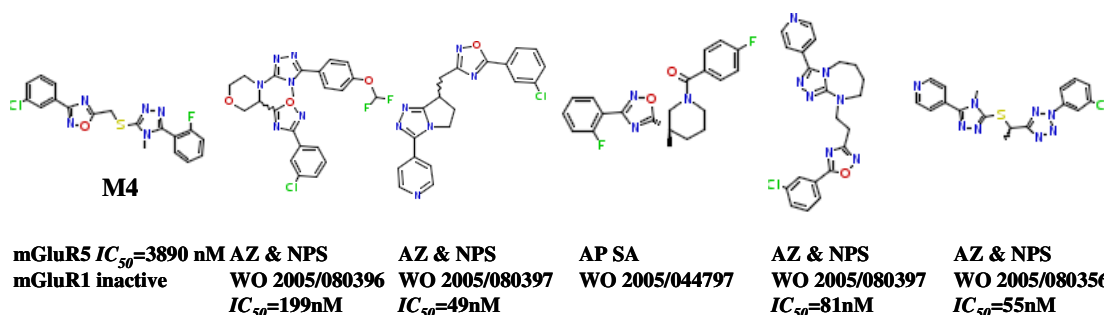
Figure 44: Molecule **M4** discovered with Bayesian classifier prediction. Structurally similar molecules included in the training data set are presented to the right of M4. AZ=AstraZeneca, AP SA=Addex Pharmaceuticals SA.

For molecule **M4** no hydrogen bonding was predicted using IFD and the modeled structure of mGluR5. Diverse binding modes resulted from IFD, therefore, the docking poses were not regarded highly reliable.

110

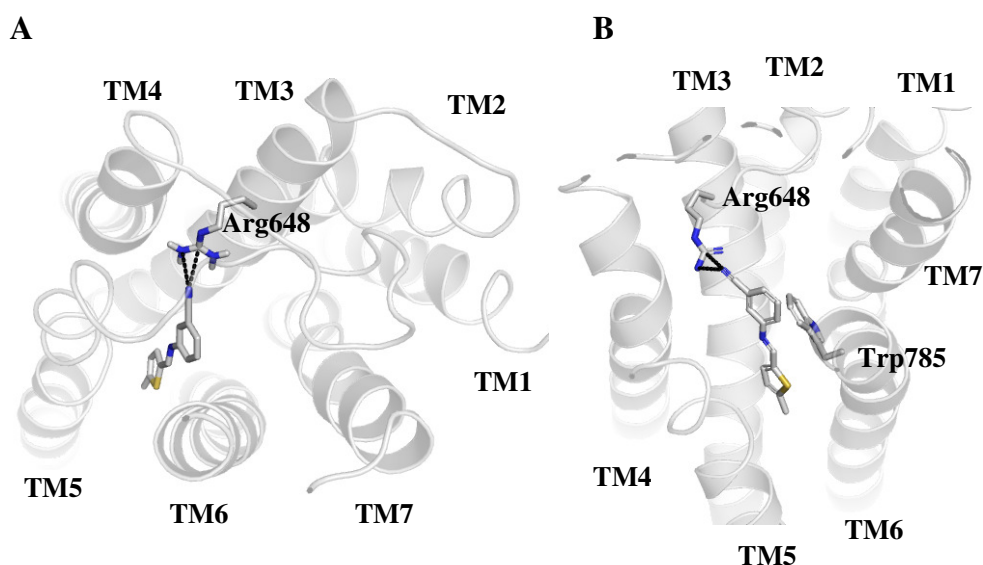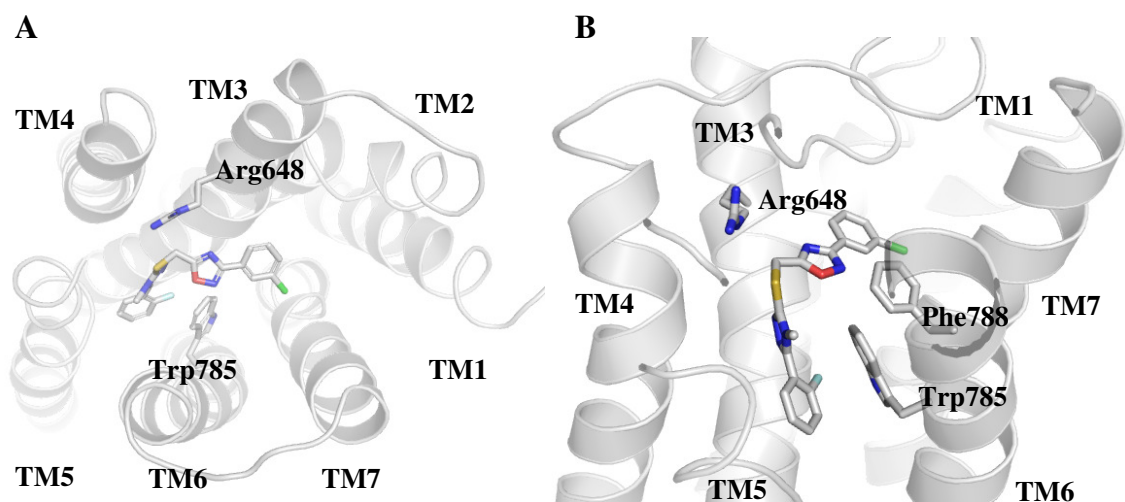**Figure 45:** Binding poses of molecule **M4** in the modeled structure of mGluR5 predicted with molecular docking (Schrödinger, LLC, New York, 2008). A) View from extracellular site B) View between TM5 and TM6.

The Bayesian approach lead to the discovery of three active NAMs down to nanomolar IC_{50} on mGluR5 but the "hit" molecules were structurally related to known molecules used in the reference data set. The reason for the high structural similarity to reference molecules can be traced back to the fingerprint calculation which focuses on substructures. The Bayesian probability calculation is sensitive to the number of reference features, in the case of the fingerprint these features were substructure frequencies. Dominating features in the reference set led to a higher scoring in classified molecules than less frequent substructures.

FCFP_4 similarity search

According to the assumption that similar molecules might exibit similar activity (Johnson and Maggiora 1990, Brown and Martin 1997, Martin *et al.* 2002), vendor molecules were compared to reference molecules measuring similarity in the space defined by a molecular descriptor. The similarity search using the FCFP_4 descriptor was performed with same reference data as the PHRFP_2 similarity search (details on procedure in Section 3.12.2) but a different molecular encoding. The goal was the prediction of novel selective mGluR5 NAMs using 615 reference NAMs. Molecules for testing were picked manually from a list ranked according to similarity to the reference mGluR5 NAMs with nanomolar activity. The selection of 25 molecules from this ranked list was done manually because of high similarity of top ranked molecules. Diverse molecule scaffolds with Tanimoto (Section 3.7.1) similarity above 0.75 according to FCFP_4 were considered for testing. **M5** (mGluR5 NAM) and **M6**

4. Results and Discussion

(mGluR5 PAM) exhibited activities of IC_{50}=8120nM and EC_{50}=288nM. The structure of **M5** will not be shown here, because it was considered for patenting. The structurally most similar molecules to **M6** which were part of the reference data set were previously described of having low nanomolar activities (Figure 46).
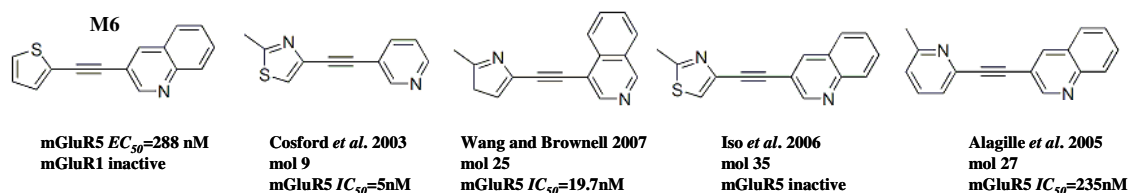
Figure 46: Active molecule **M6** discovered with a FCFP_4-based similarity search using mGluR5 binding NAMs with $IC_{50} \leq 1000$nM.

M6 is a PAM which has striking similarity to NAMs, especially to molecule 27 (Figure 46) as published by Alagille and co-workers with EC_{50}=235nM (Alagille *et al.* 2005). **M6** resembles the structures shown in Figure 46, it contains the acetylene linker and the pyridine ring on the right side of the linker, but possesses no methyl group next to the heteroatom in the thienyl or left heterocycle. Similar slight structural changes using a methyl group were reported recently by Sharma and colleagues (Sharma *et al.* 2008). The authors changed the well known ligand MPEP which is a frequently analyzed NAM possessing a simple molecule structure and derived new PAMs (Figure 47). Molecule **12a** (12a-k is the original numbering from Sharma *et al.* 2008, Figure 47) shares the scaffold with PAMs, **12h** to **12k**, and differs from **12h** only in the position of the methyl group.
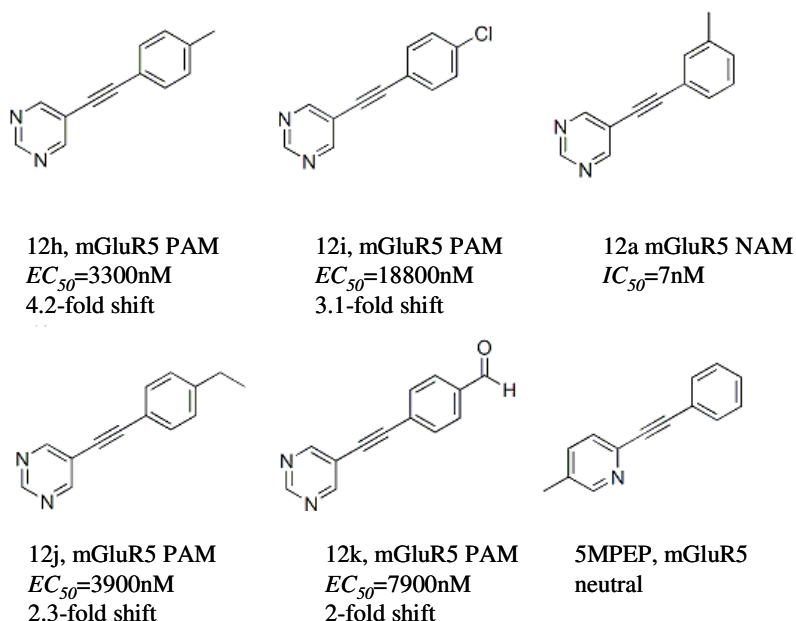
Figure 47: PAMs reported by Sharma and colleagues in 2008 (numbering as in original publication). MPEP analogues, which were converted to positive allosteric modulators of mGluR5 (Sharma *et al.* 2008).

The discovery of **M6** was interesting since this molecule suggests that the NAMs reported in literature can be converted to PAMs by addition of a single methyl group. This result is in line with studies performed by Sharma and strengthens the hypothesis that NAM and PAM can share the same binding region (Sharma *et al.* 2008).

PHRFP_2 similarity search

The pharmacophore fingerprint PHRFP_2 was applied in a similarity search (Section 3.12.2) for new selective mGluR5 NAMs. The combination of calculations for similarity of unknown molecules to mGluR5 NAMs separately from mGluR1 similarity calculations should avoid the selection of nonselective molecules and focus mainly on mGluR5 ligands. A single mGluR5 selective molecule with $IC_{50} < 10 \mu M$ was detected (**M5**, Figure 48). For this molecule **M5** it was experimentally determined that it has low micromolar activity ($IC_{50} = 1960nM$) while being inactive on mGluR1 and thereby selective.

A

M5

T5391466
IC_{50}=1960nM
Enamine screening
collection

molecule 57
mGluR5 NAM
IC_{50}=9nM
Wang and Brownell 2007

WO/2006/002981
mGluR5 NAM
IC_{50}=4nM
Grünenthal

WO 03/093236
Gharagozloo
(Euroceltique, S.A.)

B

mGluR5 PAM
EC_{50} =1.7 μM
8-fold potentiation
mol 16
unstable in DMSO

mGluR5 PAM
EC_{50}=2.7 μM
5-fold potentiation
mol 3

mGluR5 inactive
mGluR1 IC_{50}=63 μM
Renner *et al.* 2005
mol 14

mGluR inactive
Renner *et al.* 2005
mol 11

mGluR5 IC_{50}=12 μM
mGluR1 IC_{50}=17 μM
Renner *et al.* 2005
mol 24

Williams and Lindsley 2005

Figure 48: Molecule **M5** discovered with PHRFP_2-based similarity search from Enamine molecule collection. **M5** is selective regarding mGluR5 (IC_{50}=1960nM). A) Structurally similar molecules from the reference set. B) Structurally similar molecules described in literature, which were not included in the reference set.

Known reference molecules with comparable substructures were used to explain the activity of the new active molecule (Figure 48). The molecule can be regarded as a

4. Results and Discussion

chimeric structure from different molecule series. The training data sample from patent, WO 03/093236, has been reported with mGluR5 related indications as Parkinson's disease, dementia, cognitive disorder, vomiting, depression, Huntington's chorea.

Molecules from the training data set share one structurally similar part highlighted in black in Figure 48. The other molecule part (outlined in red) of **M5** was unsuccessfully developed for mGluR5 PAMs (Williams and Lindsley 2005) and independently discovered for NAMs by Renner in a virtual screening approach, but tested inactive for mGluR5 NAM activity (Renner *et al.* 2005). The combination of both substructures was combined in **M5**.

Molecule features present in the discovered molecule **M5** were analyzed as pharmacophore point pairs encoded in PHRFP_2 (Figure 49).
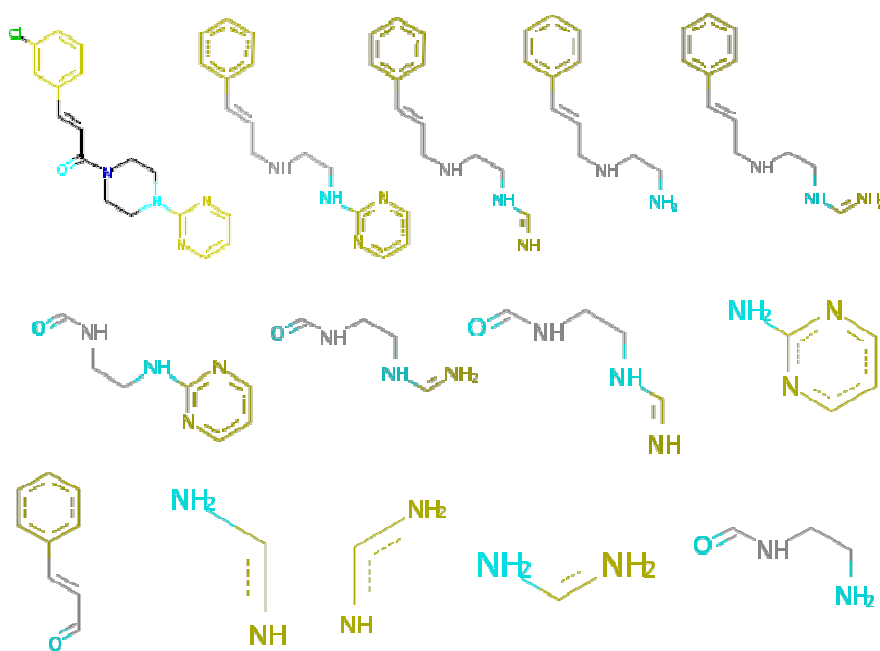
Figure 49: PHRFP_2 pharmacophore feature pairs present in the virtual hit molecule **M5**. H-bond acceptors (cyan), aromatic ring (yellow), positive ionizable (dark blue), hydrophobic (light grey) are highlighted. Starting from the top molecule on the left side, the molecule atoms are typed according to pharmacophore features and paths of variable length connecting different feature pairs are enumerated. The PHRFP_2 description captures pairs from the whole molecule as well as parts of it.

The predicted ligand binding conformation of **M5** (Figure 50) revealed a comparable binding region to **M3** (Figure 43) and a possible interaction to Arg648, however, no interaction was predicted to Gln647 as for **M3**.
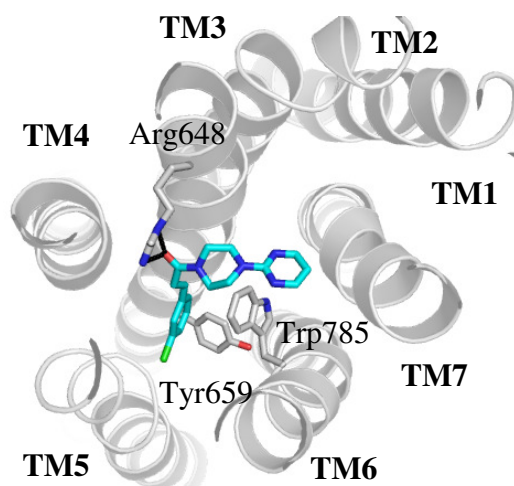
Figure 50: A predicted binding mode in the modeled structure of mGluR5 of **M5** discovered with a PHRFP_2-based similarity search. Residues in close proximity to the ligand are highlighted in stick representation.

CATS-based clustering using a self-organizing map

The Asinex Platinum Collection (version Nov2007) and the mGluR literature data set were clustered using SOMs (MOLMAP by Prof. G. Schneider, Frankfurt, Germany). The reference (literature) data were clustered in a defined area of the SOM. Based on CATS molecule description reference molecules possessed higher intra-class similarities than to most Asinex samples, which is a realistic relationship for an optimized ligand class (Figure 51). Nevertheless, 3728 of 130353 Asinex molecules which were assigned to these reference cluster neurons were scored with the Bayesian classifier referred to as the "mGluR"-model (Section 3.11.2). The "mGluR"-model determines if a molecule is more similar to active mGluR than WOMBAT molecules (other non-mGluR drugs). All molecules with "mGluR"-model score below 20 (motivated by class separation performance Figure 38) were removed and the residual ranked and selected for testing (13 molecules). "Hit" molecule **M8** (Asinex ID: ASN-17326353) reached a Bayesian "mGluR"-model score of 38.18 and was discovered at rank 26 from all tested molecules with a slightly lower "mGluR"-score of 38.18 than the highest ranked molecule with 44.58 (lowest reached 20.03). In the SOM **M8** possessed an Euclidean distance to the centre of neuron 15/10 of 0.771 and was discovered in a cluster with 269 molecules implying 86 reference compounds. The Euclidean distance of the closest reference molecule to the cluster centre equaled to 0.741 and of the farthest was 1.21.

4. Results and Discussion
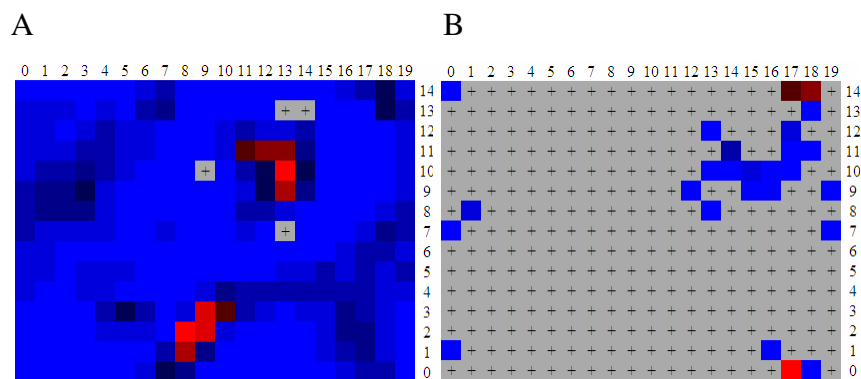
Figure 51: Self-organizing map (SOM) representation in 2D, colored by frequency of molecule samples from minimal (blue) to maximal (red) number. Training data included the Asinex Platinum Collection (version Nov2007) and mGluR literature data collections and was encoded with the chemically advanced template search (CATS) descriptor. A) SOM colored according to frequency of training data B) Selection of neurons which included selective mGluR5 molecules. "Hit" molecule **M8** (Figure 52) was discovered in neuron 15/10.

Active and inactive mGluR ligands were included in the literature data set which was clustered with the Asinex molecules by the SOM (Figure 51). It was observed that in the CATS descriptor space differently active molecules were not separated into different clusters.

A comparison of **M8** to the reference ligand collection revealed which structurally related molecules are already known (Figure 52). These molecules possess high affinity as selective NAMs of mGluR5 (Figure 52).

M8

ASN 17326353
mGluR5 IC_{50}=548 nM
mGluR1 IC_{50} >20 µM

mGluR5 IC_{50}=63nM
Alagille *et al.* 2005

K_D=44nM
WO 04108701
Hoffmann la Roche

mGluR5 pIC_{50}=8.05
Micheli *et al.* 2008

mGluR5 IC_{50}=0.94nM
Iso *et al.* 2006

mGluR5 IC_{50}=192nM
Iso *et al.* 2006
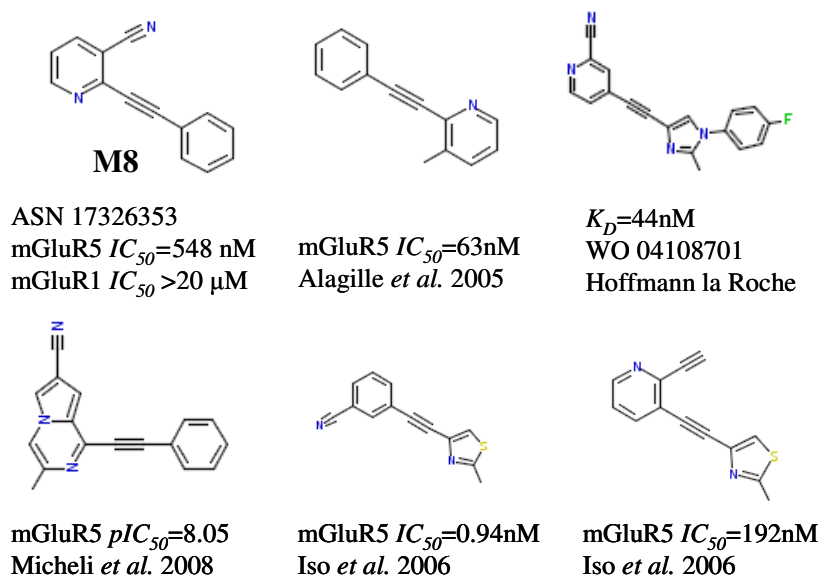
Figure 52: Hit molecule **M8**, discovered in the Asinex Platinum collection using chemically advanced template search and a self-organizing maps and structurally similar reference molecules. The IC_{50}-value of **M8** was determined in functional mGluR FLIPR assay. All molecules are subtype selective.

116

The discovered molecule **M8** exhibited activity of IC_{50}=548nM and differed from known ligands in the position of the substitution with the cyano group which lead to low nanomolar activity for acetylen linker including scaffold series (Alagille *et al.* 2005, Iso *et al.* 2006 and Micheli *et al.* 2008). A recent study by Vanejevs present a similar substitution like in **M8** as starting point and lead to the discovery of potent mGluR NAMs for mGluR1 and mGluR5 as well as PAMs for mGluR5 (Figure 53, Vanejevs *et al.* 2008).
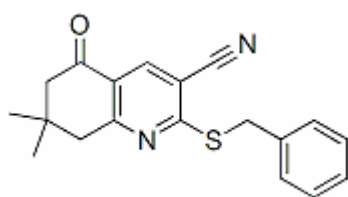
A   **B**
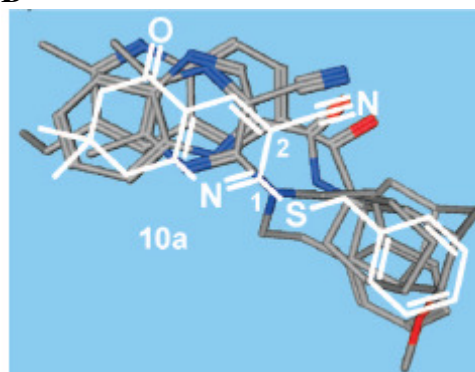
Figure 53: From Vanejevs *et al.* 2008. A) Molecule **10a** (original publication number) resulted from B) a pharmacophore alignment of **10a** with known active molecules. It shows a cyano group at a similar position on the pyridine cycle as **M8**.

Here, the single "hit" molecule **M8** was reported out of 63 molecules selected using CATS and a SOM. Besides the 13 Asinex molecules all other 50 molecules selected for experimental testing (27 Specs, 3 Maybridge and 20 Enamine pick collections), which were selected in similar manner to the Asinex molecules, all 50 were functionally inactive on mGluR1 or mGluR5. The reason might lie in the post filtering step with the Bayesian "mGluR"-model. It was applied since the SOM filtering resulted in too many molecules for experimental testing and did not allow for a linear ranking. In the present study the combination of pharmacophoric molecule ecoding with machine learning and subsequent ranking with Bayesian classifier lead to a "hit" rate of 2%.

4.2.4 Modeling of binding modes

The transmembrane ligand binding site of GPCRs was modeled and analyzed for a particular family C receptor, the human mGluR5. The binding conformations of several ligands discovered by virtual screening methods were predicted using molecular docking. Uncertainties in side chain orientations of the modeled structure were accounted for by optimization of receptor and ligand flexibility during the docking process. Selected docking poses were correlated to experimental data, gained from

4. Results and Discussion

mutations studies of family C GPCRs and crystal structures available from known complexes of family A GPCRs. Even family A binding ligands might posses an entirely different interaction profile compared to family C GPCRs both TM binding pockets originated from an ancestral transmembrane protein and both TM domains have a similar heptahelical arrangement (Pin *et al.* 2003). So far none endogenous ligands are known which bind to the TM domain of mGluRs only for the orthosteric site and therefore, it can not be expected that the binding pocket is as conserved as in family A GPCRs. The accuracy of a predicted family C GPCR structure using a family A GPCR as template remains unknown until experimental prove. Therefore the mGluR5 structure is treated in present study as hypothetical and applied for retrospective evaluation of experimental data.

Docking poses

Since for each of the mGluR5 ligands several diverse poses resulted from docking, they were evaluated manually. Selection criterion was: obligatory space occupation between TM3, TM5 and TM6 with interactions to side chains of residues facing the TM binding pocket (focusing on data from mutation studies Table A 3). Similar molecule substructures were expected to occupy similar parts of the binding pocket.
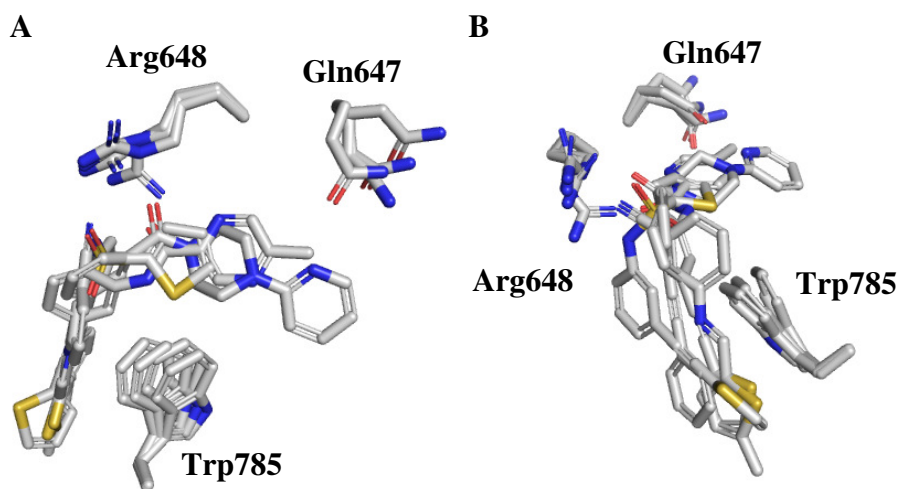
Figure 54: Docking conformation of 4 NAMs (**M1**, **M2**, **M3** and **M5**) discovered in this study in the modeled structure of mGluR5. A) View from the extracellular site showing the TM binding region. B) Side view between TM5 and TM6, residues in close proximity are highlighted.

The hypothetical binding region for acetylenic substructures was defined in parallel orientation to the membrane lipids and to Trp785 and located between TM3, TM5 and TM6. All MPEP- or MTEP-like ligands discovered with virtual screening as they are **M1**, **M2**, **M3** and **M5** were placed similarly by molecular docking with respect to the

118

acetylenic molecule part (Figure 54). Hydrogen bond interactions were mainly observed to Arg648 and Gln647, both located in TM3. For the small ligand **M1** docking poses in parallel orientation to the helices as well as parallel to the membrane were predicted (data not shown). Larger ligands, **M2**, **M3** and **M5,** occupied both oriented locations (Figure 54). The docking pose of **M1** (parallel to helix orientation) in which the acetylenic molecule part overlapped with the one in **M2**, **M3** and **M5** has been selected (Figure 54).

Binding pocket of mGluR5

Mutation studies revealed the location of the mGluR group I binding pocket between TM3, TM5, TM6 and TM7 (Litching *et al.* 1999, Pagano *et al*. 2000, Malherbe *et al.* 2003, Muehlemann *et al.* 2006, Malherbe *et al.* 2006). The docked ligand poses were predominantly placed in the binding pocket defined by these mutated residues (Figure 55).
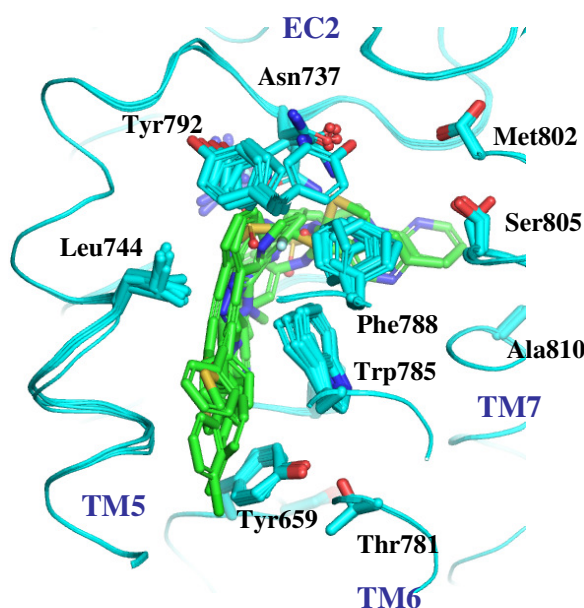
Figure 55: Binding pocket of the modeled structure of mGluR5 containing docked conformations of **M1**, **M2**, **M3**, **M4**, **M5**, **M6** and **M7**. Docking poses which achieved the highest score according to *docking_score* (IFD, Schrödinger LLC, New York, 2008). Residues involved in mGluR5 NAM binding known from mutagenesis studies are highlighted. All residues are in close distance to the ligands (< 4Å).

Amino acids Met802 (Muehlemann *et al.* 2006), Ser805 (Litching *et al.* 1999) and Ala810 (Malherbe *et al.* 2006, Pagano *et al*. 2000) at TM7 could not be correlated to the binding poses because they were pointing away from TM3 and TM6 according to the proposed receptor model. The largest molecules (**M2**, **M3**, **M4**, **M5** and **M7**) were placed in the region close to Met802, Ser805 and Ala810 but no directed interactions with them are suggested by the proposed docking poses.

4. Results and Discussion
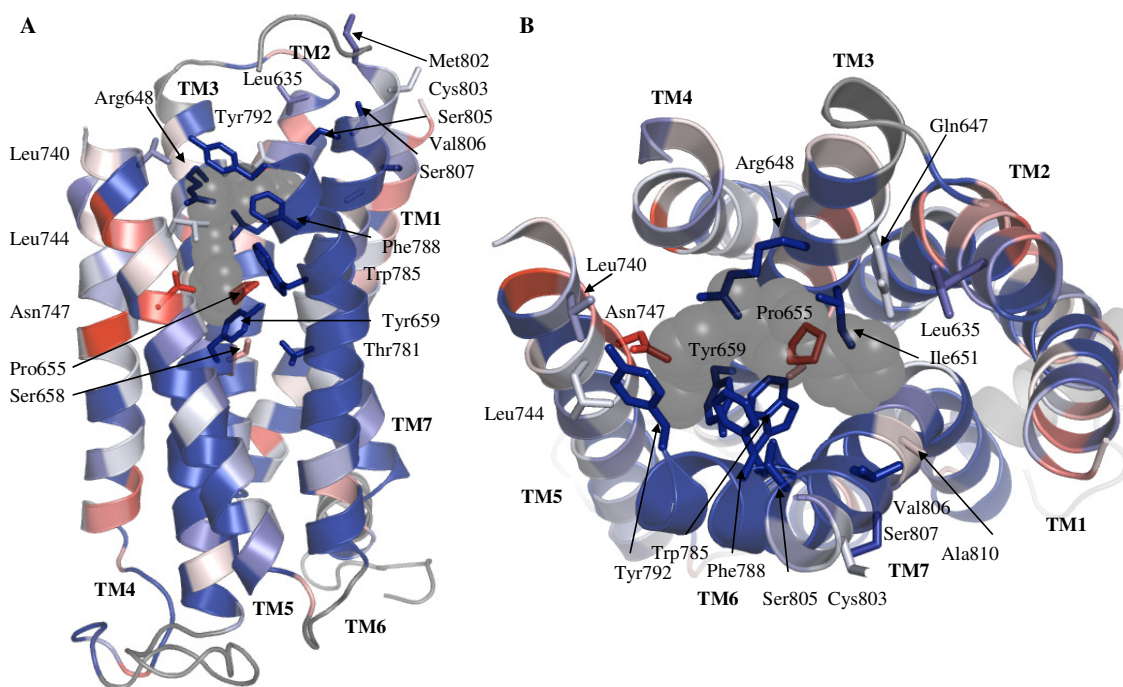
Figure 56: A modeled structure of mGluR5 with **M3** (given in sphere representation, black transparent) in **A** side view and **B** top view from extracellular side. Residues discovered with mutation studies for ligand binding in family C GPCRs are given in stick representation. Structure is colored according to conservation (Shannon entropy values) of the mGluR subfamily. Colors are applied blue (low entropy, high conservation) to red (high entropy, low conservation). Extracellular loops are omitted for clarity reasons.

The conservation level of residues involved in ligand binding of family C GPCRs was high (Figure 56 A-B, for further details see Section 4.1.7). The predicted conformation of the largest ligand, **M3**, occupied regions close to positions probed by mutations.

The selection procedure for docking poses was influenced by ligand binding conformations known from crystal structures of BR and ßAD receptors. It should be mentioned that the result of the docking is dependent on the modeled structure which changes when a different sequence alignment or structural template is applied. Since the extracellular loops could not be modeled reliably due to low conservation and the conformation of the EC2 loop (covering up the binding pocket similar to the modeling template rhodopsin), docking results for the modeled mGluR5 structure might not have provided suitable interactions to any of the loops and are restricted to the binding pocket of retinal.

120

Ligand binding of GPCRs

Several crystal structures of βAD receptors with bound inverse agonists (carazolol, cyanopindolol, and cholesterol), the BR with retinal and the human A_{2A} adenosine receptor with an antagonist have been solved so far. The bovine rhodopsin complex with retinal was used as template structure for mGluR5 modeling and belongs to familyA receptor with a covalently bound ligand. For comparison reasons known crystal structures of GPCR complexes (PDB IDs: 1U19, 2RH1, 2VT4, 3D4S) were superposed with the modeled mGluR5 structure, resulting in superposition of their ligands with the docked conformation of **M3** (Figure 58).

The superposition of the adrenergic and rhodopsin receptors revealed that the transmembrane binding sites occupied by their ligands carazolol, cyanopindolol, cholesterol and retinal overlaps in general and with the one defined by docked poses of mGluR5 NAMs (Figure 58 A and C). All three inverse agonist carazolol, cyanopindolol, cholesterol are located in the upper part of the TM cleft and bind through polar interactions to amino acids at TM3 (3.28 and 3.32), TM5 (5.42) and TM7 (7.39 and 7.43) and hydrophobic contacts to aromatic amino acids 5.32, 5.34, 5.38, 6.48, 6.52, 6.51 and 7.35 as shown for carazolol in Figure 57 (Cherezov *et al.* 2007, Hanson *et al.* 2008, Warne *et al.* 2008 and Okada *et al.* 2004).

4. Results and Discussion
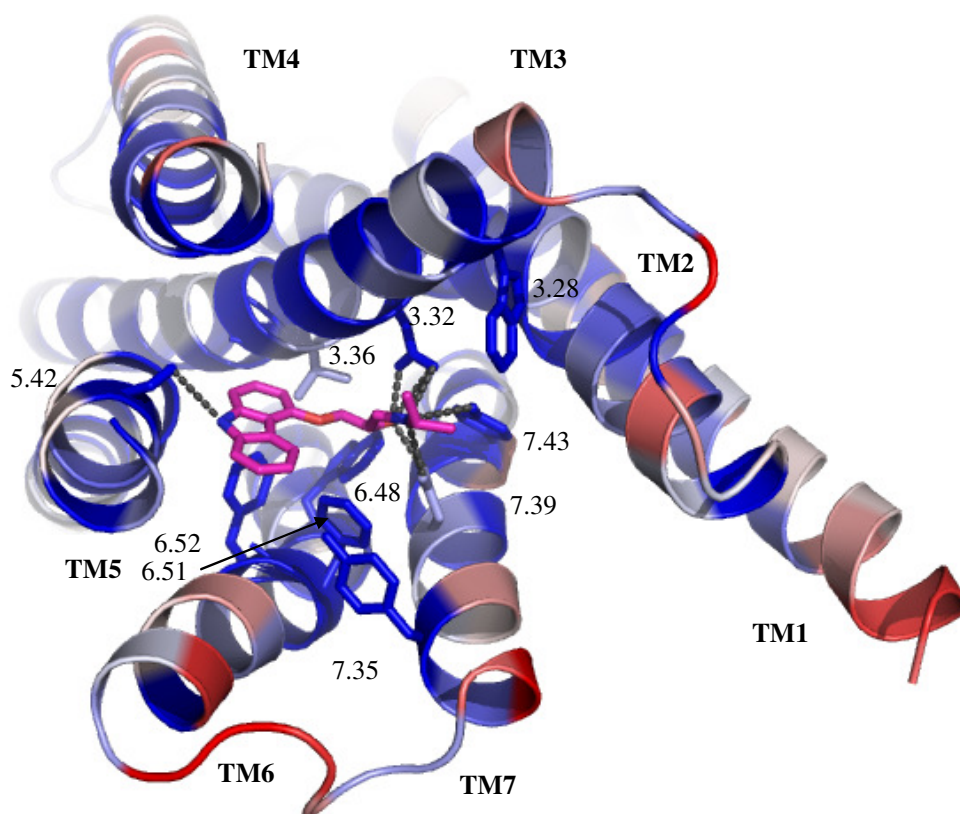
Figure 57: The structure of ß₂AD and carazolol (PDB ID: 2RH1, Cherezov *et al.* 2007). Residues in close proximity to carazolol are given in stick representation and contact indicated with dashed lines. Shannon entropy values for the β-adrenergic subfamily MSA (from GPCRDB, v.10.0, Horn *et al.* 2003) are projected onto the protein structure. Colors are applied as follows: blue (low entropy, high conservation) to red (high entropy, low conservation). EC2 is omitted for clarity.

Retinal extends deeper into the TM cleft as it has been observed for predicted conformations for NAMs of mGluR5 (Figure 58 B). The binding pocket extension in direction to TM2 and TM1 was similar for mGluR5 NAMs **M1**, **M2**, **M3** and **M5** and retinal.
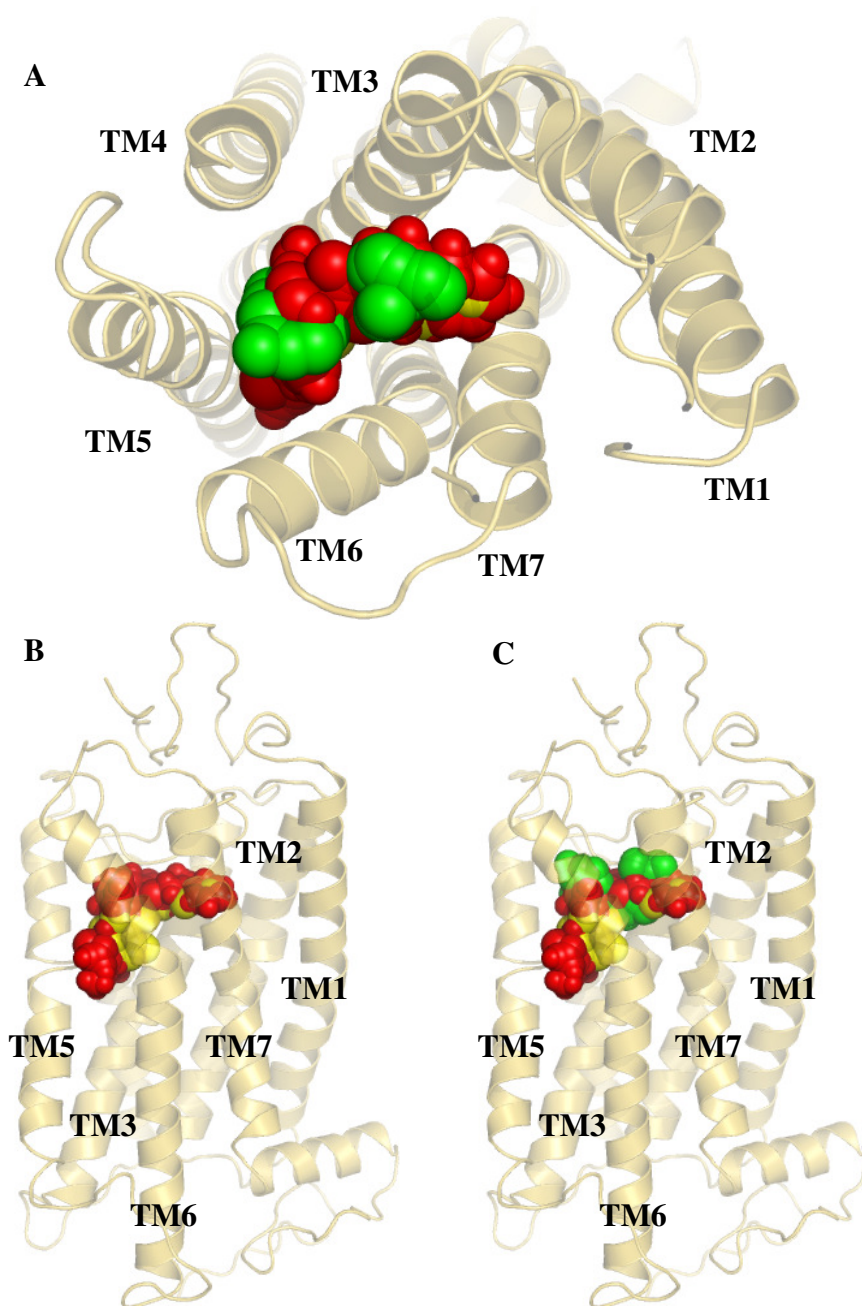
Figure 58: Superposition of crystal structures of 1U19, 2RH1, 2VT4, 3D4S and predicted mGluR5 structure with docked ligands (**M1**, **M2**, **M3** and **M5**). The rhodopsin structure (1U19) is given in cartoon and ligands in sphere representation. Ligand colors are applied as follows: βAD ligands (carazolol, cyanopindolol, cholesterol) in green, retinal in yellow and docked mGluR5 ligands (**M1**, **M2**, **M3** and **M5**) in red. Different views on the same superposition are given A) view from the extracellular site on the TM. EC2 and part of the N-terminus are omitted for clarity. B) Side view, only retinal and mGluR5 ligands present C) retinal, mGluR5 ligands and carazolol, cyanopindolol, cholesterol.

The allosteric modulation of G protein-coupling through NAMs in GPCRs is still current issue in GPCR function research. Scheerer stated that G protein-coupling at the intracellular part of the transmembrane region close to the DRY-motif common for most GPCR can induce long-range stabilization effects into the ligand binding pocket (Scheerer *et al.* 2008). In mGluR5 and 2RH1 (β₂AD with carazolol, Cherezov *et al.*

4. Results and Discussion

2007) the ligand binding pocket is in a comparable distance of 20Å and 23Å to Arg668 and Arg131.

At the time the first and until now only human A_{2A} adenosine receptor structure became available it has been suggested that no common conserved binding pocket is shared by all GPCR subfamilies (Jaakola *et al*. 2008). In agreement with the hypothesis of conformational selection by the ligand (Cozzini *et al*. 2008), meaning that a ligand can stabilize a particular conformation of the receptor which is not the only one the receptor can adopt, different binding modes for NAMs can be expected. So far no proof was provided for different agonist or antagonist binding sites. Mutations, which were collected from literature in order to define the binding pocket of agonists and antagonists, indicated overlapping binding region, as discussed in Section 4.1.8. The present study revealed that a similar location can be accommodated by mGluR5 NAMs, discovered with virtual screening strategies. Using molecular docking a common binding region for several NAM was defined which are in contact to amino acids and positions important for NAM affinity and effect.

4.2.5 Virtual Screening - conclusions

Based on the knowledge about diverse potent mGluR binding ligands several virtual screening procedures were applied in order to find novel mGluR5 selective allosteric modulators. Eight active molecules with functional activity below 10μM out of 228 tested made a "hit" rate of 3.5%. The "hit" rate per strategy was different because unequal molecule numbers were available for acquisition and further experimental testing and because the individual differences between the approaches applied; the respective results are 12% (combined Bayesian model ("mGluR"- and "selectivity"-model) with FCFP_6), 8% (FCFP_4-based similarity search), 2.4% (MPEP-based shape similarity), 1.8% (PHRFP _2-based similarity search) and 1.6% (CATS-based SOM clustering with subsequent "mGluR" Bayesian model ranking). Similarity search approaches discovered three of eight active molecules. By combination of two Bayesian models three additional molecules were retrieved. Shape-similarity search and SOM clustering with subsequent Bayesian model scoring lead to one additional active molecule each. The most potent NAMs (**M8** and **M2**) yielded IC_{50}=548nM and IC_{50}=458nM and were discovered with the CATS-based clustering using SOMS and the Bayesian two-models approaches.

124

Three molecules (**M2**, **M6** and **M8**) possessed activity in nanomolar and five (**M1**, **M3**, **M4**, **M5** and **M7**) in low micromolar range. Seven of the molecules (all but **M3**) were selective to mGluR5. The single nonselective molecule **M3** (discovered using a Bayesian model with FCFP_6) contained an acetylene linker, which so far was only reported to be part of selective mGluR5 binding ligands. The structure of the single PAM **M6** (discovered with FCFP-based similarity search) was an example similar to a recently reported series of known NAMs which were converted to PAMs by changing the position of a methyl group a pyridine ring (Sharma *et al.* 2008).

The number of tested and active molecules did not allow for comparison of prediction accuracy in prospective virtual screening. The goal to find novel mGluR5 binding modulators was achieved, even for both modulator classes, the negative and positive allosteric modulators.

For each of the "hit" molecules a binding mode was predicted using molecular docking with the modeled receptor structure of mGluR5. The binding area was predicted to overlap with the space occupied by ligands in several crystallized family A GPCRs. A hypothesis has been derived for molecules with differences in size regarding orientation in the TM region. From the predicted binding poses it was suggested that small molecules could share the binding cleft between TM3, TM5 and TM7 perpendicular to the membrane plane, the occupation of this TM region is in line with published mutation experiments. Large molecules were predicted to bind using more space parallel to the membrane plane, which is analogous to the retinal binding position. For four of the ligands (**M2**, **M3**, **M5** and **M7**) it was predicted that Arg648 at TM3 might be involved in binding in mGluR5. Ligand **M3** extended to a further polar interaction to Gln647. Published mutation studies define the same region as important for binding as it was occupied by the docking poses.

5. Outlook

This study facilitates an improved understanding of the structure and function of family C GPCRs. The binding of allosteric modulators of mGluR5 were analyzed using sequence-, ligand- and structure based approaches. Future steps towards understanding of the family C GPCR activation process can be expected from more detailed experimental results regarding structure as well as improvement of computational approaches which are used its prediction and evaluation.

In future work the results of the performed conservation analysis could be used as guidance for design of molecular biology experiments targeting particular conserved positions of family C GPCRs, at conserved positions as they were discovered at helix-helix contacts. Mutation or labeling experiments at these positions could provide further insights into interaction networks between amino acids in the TM region and their contacts to loop regions. The GPCR structures known so far differ in bending of TM helices and possess unique stabilizing contacts; pairs of residues involved in those contacts might build a set of exchangeable amino acid combinations and could be used for motivation of different amino acid grouping schemes than the scheme applied in this study. This would allow to track additional features important for TM packing and might contribute to understanding of changes taking place during activation.

Uncertainties regarding the modeled structure of mGluR5 could be evaluated more thoroughly using all available family A GPCRs as templates, since each of them could provide in some structural parts a better starting point for structure prediction. Conservation profiles of structural features might be applicable to local alignment of the target sequence to several structural templates in order to reveal a closer relation. Besides these computational efforts, the binding pocket definition for family C GPCRs will benefit most from additional crystal structures of more closely related GPCRs especially those containing PAMs or agonists.

So far dissimilarities between family A GPCR complex structures were detected in interaction of ligands with the EC loops. This raised the attention to their relevance in correct ligand pose prediction. The EC2 loop has been crystallized in β-sheet (rhodopsin, 1U19), α-helix (β_2AD receptor, 2RH1) and constrained coil conformations

(A$_{2A}$ adenosine receptor, 3EML) for different receptor types. This points out that the loop regions adopt structured conformations which are important for function. Family C GPCRs possess a cysteine-rich domain, which connects the VFT to the TM and might be structured to exhibit contacts to EC loops. This is the most uncertain part in structure prediction of all GPCRs to date and remains essential for understanding of family C receptor activation.

The preparation of docking protocols could benefit from predefined interactions and provide easier selection procedures for probable common binding poses for several diverse ligands. Different structural templates would further allow to define differently formed binding pockets. Their fold cannot be adapted with docking procedures considering receptor flexibility. Several residues predicted as possible partners for interaction to the virtually discovered molecules could be tested in mutation analysis with artificial amino acids to prove the hypothetical binding mode.

The discovery of novel mGluR binding ligands remains challenging since NAMs and PAM share the same molecular scaffolds (Sharma *et al.* 2008) and it seems more likely that substituents define receptor selectivity and modulation effects. The application of a variety of virtual screening approaches revealed several methods as capable of active molecule prediction. *De novo* design of molecules (Böhm 1992, Böhm and Schneider 2000) might contribute new ideas for novel molecule frameworks which could replace known scaffolds while still presenting interacting groups in the required orientations.

6. Summary

G protein-coupled receptors (GPCR) are the largest family of cell-membrane located receptors involved in various signaling pathways and for this reason a widely analyzed biological target molecule in the pharmaceutical research area. GPCR ligands modulate neuroleptic communication. Small molecules binding to GPCRs are aimed to cure different central neural system disorders as schizophrenia, Alzheimer's and Parkinson's diseases. Receptor based drug design is difficult in case of GPCRs due to the limited number of crystal structures. Throughout evolution the function of GPCRs were optimized to enable the recognition of diverse types of transmitting molecules. However, the fold of the transmembrane domain, the signaling processes using G proteins as well as the location of the transmembrane ligand binding side remained similar. Since this study focused on family C GPCRs and the only structurally resolved GPCRs belong to family A, the transmembrane domain conserved between both types was analyzed in comparative way. Aiming at understanding of functional and structural features of family C GPCRs and their difference to family A GPCRs experimental findings were correlated to patterns conserved in their sequences. To characterize the transmembrane binding site of family C ligands, the TM domain's structure was modeled derived from a family A GPCR as template.

Representatives from two GPCR families, bovine rhodopsin (BR, family A) and the metabotropic glutamate receptor five (mGluR5, family C), were selected to investigate common feature conservation. The structure of mGluR5 was predicted choosing a suitable template for homology modeling based on the level of sequence homology. For mGluR5 sequence identity in the TM region was optimized reaching 12% to the human β-adrenergic receptor type 2 and 13% to bovine rhodopsin, therefore BR was chosen as structural template. For the entire family C and the vertebrate rhodopsin subfamily conservation profiles were calculated using Shannon entropy and projected onto the structure of mGluR5 and BR, respectively.

The Shannon entropy analysis was applied for definition of conserved positions in the TM of family C GPCRs. This analysis was based on a multiple sequence alignment (MSA) of 96 functionally diverse sequences of family C GPCRs. The MSA proposed in this study was found to contain more conserved positions than the publicly available

alignment derived in a completely automated manner. Using a chemical amino acid typing scheme, the family C MSA was analyzed position-wise regarding the conservation of these groups. The correlation of available experimental data for family A and C GPCRs allowed to define functionally and structurally important positions as well as differences for both families. According to the modeled mGluR5 structure conserved positions were located at helix-helix contacts.

Experimental data describing functional and structural effects in diverse family C GPCRs caused by mutations in TM region were collected from literature yielding 157 mutations for 14 TM domain binding molecules. In this study the combination of site-directed mutagenesis studies with the entropy conservation analysis method has proven valuable in testing and refining hypotheses for ligand binding targeting the TM allosteric binding pocket of metabotropic glutamate receptors.

In a virtual screening approach several techniques to describe molecules and to build a predictive model were evaluated. A ligand data base of 1240 diverse mGluR binding ligand series served as knowledge base to establish predictive models for mGluR5 active molecules. Previous selectivity assignments enabled the compilation of focused sets of reference molecules for model training. 2D and 3D molecule ecoding versions were applied in combination with machine learning and similarity-based techniques in order to predict potentially active molecules. Topological pharmacophore fingerprints and molecular shape similarities were employed in similarity searches with SOMs and Bayesian networks. Out of 228 molecules selected with different methods, 8 molecules exhibited functional activity below 10μM (experiments provided by Merz Pharmaceuticals). Seven molecules were tested as negative (NAM), one as positive allosteric modulator, among the seven NAMs was a single non-selective (mGluR1/mGluR5) NAM. Similarities to reference molecules could eventually explain the activity of the selected molecules. A possible binding orientation of the mGluR binding ligands was predicted for several of these virtually discovered molecules.

Negative allosteric modulators were docked into the modeled binding pocket of the human mGluR5 using a procedure which accounts for receptor flexibility and therefore allows to conformationally adapt the interacting residues. Ligand conformations were analyzed based on the resulting docking poses resulting in a hypothesis where several

6. Summary

ligands overlap in their binding area. The volume of the binding pocket was compared to known complex structures of GPCRs. Retinal (BR) and the inverse agonists' (ß2AD) binding conformations were compared to the suggested ligand poses of mGluR NAMs. An overlap of NAMs and the crystallized ligands was detected. For the β-adrenergic receptors (three experimental protein structures with diffusible ligands available so far) and the mGluR subfamilies the conservation analysis of the binding pocket revealed that residues in close proximity to ligands are conserved in functional families besides few residues, which might facilitate ligand selectivity among subtypes.

Both the sequence and the ligand-based approaches revealed similarities and differences between functionally diverse GPCRs and provided a basis for modeling of family C binding modes using knowledge from family A receptors. The prepared data, consisting of sequences, mutations and ligands, in its combination supported the application and analysis of the employed methods.

7. Zusammenfassung

Im Rahmen dieser Arbeit wurden Konzepte zur Aufklärung struktureller und funktioneller Eigenschaften von G-Protein gekoppelten Rezeptoren (GPCR) der Familie C entwickelt und angewendet. In diesem Zusammenhang wurde anhand verfügbarer experimenteller Daten aus Mutations- und Ligandenbindungsstudien ein Vergleich konservierter Bereiche der Rezeptor-Familien A und C angefertigt. Anschließend an die retrospektive Analyse wurde eine prospektive virtuelle Vorhersage neuer mGluR5 (metabotroper Glutamatrezeptor des Typs fünf) bindender Moleküle durchgeführt und das Ergebnis durch Laborexperimente validiert. Insgesamt wurden Sequenz-, Struktur- und Liganden-basierte Methoden angewendet und ihre Ergebnisse im einem strukturellen Kontext zusammengefasst und diskutiert. Dieser wurde durch die vorhergesagte dreidimensionale Struktur des mGluR5 geschaffen. Die Struktur wurde anhand der bekannten Röntgenkristallstruktur des bovinen Rhodopsins, eines Familie A GPCRs, modelliert und einer Prüfung bezüglich der Eignung für Ligandenbindungsstudien unterzogen.

GPCRs zeichnen sich durch strukturelle und funktionale Gemeinsamkeit in der Transmembrandomäne (TM) aus. Diese durchspannt mit sieben alpha-helikalen Bereichen die Zellmembran und ermöglicht die Weiterleitung extrazellulärer Signale in das Innere der Zelle. Auslöser solcher Signale können unterschiedlichen Strukturklassen zugeordnet werden: Peptide, Ionen, Neurotransmitter und andere kleine organische Moleküle. Die Erkennung der Signalmoleküle führt zu einer Konformationsänderung des Rezeptors und wirkt sich dadurch aktivierend auf G-Proteine aus, die am intrazellulären Teil des GPCRs gebunden sind. G-Proteine lösen sich vom GPCR und leiten weitere biochemische Reaktionen innerhalb der Zelle ein. Unterschiede zwischen GPCR ergeben sich durch die Bindung verschiedener Signalmoleküle und die nachgeschalteten G-Proteine. Somit kann beides zu einer Erhöhung der Variabilität von Singaltransduktionswegen beitragen. Im humanen Genom werden bis zu 900 verschiedener GPCRs vermutet. Die Implikation dieser Rezeptorklasse in die Kontrolle wichtiger Prozesse im menschlichen Körper macht ihre Erforschung aus kommerziellen Gründen besonders attraktiv; 40% der verkäuflichen Medikamente wechselwirken mit GPCRs.

7. Zusammenfassung

Aufgrund von Ähnlichkeiten in der Proteinsequenz, dem pharmakologischen Profil sowie der Art der gebundenen Liganden gilt die Klassifikation der GPCRs in fünf Familien. Die Familie A ist die größte und am meisten untersuchte und beinhaltet die zur Zeit einzigen strukturell aufgeklärten GPCRs. Strukturell charakteristisch für Familie C GPCRs ist die extrazelluläre Domäne (VFT, engl. *venus fly trap* benannt nach dem Funktionsprinzip), die für die Bindung des körpereigenen Neurotransmitters, Glutamat, erforderlich ist. Glutamat aktiviert den Rezeptor über die orthosterische Bindestelle, von der über eine Konformationsänderung die Aktivierung auf die TM übertragen wird. Die Sequenzähnlichkeit der VFT zu einem bakteriellen Protein ermöglichte die Hypothese über die Fusion dieser extrazellulären und der transmembranen Domäne (Pin *et al.* 2003).

Die am längsten bekannte nicht bakterielle GPCR-Struktur ist die des bovinen Rhodopsins (BR, engl. *bovine rhodopsin*) im Komplex mit Retinal. Zur Vorhersage einer Struktur für mGluR5 wurde die am höchsten aufgelöste Struktur von BR als Vorlage verwendet. Diese Modellierung erfolgte mit der Methode des *comparative modeling* (auch als Homologiemodellierung bezeichnet), bei der das Proteinrückgrad der strukturellen Vorlage als Initialposition zur Platzierung von Aminosäuren der Zielstruktur verwendet wird. Die Annahme, nach der Sequenzidentität zu einer ähnlichen dreidimensionalen Faltung von Proteinen führt, stellt die Grundlage für die Vergleichbarkeit der Strukturen dar und entscheidet über die Qualität des Ergebnisses. Daher werden die Proteinsequenzen als Grundlage für die Überlagerung im Modellierungsprozess aligniert. Im Falle von mGluR5 wurde das Sequenzalignment manuell angefertigt, da die Sequenzidentität bei 13% lag und nur aufgrund konservierter Aminosäuren in den einzelnen TM optimiert werden konnte. Die Verfeinerung des mGluR5-Modells erfolgte mittels struktureller Optimierung des protonierten Rezeptors zum nächsten Energieminimum hin.

Die strukturelle Verwandtschaft von GPCRs der Familien A und C sollte anhand von Proteinsequenzen verglichen werden. Als Grundlage wurde ein multiples Sequenzalignment (MSA) von 96 Familie C GPCRs verschiedener Funktion erstellt. Das MSA wurde automatisiert für die unterschiedlichen Subfamilien generiert. Diese wurden anschließend manuell korrigiert, zusammengeführt und an BR aligniert. Die Diversität in Länge und Aminosäurekomposition machte ein verlässliches Alignment

kompletter Sequenzen schwierig und die Reduktion des MSAs auf einzelne TM-Blöcke erforderlich. So entstanden sieben MSA-Blöcke, jeweils ein Block pro TM-Helix. Bei Familie C GPCRs sind die sieben TM-Helices an der Bindung allosterischer Liganden beteiligt. Die durchschnittliche Identität der Aminosäuren in der TM wurde für die unterschiedlichen Subfamilien im Einzelnen sowie für das gesamte Familie C MSA evaluiert.

Zur weiteren Konserviertheitsanalyse (KA) wurde eine Kodierung der Aminosäuretypen in Gruppen nach ihren chemischen Eigenschaften in der Entropieberechnung nach Shannon angewendet. Diese Gruppen sollten ähnliche Aminosäuren als konserviert zusammenfassen, so dass die KA insbesondere Änderungen zwischen chemischen Klassen erkennbar machen kann. Die abstrakte Betrachtung knüpfte dabei an die unterschiedliche Tolerierbarkeit einer Aminosäuresubstitution an, wie sie in Substitutionsmatrizen für Alignments Verwendung finden. Insgesamt wurden neun Gruppen definiert und die Entropieberechnung auf diesem alternativen Alphabet auf das MSA angewendet.

Entropie ist eine Eigenschaft, die in dem Bereich der Informationstheorie untersucht wurde, um den Informationsgehalt in einem weitergeleiteten Signal zu bestimmen. Entropie wird auch als der Grad der Unordnung in einem System bezeichnet. Ein hoher Entropiewert wird deswegen als Unordnung oder Rauschen interpretiert, ein niedriger hingegen repräsentiert einen Zustand mit hohem Ordnungsgrad oder Eindeutigkeit. Im Zusammenhang mit einem MSA wurde dieses Verständnis auf die Konserviertheit einer Sequenzposition in der GPCR Familie übertragen. Für jede Position wurde die Konserviertheit anhand der Häufigkeit im Auftreten verschiedener Aminosäuregruppen festgestellt. Der Beitrag einer Sequenz richtete sich dabei nach der Diversität, die diese im Vergleich zu anderen Sequenzen beitrug. Diese Maßnahme soll verhindern, dass die ungleichen Anzahlen bekannter Sequenzen in unterschiedlichen Subfamilien mitbewertet werden. Das Ergebnis der KA enthielt sowohl die Entropiewerte als auch die am meisten konservierte Gruppe einer Sequenzposition. Folgende Bewertungskriterien wurden zur Interpretation der Werte in Anspruch genommen: 1. experimentelle Befunde aus Mutationsstudien an Rezeptoren der Familie C, 2. KA der Rhodopsin Subfamilie der Vertebraten, 3. veröffentlichte KA Studien für Familie A GPCRs und 4. Kristallstrukturen von Familie A Rezeptor-Ligand-Komplexen. Aus den

7. Zusammenfassung

in der Literatur veröffentlichten Mutationsstudien wurde eine Mutationsdatensammlung erstellt, in der die Sequenzpositionen der mutierten Aminosäuren anhand des Familie C MSA auf entsprechende Positionen in der Sequenz von mGluR5 und BR übertragen werden konnten. Auf die gleiche Weise wurden bekannte Mutationen an Familie A GPCRs über das Alignment zu BR in Zusammenhang zur Familie C GPCRs gesetzt. Da BR als Grundlage für die Modellierung der mGluR5-Struktur genutzt wurde, wurde die KA auf die Subfamilie der Vertebraten-Rhodopsine angewendet. Das dazu verwendete MSA stammte aus einer öffentlichen Datenquelle und war aufgrund der Ähnlichkeit innerhalb dieser Funktionsklasse automatisiert erstellt worden. Konservierte Bereiche beider GPCR Familien wurden korreliert und einzelne Position, deren funktionale oder strukturelle Bedeutung durch Mutationsstudien bestätigt wurde, detailliert diskutiert. Zu diesen für GPCRs charakteristischen Region gehörten Helix-Kontaktstellen, funktionale Sequenzmotive, G-Protein koppelnde Stellen, Dimerisierungs-Kontakflächen und die TM-Bindetasche für Liganden. Im Bereich der Bindetasche fanden auch bekannte GPCR-Komplexe ihre Anwendung und wurden vergleichend diskutiert. Für strukturelle Merkmale sowie Teile der Bindetasche konnten Ähnlichkeiten, die aus der KA resultierten, bestätigt werden, im Gegensatz zu hypothetischen Dimerisierung-Kontaktflächen.

Über die TM-Bindetasche der mGlu-Rezeptoren ist bekannt, dass die Anregung des Rezeptors von einem allosterischen Modulator, der in der TM-Bindetasche bindet, beeinflusst werden kann. Dieser regulierende Effekt ist entscheidend für den medizinischen Einsatz von allosterischen Modulatoren. Negative allosterische Modulatoren (NAM) hemmen die Aktivierbarkeit des Rezeptors durch Glutamat und positive allosterische Modulatoren (PAM) erhöhen diese, ohne dass sie um die Bindetasche mit diesem konkurrieren.

Mit dem Ziel, neue selektive allosterische Modulatoren für mGluR5 zu finden, wurden mehrere Liganden-basierte Ansätze zur virtuellen Vorhersage der Aktivität von Molekülen entwickelt und getestet. Die dabei angewendete Strategie basierte auf der Kenntnis bereits bekannter Liganden, deren Strukturen und Aktivitätswerte für das Erstellen von Vorhersagemodelle genutzt werden konnten. Strukturen bekannter mGluR bindender Moleküle wurden wissenschaftlichen Veröffentlichungen entnommen und gemeinsam mit Eigenschaften, die sie in ihren Effekt beschreiben, in einer

134

Datensammlung festgehalten. Die Liganden-Datensammlung diente als Referenz für Struktur-Wirkungs-Beziehungen, die mit Hilfe von computergestützten Vorhersagemethoden untersucht und zur prospektiven Molekülbewertung eingesetzt wurde.

Für die virtuelle Handhabung wurden alle Moleküle mit Kodierungsverfahren in ein untereinander vergleichbares Format überführt. Ingesamt wurden Oberflächen, Substrukturen und Pharmakophor-Eigenschaften beschreibende Kodierungen angewendet. Die *Spherical Harmonics Descriptors* (SHD), die das Volumen und die dreidimensionale Form eines Moleküls wiedergeben, und die *Chemically Advanced Template Search* (CATS) Deskriptoren, die die Häufigkeiten von Pharmakophorpunkten im Molekül kodieren, wurden zur retrospektiven Analyse der Referenzliganden eingesetzt. Dabei konnte festgestellt werden, dass selektive sowie unterschiedlich stark aktive mGluR Liganden in dem von den Deskriptoren aufgespannten Raum nur schwer unterschieden werden können. Die Unterscheidbarkeit wurde anhand von Clustern in einer selbst-oganisierenden Karte (*engl. self-organizing map,* SOM) überprüft. Für NAMs und PAMs konnten einige separate Cluster definiert werden. Die Molekülkodierung mit CATS schnitt besser ab und wurde anschließend für die prospektiven Vorhersagen weiterverwendet.

Die prospektive Vorhersage stützte sich auf unterschiedliche Methoden zur Ähnlichkeitsberechnung und Arten der Molekülkodierung. Zwei verschiedene Pharmakophortypen beschreibende Deskriptoren, CATS und PHRFP, sowie die *Functional Class Extended-connectivity Fingerprints* (FCFP) wurden zur Ähnlichkeitssuche mit Tanimoto-Bewertung eingesetzt. Die Ähnlichkeit zu bekannten hochaktiven Modulatoren des mGluR5, die nicht an mGluR1 bindenden, sollte zur Auffindung selektiver NAMs führen. Durchsucht wurden, abhängig von der Methode, bis zu 5 Millionen käuflich zugänglicher Moleküle. Zwei ergänzende Klassifikatoren nach Bayes wurden trainiert, um die Ähnlichkeit zu hochaktiven mGluR-Liganden und selektiven mGluR5-Liganden als Wahrscheinlichkeiten ausdrücken zu können. Beide Modelle wurden nach dem Dominanzprinzip von Pareto zur Selektion von potentiellen Liganden kombiniert. Die mit CATS definierte Ähnlichkeit wurde weiterhin zum Clustern der Referenzmoleküle mit unbekannten Molekülen unter Verwendung einer SOM eingesetzt. Moleküle aus Neuronen, die selektive mGluR5 Liganden enthielten,

wurden im zweiten Schritt durch die Bewertung mit dem Bayes'schen Klassifikator nachgefiltert und sortiert. Ebenfalls im Rahmen der Ähnlichkeitsbewertung wurde der NAM, MPEP, als Referenz zur Molekülform (*engl*. shape) basierten Suchen verwendet. Anhand aller sortierten Molekülisten, die als Ergebnis aus unterschiedlichen Vorhersagen resultierten, wurden Moleküle zur experimentellen Testung ausgewählt. Die Testung der Moleküle wurde von Mitarbeitern des Unternehmens Merz Pharmaceuticals durchgeführt. Die Moleküle wurden hinsichtlich ihrer modulatorischen Wirkung am mGluR5 gemessen, wobei negativer oder positiver Effekt festgestellt werden konnten. Die Art der Messung war funktional, da Änderungen des Ca^{2+}-Levels in der Zelle nachgewiesen wurden. Für NAMs wurde eine Konzentrationsreihe gemessen, welche die Berechnung des IC_{50} (Konzentration des NAMs, bei der die Aktivität des Rezeptors um 50% reduziert ist) ermöglicht. Bei PAMs wurde die Erhöhung der Aktivierung gemessen. Nachgewiesene Modulatoren wurden zur Selektivitätsbestimmung einer Testung am mGluR1 unterzogen. Insgesamt konnten 8 von 228 getesteten Molekülen im Aktivitätsbereich unter 10µM ermittelt werden, darunter befand sich ein PAM. Von den restlichen sieben NAMs waren fünf selektiv für mGluR5. Der einzige nicht selektive Ligand enthielt einen Acetylenlinker, der bislang als charakteristisch für mGluR5 selektive Liganden galt. Jeder Ligand wurde auf Ähnlichkeit zu Molekülen des Referenzdatensatzes untersucht und dessen Aktivität mit diesem verglichen.

Alle identifizierten NAMs wurden hinsichtlich möglicher Interaktion mit mGluR5 untersucht. Die Moleküle wurden in das zuvor erstellte Strukturmodel von mGluR5 mit Hilfe des molekularen Docking eingepasst. Docking ist ein Verfahren, das die Struktur des Liganden in der Bindetasche des Biomoleküls platziert. Da es sich um eine vorhergesagte Struktur handelte, wurde die Ungenauigkeit in der Seitenkettenausrichtung durch flexible Anpassung beim Moleküldocking (engl. *induced fit docking*, IFD) berücksichtigt. Für eine Reihe der Seitenketten, die in Richtung der TM-Bindetasche orientiert werden könnten, wurde ein Protokoll angewendet, das diese optimiert, um die Interaktion mit dem Liganden zu verbessern. Das IFD behandelt den Liganden und den Rezeptor flexibel und ermöglicht durch schrittweise Evaluierung mit einer Bewertungsfunktion und Anpassung ein Ergebnis zu erzielen, das weniger von der Ausgangskonformation abhängt als beim Docken mit rigiden Rezeptoren. Die Ergebnisse des IFD wiesen ähnliche Bewertung für unterschiedliche Ligandenposen auf

und wurden deswegen manuell selektiert. Diese Vorgehensweise sollte eine Hypothese ermöglichen, wie die Liganden in der Bindetasche lokalisiert werden könnten. In Abwesenheit eindeutiger Vorlagen für Familie C GPCRs wurden Komplexe von Vertretern der Familie A zum Vergleich herangezogen. Diese definierten den Bindetaschenbereich, der auch in Bezug auf die gefundenen Liganden von mGluR5 als hochwahrscheinlich angenommen wurde. Überlappende Bindeposen mit den Liganden des BR und ß$_2$AD sowie der sieben zu dockenden Liganden wurden bestimmt und Seitenketten im engen Kontakt dazu analysiert. Zusätzlich wurde die Mutationssammlung zum Vergleich hinzugezogen, die den Bindebereich eingrenzt und mögliche Interaktionspartner beschreibt. Die Bindungshypothese entsprach einer Überlagerung der gefundenen Moleküle und ihrer möglicher Interaktionspunkte, konnte jedoch ohne weitere strukturelle Information aus Experimenten nicht ausreichend validiert werden.

In dieser Studie wurden mit unterschiedlichen Methodiken der Bio- und Chemieinformatik orientiert an experimentellen Ergebnissen, Fragestellungen bezüglich des Funktionsmechanismus von GPCRs untersucht. In ihrem Verlauf wurden diverse Daten aus experimentellen Befunden zusammengefasst und im strukturellen Zusammenhang korreliert. Die Untersuchung der Sequenzen ermöglichte neue Erkenntnisse bezüglich konservierter Bereiche zweier GPCR Familien. Anhand der Entdeckung neuer Liganden mit unterschiedlichen computergestützten Methoden konnten diese Suchverfahren evaluiert und ihre Schwächen und Stärken identifiziert werden. Exemplarisch am mGluR5 konnte die Eignung einer modellierten GPCR-Struktur für eine Hypothesengenerierung bezüglich Ligandenbindung und struktureller Zusammenhänge untersucht werden.

8. References

Acharya, S., Saad, Y., and Karnik, S. S. (1997). Transducin-alpha C-terminal peptide binding site consists of C-D and E-F loops of rhodopsin. J Biol Chem *272*, 6519-6524.

Alagille, D., Baldwin, R. M., Roth, B. L., Wroblewski, J. T., Grajkowska, E., and Tamagnan, G. D. (2005). Functionalization at position 3 of the phenyl ring of the potent mGluR5 noncompetitive antagonists MPEP. Bioorg Med Chem Lett *15*, 945-949.

Alagille, D., Baldwin, R. M., Roth, B. L., Wroblewski, J. T., Grajkowska, E., and Tamagnan, G. D. (2005). Synthesis and receptor assay of aromatic-ethynyl-aromatic derivatives with potent mGluR5 antagonist activity. Bioorg Med Chem *13*, 197-209.

Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. Med Res Rev *26*, 531-568.

Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P., and Hubbell, W. L. (2008). High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proc Natl Acad Sci U S A *105*, 7439-7444.

Andreeva, A., Howorth, D., Chandonia, J., Brenner, S. E., Hubbard, T. J. P., Chothia, C., and Murzin, A. G. (2007). Data growth and its impact on the SCOP database: new developments. Nucleic Acids Research *36*, D419-425.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. Science *181*, 223-230.

Arora, S. (2005). Optimizing side-chain interactions in protein-ligand interfaces. ProQuest Dissertations & Theses (PQDT) database *MAI 44/02*, 904.

Bairoch, A., Boeckmann, B., Ferro, S., and Gasteiger, E. (2004). Swiss-Prot: juggling between evolution and stability. Brief Bioinform *5*, 39-55.

Balaban, A., (Ed.) (1976). Chemical Applications of Graph Theory. 63.

Ballesteros, J. A., H. Weinstein (1995). Integrated methods for construction three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Meth Neurosci *25*, 366-428.

Bender, A., Mussa, H. Y., and Glen, R. C. (2005). Screening for dihydrofolate reductase inhibitors using MOLPRINT 2D, a fast fragment-based method employing the naive Bayesian classifier: limitations of the descriptor and the importance of balanced chemistry in training and test sets. J Biomol Screen *10*, 658-666.

Bissantz, C., Bernard, P., Hibert, M., and Rognan, D. (2003). Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? Proteins *50*, 5-25.

Bissantz, C., Schalon, C., Guba, W., and Stahl, M. (2005). Focused library design in GPCR projects on the example of 5-HT(2c) agonists: comparison of structure-based virtual screening with ligand-based search methods. Proteins *61*, 938-952.

Bockaert, J., and Pin, J. P. (1999). Molecular tinkering of G protein-coupled receptors: an evolutionary success. Embo J *18*, 1723-1729.

Böhm, H. J. (1992). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des *6*, 61-78.

Böhm, H. J., G. Klebe, H. Kubinyi (2002). Wirkstoffdesign: Der Weg zum Arzneimittel. 600.

Böhm, H. J., G. Schneider (2003). Molecular Recognition in Protein-Ligand Interactions: From Molecular Recognition to Drug Design. *Band 19*, 262.

Böhm, H. J., and Schneider, G. (2000). Virtual Screening for Bioactive Molecules. (Weinheim, Germany: Wiley-VCH).

Bondensgaard, K., Ankersen, M., Thogersen, H., Hansen, B. S., Wulff, B. S., and Bywater, R. P. (2004). Recognition of privileged structures by G-protein coupled receptors. J Med Chem *47*, 888-899.

Bonnefous, C., Vernier, J. M., Hutchinson, J. H., Chung, J., Reyes-Manalo, G., and Kamenecka, T. (2005). Dipyridyl amides: potent metabotropic glutamate subtype 5 (mGlu5) receptor antagonists. Bioorg Med Chem Lett *15*, 1197-1200.

Bostrom, J., Hogner, A., and Schmitt, S. (2006). Do structurally similar ligands bind in a similar fashion? J Med Chem *49*, 6716-6725.

Broto, P., Moreau, G., Vandycke, C. (1984). Molecular structures: perception, autocorrelation descriptor and sar studies. Perception of molecules: topological structure and 3-dimensional structure. European journal of medicinal chemistry *19*, 61-65.

Brown, R., Martin Y. (1997). The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand-Receptor Binding. J Chemical Information and Computer Science *37*, 1-9.

Byvatov, E., and Schneider, G. (2003). Support vector machine applications in bioinformatics. Appl Bioinformatics *2*, 67-77.

Carlson, H. A., and McCammon, J. A. (2000). Accommodating protein flexibility in computational drug design. Mol Pharmacol *57*, 213-218.

Ceccarelli, S. M., Jaeschke, G., Buettelmann, B., Huwyler, J., Kolczewski, S., Peters, J. U., Prinssen, E., Porter, R., Spooren, W., and Vieira, E. (2007). Rational design, synthesis, and structure-activity relationship of benzoxazolones: new potent mglu5 receptor antagonists based on the fenobam structure. Bioorg Med Chem Lett *17*, 1302-1306.

8. References

Chen, Y., Goudet, C., Pin, J. P., and Conn, P. J. (2008). N-{4-Chloro-2-[(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)methyl]phenyl}-2-hy droxybenzamide (CPPHA) acts through a novel site as a positive allosteric modulator of group 1 metabotropic glutamate receptors. Mol Pharmacol *73*, 909-918.

Chen, Y., Nong, Y., Goudet, C., Hemstapat, K., de Paulis, T., Pin, J. P., and Conn, P. J. (2007). Interaction of novel positive allosteric modulators of metabotropic glutamate receptor 5 with the negative allosteric antagonist site is required for potentiation of receptor responses. Mol Pharmacol *71*, 1389-1398.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science *318*, 1258-1265.

Chothia, C., and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. Embo J *5*, 823-826.

Chugunov, A. O., Novoseletsky, V. N., Arseniev, A. S., and Efremov, R. G. (2007). A novel method for packing quality assessment of transmembrane alpha-helical domains in proteins. Biochemistry (Mosc) *72*, 293-300.

Conn, P. J., Christopoulos, A., and Lindsley, C. W. (2009). Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. Nat Rev Drug Discov *8*, 41-54.

Connelly, P. R., Aldape, R. A., Bruzzese, F. J., Chambers, S. P., Fitzgibbon, M. J., Fleming, M. A., Itoh, S., Livingston, D. J., Navia, M. A., Thomson, J. A., and et al. (1994). Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. Proc Natl Acad Sci U S A *91*, 1964-1968.

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. Science *221*, 709-713.

Cosford, N. D., Tehrani, L., Roppe, J., Schweiger, E., Smith, N. D., Anderson, J., Bristow, L., Brodkin, J., Jiang, X., McDonald, I.*, et al.* (2003). 3-[(2-Methyl-1,3-thiazol-4-yl)ethynyl]-pyridine: a potent and highly selective metabotropic glutamate subtype 5 receptor antagonist with anxiolytic activity. J Med Chem *46*, 204-206.

Costanzi, S. (2008). On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. J Med Chem *51*, 2907-2914.

Cozzini, P., Kellogg, G. E., Spyrakis, F., Abraham, D. J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L. A., Morris, G. M.*, et al.* (2008). Target flexibility: an emerging consideration in drug discovery and design. J Med Chem *51*, 6237-6255.

Crocker, E., Eilers, M., Ahuja, S., Hornak, V., Hirshfeld, A., Sheves, M., and Smith, S. O. (2006). Location of Trp265 in metarhodopsin II: implications for the activation mechanism of the visual receptor rhodopsin. J Mol Biol *357*, 163-172.

Cuff, A. L., Sillitoe, I., Lewis, T., Redfern, O. C., Garratt, R., Thornton, J., and Orengo, C. A. (2008). The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Research *37*, D310-D314.

de Paulis, T., Hemstapat, K., Chen, Y., Zhang, Y., Saleh, S., Alagille, D., Baldwin, R. M., Tamagnan, G. D., and Conn, P. J. (2006). Substituent effects of N-(1,3-diphenyl-1H-pyrazol-5-yl)benzamides on positive allosteric modulation of the metabotropic glutamate-5 receptor in rat cortical astrocytes. J Med Chem *49*, 3332-3344.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. on World Wide Web wwwpymolorg.

Domingos, P., M. Pazzani (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning *29*, 103-130.

Duda, R. O., P.E. Hart, D.G. Stork (2001). Pattern Classification.

Durbin, R., S. R. Eddy, A. Krogh, G. Mitchison (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids: Cambridge University Press).

Ehrlich, P. (1904). Gesammelte Arbeiten zur Immunitäts Forschung. 1-776.

Elofsson, A. (2002). A study on protein sequence alignment quality. Proteins *46*, 330-339.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci *Chapter 2*, Unit 2 9.

Evers, A., and Klabunde, T. (2005). Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. J Med Chem *48*, 1088-1097.

Evers, A., and Klebe, G. (2004). Ligand-supported homology modeling of g-protein-coupled receptor sites: models sufficient for successful virtual screening. Angew Chem Int Ed Engl *43*, 248-251.

Evers, A., and Klebe, G. (2004). Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. J Med Chem *47*, 5381-5392.

Fanelli, F., and De Benedetti, P. G. (2005). Computational modeling approaches to structure-function analysis of G protein-coupled receptors. Chem Rev *105*, 3297-3351.

Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996). Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science *274*, 768-770.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters *27*, 861-874.

8. References

Fechner, U., Franke, L., Renner, S., Schneider, P., and Schneider, G. (2003). Comparison of correlation vector methods for ligand-based similarity searching. J Comput Aided Mol Des *17*, 687-698.

Filipek, S. (2005). Organization of rhodopsin molecules in native membranes of rod cells--an old theoretical model compared to new experimental data. J Mol Model *11*, 385-391.

Filipek, S., Krzysko, K. A., Fotiadis, D., Liang, Y., Saperstein, D. A., Engel, A., and Palczewski, K. (2004). A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface. Photochem Photobiol Sci *3*, 628-638.

Filizola, M., and Weinstein, H. (2005). The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. Febs J *272*, 2926-2938.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges *27*, 2984–2993.

Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., and Schioth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol *63*, 1256-1272.

Fukami-Kobayashi, K., and Saito, N. (2002). [How to make good use of CLUSTALW]. Tanpakushitsu Kakusan Koso *47*, 1237-1239.

Gama, L., Wilt, S. G., and Breitwieser, G. E. (2001). Heterodimerization of calcium sensing receptors with metabotropic glutamate receptors in neurons. J Biol Chem *276*, 39053-39059.

Gasparini, F., Lingenhohl, K., Stoehr, N., Flor, P. J., Heinrich, M., Vranesic, I., Biollaz, M., Allgeier, H., Heckendorn, R., Urwyler, S.*, et al.* (1999). 2-Methyl-6-(phenylethynyl)-pyridine (MPEP), a potent, selective and systemically active mGlu5 receptor antagonist. Neuropharmacology *38*, 1493-1503.

Gether, U., and Kobilka, B. K. (1998). G protein-coupled receptors. II. Mechanism of agonist activation. J Biol Chem *273*, 17979-17982.

Goudet, C., Gaven, F., Kniazeff, J., Vol, C., Liu, J., Cohen-Gonsaud, M., Acher, F., Prezeau, L., and Pin, J. P. (2004). Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. Proc Natl Acad Sci U S A *101*, 378-383.

Gouldson, P. R., Kidley, N. J., Bywater, R. P., Psaroudakis, G., Brooks, H. D., Diaz, C., Shire, D., and Reynolds, C. A. (2004). Toward the active conformations of rhodopsin and the beta2-adrenergic receptor. Proteins *56*, 67-84.

GPCRDB (v.10.0 June 2006). http://www.gpcr.org/7tm/.

Grant, J. A., B. T. Pickup, A. Nicholls (2001). A smooth permittivity function for Poisson-Boltzmann solvation methods. Journal of Computational Chemistry *22*, 608 - 640.

Grishin, N. V. (2001). Fold change in evolution of protein structures. J Struct Biol *134*, 167-185.

Guo, W., Shi, L., and Javitch, J. A. (2003). The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. J Biol Chem *278*, 4385-4388.

Hann, M., Hudson, B., Lewell, X., Lifely, R., Miller, L., and Ramsden, N. (1999). Strategic pooling of compounds for high-throughput screening. J Chem Inf Comput Sci *39*, 897-902.

Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y., Velasquez, J., Kuhn, P., and Stevens, R. C. (2008). A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. Structure *16*, 897-905.

He, X., Chow, D., Martick, M. M., and Garcia, K. C. (2001). Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. Science *293*, 1657-1662.

Hebert, T. E., Moffett, S., Morello, J. P., Loisel, T. P., Bichet, D. G., Barret, C., and Bouvier, M. (1996). A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. J Biol Chem *271*, 16384-16392.

Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A *89*, 10915-10919.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595-603.

Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E., and Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res *31*, 294-297.

Hu, E., Chua, P. C., Tehrani, L., Nagasawa, J. Y., Pinkerton, A. B., Rowe, B. A., Vernier, J. M., Hutchinson, J. H., and Cosford, N. D. (2004). Pyrimidine methyl anilines: selective potentiators for the metabotropic glutamate 2 receptor. Bioorg Med Chem Lett *14*, 5071-5074.

Hu, J., McLarnon, S. J., Mora, S., Jiang, J., Thomas, C., Jacobson, K. A., and Spiegel, A. M. (2005). A region in the seven-transmembrane domain of the human Ca2+ receptor critical for response to Ca2+. J Biol Chem *280*, 5113-5120.

Hu, J., Reyes-Cruz, G., Chen, W., Jacobson, K. A., and Spiegel, A. M. (2002). Identification of acidic residues in the extracellular loops of the seven-transmembrane domain of the human Ca2+ receptor critical for response to Ca2+ and a positive allosteric modulator. J Biol Chem *277*, 46622-46631.

8. References

Hubbard, T. J., Murzin, A. G., Brenner, S. E., and Chothia, C. (1997). SCOP: a structural classification of proteins database. Nucleic Acids Res *25*, 236-239.

Iso, Y., Grajkowska, E., Wroblewski, J. T., Davis, J., Goeders, N. E., Johnson, K. M., Sanker, S., Roth, B. L., Tueckmantel, W., and Kozikowski, A. P. (2006). Synthesis and structure-activity relationships of 3-[(2-methyl-1,3-thiazol-4-yl)ethynyl]pyridine analogues as potent, noncompetitive metabotropic glutamate receptor subtype 5 antagonists; search for cocaine medications. J Med Chem *49*, 1080-1100.

Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P., and Stevens, R. C. (2008). The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science *322*, 1211-1217.

Jacoby, E., Bouhelal, R., Gerspacher, M., and Seuwen, K. (2006). The 7 TM G-protein-coupled receptor target family. ChemMedChem *1*, 761-782.

Jaroszewski, L., Slabinski, L., Wooley, J., Deacon, A. M., Lesley, S. A., Wilson, I. A., and Godzik, A. (2008). Genome pool strategy for structural coverage of protein families. Structure *16*, 1659-1667.

Java, Sun Microsystems, Inc. 4150 Network CircleSanta Clara, CA 95054 USA.

Jiang, P., Cui, M., Zhao, B., Liu, Z., Snyder, L. A., Benard, L. M., Osman, R., Margolskee, R. F., and Max, M. (2005). Lactisole interacts with the transmembrane domains of human T1R3 to inhibit sweet taste. J Biol Chem *280*, 15238-15246.

Jiang, P., Cui, M., Zhao, B., Snyder, L. A., Benard, L. M., Osman, R., Max, M., and Margolskee, R. F. (2005). Identification of the cyclamate interaction site within the transmembrane domain of the human sweet taste receptor subunit T1R3. J Biol Chem *280*, 34296-34305.

Jiang, P., Ji, Q., Liu, Z., Snyder, L. A., Benard, L. M., Margolskee, R. F., and Max, M. (2004). The cysteine-rich region of T1R3 determines responses to intensely sweet proteins. J Biol Chem *279*, 45068-45075.

Johnson, M., Maggiora, GM (1990). Concepts and applications of molecular similarity.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994). A mutation data matrix for transmembrane proteins. FEBS Lett *339*, 269-275.

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic

Liquids. J Am Chem Soc *118*, 11225-11236.

Kelly, K. (1996). Multiple sequence and structural alignment in MOE. Chemical Computing Group Inc. http://wwwchemcompcom/journal/alignhtm.

Kew, J. N. (2004). Positive and negative allosteric modulation of metabotropic glutamate receptors: emerging therapeutic potential. Pharmacol Ther *104*, 233-244.

144

Kew, J. N., and Kemp, J. A. (2005). Ionotropic and metabotropic glutamate receptor structure and pharmacology. Psychopharmacology (Berl) *179*, 4-29.

Klabunde, T., and Hessler, G. (2002). Drug design strategies for targeting G-protein-coupled receptors. Chembiochem *3*, 928-944.

Klebe, G. (2006). Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today *11*, 580-594.

Knegtel, R. M., Kuntz, I. D., and Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures. J Mol Biol *266*, 424-440.

Knoflach, F., Woltering, T., Adam, G., Mutel, V., and Kemp, J. A. (2001). Pharmacological properties of native metabotropic glutamate receptors in freshly dissociated Golgi cells of the rat cerebellum. Neuropharmacology *40*, 163-169.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics *43*, 59-69.

Kollmann, P. A. (1993). Free Energy Calculations-Applications To Chemical And Biochemical Phenomena. Chem Rev *93*, 2395-2417.

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci U S A *44*, 98-104.

Kratochwil, N. A., Malherbe, P., Lindemann, L., Ebeling, M., Hoener, M. C., Muhlemann, A., Porter, R. H., Stahl, M., and Gerber, P. R. (2005). An automated system for the analysis of G protein-coupled receptor transmembrane binding pockets: alignment, receptor-based pharmacophores, and their application. J Chem Inf Model *45*, 1324-1336.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., and Morikawa, K. (2000). Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature *407*, 971-977.

Leach, A. R. (1994). Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol *235*, 345-356.

Leach, A. R., Shoichet, B. K., and Peishoff, C. E. (2006). Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem *49*, 5851-5855.

Ligprep (v.2.0). Distributed by Schrödinger, LLC, New York http://www.schrodinger.com.

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Rev *23*, 3-25.

8. References

Litschig, S., Gasparini, F., Rueegg, D., Stoehr, N., Flor, P. J., Vranesic, I., Prezeau, L., Pin, J. P., Thomsen, C., and Kuhn, R. (1999). CPCCOEt, a noncompetitive metabotropic glutamate receptor 1 antagonist, inhibits receptor signaling without affecting glutamate binding. Mol Pharmacol *55*, 453-461.

Livnah, O., Stura, E. A., Middleton, S. A., Johnson, D. L., Jolliffe, L. K., and Wilson, I. A. (1999). Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. Science *283*, 987-990.

Lu, Z. L., Coetsee, M., White, C. D., and Millar, R. P. (2007). Structural determinants for ligand-receptor conformational selection in a peptide G protein-coupled receptor. J Biol Chem *282*, 17921-17929.

MacKerell, J. A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha*, et al.* (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B *102*, 3586-3616.

Madabushi, S., Gross, A. K., Philippi, A., Meng, E. C., Wensel, T. G., and Lichtarge, O. (2004). Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. J Biol Chem *279*, 8126-8132.

Malherbe, P., Kratochwil, N., Knoflach, F., Zenner, M. T., Kew, J. N., Kratzeisen, C., Maerki, H. P., Adam, G., and Mutel, V. (2003). Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. J Biol Chem *278*, 8340-8347.

Malherbe, P., Kratochwil, N., Muhlemann, A., Zenner, M. T., Fischer, C., Stahl, M., Gerber, P. R., Jaeschke, G., and Porter, R. H. (2006). Comparison of the binding pockets of two chemically unrelated allosteric antagonists of the mGlu5 receptor and identification of crucial residues involved in the inverse agonism of MPEP. J Neurochem *98*, 601-615.

Malherbe, P., Kratochwil, N., Zenner, M. T., Piussi, J., Diener, C., Kratzeisen, C., Fischer, C., and Porter, R. H. (2003). Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. Mol Pharmacol *64*, 823-832.

Marti-Renom, M. A., Madhusudhan, M. S., Fiser, A., Rost, B., and Sali, A. (2002). Reliability of assessment of protein structure prediction methods. Structure *10*, 435-440.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct *29*, 291-325.

Martin, Y. C., Kofron, J. L., and Traphagen, L. M. (2002). Do structurally similar molecules have similar biological activity? J Med Chem *45*, 4350-4358.

Mason, J. S., Good, A. C., and Martin, E. J. (2001). 3-D pharmacophores in drug discovery. Curr Pharm Des *7*, 567-597.

MATLAB The Math-Works, v2006b, http://www.mathworks.com.

McCammon, J. A. (2005). Target flexibility in molecular recognition. Biochim Biophys Acta *1754*, 221-224.

Merz GmbH & Co. KGaA, Eckenheimer Landstraße 100, D-60318 Frankfurt am Main, Germany.

Micheli, F., Bertani, B., Bozzoli, A., Crippa, L., Cavanni, P., Di Fabio, R., Donati, D., Marzorati, P., Merlo, G., Paio, A.*, et al.* (2008). Phenylethynyl-pyrrolo[1,2-a]pyrazine: a new potent and selective tool in the mGluR5 antagonists arena. Bioorg Med Chem Lett *18*, 1804-1809.

Miedlich, S. U., Gama, L., Seuwen, K., Wolf, R. M., and Breitwieser, G. E. (2004). Homology modeling of the transmembrane domain of the human calcium sensing receptor and localization of an allosteric binding site. J Biol Chem *279*, 7254-7263.

Mirzadegan, T., Benko, G., Filipek, S., and Palczewski, K. (2003). Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry *42*, 2759-2767.

MOE (release 2006.08). Molecular Operating Environment. Chemical Computing Group, Montreal, Quebec, Canada *www.chemcomp.com*.

Morgan, H. L. (1965). The generation of unique machine description for chemical structures - a technique developped at chemical abstracts service. Journal of Chemical Documentation, 107-113.

Muehlemann, A., Ward, N. A., Kratochwil, N., Diener, C., Fischer, C., Stucki, A., Jaeschke, G., Malherbe, P., and Porter, R. H. (2006). Determination of key amino acids implicated in the actions of allosteric modulation by 3,3'-difluorobenzaldazine on rat mGlu5 receptors. Eur J Pharmacol *529*, 95-104.

Murakami, M., and Kouyama, T. (2008). Crystal structure of squid rhodopsin. Nature *453*, 363-367.

Murzin, A. G. (1998). How far divergent evolution goes in proteins. Curr Opin Struct Biol *8*, 380-387.

Nicoletti, F., Wroblewski, J. T., Novelli, A., Alho, H., Guidotti, A., and Costa, E. (1986). The activation of inositol phospholipid metabolism as a signal-transducing system for excitatory amino acids in primary cultures of cerebellar granule cells. J Neurosci *6*, 1905-1911.

Noeske, T., Gutcaits A., Parsons, Christopher G., Weil, T. (2005). Allosteric Modulation of Family 3 GPCRs. QSAR & Combinatorial Science *25*, 134-146.

Noeske, T., Sasse, B. C., Stark, H., Parsons, C. G., Weil, T., and Schneider, G. (2006). Predicting compound selectivity by self-organizing maps: cross-activities of metabotropic glutamate receptor antagonists. ChemMedChem *1*, 1066-1068.

8. References

Okada, T., Le Trong, I., Fox, B. A., Behnke, C. A., Stenkamp, R. E., and Palczewski, K. (2000). X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. J Struct Biol *130*, 73-80.

Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P., and Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. J Mol Biol *342*, 571-583.

Olah, M. M., Bologa, C. G., and Oprea, T. I. (2004). Strategies for compound selection. Curr Drug Discov Technol *1*, 211-220.

Oldham, W. M., Van Eps, N., Preininger, A. M., Hubbell, W. L., and Hamm, H. E. (2006). Mechanism of the receptor-catalyzed activation of heterotrimeric G proteins. Nat Struct Mol Biol *13*, 772-777.

Oliveira, L., A.C.M. Paiva, G. Vriend (1993). A common motif in G-protein-coupled seven transmembrane helix receptors. Journal of Computer-Aided Molecular Design *7*, 649-658.

Oliveira, L., Paiva, A. C., and Vriend, G. (2002). Correlated mutation analyses on very large sequence families. Chembiochem *3*, 1010-1017.

Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? Nat Rev Drug Discov *5*, 993-996.

Pagano, A., Ruegg, D., Litschig, S., Stoehr, N., Stierlin, C., Heinrich, M., Floersheim, P., Prezeau, L., Carroll, F., Pin, J. P.*, et al.* (2000). The non-competitive antagonists 2-methyl-6-(phenylethynyl)pyridine and 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester interact with overlapping binding pockets in the transmembrane region of group I metabotropic glutamate receptors. J Biol Chem *275*, 33750-33758.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E.*, et al.* (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. Science *289*, 739-745.

Pareto, V. (1896). Cours d'Economie Politique. Rouge, Lausanne, France.

Pauling, L. (1946). Molecular Architecture And Biological Reactions. Chemical & Engineering News *24*, 1375-1377.

Pei, J., and Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics *17*, 700-712.

Petrel, C., Kessler, A., Dauban, P., Dodd, R. H., Rognan, D., and Ruat, M. (2004). Positive and negative allosteric modulators of the Ca2+-sensing receptor interact within overlapping but not identical binding sites in the transmembrane domain. J Biol Chem *279*, 18990-18997.

Petrel, C., Kessler, A., Maslah, F., Dauban, P., Dodd, R. H., Rognan, D., and Ruat, M. (2003). Modeling and mutagenesis of the binding site of Calhex 231, a novel negative allosteric modulator of the extracellular Ca(2+)-sensing receptor. J Biol Chem *278*, 49487-49494.

Phase (v.2.5). Schrödinger, LLC, New York, 2008.

Pin, J. P., and Acher, F. (2002). The metabotropic glutamate receptors: structure, activation mechanism and pharmacology. Curr Drug Targets CNS Neurol Disord *1*, 297-317.

Pin, J. P., Galvez, T., and Prezeau, L. (2003). Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacol Ther *98*, 325-354.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J Mol Biol *7*, 95-99.

Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F.*, et al.* (2007). Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature *450*, 383-387.

Ray, K., and Hauschild, B. C. (2000). Cys-140 is critical for metabotropic glutamate receptor-1 dimerization. J Biol Chem *275*, 34245-34251.

Ray, K., Hauschild, B. C., Steinbach, P. J., Goldsmith, P. K., Hauache, O., and Spiegel, A. M. (1999). Identification of the cysteine residues in the amino-terminal extracellular domain of the human Ca(2+) receptor critical for dimerization. Implications for function of monomeric Ca(2+) receptor. J Biol Chem *274*, 27642-27650.

Rechenberg, I. (1973). Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen

Evolution. (Stuttgart: Frommann-Holzboog).

Reggio, P. H. (2006). Computational methods in drug design: modeling G protein-coupled receptor monomers, dimers, and oligomers. Aaps J *8*, E322-336.

Remelli, R., Robbins, M. J., and McIlhinney, R. A. (2008). The C-terminus of the metabotropic glutamate receptor 1b regulates dimerization of the receptor. J Neurochem *104*, 1020-1031.

Renner, S., Noeske, T., Parsons, C. G., Schneider, P., Weil, T., and Schneider, G. (2005). New allosteric modulators of metabotropic glutamate receptor 5 (mGluR5) found by ligand-based virtual screening. Chembiochem *6*, 620-625.

Renner, S., Hechenberger, M., Noeske, T., Böcker, A., Jatzke, C., Schmuker, M., Parsons, C.G., Weil, T. and Schneider, G. (2007). Searching for Drug Scaffolds with 3D Pharmacophores and Neural Network Ensembles. Angew. Chem. Int. Ed. 2007, *46*, 5336 –5339.

8. References

Robbins, M. J., Ciruela, F., Rhodes, A., and McIlhinney, R. A. (1999). Characterization of the dimerization of metabotropic glutamate receptors using an N-terminal truncation of mGluR1alpha. J Neurochem *72*, 2539-2547.

Rogers, D., Brown, R. D., and Hahn, M. (2005). Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. J Biomol Screen *10*, 682-686.

Romano, C., Yang, W. L., and O'Malley, K. L. (1996). Metabotropic glutamate receptor 5 is a disulfide-linked dimer. J Biol Chem *271*, 28612-28616.

Roppe, J., Smith, N. D., Huang, D., Tehrani, L., Wang, B., Anderson, J., Brodkin, J., Chung, J., Jiang, X., King, C.*, et al.* (2004). Discovery of novel heteroarylazoles that are metabotropic glutamate subtype 5 receptor antagonists with anxiolytic activity. J Med Chem *47*, 4645-4648.

Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Eng *12*, 85-94.

Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C., and Schertler, G. F. (2004). Electron crystallography reveals the structure of metarhodopsin I. Embo J *23*, 3609-3620.

Russ, W. P., and Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. J Mol Biol *296*, 911-919.

Sadowski, J., and Kubinyi, H. (1998). A scoring scheme for discriminating between drugs and nondrugs. J Med Chem *41*, 3325-3329.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol *4*, 406-425.

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol *234*, 779-815.

Salo, O. M., Lahtela-Kakkonen, M., Gynther, J., Jarvinen, T., and Poso, A. (2004). Development of a 3D model for the human cannabinoid CB1 receptor. J Med Chem *47*, 3048-3057.

Salom, D., Lodowski, D. T., Stenkamp, R. E., Le Trong, I., Golczak, M., Jastrzebska, B., Harris, T., Ballesteros, J. A., and Palczewski, K. (2006). Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci U S A *103*, 16123-16128.

Sanchez, R., and Sali, A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. Proteins *Suppl 1*, 50-58.

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56-68.

Schaffhauser, H., Rowe, B. A., Morales, S., Chavez-Noriega, L. E., Yin, R., Jachec, C., Rao, S. P., Bain, G., Pinkerton, A. B., Vernier, J. M.*, et al.* (2003). Pharmacological characterization and identification of amino acids involved in the positive modulation of metabotropic glutamate receptor subtype 2. Mol Pharmacol *64*, 798-810.

Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H. W., Hofmann, K. P., and Ernst, O. P. (2008). Crystal structure of opsin in its G-protein-interacting conformation. Nature *455*, 497-502.

Schneider, G., Baringhaus, KH (2008). Molecular Design: Concepts and Applications. 277.

Schneider, G., and Bohm, H. J. (2002). Virtual screening and fast automated docking methods. Drug Discov Today *7*, 64-70.

Schneider, G., and Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov *4*, 649-663.

Schneider, G., Neidhart, W., Giller, T., and Schmid, G. (1999). "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. Angew Chem Int Ed Engl *38*, 2894-2896.

Schneider, G., and Wrede, P. (1998). Artificial neural networks for computer-based molecular design. Prog Biophys Mol Biol *70*, 175-222.

Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res *18*, 6097-6100.

Schoepp, D. D., Jane, D. E., and Monn, J. A. (1999). Pharmacological agents acting at subtypes of metabotropic glutamate receptors. Neuropharmacology *38*, 1431-1476.

Schrödinger LLC, New York, 2008.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

SciTegic Inc. 10188 Telesis Court, Suite 100, San Diego, Ca 92121, USA.

Selkirk, J. V., Challiss, R. A., Rhodes, A., and McIlhinney, R. A. (2002). Characterization of an N-terminal secreted domain of the type-1 human metabotropic glutamate receptor produced by a mammalian cell line. J Neurochem *80*, 346-353.

Shacham, S., Marantz, Y., Bar-Haim, S., Kalid, O., Warshaviak, D., Avisar, N., Inbal, B., Heifetz, A., Fichman, M., Topf, M.*, et al.* (2004). PREDICT modeling and in-silico screening for G-protein coupled receptors. Proteins *57*, 51-86.

Shanno, D. F. a. P., K. H (1980). Minimization of unrestrained multivariate functions. ACM Trans Math Soft *6*, 618-622.

Shanno, D. F. a. P., K. H (1982). Remark on algorithm 500. Collected algorithms from ACM Trans Math Software *2*.

8. References

Shannon, C. (1948). A mathematical theory of communication. Bell System Technical Journal *27*, 623-656.

Sharma, S., Rodriguez, A. L., Conn, P. J., and Lindsley, C. W. (2008). Synthesis and SAR of a mGluR5 allosteric partial antagonist lead: unexpected modulation of pharmacology with slight structural modifications to a 5-(phenylethynyl)pyrimidine scaffold. Bioorg Med Chem Lett *18*, 4098-4101.

Sheridan, R. P., and Kearsley, S. K. (2002). Why do we need so many chemical similarity search methods? Drug Discov Today *7*, 903-911.

Sheridan, R. P., Rusinko, A., 3rd, Nilakantan, R., and Venkataraghavan, R. (1989). Searching for pharmacophores in large coordinate data bases and its use in drug design. Proc Natl Acad Sci U S A *86*, 8165-8169.

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem *49*, 534-553.

Sladeczek, F., Pin, J. P., Recasens, M., Bockaert, J., and Weiss, S. (1985). Glutamate stimulates inositol phosphate formation in striatal neurones. Nature *317*, 717-719.

Snarey, M., Terrett, N. K., Willett, P., and Wilton, D. J. (1997). Comparison of algorithms for dissimilarity-based compound selection. J Mol Graph Model *15*, 372-385.

Spijker, P., Vaidehi, N., Freddolino, P. L., Hilbers, P. A., and Goddard, W. A., 3rd (2006). Dynamic behavior of fully solvated beta2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. Proc Natl Acad Sci U S A *103*, 4882-4887.

Sugiyama, H., Ito, I., and Hirono, C. (1987). A new type of glutamate receptor linked to inositol phospholipid metabolism. Nature *325*, 531-533.

Surgand, J. S., Rodrigo, J., Kellenberger, E., and Rognan, D. (2006). A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. Proteins *62*, 509-538.

Tanimoto, T. T. (1957). IBM Internal Report.

Tanrikulu, Y., and Schneider, G. (2008). Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. Nat Rev Drug Discov *7*, 667-677.

Taylor, R. D., Jewsbury, P. J., and Essex, J. W. (2002). A review of protein-small molecule docking methods. J Comput Aided Mol Des *16*, 151-166.

Teller, D. C., Stenkamp, R. E., and Palczewski, K. (2003). Evolutionary analysis of rhodopsin and cone pigments: connecting the three-dimensional structure with spectral tuning and signal transfer. FEBS Lett *555*, 151-159.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Thornton, J. M., Orengo, C. A., Todd, A. E., and Pearl, F. M. (1999). Protein folds, functions and evolution. J Mol Biol *293*, 333-342.

Todeshini, R., and Consonni, V. (2000). Methods and Principles in Medicinal Chemistry, Vol.11, In Handbook of Molecular Descriptors (Weinheim: VCH).

Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H., and Morikawa, K. (2002). Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+. Proc Natl Acad Sci U S A *99*, 2660-2665.

van de Waterbeemd, H., and Gifford, E. (2003). ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov *2*, 192-204.

Van Drie, J. H. (2007). Computer-aided drug design: the next 20 years. J Comput Aided Mol Des *21*, 591-601.

Vanejevs, M., Jatzke, C., Renner, S., Muller, S., Hechenberger, M., Bauer, T., Klochkova, A., Pyatkin, I., Kazyulkin, D., Aksenova, E.*, et al.* (2008). Positive and negative modulation of group I metabotropic glutamate receptors. J Med Chem *51*, 634-647.

Vapnik, V. (1998). Statistical Learning Theory (New York: Wiley).

Varney, M. A., Cosford, N. D., Jachec, C., Rao, S. P., Sacaan, A., Lin, F. F., Bleicher, L., Santori, E. M., Flor, P. J., Allgeier, H.*, et al.* (1999). SIB-1757 and SIB-1893: selective, noncompetitive antagonists of metabotropic glutamate receptor type 5. J Pharmacol Exp Ther *290*, 170-181.

Walters, P. W., Stahl, M., and Murcko, M. A. (1998). Virtual screening - an overview. Drug Discov Today *3*, 160-178.

Wang, J.-Q., and Brownell, A.-L. (2007). Development of Metabotropic Glutamate Receptor Ligands for Neuroimaging. Current Medical Imaging Reviews *3*, 186-205.

Wang, J., Brownell, AL (2007). Development of Metabotropic Glutamate Receptor Ligands for Neuroimaging. Current Medical Imaging Reviews *3*, 186-205.

Wang, Q. (2008) Comparison of Spatial Distributions for Virtual Screening using Support Vector Machines, Johann Wolfgang Goethe-University, Frankfurt am Main, Germany.

Wang, R., Lu, Y., and Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem *46*, 2287-2303.

8. References

Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G., and Schertler, G. F. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. Nature *454*, 486-491.

Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S.*, et al.* (2006). A critical assessment of docking programs and scoring functions. J Med Chem *49*, 5912-5931.

Wermuth, C. G. (2008). The Practice of Medicinal Chemistry, Vol Third edition (London, UK: Elsevier Ltd.).

Westbrook, J., Feng, Z., Jain, S., Bhat, T. N., Thanki, N., Ravichandran, V., Gilliland, G. L., Bluhm, W., Weissig, H., Greer, D. S.*, et al.* (2002). The Protein Data Bank: unifying the archive. Nucleic Acids Res *30*, 245-248.

Willett, P. (1998). Chemical Similarity Searching. J Chem Inf Comput Sci *38*, 983-996.

Williams, D. L., Jr., and Lindsley, C. W. (2005). Discovery of positive allosteric modulators of metabotropic glutamate receptor subtype 5 (mGluR5). Curr Top Med Chem *5*, 825-846.

Winnig, M., Bufe, B., and Meyerhof, W. (2005). Valine 738 and lysine 735 in the fifth transmembrane domain of rTas1r3 mediate insensitivity towards lactisole of the rat sweet taste receptor. BMC Neurosci *6*, 22.

WOMBAT Sunset Molecular Discovery LLC, 1704 B Llano St., Suite 140, Santa Fe, NM 87505, USA.

Wrabl, J. O., and Grishin, N. V. (2005). Grouping of amino acid types and extraction of amino acid properties from multiple sequence alignments using variance maximization. Proteins *61*, 523-534.

Xia, X., Maliski, E. G., Gallant, P., and Rogers, D. (2004). Classification of kinase inhibitors using a Bayesian model. J Med Chem *47*, 4463-4470.

Xu, H., Staszewski, L., Tang, H., Adler, E., Zoller, M., and Li, X. (2004). Different functional roles of T1R subunits in the heteromeric taste receptors. Proc Natl Acad Sci U S A *101*, 14258-14263.

Yanamala, N., Tirupula, K. C., and Klein-Seetharaman, J. (2008). Preferential binding of allosteric modulators to active and inactive conformational states of metabotropic glutamate receptors. BMC Bioinformatics *9 Suppl 1*, S16.

Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P., and Bowie, J. U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci U S A *101*, 959-963.

Yohannan, S., Yang, D., Faham, S., Boulting, G., Whitelegge, J., and Bowie, J. U. (2004). Proline substitutions are not easily accommodated in a membrane protein. J Mol Biol *341*, 1-6.

Zhan, W., Stein, E. A., and Yang, Y. (2006). A rotation-invariant spherical harmonic decomposition method for mapping intravoxel multiple fiber structures. Neuroimage *29*, 1212-1223.

Zou, K. H., O'Malley, A. J., and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation *115*, 654-657.

9. Abbreviations

2D and 3D	Two dimensional and three dimensional
5HT2c	5-hydroxytryptamine (serotonin) receptor 2C
Å	Ångström
Ala, A	Alanine
AP	Addex Pharmaceuticals
Arg, R	Arginine
ASN	Asinex serial number
Asn, N	Asparagine
Asp, D	Aspartic acid
AUC	Area under curve
AZ	Astra Zeneca
$\text{ß}_2\text{AD}$	Beta adrenergic receptor type two
Blosum45	BLOcks SUbstitution Matrix
BOSS	Bride of Sevenless Proteins
BR	Bovine rhodopsin
BW	Generic numbering according to Ballesteros and Weinstein (Ballesteros and Weinstein 1985)
Ca^{2+}	Calcium
Calhex231	4-Chloro-N-[2-(1-naphthalen-1-yl-ethylamino)-cyclohexyl]-benzamide
CaSR	Calcium sensing receptor
CATS	Chemically Advanced Template Search
CHARMM	Chemistry at HARvard Molecular Mechanics, a force-field and macromolecular simulations program
ClustalW	General purpose multiple sequence alignment program for DNA or protein sequences
CNS	Central nervous system
CPCCOEt	2-Dihydro-1H-7-oxacyclopropa[b]naphthalene-7a-carboxylic Acid Ethyl Ester
CRD	Cystein-rich domain
Cys, C	Cystein
D_3	Dopamine receptor 3
DEER	double electron-electron resonance
DFB	N,N'-Bis-(3-fluoro-benzylidene)-hydrazine
DMSO	Dimethylsulfoxid
DRC	Dose response curve
EC	Extracellular
EC_{50}	Effective Concentration, a doses which leads to 50% of measured effect
EM-TBPC	(1-Ethyl-2-methyl-6-oxo-4-(1,2,4,5-tetrahydro-benzo[d]azepin-3-yl)-1,6-dihydro-pyrimidine-5-carbonitrile

EPR	electron paramagnetic resonance
HV	Entropy value
FCFP	Functional class extended-connectivity fingerprint
FLIPR	Fluorescent Imaging Plate Reader
FN	False negatives
FP	False positives
GABA-B, GABR	Gamma-aminobutyric acid type B receptor
GDP	guanosine diphosphate
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
GPCR	G-protein coupled receptor
GPCR5	G-protein coupled receptor family C group 5
GPCR6	G-protein coupled receptor family C group 6
GPCRDB	GPCR data base, information system for G protein-coupled receptors
G-Protein	Guanine nucleotide-binding proteins
H	Helix
HD	Heptahelical domain
hCaSR	Human extracellular calcium-sensing receptor precursor (CaSR)
His, H	Histidine
hT1R3	Human taste receptor type 1 member 3 precursor (Sweet taste receptor T1R3)
IC	Intracellular loop
IC_{50}	Inhibitory concentration, a measure of the effectiveness of a compound in inhibiting biological or biochemical function
ID	Identification
IFD	Induced fit docking
Ile, I	Isoleucine
IP	Inositol phosphate
IUPAC	International Union of Pure and Applied Chemistry
KA	Konserviertheitsanalyse
K_i	The binding affinity of the ligand
Leu, L	Leucine
Ligprep	Ligand preparation module Schrödinger, LLC, New York, 2008
LISS	Ligand-induced selective signaling
LPV	ligand binding pocket vector
LY 487379	N-(4-(2-methoxyphenoxy)phenyl)-N-(2,2,2-trifluoroethylsulfonyl)-pyrid-3-yl-methylamine
Lys, K	Lysine

9. Abbreviations

M	Molecule
M_1	Muscarinic receptor 1
MATLAB	"Matrix laboratory", a numerical computing environment and programming language
max	Maximum
MCPG	(S)-α-methyl-4-carboxyphenylglycine
Met, M	Methionine
mGluR or mGR	Metabotropic glutamate receptor
mGluR1, mGluR5	Metabotropic glutamate receptor subtype one, five
min	Minimum
M-MPEP	[3H]2-methyl-6-(3-methoxyphenyl) ethynyl pyridine
MOE	Molecular Operating Environment
MPEP	2-Methyl-6-(phenylethynyl)-pyridine
MSA	Multiple sequence alignment
MSF	Multiple sequence files
MTEP	3-(2-Methyl-thiazol-4-ylethynyl)-pyridine
NAM	Negative allosteric modulator
nM	nanomol
NMR	nuclear magnetic resonance
NPS R-568	[3- (2-Chloro-phenyl)-propyl]-[1-(3-methoxy-phenyl)-ethyl]-amine
NPS2143	2-Chloro-6-[3-(1,1-dimethyl-2-naphthalen-2-yl-ethylamino)-2-hydroxy-propoxy]-benzonitrile
NA	Nucleic Acids
PAM	Positive allosteric modulator
PDB	Protein data bank
pdf	Probability density function
pH	"Power of hydrogen", a measure of the acidity or basicity of a solution
PHRFP	Chemically advanced Template search (CATS) pharmacophore fingerprints
PML	PyMOL script
Pro, P	Proline
QSAR	Quantitative structure activity relationship
RMSD	Root mean square deviation
Ro 67-7476	2-(4-Fluoro-phenyl)-1-(toluene-4-sulfonyl)-pyrrolidine
ROC	Receiver Operating Characteristic
SDF	Structure-data file
SDSL	site-directed spin labelling
H	Shannon entropy
Ser, S	Serine
HV	Shannon entropy value
SHD	Spherical harmonics descriptors

SMILES	Simplified Molecular Input Line Entry Specification
SOM	Self-organizing map
SP	Standard precision
SwissProt	Biological database of protein sequences
Thr, T	Threonine
TM	Transmembrane
TMHMM	Transmembrane hidden Markov models
TN	True negatives
TP	True positives
TR	Taste receptor
Trp, W	Tryptophane
Tyr, Y	Tyrosine
Val,V	Vvaline
V_{1A}	Vasopressin receptor
VR	Vertebrate rhodopsin
VTFM	Variable target function method
ZM241385	4-{2-[(7-amino-2-furan-2-yl[1,2,4]triazolo[1,5-a][1,3,5]triazin-5-yl)amino]ethyl}phenol
µM	Micromol

10. Appendix

10.1 Schroedinger IFD Protocol in .inp File Format

mGluR5

```
# Global Variables
#  These variables affect the entire job, and must all appear
#  before the first STAGE declaration.  Multiple INPUT_FILE
#  entries are supported, as are files containing multiple
#  receptor structures.
#
#  If beginning with an existing Pose Viewer file, simply specify
#  it as the INPUT_FILE (making sure the name ends in "_pv.mae")
#  and ensure that the first GLIDE_DOCKING stage is commented out.
#  The ligand used in producing the Pose Viewer file must also be
#  provided to the second GLIDE_DOCKING stage, using the LIGAND_FILE
#  keyword.
INPUT_FILE  mGluR5_rec.mae
SUBJOB_HOST new.q
NUM_PRIME_CPUS    1
NUM_GLIDE_CPUS    1


# Protein Preparation
#  Run a simple constrained minimization of the receptor
#  structure(s).
STAGE PPREP
  RMSD      2.2


# Prime Loop Prediction
#  Perform a loop prediction on the specified loop, including
#  side chains within the given distance.  Only return
#  structures within the specified energy range from the
#  lowest energy prediction, up to the maximum number of
#  conformations given.
#
#  Note: This stage is disabled by default.  Uncomment the
#   lines below and edit the fields appropriately to enable it.
#STAGE PRIME_LOOP
#  START_RESIDUE A:11
#  END_RESIDUE A:16
#  DISTANCE_CUTOFF 5.0
#  MAX_ENERGY_GAP 30.0
#  MAX_STRUCTURES 5
#  USE_MEMBRANE no


# In order to temporarily remove the side chains of residues
# (i.e., mutate to Ala) that are blocking the binding site,
# uncomment the following STAGE line, and then specify the
# sidechains to be removed using either one of the two Methods
# described below.
#
STAGE TRIM_SIDECHAINS
#
# Method 1: Manual specification.  Uncomment the following
# line and list the desired residues in the format indicated.
#
  RESIDUES  _:722,_:726,_:732,_:734,_:735
#
# Method 2: Automatic determination based on predicted
# flexibility.  Uncomment the following lines, and specify
```

```
# the residues, ligands or other species which define the
# binding site.  All sidechains within DISTANCE_CUTOFF
# of these species will be tested for flexibility and
# trimmed if deemed necessary.
#
# RESIDUES  AUTO
# BINDING_SITE
# DISTANCE_CUTOFF 5.0

# Glide Docking
#  Perform the initial Glide docking, producing a
#  ligand-receptor complex for each pose requested/found.
#  If multiple receptor structures are used, the requested
#  number of poses will be generated for each structure.
STAGE GLIDE_DOCKING
  RECEPTOR_CCUT   0.25
  LIGAND_FILE     ligands.maegz
  LIGANDS_TO_DOCK all
  LIGAND_CCUT     0.15
  CV_CUTOFF 100.0
  HBOND_CUTOFF    -0.05
  INNER_BOX 10.0
  MINIMUM_POSES   1
  BINDING_SITE    residues
:624,:628,:629,:631,:632,:641,:643,:644,:645,:647,:648,:649,:650,:651,
:652,:653,:654,:655,:656,:657,:658,:659,:662,:710,:713,:714,:716,:717,
:718,:720,:722,:724,:730,:731,:732,:733,:734,:735,:736,:737,:739,:740,
:743,:744,:747,:748,:752,:777,:778,:781,:782,:784,:785,:787,:788,:789,
:791,:792,:808,:809,:811,:812,:813,:815,:816
  OUTER_BOX auto
  RECEPTOR_SCALE  0.70
  LIGAND_SCALE    0.50
  MAX_POSESPERLIG 10
  PRECISION SP

# Determine Residue to Refine
#  Compile a list of all residues within the specified
#  distance of any pose of the ligand.
STAGE COMPILE_RESIDUE_LIST
  DISTANCE_CUTOFF 5.0

# Prime Refinement
#  Optimize the side chains of the residue list compiled
#  previously, then minimize them along with the ligand.
STAGE PRIME_REFINEMENT
  NUMBER_OF_PASSES      1
  USE_MEMBRANE no

# Sort and Filter
#  Only retain poses with Prime Energies within the
#  specified range from the lowest energy pose.
STAGE SORT_AND_FILTER
  POSE_FILTER     r_psp_Prime_Energy
  POSE_KEEP 30.0

# Sort and Filter
#  Only retain the top number of poses specified.
STAGE SORT_AND_FILTER
  POSE_FILTER     r_psp_Prime_Energy
  POSE_KEEP 20#

# Glide Docking
#  Redock the ligand back into the newly optimized receptor,
```

10. Appendix

```
#  using default Glide settings.
STAGE GLIDE_DOCKING
  BINDING_SITE    ligand Z:999
  RECEPTOR_SCALE  1.00
  RECEPTOR_CCUT   0.25
  LIGAND_FILE     ligands.maegz
  LIGANDS_TO_DOCK self
  LIGAND_SCALE    0.80
  LIGAND_CCUT     0.15
  CV_CUTOFF 0.0
  HBOND_CUTOFF    0.0
  INNER_BOX 10.0
  MAX_POSESPERLIG 1
  OUTER_BOX auto
  PRECISION SP

# Scoring
#  Compile the IFD Score, consisting of the GlideScore for
#  the Glide Redocking plus 5% of the Prime Energy from the
#  Prime Refinement.
STAGE SCORING
  SCORE_NAME  r_psp_IFDScore
  TERM 1.0,r_i_glide_gscore,0
  TERM 0.05,r_psp_Prime_Energy,1
  REPORT_FILE report.csv
```

10.2 Homology Modelling Data

10.2.1 Alignment in PIR Format

The following alignment is referred to as 'alignment.ali' in the Modeller script file.

```
>P1;1U19
structureX:1U19: : A: : A:::2.2:0.200
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------MNGTEGPNF-----YVPFSNK
TGVV-----RSP-F-EAP-QYYL-AE--PWQF----SMLAAYMFLLIMLGFPINFLTLYV
TVQHKKLRTP-LNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERYVVV-------C--KP--MS-NFRFGENHAIMGVAFTWVMALACA
APPLVGWSRYIPEGMQ--CSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQL
VFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPI
FMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCG--KNPLGDDEASTTVSKTETSQ
VAPA*
>P1;H5_straight
structureX:H5_straight:::::::::
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

```
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
------------------------------LGVVTPLGYDGLLILSCTFYAFKTR
N*
>P1;H7_straight
structureX:H7_straight:::::::::
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
--------------------------------------------------------
-------------------------------------------------------MC
FSVSLSATVALGCMFVPKVYIIL*
>P1;mGluR5
sequence:mGluR:::::::::
MVLLLILSVLLLKEDVRGSAQSSERRVVAHMPGDIIIGALFSVHHQPTVDKVHERKCGAV
REQYGIQRVEAMLHTLERINSDPTLLPNITLGCEIRDSCWHSAVALEQSIEFIRDSLISS
EEEEGLVRCVDGSSSSFRSKKPIVGVIGPGSSSVAIQVQNLLQLFNIPQIAYSATSMDLS
DKTLFKYFMRVVPSDAQQARAMVDIVKRYNWTYVSAVHTEGNYGESGMEAFKDMSAKEGI
CIAHSYKIYSNAGEQSFDKLLKKLTSHLPKARVVACFCEGMTVRGLLMAMRRLGLAGEFL
LLGSDGWADRYDVTDGYQREAVGGITIKLQSPDVKWFDDYYLKLRPETNHRNPWFQEFWQ
HRFQCRLEGFPQENSKYNKTCNSSLTLKTHHVQDSKMGFVINAIYSMAYGLHNMQMSLCP
GYAGLCDAMKPIDGRKLLESLMKTNFTGVSGDTILFDENGDSPGRYEIMNFKEMGKDYFD
YINVGSWDNGELKMDDDEVWSKKSNIIRSVCSEPCEKGQIKVIRKGEVSCCWTCTPCKEN
EYVFDEYTCKACQLGSWPTDDLTGCDLIPVQYLRWGDPEPIAAVVFACLGLLATLFVTVV
FIIYRD--TPVVKSSSRELCYIILAGICLGYLCTFCLIAK----PKQIYCYLQRIGIGLS
PAMSYSALVTKTNRIARILAGSKKKICTKKPRFMSACAQLVIAFILICIQLGIIVALFIM
EPPDIM--HDYPSIREVYLIC--N--T----T-N-LGVVTPLGYNGLLILSCTFYAFKTR
NVP--------------ANFNEAKYIAFTMYTTCIIWLAFVPIYFGSNY---KIITMC
FSVSLSATVALGCMFVPKVYIILAKPERNVRSAFTTSTVVRMHVGDGKSSSAASRSSSLV
NLWKRRGSSGETLRYKDRRLAQHKSEIECFTPKGSMGNGGRATMSSSNGKSVTWAQNEKS
SRGQHLWQRLSIHINKKENPNQTAVIKPFPKSTESRGLGAGAGAGGSAGGVGATGGAGCA
GAGPGGPESPDAGPKALYDVAEAEEHFPAPARPRSPSPISTLSHRAGSASRTDDDVPSLH
SEPVARSSSSQGSLMEQISSVVTRFTANISELNSMMLSTAAPSPGVGAPLCSSYLIPKEI
QLPTTMTTFAEIQPLPAIEVTGGAQPAAGAQAAGDAARESPAAGPEAAAAKPDLEELVAL
TPPSPFRDSVDSGSTTPNSPVSESALCIPSSPKYDTLIIRDYTQSSSSL*
```

10.2.2 Modeller script

```
# Addition of restraints to the default ones
from modeller import *
from modeller.automodel import *    # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'
# directories for input atom files
env.io.atom_files_directory = 'directory'

class mymodel(automodel):
    def select_atoms(self):
#           All residues from 1 to 5:
```

10. Appendix

```
        return selection(self.residue_range('577', '860'))
#   def select_atoms(self):
#       return selection(self) - selection(self.residue_range('1',
'5'))
    def special_restraints(self, aln):
        rsr = self.restraints
#       Add some restraints from a file:
#       rsr.append(file='my_rsrs1.rsr')

#       Residues 580-604 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('580',
'604')))
#       Residues 608-638 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('608',
'638')))
#       Residues 639-675 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('639',
'675')))
#       Residues 685-718 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('685',
'718')))
#       Residues 738-762 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('738',
'762')))
#       Residues 766-797 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('766',
'797')))
#       Residues 801-839 should be an alpha helix:
        rsr.add(secondary_structure.alpha(self.residue_range('801',
'839')))

    def special_patches(self, aln):
        # A disulfide between residues 644 and 733:
        self.patch(residue_type='DISU',residues=(self.residues['644'],

self.residues['733']))

a = automodel(env,
              # file with template codes and target sequence
              alnfile  = 'alignment.ali',
              # PDB codes of the templates
              knowns   = ('1U19', 'H5_straight', 'H7_straight'),
              # code of the target
              sequence = 'mGluR5')
a.starting_model=1
a.ending_model=5
a.md_level = refine.slow
a.make()                            # do homology modelling
```

10.3 Multiple Sequence Alignment

TM1

```
BW position                      1.4       1.5       1.6
                                  |         |         |
1U19.A                  SMLAAYMFLLIMLGFPINFLTLYVTVQHKK
P22815|BOSS_DROME       DTWVATGLTAAILGLIATLAILVFIVVRIS
Q24738|BOSS_DROVI       DTWVATGLTAAILGLIATLAILVFIVVRIS
Q24265|Q24265_DROME     DTWVATGLTAAILGLIATLAILVFIVVRIS
Q8MSJ2|Q8MSJ2_DROME     DTWVAAGLTAAILGLIATLAILVFIVVRIS
Q90WL6|Q90WL6_SPAAU     EPFGIALAICAVLGVVLTAFVMGVFVRFRN
Q8JI04|Q8JI04_SQUAC     EPFGIALTIFAVLGILITSFVLGVFIKFRN
Q6XAF1|Q6XAF1_SALSA     EPFGIALALCSVLGVFLTAFVMGVFIKFRN
Q6XAF2|Q6XAF2_SALSA     EPFGIALALCSVLGVFLTAFVMGVFIKFRN
P41180|CASR_HUMAN       EPFGIALTLFAVLGIFLTAFVLGVFIKFRN
Q80ZA8|Q80ZA8_RAT       EPFGIALTLFAVLGIFLTAFVLGVFIKFRN
P48442|CASR_RAT         EPFGIALTLFAVLGIFLTAFVLGVFIKFRN
P35384|CASR_BOVIN       EPFGIALTLFAVLGIFLTAFVLGVFIKFRN
Q9QY96|CASR_MOUSE       EPFGIALTLFAVLGIFLTAFVLGVFIKFRN
Q6XAF3|Q6XAF3_SALSA     EPFGIALALCSVLGVFLTAFVMGVFIKFRN
Q5YEV6|Q5YEV6_OREMO     EPFGIALAICAVLGVVLTAFVIGVFVRFRN
O73639|O73639_FUGRU     DTIGIALLVVSLIGSFLTCAVALVFFYHRT
O73638|O73638_FUGRU     ETMGALLAAVSLFGAALTSLVFCVFFRFRH
O73637|O73637_FUGRU     EPLGICLTAASLLGTVISVVVLGIFIHHRS
O73636|O73636_FUGRU     EVLGIILAVFSVGGACLAVITAAVFFHHRT
O73635|O73635_FUGRU     EPFGIALAICAVLGVLLTAFVMGVFVRFRN
O75899|GABR2_HUMAN      LPLYSILSALTILGMIMASAFLFFNIKNRN
O88871|GABR2_RAT        LPLYSILSALTILGMIMASAFLFFNIKNRN
Q9BML6|Q9BML6_DROME     PTIYIVSASASVIGVIIATVFLAFNIKYRN
Q9Y133|Q9Y133_DROME     PTIYIVSASASVIGVIIATVFLAFNIKYRN
Q8IN24|Q8IN24_DROME     PTIYIVSASASVIGVIIATVFLAFNIKYRN
Q5TRW6|Q5TRW6_ANOGA     ITIFVVLASTSCVGIIMATVFLAVNITFRN
Q9V3Q9|Q9V3Q9_DROME     LPLFVCMCTISSCGIFVAFALIIFNIWNKH
Q9BML7|Q9BML7_DROME     LPLFVCMCTISSCGIFVAFALIIFNIWNKH
Q8NHA5|Q8NHA5_HUMAN     QKLFISVSVLSSLGIVLAVVCLSFNIYNSH
Q8IW08|Q8IW08_HUMAN     QKLFISVSVLSSLGIVLAVVCLSFNIYNSH
Q6PGJ2|Q6PGJ2_MOUSE     QKLFISVSVLSSLGIVLAVVCLSFNIYNSH
Q9WV18|GABR1_MOUSE      QKLFISVSVLSSLGIVLAVVCLSFNIYNSH
Q9UBS5|GABR1_HUMAN      QKLFISVSVLSSLGIVLAVVCLSFNIYNSH
Q8K451|GP156_RAT        PALLGVIWTFLSCGLLLVLFFLAFTIRCRK
Q8NFN8|GP156_HUMAN      PVLLGIVWTFLSCGLLLILFFLAFTIHCRK
Q6PCP7|GP156_MOUSE      PALLGIMWTFLSCGLLLVLFFLAFTIRCRK
Q9VPS7|Q9VPS7_DROME     PLAFYTIATLSSVGIALAIAFLAFNLHFRK
Q8IPW4|Q8IPW4_DROME     PLAFYTIATLSSVGIALAIAFLAFNLHFRK
Q8MSP9|Q8MSP9_DROME     PLAFYTIATLSSVGIALAIAFLAFNLHFRK
O96954|O96954_GEOCY     VPLTVVYVALAVGGLVFAIVCVFFTVIFRK
Q8BHL4|RAI3_MOUSE       EGWGIALETLAAVGAVATVACMFALVFLIC
Q8NFJ5|RAI3_HUMAN       EAWGIVLETVATAGVVTSVAFMLTLPILVC
Q9JIL6|GPC5D_MOUSE      GPWAIVLESLAVIGIVVTILLLLAFLFLMR
Q9NZD1|GPC5D_HUMAN      GPWGIILESLAILGIVVTILLLLAFLFLMR
Q8K3J9|GPC5C_MOUSE      GAWGIVSEAVAGAGIITTFVLTIILVASLP
Q9NQ84|GPC5C_HUMAN      GAWGIVLEAVAGAGIVTTFVLTIILVASLP
Q923Z0|GPC5B_MOUSE      AIWGIVVEAVAGAGALITLLLMLILLVRLP
Q9NZH0|GPC5B_HUMAN      AIWGIVVEAVAGAGALITLLLMLILLVRLP
Q6PA25|Q6PA25_XENLA     AAWGIVLETLAAAGIVFSIILILALLIMMP
Q5T6X5|GPC6A_HUMAN      DSLAILLLILSLLGIIFVLVVGIIFTRNLN
Q5U9X3|GPC6A_BRARE      SGFAIVLLILAALGVLLLFFMSALFFWQRH
Q8K4Z6|GPC6A_MOUSE      DSLALLLIALSLLGIAFVLAIGIIFTRNLK
Q9PW88|GPC6A_CARAU      SGFAIALLTLAALGILLLISMSALFFWQRN
Q9Z0R8|TS1R1_RAT        EPISLVLIAANTLLLLLLVGTAGLFAWHFH
Q99PG6|TS1R1_MOUSE      EPISLVLLAANTLLLLLLIGTAGLFAWRLH
Q7RTX1|TS1R1_HUMAN      EHTSWVLLAANTLLLLLLLGTAGLFAWHLD
```

10. Appendix

```
Q9Z0R7|TS1R2_RAT      EVPTIVVAILAALGFFSTLAILFIFWRHFQ
Q925I4|TS1R2_MOUSE    EVPTIVVTILAALGFISTLAILLIFWRHFQ
Q8TE23|TS1R2_HUMAN    EAPTIAVALLAALGFLSTLAILVIFWRHFQ
Q7RTX0|TS1R3_HUMAN    EPAVLLLLLLLSLALGLVLAALGLFVHHRD
Q717C1|TS1R3_GORGO    EPAVLLLLLLLSLALGLVLAALGLFVHHRD
Q717C2|TS1R3_PANTR    EPAVLLLLLLLSLALGLVLAALGLFIHHRD
Q923K1|TS1R3_RAT      EPAVLSLLLLLCLVLGLTLAALGLFVHYWD
Q925D8|TS1R3_MOUSE    EPVVLSLLLLLCLVLGLALAALGLSVHHWD
P41594|MGR5_HUMAN     DPEPIAAVVFACLGLLATLFVTVVFIIYRD
Q75QW7|Q75QW7_APIME   SAFAIAPAVISCLGIVATMAVACLLFHHRD
P91685|MGR_DROME      SLFALIPMAIAIFGIALTSIVIVLFAKNHD
Q8CFQ7|Q8CFQ7_MOUSE   SPWAALPLLLAVLGIMATTTIIATFMRHND
Q863I4|MGR6_RABIT     SPWAAPPLLLAVLGIMATTTVVGTFVRHNN
Q68ED2|MGR7_MOUSE     SPWAVIPVFLAMLGIIATIFVMATFIRYND
P70579|MGR8_RAT       SPWAVVPVFIAILGIIATTFVIVTFVRYND
P47743|MGR8_MOUSE     SPWAVVPVFIAILGIIATTFVIVTFVRYND
O00222|MGR8_HUMAN     SPWAVVPVFVAILGIIATTFVIVTFVRYND
P35400|MGR7_RAT       SPWAVIPVFLAMLGIIATIFVMATFIRYND
Q14831|MGR7_HUMAN     SPWAVIPVFLAMLGIIATIFVMATFIRYND
P35349|MGR6_RAT       SPWAALPLLLAVLGIMATTTIMATFMRHND
O15303|MGR6_HUMAN     SPWAAPPLLLAVLGIVATTTVVATFVRYNN
P31424|MGR5_RAT       DPEPIAAVVFACLGLLATLFVTVIFIIYRD
P31423|MGR4_RAT       SPWAVLPLFLAVVGIAATLFVVVTFVRYND
Q14833|MGR4_HUMAN     SPWAVLPLFLAVVGIAATLFVVITFVRYND
P31422|MGR3_RAT       DAWAIGPVTIACLGFLCTCIVITVFIKHNN
Q9QYS2|MGR3_MOUSE     DAWAIGPVTIACLGFMCTCIVITVFIKHNN
Q14832|MGR3_HUMAN     DAWAIGPVTIACLGFMCTCMVVTVFIKHNN
P31421|MGR2_RAT       DAWAVGPVTIACLGALATLFVLGVFVRHNA
Q14416|MGR2_HUMAN     DAWAVGPVTIACLGALATLFVLGVFVRHNA
P23385|MGR1_RAT       DIESIIAIAFSCLGILVTLFVTLIFVLYRD
P97772|MGR1_MOUSE     DIESIIAIAFSCLGILVTLFVTLIFVLYRD
Q13255|MGR1_HUMAN     NIESIIAIAFSCLGILVTLFVTLIFVLYRD
Q90ZF3|Q90ZF3_ONCMA   NPESIVQVVFACLGILVTSFVTFIFVLYRD
Q9V4U4|Q9V4U4_DROME   SAWAIGAMAFSATGILVTLFVMGVFVRHND
Q9V4U3|Q9V4U3_DROME   SAWAIGAMAFSATGILVTLFVMGVFVRHND
Q7KQS9|Q7KQS9_DROME   SAWAIGAMAFSATGILVTLFVMGVFVRHND
Q75QW6|Q75QW6_APIME   SGWAIGAMSFSATGILITLFVCGVFLKHND
Q70GQ8|Q70GQ8_DROME   SAWAIGAMAFSATGILVTLFVMGVFVRHND
Q62916|Q62916_RAT     SPWAVLPLFLAVVGIAATLFVVVTFVRYND
```

TM2

```
BW position          2.40       2.5       2.6
                     |          |         |
1U19.A               LNYILLNLAVADLFMVFGGFTTTLYTSLHG
P22815|BOSS_DROME    VFEGNPTTSILLLLSLILVFCSFVPYSIEY
Q24738|BOSS_DROVI    VFEGNPVTSILLLLSLILVFCSFVPFSMEY
Q24265|Q24265_DROME  VFEGNPTTSILLLLSLILVFCSFVPYSIEY
Q8MSJ2|Q8MSJ2_DROME  VFEGNPTTSILLLLSLILVFCSFVPYSIEY
Q90WL6|Q90WL6_SPAAU  VKATNRELSYVLLFSLICCFSSSLIFIGQP
Q8JI04|Q8JI04_SQUAC  VKATNRELSYLLLFSLICCFSSSLIFIGEP
Q6XAF1|Q6XAF1_SALSA  VKATNRELSYLLLFSLICCFSSSLIFIGEP
Q6XAF2|Q6XAF2_SALSA  VKATNRELSYLLLFSLICCFSSSLIFIGEP
P41180|CASR_HUMAN    VKATNRELSYLLLFSLLCCFSSSLFFIGEP
Q80ZA8|Q80ZA8_RAT    VKATNRELSYLLLFSLLCCFSSSLFFIGEP
P48442|CASR_RAT      VKATNRELSYLLLFSLLCCFSSSLFFIGEP
P35384|CASR_BOVIN    VKATNRELSYLLLFSLLCCFSSSLFFIGEP
Q9QY96|CASR_MOUSE    VKATNRELSYLLLFSLLCCFSSSLFFIGEP
Q6XAF3|Q6XAF3_SALSA  VKATNRELSYLLLFSLICCFSSSLIFIGEP
Q5YEV6|Q5YEV6_OREMO  VKATNRELSYVLLFSLICCFSSSLIFIGEP
O73639|O73639_FUGRU  VRANNSDLSFLLLFSLTLCFLCSLTFISPP
O73638|O73638_FUGRU  VKASNSELSFLLLFSLTLCFLCSLTFIGRP
O73637|O73637_FUGRU  VRANNSELSFLLLVSLKLCFLCSLLFIGRP
```

```
O73636|O73636_FUGRU    VRANNSELSFLLLFSLTLCFLCSLTFIGAP
O73635|O73635_FUGRU    VKASNRELSYVLLLSLICCFSSSLIFIGEP
O75899|GABR2_HUMAN     IKMSSPYMNNLIILGGMLSYASIFLFGLDG
O88871|GABR2_RAT       IKMSSPYMNNLIILGGMLSYASIFLFGLDG
Q9BML6|Q9BML6_DROME    IKMSSPHLNNLIIVGCMITYLSIIFLGLDT
Q9Y133|Q9Y133_DROME    IKMSSPHLNNLIIVGCMMTYLSIIFLGLDT
Q8IN24|Q8IN24_DROME    IKMSSPHLNNLIIVGCMMTYLSIIFLGLDT
Q5TRW6|Q5TRW6_ANOGA    IKMSSPHLNNLIIIGCILTYLSVIFLGLDS
Q9V3Q9|Q9V3Q9_DROME    IQSSHPVCNTIMLFGVIICLISVILLGIDG
Q9BML7|Q9BML7_DROME    IQSSHPVCNTIMLFGVIICLISVILLGIDG
Q8NHA5|Q8NHA5_HUMAN    IQNSQPNLNNLTAVGCSLALAAVFPLGLDG
Q8IW08|Q8IW08_HUMAN    IQNSQPNLNNLTAVGCSLALAAVFPLGLDG
Q6PGJ2|Q6PGJ2_MOUSE    IQNSQPNLNNLTAVGCSLALAAVFPLGLDG
Q9WV18|GABR1_MOUSE     IQNSQPNLNNLTAVGCSLALAAVFPLGLDG
Q9UBS5|GABR1_HUMAN     IQNSQPNLNNLTAVGCSLALAAVFPLGLDG
Q8K451|GP156_RAT       VKMSSPNLNIVTLLGSCLTYSSAYLFGIQD
Q8NFN8|GP156_HUMAN     VKMSSPNLNIVTLLGSCLTYSSAYLFGIQD
Q6PCP7|GP156_MOUSE     VKMSSPNLNVVTLLGSCLTYISAYLFGIQD
Q9VPS7|Q9VPS7_DROME    IKLSSPKLSNITAVGCIFVYATVILLGLDH
Q8IPW4|Q8IPW4_DROME    IKLSSPKLSNITAVGCIFVYATVILLGLDH
Q8MSP9|Q8MSP9_DROME    IKLSSPKLSNITAVGCIFVYATVILLGLDH
O96954|O96954_GEOCY    IRLSSPNLNYLIGLGAIILYFNVITLVIPT
Q8BHL4|RAI3_MOUSE      NKRKMLPAQFLFLLGVLGVFGLTFAFIIKL
Q8NFJ5|RAI3_HUMAN      NRRKMLPTQFLFLLGVLGIFGLTFAFIIGL
Q9JIL6|GPC5D_MOUSE     SQWNVLPTQFLFLLAVLGLFGLTFAFIIQL
Q9NZD1|GPC5D_HUMAN     SQWNVLPTQLLFLLSVLGLFGLAFAFIIEL
Q8K3J9|GPC5C_MOUSE     KKRSLLGTQVFFLLGTLGLFCLVFACVVKP
Q9NQ84|GPC5C_HUMAN     KKRSLLGTQVFFLLGTLGLFCLVFACVVKP
Q923Z0|GPC5B_MOUSE     ERKRPVCLHFLFLLGTLGLFGLTFAFIIQM
Q9NZH0|GPC5B_HUMAN     EKKSPVGLHFLFLLGTLGLFGLTFAFIIQE
Q6PA25|Q6PA25_XENLA    AKRAVSPVQLIFLIGTFGIFGLTFAFIVEL
Q5T6X5|GPC6A_HUMAN     KSSGGLRVCYVILLCHFLNFASTSFFIGEP
Q5U9X3|GPC6A_BRARE     VKAAGGPLCHLILVSLLGSFISVVFFVGEP
Q8K4Z6|GPC6A_MOUSE     KSSGGLVVCYVMLICHALNFASTGFFIGEP
Q9PW88|GPC6A_CARAU     VKAAGGPLCHLILFSLLGSFISVIFFVGEP
Q9Z0R8|TS1R1_RAT       VRSAGGRLCFLMLGSLVAGSCSFYSFFGEP
Q99PG6|TS1R1_MOUSE     VRSAGGRLCFLMLGSLVAGSCSLYSFFGKP
Q7RTX1|TS1R1_HUMAN     VRSAGGRLCFLMLGSLAAGSGSLYGFFGEP
Q9Z0R7|TS1R2_RAT       VRSAGGPMCFLMLVPLLLAFGMVPVYVGPP
Q925I4|TS1R2_MOUSE     VRSAGGPMCFLMLVPLLLAFGMVPVYVGPP
Q8TE23|TS1R2_HUMAN     VRSAGGPMCFLMLTLLLVAYMVVPVYVGPP
Q7RTX0|TS1R3_HUMAN     VQASGGPLACFGLVCLGLVCLSVLLFPGQP
Q717C1|TS1R3_GORGO     VQASGGPLACFGLVCLGLVCLSVLLFPGQP
Q717C2|TS1R3_PANTR     VQASGGPLACFGLVCLGLVCLSVLLFPGQP
Q923K1|TS1R3_RAT       VQASGGSLFCFGLICLGLFCLSVLLFPGRP
Q925D8|TS1R3_MOUSE     VQASGGSQFCFGLICLGLFCLSVLLFPGRP
P41594|MGR5_HUMAN      VKSSSRELCYIILAGICLGYLCTFCLIAKP
Q75QW7|Q75QW7_APIME    VRASGRELTIILLAGVLVCYLNTFLLLATP
P91685|MGR_DROME       VRASGRELSYTLLFGILVCYCNTFALIAKP
Q8CFQ7|Q8CFQ7_MOUSE    VRASGRELSYVLLTGIFLIYAITFLMVAEP
Q863I4|MGR6_RABIT      VRASGRELSYVLLTGIFLIYAVTFLMVAEP
Q68ED2|MGR7_MOUSE      VRASGRELSYVLLTGIFLCYIITFLMIAKP
P70579|MGR8_RAT        VRASGRELSYVLLTGIFLCYSITFLMIAAP
P47743|MGR8_MOUSE      VRASGRELSYVLLTGIFLCYSITFLMIAAP
O00222|MGR8_HUMAN      VRASGRELSYVLLTGIFLCYSITFLMIAAP
P35400|MGR7_RAT        VRASGRELSYVLLTGIFLCYIITFLMIAKP
Q14831|MGR7_HUMAN      VRASGRELSYVLLTGIFLCYIITFLMIAKP
P35349|MGR6_RAT        VRASGRELSYVLLTGIFLIYAITFLMVAEP
O15303|MGR6_HUMAN      VRASGRELSYVLLTGIFLIYAITFLMVAEP
P31424|MGR5_RAT        VKSSSRELCYIILAGICLGYLCTFCLIAKP
P31423|MGR4_RAT        VKASGRELSYVLLAGIFLCYATTFLMIAEP
Q14833|MGR4_HUMAN      VKASGRELSYVLLAGIFLCYATTFLMIAEP
P31422|MGR3_RAT        VKASGRELCYILLFGVSLSYCMTFFFIAKP
```

```
Q9QYS2|MGR3_MOUSE        VKASGRELCYILLFGVSLSYCMTFFFIAKP
Q14832|MGR3_HUMAN        VKASGRELCYILLFGVGLSYCMTFFFIAKP
P31421|MGR2_RAT          VKASGRELCYILLGGVFLCYCMTFVFIAKP
Q14416|MGR2_HUMAN        VKASGRELCYILLGGVFLCYCMTFIFIAKP
P23385|MGR1_RAT          VKSSSRELCYIILAGIFLGYVCPFTLIAKP
P97772|MGR1_MOUSE        VKSSSRELCYIILAGIFLGYVCPFTLIAKP
Q13255|MGR1_HUMAN        VKSSSRELCYIILAGIFLGYVCPFTLIAKP
Q90ZF3|Q90ZF3_ONCMA      VKSSSRELCYIILAGIFLGYICPFTLIAQP
Q9V4U4|Q9V4U4_DROME      VRASGRELSYILLAGIFMCYGVTFALVLKP
Q9V4U3|Q9V4U3_DROME      VRASGRELSYILLAGIFMCYGVTFALVLKP
Q7KQS9|Q7KQS9_DROME      VRASGRELSYILLAGIFMCYGVTFALVLKP
Q75QW6|Q75QW6_APIME      VRASGRELSYVLLSGILLCYLVTFALVLRP
Q70GQ8|Q70GQ8_DROME      VRASGRELSYILLAGIFMCYGVTFALVLKP
Q62916|Q62916_RAT        VKASGRELSYVLLAGIFLCYATTFLMIAEP
```

TM3

```
BW position               3.3       3.4       3.5
                           |         |         |
1U19.A                   CNLEGFFATLGGEIALWSLVVLAIERYVVV
P22815|BOSS_DROME        CAVRVFIMTLVYCFVFSLLLCRAVMLASIG
Q24738|BOSS_DROVI        CGVRVFIMTLVYCFVFSLLLCRAVMLASIG
Q24265|Q24265_DROME      CAVRVFIMTLVYCFVFSLLLCRAVMLASIG
Q8MSJ2|Q8MSJ2_DROME      CAVRVFIMTLVYCFVFSLLLCRAVMLASIG
Q90WL6|Q90WL6_SPAAU      CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q8JI04|Q8JI04_SQUAC      CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q6XAF1|Q6XAF1_SALSA      CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q6XAF2|Q6XAF2_SALSA      CRLRQPAFGISFVLCISCILVKTNRVLLVF
P41180|CASR_HUMAN        CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q80ZA8|Q80ZA8_RAT        CRLRQPAFGISFVLCISCILVKTNRVLLLF
P48442|CASR_RAT          CRLRQPAFGISFVLCISCILVKTNRVLLVF
P35384|CASR_BOVIN        CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q9QY96|CASR_MOUSE        CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q6XAF3|Q6XAF3_SALSA      CRLRQPAFGISFVLCISCILVKTNRVLLVF
Q5YEV6|Q5YEV6_OREMO      CRLRQPAFGVSFVLCISCILVKTNRVLLVF
O73639|O73639_FUGRU      CMLRHTAFGITFVLCISCILGKTIVVLMAF
O73638|O73638_FUGRU      CVLRHTAFGITFALCMSCVLAKTVAVLFAF
O73637|O73637_FUGRU      CQLRHAAFGISFVLCVSCILVKTMVVLAVF
O73636|O73636_FUGRU      CMLRHTAFGITFVLCISCVLGKTVVVLMAF
O73635|O73635_FUGRU      CRLRQPAFGISFVLCISCILVKTNRVLLVF
O75899|GABR2_HUMAN       CTVRTWILTVGYTTAFGAMFAKTWRVHAIF
O88871|GABR2_RAT         CTVRTWILTVGYTTAFGAMFAKTWRVHAIF
Q9BML6|Q9BML6_DROME      CTARAWILMAGFSLSFGAMFSKTWRVHSIF
Q9Y133|Q9Y133_DROME      CTARAWILMAGFSLSFGAMFSKTWRVHSIF
Q8IN24|Q8IN24_DROME      CTARAWILMAGFSLSFGAMFSKTWRVHSIF
Q5TRW6|Q5TRW6_ANOGA      CTARAWLLMAGFSLAFGAMFSKTWRVHSIF
Q9V3Q9|Q9V3Q9_DROME      CQARAWLLSTGFTLAYGAMFSKVWRVHRFT
Q9BML7|Q9BML7_DROME      CQARAWLLSTGFTLAYGAMFSKVWRVHRFT
Q8NHA5|Q8NHA5_HUMAN      CQARLWLLGLGFSLGYGSMFTKIWWVHTVF
Q8IW08|Q8IW08_HUMAN      CQARLWLLGLGFSLGYGSMFTKIWWVHTVF
Q6PGJ2|Q6PGJ2_MOUSE      CQARLWLLGLGFSLGYGSMFTKIWWVHTVF
Q9WV18|GABR1_MOUSE       CQARLWLLGLGFSLGYGSMFTKIWWVHTVF
Q9UBS5|GABR1_HUMAN       CQARLWLLGLGFSLGYGSMFTKIWWVHTVF
Q8K451|GP156_RAT         IQTRLSLLCIGTTLVFGPILGKSWRLYKVF
Q8NFN8|GP156_HUMAN       IQTRLSMLCIGTSLVFGPILGKSWRLYKVF
Q6PCP7|GP156_MOUSE       IQTRLSLLCIGTSLVFGPILGKSWRLYKVF
Q9VPS7|Q9VPS7_DROME      CTARVYLLSAGFSLAFGSMFAKTYRVHRIF
Q8IPW4|Q8IPW4_DROME      CTARVYLLSAGFSLAFGSMFAKTYRVHRIF
Q8MSP9|Q8MSP9_DROME      CTARVYLLSAGFSLAFGSMFAKTYRVHRIF
O96954|O96954_GEOCY      CNINPWLTSLGYSLCYGTILAKTIRIWFIF
Q8BHL4|RAI3_MOUSE        GPTRFFLFGVLFAICFSCLLAHAFNLIKLV
Q8NFJ5|RAI3_HUMAN        GPTRFFLFGILFSICFSCLLAHAVSLTKLV
Q9JIL6|GPC5D_MOUSE       APVRYFLFGVLFAICFSCLLAHASNLVKLV
```

```
Q9NZD1|GPC5D_HUMAN     APVRYFLFGVLFALCFSCLLAHASNLVKLV
Q8K3J9|GPC5C_MOUSE     CASRRFLFGVLFAICFSCLVAHVLSLNFLT
Q9NQ84|GPC5C_HUMAN     CASRRFLFGVLFAICFSCLAAHVFALNFLA
Q923Z0|GPC5B_MOUSE     CSIRRFLWGVLFALCFSCLLSQAWRVRRLV
Q9NZH0|GPC5B_HUMAN     CSVRRFLWGVLFALCFSCLLSQAWRVRRLV
Q6PA25|Q6PA25_XENLA    CPTRFFLFGVLFAICFSCLLAHASKLVRLV
Q5T6X5|GPC6A_HUMAN     CKTRQTMFGVSFTLCISCILTKSLKILLAF
Q5U9X3|GPC6A_BRARE     CRARQVIFGFSFTLCVSCILVKSLKILLAF
Q8K4Z6|GPC6A_MOUSE     CKTRQTLFGVSFTLCVSCILTKSLKILLAF
Q9PW88|GPC6A_CARAU     CRVRQVIFGLSFTLCVSCILVKSLKILLAF
Q9Z0R8|TS1R1_RAT       CLLRQPLFSLGFAIFLSCLTIRSFQLVIIF
Q99PG6|TS1R1_MOUSE     CLLRQPLFSLGFAIFLSCLTIRSFQLVIIF
Q7RTX1|TS1R1_HUMAN     CLLRQALFALGFTIFLSCLTVRSFQLIIIF
Q9Z0R7|TS1R2_RAT       CFCRQAFFTVCFSICLSCITVRSFQIVCVF
Q925I4|TS1R2_MOUSE     CFCRQAFFTVCFSVCLSCITVRSFQIVCVF
Q8TE23|TS1R2_HUMAN     CLCRQALFPLCFTICISCIAVRSFQIVCAF
Q7RTX0|TS1R3_HUMAN     CLAQQPLSHLPLTGCLSTLFLQAAEIFVES
Q717C1|TS1R3_GORGO     CLAQQPLSHLPLTGCLSTLFLQAAEIFVES
Q717C2|TS1R3_PANTR     CLAQQPLSHLPLTGCLSTLFLQAAEIFVES
Q923K1|TS1R3_RAT       CLAQQPMAHLPLTGCLSTLFLQAAEIFVES
Q925D8|TS1R3_MOUSE     CLAQQPMAHLPLTGCLSTLFLQAAETFVES
P41594|MGR5_HUMAN      CYLQRIGIGLSPAMSYSALVTKTNRIARIL
Q75QW7|Q75QW7_APIME    CILQRFGVGVSFSAVYGALLTKTNRIARIF
P91685|MGR_DROME       CVLQRFGIGVGFSIIYSALLTKTNRISRIF
Q8CFQ7|Q8CFQ7_MOUSE    CASRRLLLGLGTTLSYSALLTKTNRIYRIF
Q863I4|MGR6_RABIT      CATRRLFLGLGTTLSYSALLTKTNRIYRIF
Q68ED2|MGR7_MOUSE      CSFRRVFLGLGMCISYAALLTKTNRIYRIF
P70579|MGR8_RAT        CSFRRIFLGLGMCFSYAALLTKTNRIHRIF
P47743|MGR8_MOUSE      CSFRRIFLGLGMCFSYAALLTKTNRIHRIF
O00222|MGR8_HUMAN      CSFRRVFLGLGMCFSYAALLTKTNRIHRIF
P35400|MGR7_RAT        CSFRRVFLGLGMCISYAALLTKTNRIYRIF
Q14831|MGR7_HUMAN      CSFRRVFLGLGMCISYAALLTKTNRIYRIF
P35349|MGR6_RAT        CAARRLLLGLGTTLSYSALLTKTNRIYRIF
O15303|MGR6_HUMAN      CAARRLFLGLGTTLSYSALLTKTNRIYRIF
P31424|MGR5_RAT        CYLQRIGIGLSPAMSYSALVTKTNRIARIL
P31423|MGR4_RAT        CSLRRIFLGLGMSISYAALLTKTNRIYRIF
Q14833|MGR4_HUMAN      CSLRRIFLGLGMSISYAALLTKTNRIYRIF
P31422|MGR3_RAT        CALRRLGLGTSFAICYSALLTKTNCIARIF
Q9QYS2|MGR3_MOUSE      CALRRLGLGTSFAICYSALLTKTNCIARIF
Q14832|MGR3_HUMAN      CALRRLGLGSSFAICYSALLTKTNCIARIF
P31421|MGR2_RAT        CTLRRLGLGTAFSVCYSALLTKTNRIARIF
Q14416|MGR2_HUMAN      CTLRRLGLGTAFSVCYSALLTKTNRIARIF
P23385|MGR1_RAT        CYLQRLLVGLSSAMCYSALVTKTNRIARIL
P97772|MGR1_MOUSE      CYLQRLLVGLSSAMCYSALVTKTNRIARIL
Q13255|MGR1_HUMAN      CYLQRLLVGLSSAMCYSALVTKTNRIARIL
Q90ZF3|Q90ZF3_ONCMA    CYLQRLLVGLSATMCYSALVTKTNRIARIL
Q9V4U4|Q9V4U4_DROME    CAIQRFGVGFCFTVVYAALLTKTNRIARIF
Q9V4U3|Q9V4U3_DROME    CAIQRFGVGFCFTVVYAALLTKTNRIARIF
Q7KQS9|Q7KQS9_DROME    CAIQRFGVGFCFTVVYAALLTKTNRIARIF
Q75QW6|Q75QW6_APIME    CGIQRFAAGFCFTVVYAALLTKTNRISRIF
Q70GQ8|Q70GQ8_DROME    CAIQRFGVGFCFTVVYAALLTKTNRIARIF
Q62916|Q62916_RAT      CSLRRIFLGLGMSISYAALLTKTNRIYRIF
```

TM4

```
BW position                     4.4       4.5       4.6
                                 |         |         |
1U19.A                 KPMSNFRFGENHAIMGVAFTWVMALACAAPP
P22815|BOSS_DROME      HVNGYIQAVICAFSVVAQVGMSVQLLVVMHV
Q24738|BOSS_DROVI      HVNGYIQAIICVLSVFVQVGMSVQLLVVMHL
Q24265|Q24265_DROME    HVNGYIQAVICAFSVVAQVGMSVQLLVVMHV
Q8MSJ2|Q8MSJ2_DROME    HVNGYIQAVICAFSVVAQVGMSVQLLVVMHV
Q90WL6|Q90WL6_SPAAU    WWGLNLQFLLVFLCTFVQVMICVVWLYNAPP
```

10. Appendix

```
Q8JI04|Q8JI04_SQUAC    WVGLNLQFLLVFLCILVQIVTCIIWLYTAPP
Q6XAF1|Q6XAF1_SALSA    WWGLNLQFLLVFLFTFVQVMICVVWLYNAPP
Q6XAF2|Q6XAF2_SALSA    WWGLNLQFLLVFLFTFVQVMICVVWLYNAPP
P41180|CASR_HUMAN      WWGLNLQFLLVFLCTFMQIVICVIWLYTAPP
Q80ZA8|Q80ZA8_RAT      WWGLNLQFLLVFLCTFMQILICIIWLYTAPP
P48442|CASR_RAT        WWGLNLQFLLVFLCTFMQILICIIWLYTAPP
P35384|CASR_BOVIN      WWGLNLQFLLVFLCTFMQIVICAIWLNTAPP
Q9QY96|CASR_MOUSE      WWGLNLQFLLVFLCTFMQIVICIIWLYTAPP
Q6XAF3|Q6XAF3_SALSA    WWGLNLQFLLVFLFTFVQVMICVVWLYNAPP
Q5YEV6|Q5YEV6_OREMO    WWGLNLQFLLVFLCTFVQVMICVVWLYNAPP
O73639|O73639_FUGRU    WFGPGKQKAIITFSTLVQVVICTVWLVVAPP
O73638|O73638_FUGRU    YCSVPLQRTSVFACITLQVIICVLWLTLAPP
O73637|O73637_FUGRU    WFGAVQQRGTVLGLTSIQAAICFAWLLSSSP
O73636|O73636_FUGRU    WFGPPQQRMTVVTFTSIQVLICIVWLVVNPP
O73635|O73635_FUGRU    WWGLNLQFLLVFLCTFVQVMICVVWLYNAPP
O75899|GABR2_HUMAN     IIKDQKLLVIVGGMLLIDLCILICWQAVDPL
O88871|GABR2_RAT       IIKDQKLLVIVGGMLLIDLCILICWQAVDPL
Q9BML6|Q9BML6_DROME    VIKDYQLFMVVGVLLAIDIAIITTWQIADPF
Q9Y133|Q9Y133_DROME    VIKDYQLFMVVGVLLAIDIAIITTWQIADPF
Q8IN24|Q8IN24_DROME    VIKDYQLFMVVGVLLAIDIAIITTWQIADPF
Q5TRW6|Q5TRW6_ANOGA    VIKDYQLFIVVGVLLAIDLAIMTTWQIADPF
Q9V3Q9|Q9V3Q9_DROME    KVEPWKLYTMVSGLLSIDLVILLSWQIFDPL
Q9BML7|Q9BML7_DROME    KVEPWKLYTMVSGLLSIDLVILLSWQIFDPL
Q8NHA5|Q8NHA5_HUMAN    TLEPWKLYATVGLLVGMDVLTLAIWQIVDPL
Q8IW08|Q8IW08_HUMAN    TLEPWKLYATVGLLVGMDVLTLAIWQIVDPL
Q6PGJ2|Q6PGJ2_MOUSE    TLEPWKLYATVGLLVGMDILTLAIWQIVDPL
Q9WV18|GABR1_MOUSE     TLEPWKLYATVGLLVGMDILTLAIWQIVDPL
Q9UBS5|GABR1_HUMAN     TLEPWKLYATVGLLVGMDVLTLAIWQIVDPL
Q8K451|GP156_RAT       IIKDLQLLGLVAALVVADVILLVTWVLTDPI
Q8NFN8|GP156_HUMAN     IIKDLQLLGLVAALLMADVILLMTWVLTDPI
Q6PCP7|GP156_MOUSE     IIKDLQLLGLVAALVVADVILLVTWVLTDPI
Q9VPS7|Q9VPS7_DROME    MLQDIQLILLVGGLLLVDALLVTLWVVTDPM
Q8IPW4|Q8IPW4_DROME    MLQDIQLILLVGGLLLVDALLVTLWVVTDPM
Q8MSP9|Q8MSP9_DROME    MLQDIQLILLVGGLLLVDALLVTLWVVTDPM
O96954|O96954_GEOCY    VIKDYALALFVVSLVVIDVIILGIFAIVEGL
Q8BHL4|RAI3_MOUSE      PLSWLVILSLAVGFSLVQDVIAIEYLVLTMN
Q8NFJ5|RAI3_HUMAN      PLSLLVILGLAVGFSLVQDVIAIEYIVLTMN
Q9JIL6|GPC5D_MOUSE     SFCWTTILFIAIGVSLLQTIIAIEYVTLIMT
Q9NZD1|GPC5D_HUMAN     SFSWTTILCIAIGCSLLQIIIATEYVTLIMT
Q8K3J9|GPC5C_MOUSE     GPRGWVIFTVALLLTLVEVIINTEWLIITLV
Q9NQ84|GPC5C_HUMAN     GPRGWVIFTVALLLTLVEVIINTEWLIITLV
Q923Z0|GPC5B_MOUSE     SPASWQLVSLALCLMLVQVIIATEWLVLTVL
Q9NZH0|GPC5B_HUMAN     GPAGWQLVGLALCLMLVQVIIAVEWLVLTVL
Q6PA25|Q6PA25_XENLA    GICWWMMLLMALFLPLVQVVIAILYIVLGLV
Q5T6X5|GPC6A_HUMAN     LKCLYRPILIIFTCTGIQVVICTLWLIFAAP
Q5U9X3|GPC6A_BRARE     LCMLYKPYMIVSVGMGVQIIICTVWLTLYKP
Q8K4Z6|GPC6A_MOUSE     LKCLYRPVPIVLTCTGIQVVICTLWLVLAAP
Q9PW88|GPC6A_CARAU     LRKLYKPYVIVCMCMGLQVTICTLWLTLHRP
Q9Z0R8|TS1R1_RAT       WAQNHGAGLFVIVSSTVHLLICLTWLVMWTP
Q99PG6|TS1R1_MOUSE     WAQNHGAGIFVIVSSTVHLFLCLTWLAMWTP
Q7RTX1|TS1R1_HUMAN     WVQNHGAGLFVMISSAAQLLICLTWLVVWTP
Q9Z0R7|TS1R2_RAT       WMRYHGPYVFVAFITAIKVALVVGNMLATTI
Q925I4|TS1R2_MOUSE     WMRYHGPYVFVAFITAVKVALVAGNMLATTI
Q8TE23|TS1R2_HUMAN     WVRYQGPYVSMAFITVLKMVIVVIGMLATGL
Q7RTX0|TS1R3_HUMAN     CLRGPWAWLVVLLAMLVEVALCTWYLVAFPP
Q717C1|TS1R3_GORGO     CLRGPWAWLVVLLAMLVEVALCTWYLVAFPP
Q717C2|TS1R3_PANTR     CLRGPWAWLVVLLAMLVEVALCTWYLVAFPP
Q923K1|TS1R3_RAT       YLRGPWAWLVVLLATLVEAALCAWYLMAFPP
Q925D8|TS1R3_MOUSE     YLRGLWAWLVVLLATFVEAALCAWYLIAFPP
P41594|MGR5_HUMAN      FMSACAQLVIAFILICIQLGIIVALFIMEPP
Q75QW7|Q75QW7_APIME    YISPASQVCIAAALIALQIVLTLVWMIIEPP
P91685|MGR_DROME       YISPQSQVVITTSLIAIQVLITMIWMVVEPP
Q8CFQ7|Q8CFQ7_MOUSE    FISPTSQLVITFGLTSLQVVGVIAWLGAQPP
```

```
Q863I4|MGR6_RABIT     FISPTSQLVITFSLTSLQVVGVIAWLGAQPP
Q68ED2|MGR7_MOUSE     LISPTSQLAITSSLISVQLLGVFIWFGVDPP
P70579|MGR8_RAT       FISPASQLVITFSLISVQLLGVFVWFVVDPP
P47743|MGR8_MOUSE     FISPASQLVITFSLISVQLLGVFVWFVVDPP
O00222|MGR8_HUMAN     FISPASQLVITFSLISVQLLGVFVWFVVDPP
P35400|MGR7_RAT       LISPTSQLAITSSLISVQLLGVFIWFGVDPP
Q14831|MGR7_HUMAN     LISPTSQLAITSSLISVQLLGVFIWFGVDPP
P35349|MGR6_RAT       FISPTSQLVITFGLTSLQVVGVIAWLGAQPP
O15303|MGR6_HUMAN     FISPTSQLVITFSLTSLQVVGMIAWLGARPP
P31424|MGR5_RAT       FMSACAQLVIAFILICIQLGIIVALFIMEPP
P31423|MGR4_RAT       FISPASQLAITFILISLQLLGICVWFVVDPS
Q14833|MGR4_HUMAN     FISPASQLAITFSLISLQLLGICVWFVVDPS
P31422|MGR3_RAT       FISPSSQVFICLGLILVQIVMVSVWLILETP
Q9QYS2|MGR3_MOUSE     FISPSSQVFICLGLILVQIVMVSVWLILETP
Q14832|MGR3_HUMAN     FISPSSQVFICLGLILVQIVMVSVWLILEAP
P31421|MGR2_RAT       FISPASQVAICLALISGQLLIVAAWLVVEAP
Q14416|MGR2_HUMAN     FISPASQVAICLALISGQLLIVVAWLVVEAP
P23385|MGR1_RAT       FMSAWAQVIIASILISVQLTLVVTLIIMEPP
P97772|MGR1_MOUSE     FMSAWAQVIIASILISVQLTLVVTLIIMEPP
Q13255|MGR1_HUMAN     FMSAWAQVIIASILISVQLTLVVTLIIMEPP
Q90ZF3|Q90ZF3_ONCMA   FMSAWAQLVIAGLLVSVQLTLEVTLIILEPP
Q9V4U4|Q9V4U4_DROME   FISPKSQLVICACLVSVQILINGVWMVIAPS
Q9V4U3|Q9V4U3_DROME   FISPKSQLVICACLVSVQILINGVWMVIAPS
Q7KQS9|Q7KQS9_DROME   FISPKSQLVICACLVSVQILINGVWMVIAPS
Q75QW6|Q75QW6_APIME   FISPRSQLIICSGLVFVQILINGVWMIIDPA
Q70GQ8|Q70GQ8_DROME   FISPKSQLVICACLVSVQILINGVWMVIAPS
Q62916|Q62916_RAT     FISPASQLAITFILISLQLLGICVWFVVDPS
```

TM5

```
BW position                  5.4       5.5       5.6
                             |         |         |
1U19.A                TNNESFVIYMFVVHFIIPLIVIFFCYGQLV
P22815|BOSS_DROME     RWLWGLLAYDFALLCCVGALIPSIYRSQRN
Q24738|BOSS_DROVI     RWLWGLLAYDFLLLCSLVSLVPFIYRSQRN
Q24265|Q24265_DROME   RWLWGLLAYDFALLCCVGALIPSIYRSQRN
Q8MSJ2|Q8MSJ2_DROME   RWLWGLLAYDFALLCCVGALIPSIYRSQRN
Q90WL6|Q90WL6_SPAAU   EGSVMALGFLIGYTCLLAAICFFFAFKSRK
Q8JI04|Q8JI04_SQUAC   EGSLMALGFLIGYTCLLAAICFFFAFKSRK
Q6XAF1|Q6XAF1_SALSA   EGSMMALGFLIGYTCLLAAICFFFAFKSRK
Q6XAF2|Q6XAF2_SALSA   EGSMMALGFLIGYTCLLAAICFFFAFKSRK
P41180|CASR_HUMAN     EGSLMALGFLIGYTCLLAAICFFFAFKSRK
Q80ZA8|Q80ZA8_RAT     EGSLMALGSLIGYTCLLAAICFFFAFKSRK
P48442|CASR_RAT       EGSLMALGSLIGYTCLLAAICFFFAFKSRK
P35384|CASR_BOVIN     EGSLMALGFLIGYTCLLAAICFFFAFKSRK
Q9QY96|CASR_MOUSE     EGSLMALGSLIGYTCLLAAICFFFAFKSRK
Q6XAF3|Q6XAF3_SALSA   EGSMMALGFLIGYTCLLAAICFFFAFKSRK
Q5YEV6|Q5YEV6_OREMO   EGSVMALGFLIGYTCILAAICFFFAFKSRK
O73639|O73639_FUGRU   EGSTIAFSLVLGYIGVLACMCFLLAFLARK
O73638|O73638_FUGRU   LGSPVWFWVVLGYIGLLAVICFILAFLARK
O73637|O73637_FUGRU   VGSTVGFAVLLSYIGLLAILSFLLAFLARN
O73636|O73636_FUGRU   LGSSVGFWAVLGYIGLLAAVCLVLAVLARK
O73635|O73635_FUGRU   EGSVMALGFLIGYTCLLAAICFFFAFKSRK
O75899|GABR2_HUMAN    THMTIWLGIVYAYKGLLMLFGCFLAWETRN
O88871|GABR2_RAT      THMTIWLGIVYAYKGLLMLFGCFLAWETRN
Q9BML6|Q9BML6_DROME   EHMTIFVSIIYAYKGLLLVFGAFLAWETRH
Q9Y133|Q9Y133_DROME   EHMTIFVSIIYAYKGLLLVFGAFLAWETRH
Q8IN24|Q8IN24_DROME   EHMTIFVSIIYAYKGLLLVFGAFLAWETRH
Q5TRW6|Q5TRW6_ANOGA   SKMTIFIGVIYAYKGLLLIFGAFLAWETRH
Q9V3Q9|Q9V3Q9_DROME   QRNSMWLGLVYGFKGLILVFGLFLAYETRS
Q9BML7|Q9BML7_DROME   QRNSMWLGLVYGFKGLILVFGLFLAYETRS
Q8NHA5|Q8NHA5_HUMAN   AIRALGLCIFYGYKGLLLLLGIFLAYETKS
Q8IW08|Q8IW08_HUMAN   RKMNTWLGIFYGYKGLLLLLGIFLAYETKS
```

10. Appendix

```
Q6PGJ2|Q6PGJ2_MOUSE    KKMNTWLGIFYGYKGLLLLLGIFLAYETKS
Q9WV18|GABR1_MOUSE     KKMNTWLGIFYGYKGLLLLLGIFLAYETKS
Q9UBS5|GABR1_HUMAN     RKMNTWLGIFYGYKGLLLLLGIFLAYETKS
Q8K451|GP156_RAT       RYSDVWIALVLGCKGLLLLYGAYLAGLTNH
Q8NFN8|GP156_HUMAN     RYSDVWIALIWGCKGLLLLYGAYLAGLTGH
Q6PCP7|GP156_MOUSE     RYSDVWIALVLGCKGLLLLYGAYLAGLTNH
Q9VPS7|Q9VPS7_DROME    QHTQTWLSVLYAYKGLLLVVGVYMAWETRH
Q8IPW4|Q8IPW4_DROME    QHTQTWLSVLYAYKGLLLVVGVYMAWETRH
Q8MSP9|Q8MSP9_DROME    QHTQTWLSVLYAYKGLLLVVGVYMAWETRH
O96954|O96954_GEOCY    KGQVALFTVLFGYKGLLQVTALILAFNTRK
Q8BHL4|RAI3_MOUSE      PRRNEDFVMLLIYVLVLMVLTFFTSFLVFC
Q8NFJ5|RAI3_HUMAN      PRRNEDFVLLLTYVLFLMALTFLMSSFTFC
Q9JIL6|GPC5D_MOUSE     YQLNVDFVCLLIYVLFLMALTFFVSKATFC
Q9NZD1|GPC5D_HUMAN     CQLNVDFVVLLVYVLFLMALTFFVSKATFC
Q8K3J9|GPC5C_MOUSE     AIANMDFVMALIYVMLLLLTAFLGAWPTLC
Q9NQ84|GPC5C_HUMAN     AIANMDFVMALIYVMLLLLGAFLGAWPALC
Q923Z0|GPC5B_MOUSE     AYEPMDFVMALIYDMVLLAITLAQSLFTLC
Q9NZH0|GPC5B_HUMAN     AYEPMDFVMALIYDMVLLVVTLGLALFTLC
Q6PA25|Q6PA25_XENLA    HQLNQDFVLILIYVFLLMAITFLVSLISLC
Q5T6X5|GPC6A_HUMAN     EGSILAFGTMLGYIAILAFICFIFAFKGKY
Q5U9X3|GPC6A_BRARE     EGFYVMFWLMLGYIALLALFCFTFAYIGRK
Q8K4Z6|GPC6A_MOUSE     EGSALAFGTMLGYITVLAFICFVFAFKGRK
Q9PW88|GPC6A_CARAU     EGSDLMFGLMLGYIVLLALICFTFAYKGRK
Q9Z0R8|TS1R1_RAT       EVNSVGFLLAFTHNILLSISTFVCSYLGKE
Q99PG6|TS1R1_MOUSE     EVNSVGFLVAFAHNILLSISTFVCSYLGKE
Q7RTX1|TS1R1_HUMAN     ETNSLGFILAFLYNGLLSISAFACSYLGKD
Q9Z0R7|TS1R2_RAT       PNYRNGLLFNTSMDLLLSVLGFSFAYMGKE
Q925I4|TS1R2_MOUSE     PNYRNGLLFNTSMDLLLSVLGFSFAYVGKE
Q8TE23|TS1R2_HUMAN     PNYRNSLLFNTSLDLLLSVVGFSFAYMGKE
Q7RTX0|TS1R3_HUMAN     TRSWVSFGLAHATNATLAFLCFLGTFLVRS
Q717C1|TS1R3_GORGO     TRSWVSFGLAHATNATLAFLCFLGTFLVRS
Q717C2|TS1R3_PANTR     TRSWVSFGLAHATNATLAFLCFLGTFLVRS
Q923K1|TS1R3_RAT       MRSWVSLGLVHITNAVLAFLCFLGTFLVQS
Q925D8|TS1R3_MOUSE     VRSWVSLGLVHITNAMLAFLCFLGTFLVQS
P41594|MGR5_HUMAN      NTTNLGVVTPLGYNGLLILSCTFYAFKTRN
Q75QW7|Q75QW7_APIME    NIQDMSFLFSQLYNALLILISTVYAVKTRK
P91685|MGR_DROME       KIQDMSFLFSQLYNMILITICTIYAIKTRK
Q8CFQ7|Q8CFQ7_MOUSE    DMSDLSLIGCLGYSLLLMVTCTVYAIKARG
Q863I4|MGR6_RABIT      DMSDLSLIGCLGYSLLLMVTCTVYAIKARG
Q68ED2|MGR7_MOUSE      DITDLQIICSLGYSILLMVTCTVYAIKTRG
P70579|MGR8_RAT        DISDLSLICSLGYSILLMVTCTVYAIKTRG
P47743|MGR8_MOUSE      DISDLSLICSLGYSILLMVTCTVYAIKTRG
O00222|MGR8_HUMAN      DISDLSLICSLGYSILLMVTCTVYAIKTRG
P35400|MGR7_RAT        DITDLQIICSLGYSILLMVTCTVYAIKTRG
Q14831|MGR7_HUMAN      DITDLQIICSLGYSILLMVTCTVYAIKTRG
P35349|MGR6_RAT        DMSDLSLIGCLGYSLLLMVTCTVYAIKARG
O15303|MGR6_HUMAN      DMSDLSLIGCLGYSLLLMVTCTVYAIKARG
P31424|MGR5_RAT        NTTNLGVVTPLGYNGLLILSCTFYAFKTRN
P31423|MGR4_RAT        DISDLSLICLLGYSMLLMVTCTVYAIKTRG
Q14833|MGR4_HUMAN      DISDLSLICLLGYSMLLMVTCTVYAIKTRG
P31422|MGR3_RAT        NVKDSSMLISLTYDVVLVILCTVYAFKTRK
Q9QYS2|MGR3_MOUSE      NVKDSSMLISLTYDVVLVILCTVYAFKTRK
Q14832|MGR3_HUMAN      NVKDSSMLISLTYDVILVILCTVYAFKTRK
P31421|MGR2_RAT        NHRDASMLGSLAYNVLLIALCTLYAFKTRK
Q14416|MGR2_HUMAN      NHRDASMLGSLAYNVLLIALCTLYAFKTRK
P23385|MGR1_RAT        NTSNLGVVAPVGYNGLLIMSCTYYAFKTRN
P97772|MGR1_MOUSE      NTSNLGVVAPVGYNGLLIMSCTYYAFKTRN
Q13255|MGR1_HUMAN      NTSNLGVVAPLGYNGLLIMSCTYYAFKTRN
Q90ZF3|Q90ZF3_ONCMA    NTSTVGMVAPLGYNGLLIMSCTYYAFKTRN
Q9V4U4|Q9V4U4_DROME    SYIDASYMIAFSYPIFLIVICTVYAVLTRK
Q9V4U3|Q9V4U3_DROME    SYIDASYMIAFSYPIFLIVICTVYAVLTRK
Q7KQS9|Q7KQS9_DROME    SYIDASYMIAFSYPIFLIVICTVYAVLTRK
Q75QW6|Q75QW6_APIME    SYVDASYMIAFAYPIMLIVVCTVYAVLTRK
```

```
Q70GQ8|Q70GQ8_DROME     SYIDASYMIAFSYPIFLIVICTVYAVLTRK
Q62916|Q62916_RAT       DISDLSLICLLGYSMLLMVTCTVYAIKTRG
```

TM6
```
BW position                    6.4       6.5
                                |         |
1U19.A                  EVTRMVIIMVIAFLICWLPYAGVAFYI
P22815|BOSS_DROME       REGILIVIGSVLIMVIWVAWIALSLFG
Q24738|BOSS_DROME       REGILIVIGAVLILIIWSVWIALSMFG
Q24265|Q24265           REGILIVIGSVLIMVIWVAWIALSLFG
Q8MSJ2|Q8MSJ2           REGILIVIGSVLIMVIWVAWIALSLFG
Q90WL6|Q90WL6_SPAAU     TEAKFITFSMLIFFIVWISFIPAYFST
Q8JI04|Q8JI04_SQUAC     NEAKFITFSMLIFFIVWISFIPAYVST
Q6XAF1|Q6XAF1_SALSA     TEAKFITFSMLIFFIVWISFIPAYFST
Q6XAF2|Q6XAF2_SALSA     TEAKFITFSMLIFFIVWISFIPAYFST
P41180|CASR_HUMAN       NEAKFITFSMLIFFIVWISFIPAYAST
Q80ZA8|Q80ZA8_RAT       NEAKFITFSMLIFFIVWISFIPAYAST
P48442|CASR_RAT         NEAKFITFSMLIFFIVWISFIPAYAST
P35384|CASR_BOVIN       NEAKFITFSMLIFFIVWISFIPAYAST
Q9QY96|CASR_MOUSE       NEAKFITFSMLIFFIVWISFIPAYAST
Q6XAF3|Q6XAF3_SALSA     TEAKFITFSMLIFFIVWISFIPAYFST
Q5YEV6|Q5YEV6_OREMO     TEAKFITFSMLIFFIVWISFIPAYFST
O73639|O73639_FUGRU     NEARLIAFSMLIFCAVWVAFVPAYISS
O73638|O73638_FUGRU     NEAKFITFSMLIFCAVWVTFIPAYVSS
O73637|O73637_FUGRU     NEAKLITFSMLIFCAVWVAFVPAYINS
O73636|O73636_FUGRU     NEAKMITFSMLIFCAVWITFIPAYVSS
O73635|O73635_FUGRU     TEAKFITFCMLIFFIVWISFIPAYFST
O75899|GABR2_HUMAN      NDSKYIGMSVYNVGIMCIIGAAVSFLT
O88871|GABR2_RAT        NDSKYIGMSVYNVGIMCIIGAAVSFLT
Q9BML6|Q9BML6_DROME     NDSKHIGFSVYNVFITCLAGAAISLVL
Q9Y133|Q9Y133_DROME     NDSKHIGFSVYNVFITCLAGAAISLVL
Q8IN24|Q8IN24_DROME     NDSKHIGFSVYNVFITCLAGAAISLVL
Q5TRW6|Q5TRW6_ANOGA     NDSKHVGLSVYNCVIMCVMGAAIALVL
Q9V3Q9|Q9V3Q9_DROME     NDSRYVGMSIYNVVVLCLITAPVGMVI
Q9BML7|Q9BML7_DROME     NDSRYVGMSIYNVVVLCLITAPVGMVI
Q8NHA5|Q8NHA5_HUMAN     NDHRAVGMAIYNVAVLCLITAPVTMIL
Q8IW08|Q8IW08_HUMAN     NDHRAVGMAIYNVAVLCLITAPVTMIL
Q6PGJ2|Q6PGJ2_MOUSE     NDHRAVGMAIYNVAVLCLITAPVTMIL
Q9WV18|GABR1_MOUSE      NDHRAVGMAIYNVAVLCLITAPVTMIL
Q9UBS5|GABR1_HUMAN      NDHRAVGMAIYNVAVLCLITAPVTMIL
Q8K451|GP156_RAT        NQSLTIMVGVNLLLLTAGLLFVVTRYL
Q8NFN8|GP156_HUMAN      NQSLTIMVGVNLLVLAAGLLFVVTRYL
Q6PCP7|GP156_MOUSE      NQSLTIMVGVNLLLLTAGLLFVVTRYL
Q9VPS7|Q9VPS7_DROME     NDSQYIGVSVYSVVITSAIVVVLANLI
Q8IPW4|Q8IPW4_DROME     NDSQYIGVSVYSVVITSAIVVVLANLI
Q8MSP9|Q8MSP9_DROME     NDSQYIGVSVYSVVITSAIVVVLANLI
O96954|O96954_GEOCY     DDSKYIAAAIYVTSIVLAVAAISTYTL
Q8BHL4|RAI3_MOUSE       RHGFHICFTSFLSIAIWVAWIVLLLIP
Q8NFJ5|RAI3_HUMAN       RHGAHIYLTMLLSIAIWVAWITLLMLP
Q9JIL6|GPC5D_MOUSE      QHGRLIFATVLVSIIIWVVWISMLLRG
Q9NZD1|GPC5D_HUMAN      QHGRLIFITVLFSIIIWVVWISMLLRG
Q8K3J9|GPC5C_MOUSE      KHGVFVLLTTVISIAIWVVWIVMYTYG
Q9NQ84|GPC5C_HUMAN      KHGVFVLLTTATSVAIWVVWIVMYTYG
Q923Z0|GPC5B_MOUSE      VNGAFILVTTFLSALIWVVWMTMYLFG
Q9NZH0|GPC5B_HUMAN      LNGAFLLITAFLSVLIWVAWMTMYLFG
Q6PA25|Q6PA25_XENLA     RHGAHIYVTMFFSIGIWVAWICMLLRG
Q5T6X5|GPC6A_HUMAN      NEAKFITFGMLIYFIAWITFIPIYATT
Q5U9X3|GPC6A_BRARE      NEAKFITFSMVICLMAWIIFIPIHVTT
Q8K4Z6|GPC6A_MOUSE      NEAKFLTFGMLIYFIAWITFIPVYTTT
Q9PW88|GPC6A_CARAU      NEAKFITFGMLIYLMAWVIFIPVHVTT
Q9Z0R8|TS1R1_RAT        NEAKCVTFSLLLNFVSWIAFFTMASIY
Q99PG6|TS1R1_MOUSE      NEAKCVTFSLLLHFVSWIAFFTMSSIY
```

10. Appendix

```
Q7RTX1|TS1R1_HUMAN      NEAKCVTFSLLFNFVSWIAFFTTASVY
Q9Z0R7|TS1R2_RAT        NEAKFITLSMTFSFTSSISLCTFMSVH
Q925I4|TS1R2_MOUSE      NEAKFITLSMTFSFTSSISLCTFMSVH
Q8TE23|TS1R2_HUMAN      NEAKFITLSMTFYFTSSVSLCTFMSAY
Q7RTX0|TS1R3_HUMAN      NRARGLTFAMLAYFITWVSFVPLLANV
Q717C1|TS1R3_GORGO      NRARGLTFAMLAYFITWVSFVPLLANV
Q717C2|TS1R3_PANTR      NRARGLTFAMLAYFITWVSFVPLLANV
Q923K1|TS1R3_RAT        NRARGLTFAMLAYFIIWVSFVPLLANV
Q925D8|TS1R3_MOUSE      NRARGLTFAMLAYFITWVSFVPLLANV
P41594|MGR5_HUMAN       NEAKYIAFTMYTTCIIWLAFVPIYFG.
Q75QW7|Q75QW7_APIME     NESKFIGFTMYTTCIIWLAFVPIYFG.
P91685|MGR_DROME        NESKFIGFTMYTTCIIWLAFVPIYFG.
Q8CFQ7|Q8CFQ7_MOUSE     NEAKPIGFTMYTTCIIWLAFVPIFFG.
Q863I4|MGR6_RABIT       NEAKPIGFTMYTTCIVWLAFVPIFFG.
Q68ED2|MGR7_MOUSE       NEAKPIGFTMYTTCIVWLAFIPIFFG.
P70579|MGR8_RAT         NEAKPIGFTMYTTCIIWLAFIPIFFG.
P47743|MGR8_MOUSE       NEAKPIGFTMYTTCIIWLAFIPIFFG.
O00222|MGR8_HUMAN       NEAKPIGFTMYTTCIIWLAFIPIFFG.
P35400|MGR7_RAT         NEAKPIGFTMYTTCIVWLAFIPIFFG.
Q14831|MGR7_HUMAN       NEAKPIGFTMYTTCIVWLAFIPIFFG.
P35349|MGR6_RAT         NEAKPIGFTMYTTCIIWLAFVPIFFG.
O15303|MGR6_HUMAN       NEAKPIGFTMYTTCIIWLAFVPIFFG.
P31424|MGR5_RAT         NEAKYIAFTMYTTCIIWLAFVPIYFG.
P31423|MGR4_RAT         NEAKPIGFTMYTTCIVWLAFIPIFFG.
Q14833|MGR4_HUMAN       NEAKPIGFTMYTTCIVWLAFIPIFFG.
P31422|MGR3_RAT         NEAKFIGFTMYTTCIIWLAFLPIFYV.
Q9QYS2|MGR3_MOUSE       NEAKFIGFTMYTTCIIWLAFLPIFYV.
Q14832|MGR3_HUMAN       NEAKFIGFTMYTTCIIWLAFLPIFYV.
P31421|MGR2_RAT         NEAKFIGFTMYTTCIIWLAFLPIFYV.
Q14416|MGR2_HUMAN       NEAKFIGFTMYTTCIIWLAFLPIFYV.
P23385|MGR1_RAT         NEAKYIAFTMYTTCIIWLAFVPIYFG.
P97772|MGR1_MOUSE       NEAKYIAFTMYTTCIIWLAFVPIYFG.
Q13255|MGR1_HUMAN       NEAKYIAFTMYTTCIIWLAFVPIYFG.
Q90ZF3|Q90ZF3_ONCMA     NEAKYIAFTMYTTCIIWLAFVPIYFG.
Q9V4U4|Q9V4U4_DROME     NESKHIGFTMYTTCVIWLAFVPLYFG.
Q9V4U3|Q9V4U3_DROME     NESKHIGFTMYTTCVIWLAFVPLYFG.
Q7KQS9|Q7KQS9_DROME     NESKHIGFTMYTTCVIWLAFVPLYFG.
Q75QW6|Q75QW6_APIME     NESKHIGFTMYTTCVIWLAFVPLYFG.
Q70GQ8|Q70GQ8_DROME     NESKHIGFTMYTTCVIWLAFVPLYFG.
Q62916|Q62916_RAT       NEAKPIGFTMYTTCIVWLAFIPIFFG.
```

TM7

```
BW position                       7.4       7.5
                                   |         |
1U19.A                  PIFMTIPAFFAKTSAVYNPVIYIMMNKQ
P22815|BOSS_DROME       IPLGLQASGWAVLVGILIPRTFLIVRGI
Q24738|BOSS_DROVI       IPLGMQASGWAVLVGILIPRTFLIVRGI
Q24265|Q24265_DROME     IPLGLQASGWAVLVGILIPRTFLIVRGI
Q8MSJ2|Q8MSJ2_DROME     IPLGLQASGWAVLVGILIPRTFLIVRGI
Q90WL6|Q90WL6_SPAAU     EAIAILASSFGMLACIFFNKVYIILFKP
Q8JI04|Q8JI04_SQUAC     EVIAILASSFGLLGCIYFNKCYIILFKP
Q6XAF1|Q6XAF1_SALSA     EVIAILASSFGLLACIFFNKVYIILFKP
Q6XAF2|Q6XAF2_SALSA     EVIAILASSFGLLACIFFNKVYIIHQP.
P41180|CASR_HUMAN       EVIAILAASFGLLACIFFNKIYIILFKP
Q80ZA8|Q80ZA8_RAT       EVIAILAASFGLLACIFFNKVYIILFKP
P48442|CASR_RAT         EVIAILAASFGLLACIFFNKVYIILFKP
P35384|CASR_BOVIN       EVIAILAASFGLLACIFFNKVYIILFKP
Q9QY96|CASR_MOUSE       EVIAILAASFGLLACIFFNKVYIILFKP
Q6XAF3|Q6XAF3_SALSA     EVIAILASSFGLLACIFFNKVYIILFKP
Q5YEV6|Q5YEV6_OREMO     EVIAILASSFGMLACIFFNKVYIILFKP
O73639|O73639_FUGRU     EIFAILASSYGLLGCIFAPKCYIILMKS
O73638|O73638_FUGRU     EIFAILASSFGLLFCIFAPKCYILILKP
```

174

```
O73637|O73637_FUGRU    EVFAILTSSFGLLVALFGPKCYIILFRP
O73636|O73636_FUGRU    EIFAILASSFGLILCIFAPKCFIILFKP
O73635|O73635_FUGRU    EAIAILASSYGMLACIFFNKVYIILFKP
O75899|GABR2_HUMAN     VALVIIFCSTITLCLVFVPKLITLRTNP
O88871|GABR2_RAT       VALVIIFCSTITLCLVFVPKLITLRTNP
Q9BML6|Q9BML6_DROME    LSFFIIFCTTATLCLVFVPKLVELKRNP
Q9Y133|Q9Y133_DROME    LSFFIIFCTTATLCLVFVPKLVELKRNP
Q8IN24|Q8IN24_DROME    LSFFIIFCTTATLCLVFVPKLVELKRNP
Q5TRW6|Q5TRW6_ANOGA    ISVFIIFCTTATLCLVFVPKLVELKRNP
Q9V3Q9|Q9V3Q9_DROME    VALAVIFCCFLSMLLIFVPKVIEVIRHP
Q9BML7|Q9BML7_DROME    VALAVIFCCFLSMLLIFVPKVIEVIRHP
Q8NHA5|Q8NHA5_HUMAN    ASLAIVFSSYITLVVLFVPKMRRLITRG
Q8IW08|Q8IW08_HUMAN    ASLAIVFSSYITLVVLFVPKMRRLITRG
Q6PGJ2|Q6PGJ2_MOUSE    ASLAIVFSSYITLVVLFVPKMRRLITRG
Q9WV18|GABR1_MOUSE     ASLAIVFSSYITLVVLFVPKMRRLITRG
Q9UBS5|GABR1_HUMAN     ASLAIVFSSYITLVVLFVPKMRRLITRG
Q8K451|GP156_RAT       TSGGIFVCTTTVNCCVFLPQLRQRKAFE
Q8NFN8|GP156_HUMAN     TSGGIFVCTTTINCFIFIPQLKQWKAFE
Q6PCP7|GP156_MOUSE     TSGGIFVCTTTVNCCVFIPQLKQWKAFE
Q9VPS7|Q9VPS7_DROME    ITALILTSTTATLCLLFIPKLHDIWARN
Q8IPW4|Q8IPW4_DROME    ITALILTSTTATLCLLFIPKLHDIWARN
Q8MSP9|Q8MSP9_DROME    ITALILTSTTATLLSAFHPKTP......
O96954|O96954_GEOCY    VGIGFLLGTTMILGLVFVPRMVGLYQDP
Q8BHL4|RAI3_MOUSE      LSTALVANGWVFLAFYILPEFRQLPRQR
Q8NFJ5|RAI3_HUMAN      LSSALAANGWVFLLAYVSPEFWLLTKQR
Q9JIL6|GPC5D_MOUSE     ICIGLVTNAWVFLLIYIIPELSILYRSC
Q9NZD1|GPC5D_HUMAN     VCIALVTNAWVFLLLYIVPELCILYRSC
Q8K3J9|GPC5C_MOUSE     LAIALAANAWTFVLFYVIPEVSQVTKPS
Q9NQ84|GPC5C_HUMAN     LAIALAANAWAFVLFYVIPEVSQVTKSS
Q923Z0|GPC5B_MOUSE     LAITLAASGWVFVIFHAIPEIHYTLLPP
Q9NZH0|GPC5B_HUMAN     LAITLAASGWVFVIFHAIPEIHCTLLPA
Q6PA25|Q6PA25_XENLA    LSIALVANGWVFLMMYMVPELCLMTRCQ
Q5T6X5|GPC6A_HUMAN     EIIVILISNYGILYCTFIPKCYVIICKQ
Q5U9X3|GPC6A_BRARE     EMVVILISNYGILSCHFLPKSYIILFKK
Q8K4Z6|GPC6A_MOUSE     EIIVILISNYGILCCIFFPKCYIILCKQ
Q9PW88|GPC6A_CARAU     EVVVILISNYGILSCHFLPKCYIIIFKK
Q9Z0R8|TS1R1_RAT       NVLAGLTTLSGGFSGYFLPKCYVILCRP
Q99PG6|TS1R1_MOUSE     NVLAGLATLSGGFSGYFLPKCYVILCRP
Q7RTX1|TS1R1_HUMAN     NMMAGLSSLSSGFGGYFLPKCYVILCRP
Q9Z0R7|TS1R2_RAT       DLLVTVLNFLAIGLGYFGPKCYMILFYP
Q925I4|TS1R2_MOUSE     DLLVTVLNFLAIGLGYFGPKCYMILFYP
Q8TE23|TS1R2_HUMAN     DLLVTVLNLLAISLGYFGPKCYMILFYP
Q7RTX0|TS1R3_HUMAN     QMGALLLCVLGILAAFHLPRCYLLMRQP
Q717C1|TS1R3_GORGO     QMGALLLCVLGILAAFHLPRCYLLIRQP
Q717C2|TS1R3_PANTR     QMGALLLCVLGILAAFHLPRCYLLMWQP
Q923K1|TS1R3_RAT       QMGAILFCALGILATFHLPKCYVLLWLP
Q925D8|TS1R3_MOUSE     QMGAILVCALGILVTFHLPKCYVLLWLP
P41594|MGR5_HUMAN      MCFSVSLSATVALGCMFVPKVYIILAKP
Q75QW7|Q75QW7_APIME    LCVAISLSATVTLVCLYSPKIYIILFQP
P91685|MGR_DROME       LCISISLSASVALVCLYSPKVYILVFHP
Q8CFQ7|Q8CFQ7_MOUSE    LTVSLSLSASVSLGMLYVPKTYVILFHP
Q863I4|MGR6_RABIT      LTVSLSLSASVSLGMLYVPKTYVILFHP
Q68ED2|MGR7_MOUSE      LTISMNLSASVALGMLYMPKVYIIIFHP
P70579|MGR8_RAT        LTVSMSLSASVSLGMLYMPKVYIIIFHP
P47743|MGR8_MOUSE      LTVSMSLSASVSLGMLYMPKVYIIIFHP
O00222|MGR8_HUMAN      LTVSMSLSASVSLGMLYMPKVYIIIFHP
P35400|MGR7_RAT        LTISMNLSASVALGMLYMPKVYIIIFHP
Q14831|MGR7_HUMAN      LTISMNLSASVALGMLYMPKVYIIIFHP
P35349|MGR6_RAT        LTVSLSLSASVSLGMLYVPKTYVILFHP
O15303|MGR6_HUMAN      LTVSLSLSASVSLGMLYVPKTYVILFHP
P31424|MGR5_RAT        MCFSVSLSATVALGCMFVPKVYIILAKP
P31423|MGR4_RAT        LTVSVSLSASVSLGMLYMPKVYIILFHP
Q14833|MGR4_HUMAN      LTVSVSLSASVSLGMLYMPKVYIILFHP
```

10. Appendix

```
P31422|MGR3_RAT        MCISVSLSGFVVLGCLFAPKVHIVLFQP
Q9QYS2|MGR3_MOUSE      MCISVSLSGFVVLGCLFAPKVHIVLFQP
Q14832|MGR3_HUMAN      MCISVSLSGFVVLGCLFAPKVHIILFQP
P31421|MGR2_RAT        MCVSVSLSGSVVLGCLFAPKLHIILFQP
Q14416|MGR2_HUMAN      MCVSVSLSGSVVLGCLFAPKLHIILFQP
P23385|MGR1_RAT        TCFAVSLSVTVALGCMFTPKMYIIIAKP
P97772|MGR1_MOUSE      TCFAVSLSVTVALGCMFTPKMYIIIAKP
Q13255|MGR1_HUMAN      TCFAVSLSVTVALGCMFTPKMYIIIAKP
Q90ZF3|Q90ZF3_ONCMA    TSFSVSLSVTVALGCMFSPKIYIILAKP
Q9V4U4|Q9V4U4_DROME    MSVTISLSASVTIACLFSPKLYIILIRP
Q9V4U3|Q9V4U3_DROME    MSVTISLSASVTIACLFSPKLYIILIRP
Q7KQS9|Q7KQS9_DROME    MSVTISLSASVTIACLFSPKLYIILIRP
Q75QW6|Q75QW6_APIME    MSVTISLSASVTIACLFSPKLYIILIRP
Q70GQ8|Q70GQ8_DROME    MSVTISLSASVTIACLFSPKLYIILIRP
Q62916|Q62916_RAT      LTVSVSLSASVSLGMLYMPKVYIILFHI
```

Figure A 1: Multiple sequence alignment of 96 family C GPCRs. Positions are assigned numbering according to Ballesteros-Weinstein (BW) numbering scheme. Sequences are named using Swiss-Prot accession codes.

10.4 Ligand data collection

Table A 1: Ligand data collection of 1240 mGluR binding ligands. O represents the organism, the ligand was tested, and sel, whether the ligand is selective (+) or not selective (-) for the given receptor. If nothing is assigned, then data is missing.

SMILES	ligand type	receptor	pIC_{50}	O	sel
O(C)c1cc(ccc1)C#Cc1ncc(cc1)C	partial	mGluR5	6.84	R	
Brc1cc(cnc1)C#Cc1ncc(cc1)C	partial	mGluR5	6.74	R	
O(C)c1cc(ccc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.89	R	
FC(F)(F)c1cc(ccc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.78	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1cc([N+](=O)[O-])ccc1	PAM	mGluR5	5.65	R	
O(C)c1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.38	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc(cc1)C#N	PAM	mGluR5	5.81	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc([N+](=O)[O-])cc1	PAM	mGluR5	6.60	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc(cc1)C(=O)N	PAM	mGluR5	5.00	R	
Oc1ccccc1C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.00	R / H	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccccc1C	PAM	mGluR5	5.00	R / H	
OC(=O)c1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.00	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1cc([N+](=O)[O-])cc([N+](=O)[O-])c1	PAM	mGluR5	7.19	R	
O(C)c1cc(ccc1OC)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.00	R	
O(C)c1cc(cc(OC)c1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.00	R	-
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1cc(C)c(cc1)C	PAM	mGluR5	6.37	R	
Clc1cc(ccc1Cl)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.80	R	
Clc1ccccc1-c1nn(-c2ccccc2)c(NC(=O)c2ccccc2)c1	PAM	mGluR5	5.00	R	
Fc1ccccc1-n1nc(cc1NC(=O)c1ccccc1)-c1ccccc1	PAM	mGluR5	5.57	R	
Clc1ccccc1-n1nc(cc1NC(=O)c1ccccc1)-c1ccccc1	PAM	mGluR5	6.30	R	
Brc1ccccc1-n1nc(cc1NC(=O)c1ccccc1)-c1ccccc1	PAM	mGluR5	5.59	R	
Brc1cc(-n2nc(cc2NC(=O)c2cc(ccc2)C#N)-c2ccccc2)ccc1	PAM	mGluR5	5.00	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1cc(ccc1)C#N)c1ccccc1	PAM	mGluR5	5.00	R	
Brc1ccccc1-n1nc(cc1NC(=O)c1cc(ccc1)C#N)-c1ccccc1	PAM	mGluR5	6.64	R	
O=C(N(C(=O)c1ccc(cc1)C#N)c1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc(cc1)C#N	PAM	mGluR5	5.52	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccccc1[N+](=O)[O-]	PAM	mGluR5	5.00	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc(nc1)C	PAM	mGluR5	5.46	R	-
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccnc1	PAM	mGluR5	5.00	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)C=1CCCCC=1	PAM	mGluR5	5.35	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)C1CCCCC1	PAM	mGluR5	4.99	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)C1CCCC1	PAM	mGluR5	4.94	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)CC1CCCC1	PAM	mGluR5	5.00	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)\C=C\c1ccc([N+](=O)[O-	PAM	mGluR5	5.74	R	-

176

])cc1

n1n(-c2ccccc2)c(-n2nnc(c2)-c2ccccc2)cc1-c1ccccc1	PAM	mGluR5	5.23	R	
O=C(Nc1n(nc(c1)-c1ccccc1)Cc1ncccc1	PAM	mGluR5	5.00	H	
Fc1ccccc1\C=N\N=C\c1ccccc1F	PAM	mGluR5	4.85	H	
N(C(c1ccccc1)c1ccccc1)CCNC(c1ccccc1)c1ccccc1	PAM	mGluR1	7.82	R	
Fc1ccccc1-n1nc(cc1NC(=O)c1ccc([N+](=O)[O-])cc1)-c1ccccc1	PAM	mGluR5	6.80	R	
Clc1ccccc1-n1nc(cc1NC(=O)c1ccc([N+](=O)[O-])cc1)-c1ccccc1	PAM	mGluR5	6.96	R	
Brc1ccccc1-n1nc(cc1NC(=O)c1ccc([N+](=O)[O-])cc1)-c1ccccc1	PAM	mGluR5	6.77	R	-
Brc1ccccc1-n1nc(cc1NC(=O)c1cc([N+](=O)[O-])ccc1)-c1ccccc1	PAM	mGluR5	6.64	R	
Fc1ccccc1-c1nn(-c2ccccc2)c(NC(=O)c2ccc([N+](=O)[O-])cc2)c1	PAM	mGluR5	6.80	R	
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)CC(C)C)c(O)c1CCC	PAM	mGluR2	6.38	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)C)c(O)c1CCC	PAM	mGluR2	6.46	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)C)c(O)c1CCCCC	PAM	mGluR2	5.00	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)C)c(O)c1C	PAM	mGluR2	5.47	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)CC)c(O)c1C	PAM	mGluR2	6.42	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)CCC)c(O)c1C	PAM	mGluR2	6.52	H	+
O(CCCCOc1ccc(cc1)-c1nn[nH]n1)c1ccc(C(=O)CC(C)C)c(O)c1C	PAM	mGluR2	6.64	H	+
Clc1cc(Cl)ccc1CN[C@H]1CCN(C1)c1ncc(Cl)cn1	PAM	mGluR4	7.48	R	
o1nc(nc1C(C)(C)C)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR7	7.47		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)C(C)(C)C	PAM	mGluR1	7.47	R	
O(CCC)C(=O)NC(=O)C(c1ccccc1)c1cccccc1	PAM	mGluR1	7.47	R	
O(CCCC)C(=O)NC(=O)C(c1ccccc1)c1ccccc1	PAM	mGluR1	7.47	R	
O(CCCCC)C(=O)NC(=O)C(c1ccccc1)c1ccccc1	PAM	mGluR1	7.30	R	
O(C(C)C)C(=O)NC(=O)C(c1ccccc1)c1ccccc1	PAM	mGluR1	6.03	R	
O1c2c(cccc2)C(c2c1cccc2)C(=O)NC(OCCC)=O	PAM	mGluR1	7.47	R	
O1c2c(cccc2)C(c2c1cccc2)C(=O)NC(OCCCC)=O	PAM	mGluR1	7.47	R	
S1c2c(cccc2)C(c2c1cccc2)C(=O)NC(OCC)=O	PAM	mGluR1	6.85	R	
Fc1ccc(cc1)C(=O)N1C[C@@](CCC1)(C)c1onc(n1)-c1ncccc1	PAM	mGluR5	7.41	H	
Fc1ccccc1-c1nc(on1)[C@]1(CCCN(C1)C(=O)c1ccc(F)cc1)C	PAM	mGluR5	7.41	H	
O=C(Nc1n(ncc1-c1ccccc1)-c1ccccc1)c1ccc([N+](=O)[O-])cc1	PAM	mGluR5	5.00	R	-
O=C(N1N(\C(=N\C(=O)c2ccc(cc2)C#N)\C=C1c1ccccc1)c1ccccc1)c1ccc(cc1)C#N	PAM	mGluR5	5.28	R	
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)C(=O)c1ccccc1)CC	PAM	mGluR2	5.38		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)C(=O)c1ccccc1)CC	PAM	mGluR2	5.96		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)C(=C)c1ccccc1)CC	PAM	mGluR2	6.11		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)C(=C)c1ccccc1)CC	PAM	mGluR2	6.70		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)[C@@H](C)c1ccccc1)CC	PAM	mGluR2	5.85		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)[C@@H](C)c1ccccc1)CC	PAM	mGluR2	7.14		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)Cc1ccccc1)CC	PAM	mGluR2	5.80		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)Cc1ccccc1)CC	PAM	mGluR2	6.51		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)[C@@H](F)c1ccccc1)CC	PAM	mGluR2	5.00		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)[C@@H](F)c1ccccc1)CC	PAM	mGluR2	6.74		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)COc1ccccc1)CC	PAM	mGluR2	5.68		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)COc1ccccc1)CC	PAM	mGluR2	6.80		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(Oc2ccccc2)ccc1)CC	PAM	mGluR2	6.44		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OCc2ccccc2)ccc1)CC(F)(F)F	PAM	mGluR2	6.21		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OCc2ccccc2)ccc1)CC	PAM	mGluR2	6.36		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC(F)(F)F)ccc1)CC(F)(F)F	PAM	mGluR2	6.85		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC(F)(F)F)ccc1)CC	PAM	mGluR2	5.80		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC)ccc1)CC(F)(F)F	PAM	mGluR2	5.68		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OCC(F)(F)F)ccc1)CC(F)(F)F	PAM	mGluR2	7.00		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC(C)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OCC(C)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(O[C@@H](CC)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC(CC)CC)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC[C@@H](CC)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC2CCC2)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC2CCCC2)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC2CCCCC2)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OCC(OCC)=O)ccc1)CC(F)(F)F	PAM	mGluR2	6.19		+

10. Appendix

SMILES					
S(=O)(=O)(N(Cc1cccnc1)c1cc(O[C@@H](C(OCC)=O)C)ccc1)CC(F)(F)F	PAM	mGluR2	6.64		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(OC(C(OCC)=O)(C)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.06		+
S(=O)(=O)(N(Cc1cccnc1)c1ccc(cc1)[C@@H](O)c1ccccc1)CC	PAM	mGluR2	5.00		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(ccc1)[C@@H](O)c1ccccc1)CC	PAM	mGluR2	5.60		+
FC(F)(F)Oc1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.58	R	+
O(C)c1cc(ccc1)-c1nn(-c2ccccc2)c(NC(=O)c2ccccc2)c1	PAM	mGluR5	5.49	R	
O(C)c1cc(-n2nc(cc2NC(=O)c2ccccc2)-c2ccccc2)ccc1	PAM	mGluR5	5.68	R	
S(=O)(=O)(N(Cc1cccnc1)c1ccc(Oc2ccccc2)cc1)CC	PAM	mGluR2	5.82		+
S(=O)(=O)(N(Cc1cccnc1)c1cc(O[C@@H](CCC)C)ccc1)CC(F)(F)F	PAM	mGluR2	7.27		+
S(=O)(=O)(N1CCC[C@@H]1c1ccc(F)cc1)c1ccc(cc1)C	PAM	mGluR1	7.22	R	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ncccc1)c1ccccc1	PAM	mGluR5	5.00	R	
Fc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	5.82	H	
O(C)c1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	5.85	H	
Clc1cc(CN2C(=O)c3c(cc(C)c(c3)C)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	5.68	H	
Clc1cc(CN2C(=O)c3c(ccc(F)c3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	6.48	H	
Clc1cc(CN2C(=O)c3c(ccc(OCC)c3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	6.55	H	
Clc1cc2c(cc1Cl)C(=O)N(Cc1cc(Cl)ccc1NC(=O)c1ncccc1)C2=O	PAM	mGluR5	5.55	H	
Clc1cc(CN2C(=O)c3c(ccc(c3)C)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	5.00	H	
Clc1cc(CN2C(=O)c3c(cccc3F)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	6.21	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2cccnc2)cc1	PAM	mGluR5	4.46	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ccncc2)cc1	PAM	mGluR5	5.00	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2occc2)cc1	PAM	mGluR5	5.64	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2nsnc2)cc1	PAM	mGluR5	5.11	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2nc3SC=Cn3c2)cc1	PAM	mGluR5	5.60	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncccc2O)cc1	PAM	mGluR5	6.73	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncsc2)cc1	PAM	mGluR5	5.72	H	
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2cc(OC)ccc2)cc1	PAM	mGluR5	5.82	H	
Brc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	6.12	H	
O=C1N(Cc2ccccc2NC(=O)c2ncccc2)C(=O)c2c1cccc2	PAM	mGluR5	7.10	R / H	
Fc1ccc(cc1)-c1nc(on1)[C@]1(CCCN(C1)C(=O)c1ccc(F)cc1)C	PAM	mGluR5	6.37	H	
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1cc(ccc1)C#N	PAM	mGluR5	5.42	R	+
O(CC)C(=O)NC(=O)C(c1ccccc1)c1ccccc1	PAM	mGluR1	6.77		
O1c2c(cccc2)C(c2c1cccc2)C(=O)NC(OCC)=O	PAM	mGluR1	7.47		
S(=O)(=O)(N1CCC[C@@H]1c1ccc(cc1)C)c1ccc(cc1)C	PAM	mGluR1	6.70		
o1nc(nc1C)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.28		
o1nc(nc1CC)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.47		
o1nc(nc1C(C)C)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.47		
o1nc(nc1CCC)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.47		
o1nc(nc1C1CC1)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.47		
o1nc(nc1CC(C)C)NC(=O)C1c2c(Oc3c1cccc3)cccc2	PAM	mGluR1	7.47		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)C	PAM	mGluR1	6.74		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)CC	PAM	mGluR1	7.19		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)CCC	PAM	mGluR1	7.47		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)C(C)C	PAM	mGluR1	7.35		
O1c2c(cccc2)C(c2c1cccc2)C(=O)Nc1nn(nn1)CC(C)C	PAM	mGluR1	7.47		
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ccccc2O)cc1	PAM	mGluR5	6.82		
Clc1cc(CN2C(=O)c3c(cccc3)C2=O)c(NC(=O)c2ncccc2)cc1	PAM	mGluR5	6.60		
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccccc1	PAM	mGluR5	5.38		
Clc1ccccc1C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	5.21		
Clc1cc(ccc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.39		
Clc1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.77		
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1cc(ccc1)C	PAM	mGluR5	7.00		
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1ccc(cc1)C	PAM	mGluR5	6.11		
Fc1ccccc1C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	7.00		
Fc1cc(ccc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	7.15		
Fc1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	7.30		
Fc1cc(ccc1F)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	7.70		
FC(F)(F)c1ccc(cc1)C(=O)Nc1n(nc(c1)-c1ccccc1)-c1ccccc1	PAM	mGluR5	6.00		
O=C(Nc1n(nc(c1)-c1ccccc1)-c1ccccc1)c1n[nH]cc1	PAM	mGluR5	8.08		

178

Oc1cc(ccc1)\C=N\N=C\c1cc(O)ccc1	neutral	mGluR5	5.00	H
n1cc(ccc1C#Cc1ccccc1)C	neutral	mGluR5	6.41	
O=C(Nc1nc(ccc1)C)c1nc(ccc1CNc1cccnc1)C	NAM	mGluR5	7.44	
O(c1ccccc1-n1nc(nn1)-c1ccccc1)c1cccnc1	NAM	mGluR5	5.80	
s1cc(nc1C)C#Cc1cc(F)ccc1	NAM	mGluR5	7.64	R
s1cc(nc1C)C#Cc1ccc(F)cc1	NAM	mGluR5	5.53	R
s1cc(nc1C)C#Cc1cc(OC)ccc1	NAM	mGluR5	7.31	R
s1cc(nc1C)C#Cc1ccccc1F	NAM	mGluR5	6.85	R
s1cc(nc1C)C#Cc1ccccc1OC	NAM	mGluR5	7.44	R
s1cc(nc1C)C#Cc1cc(ccc1)C	NAM	mGluR5	7.08	R
Clc1cc(ccc1)C#Cc1nc(sc1)C	NAM	mGluR5	6.89	R
s1cc(nc1C)C#Cc1cc(ccc1)C(F)(F)F	NAM	mGluR5	6.38	R
s1cc(nc1C)C#Cc1cc(OC(F)(F)F)ccc1	NAM	mGluR5	6.58	R
s1cc(nc1C)C#Cc1cc(ccc1)C#N	NAM	mGluR5	7.89	R
s1cc(nc1C)C#Cc1cc(NC(=O)C)ccc1	NAM	mGluR5	5.69	R
s1cc(nc1C)C#Cc1cc(F)cc(F)c1	NAM	mGluR5	7.60	R
s1c(C)c(nc1C)C#Cc1cccnc1	NAM	mGluR5	6.16	R
s1c(CC)c(nc1C)C#Cc1cccnc1	NAM	mGluR5	5.91	R
s1cc(nc1C)C#Cc1ccc(OC)nc1	NAM	mGluR5	6.37	R
s1cc(nc1C)C#Cc1ncc(F)cc1	NAM	mGluR5	5.96	R
Brc1cc(cnc1)C#Cc1nc(sc1)C	NAM	mGluR5	7.22	R
s1cc(nc1C)C#Cc1ccc(F)nc1	NAM	mGluR5	6.08	R
s1cc(nc1C)C#Cc1cncnc1	NAM	mGluR5	6.02	R
s1cc(nc1C)C#Cc1nccnc1	NAM	mGluR5	5.74	R
s1cccc1C#Cc1nc(sc1)C	NAM	mGluR5	6.32	R
s1cc(nc1C)C#Cc1ccsc1	NAM	mGluR5	6.47	R
s1cc(nc1C)C#Cc1cc(O)ccc1	NAM	mGluR5	6.18	R
s1cc(nc1C)C#Cc1cc(ccc1)C(=O)N	NAM	mGluR5	6.00	R
s1cc(nc1N)C#Cc1cc(F)ccc1	NAM	mGluR5	7.10	R
s1cc(nc1N)C#Cc1cccnc1	NAM	mGluR5	6.10	R
s1cc(nc1NC(=O)C)C#Cc1cccnc1	NAM	mGluR5	5.72	R
Brc1scc(n1)C#Cc1cc(F)ccc1	NAM	mGluR5	6.28	R
s1cc(nc1C)C#Cc1cc(cnc1)-c1ccc(F)cc1	NAM	mGluR5	6.79	R
s1cc(nc1C)C#Cc1cc(cnc1)-c1ccc(OC)cc1	NAM	mGluR5	6.65	R
s1cc(nc1C)C#Cc1cc(cnc1)C#CCO	NAM	mGluR5	7.16	R
s1cc(nc1C)C#Cc1cc(cnc1)C#C	NAM	mGluR5	7.72	R
s1cc(nc1C)C#Cc1cc(cnc1)C=C	NAM	mGluR5	7.89	R
s1cc(nc1C)C#Cc1ccc(OS(=O)(=O)C)nc1	NAM	mGluR5	5.63	R
s1cc(nc1C)C#Cc1ccc(nc1)-c1ccc(F)cc1	NAM	mGluR5	7.89	H
s1cc(nc1C)C#Cc1ccc(nc1)C#C	NAM	mGluR5	5.53	R
s1cc(nc1C)C#Cc1cccnc1OC	NAM	mGluR5	6.64	R
Clc1ncccc1C#Cc1nc(sc1)C	NAM	mGluR5	7.34	R
s1cc(nc1C)\C=C\c1cccnc1	NAM	mGluR5	5.51	R
Clc1ccc(Sc2nc(nc(c2)C)-c2ncccc2)cc1	NAM	mGluR5	5.90	
Clc1ccc(cc1)[C@@H](O)c1nc(ncc1)-c1ncccc1	NAM	mGluR5	5.52	
Clc1cc(Sc2nc(ncc2)-c2nc(ccc2)C)ccc1Cl	NAM	mGluR5	6.41	
Clc1cc(Sc2nc(ncc2)-c2ncc(F)cc2)ccc1Cl	NAM	mGluR5	7.10	
Clc1cc(Sc2cc(ncc2)-c2nccccc2)ccc1Cl	NAM	mGluR5	5.63	
Clc1ccc(Sc2nc(ncc2)-c2ncccc2)cc1	NAM	mGluR5	6.49	
Clc1cc(Sc2nc(ncc2)-c2ncccc2)ccc1	NAM	mGluR5	6.21	
S(c1cc(ccc1)C)c1nc(ncc1)-c1ncccc1	NAM	mGluR5	5.64	
S(c1ccc(cc1)C(F)(F)F)c1nc(ncc1)-c1ncccc1	NAM	mGluR5	6.68	
S(c1cc(F)c(ccc1)c1nc(ncc1)-c1ncccc1	NAM	mGluR5	5.82	
Clc1cc(Sc2nc(ncc2)-c2ncccc2)ccc1Cl	NAM	mGluR5	7.05	
Brc1cc(Sc2nc(ncc2)-c2ncccc2)ccc1	NAM	mGluR5	5.54	
s1cc(nc1C)C#Cc1ccccc1C	NAM	mGluR5	7.82	
s1cc(nc1C)C#Cc1ccc(cc1)C	NAM	mGluR5	6.07	
s1cc(nc1C)C#Cc1ccccc1-c1ccccc1	NAM	mGluR5	6.37	
s1cc(nc1C)C#Cc1cc(ccc1)-c1ccccc1	NAM	mGluR5	7.39	
s1cc(nc1C)C#Cc1ccc(cc1)-c1ccccc1	NAM	mGluR5	7.66	
s1cc(nc1C)C#Cc1cc(cnc1)-c1ccccc1	NAM	mGluR5	7.89	
s1cc(nc1C)C#Cc1cc(cnc1)-c1ncccc1	NAM	mGluR5	7.66	
s1cc(nc1C)C#Cc1cc(cnc1)-c1cccnc1	NAM	mGluR5	6.23	R / H

10. Appendix

s1cc(nc1C)C#Cc1ccc(nc1)-c1ccccc1	NAM	mGluR5	7.89		
s1cc(nc1C)C#Cc1ccc(nc1)-c1ncccc1	NAM	mGluR5	7.80	H	
s1cc(nc1C)C#Cc1ccc(nc1)-c1cccnc1	NAM	mGluR5	7.89	H	
s1cc(nc1C)C#Cc1ccc(nc1)-c1ccncc1	NAM	mGluR5	7.54		
O(C)c1cc(NC(=O)\N=C\2/NC(=O)CN/2C)ccc1	NAM	mGluR5	5.55	H	
O=C1N\C(=N/C(=O)Nc2nccc(c2)C)\N(C1)C	NAM	mGluR5	5.51	H	
Clc1cc(NC(=O)\N=C\2/NC(=O)[C@@H](N/2C)C)ccc1	NAM	mGluR5	5.89	R / H	
Clc1cc(NC(=O)\N=C\2/NC(=O)CN/2CC)ccc1	NAM	mGluR5	6.60	R / H	
Clc1cc(NC(=O)\N=C\2/NC(=O)[C@@H](N/2CC)C)ccc1	NAM	mGluR5	6.39	R / H	
Clc1cc(NC(=O)\N=C\2/NC(=O)CN/2CCCC)ccc1	NAM	mGluR5	5.89	R / H	
Clc1cc(NC(=O)\N=C\2/NC(=O)[C@@H]3N/2CCC3)ccc1	NAM	mGluR5	5.80	R / H	
O=C1N\C(\N2[C@@H]1CCC2)=N/C(=O)Nc1ccccc1	NAM	mGluR5	5.38	H	
O=C1N\C(=N/C(=O)Nc2cc(ccc2)C#N)\N(C1)C	NAM	mGluR5	5.41	R / H	
Clc1cc(NC(=O)\N=C\2/SCCN/2C)ccc1	NAM	mGluR5	5.92	R / H	
Clc1cc(NC(=O)\N=C/2\N(CCC\2)C)ccc1	NAM	mGluR5	5.70	R / H	
O=C1N\C(=N/C(=O)Nc2cc(ccc2)C)\N(C1)C	NAM	mGluR5	6.26	H	
O=C1N\C(=N/C(=O)Nc2ccccc2)\N(C1)C	NAM	mGluR5	5.31	R / H	
Brc1cc(NC(=O)\N=C\2/NC(=O)CN/2C)ccc1	NAM	mGluR5	6.42	H	
s1cccc1NC(=O)\N=C/1\NC(=O)CN\1C	NAM	mGluR5	5.31		
O(C)c1cc(ccc1-c1ncccc1)-c1nc2n(c1)C=CC=C2	NAM	mGluR5	7.66		
S1C=Cn2cc(nc12)-c1cc(OC)c(cc1)-c1ncccc1	NAM	mGluR5	7.38		
S1CCn2cc(nc12)-c1cc(OC)c(cc1)-c1ncccc1	NAM	mGluR5	6.49		
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.53	R/H	+
O=C(N1CCN(CC1)c1ncccc1)C#Cc1ccccc1	NAM	mGluR5	5.99		+
O(C)c1cc(cnc1Nc1nc(ccc1)C)-c1cccnc1	NAM	mGluR5	8.07		+
O(Cc1cc(ccc1)C#Cc1nc(ccc1)C)C	NAM	mGluR5	7.72	H	
O(C)c1cc(cnc1)C#Cc1ncccc1	NAM	mGluR5	7.80	H	
Fc1cc(cc(c1)C#N)C#Cc1ncccc1	NAM	mGluR5	7.40	H	
s1cccc1CC(=O)N1CCN(CC1)CC	NAM	mGluR5	5.00	R	-
s1c(S(=O)(=O)N)c(nc1NC(=O)c1c2n(nc1)C(=CC(=N2)c1ccc(cc1)C(F)(F)F)C(F)(F)F)C	NAM	mGluR2	8.07		
O(C(=O)c1c(C)c([nH]c1CNCc1ccccc1)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c([nH]c1CN)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
S(C(=O)c1[nH]c(C(OCCC)=O)c(C)c1C(OC(C)(C)C)=O)c1ccncc1	NAM	mGluR1	5.49	R	+
O(C(=O)c1c[nH]c(C(O)=O)c1C)C(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c([nH]c1CO)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.10	R	+
O(C(=O)c1c(C)c([nH]c1COC(=O)CC)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.57	R	+
O(C(=O)c1c(C)c([nH]c1C(O)=O)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCC)=O)CC	NAM	mGluR1	4.80	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCc1ccccc1)=O)Cc1ccccc1	NAM	mGluR1	5.00	R	+
o1cccc1COC(=O)c1[nH]c(C)c(C(OC(C)(C)C)=O)c1C	NAM	mGluR1	6.70	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCCCCCCCC)=O)C(C)(C)C	NAM	mGluR1	5.20	R	+
O(C(=O)c1[nH]c(C)c(C(OC(C)(C)C)=O)c1C)c1ccccc1	NAM	mGluR1	5.52	R	+
[Si](CCOC(=O)c1[nH]c(C)c(C(O)=O)c1C)(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1[nH]c(C)c(C(O[C@@H](C(C)C)C)=O)c1C)CCC	NAM	mGluR1	7.40	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCCC)=O)[C@H]1C[C@@H](CC1)C	NAM	mGluR1	6.59	R	+
O(C(=O)c1[nH]c(C)c(C(O[C@@H](C(C)(C)C)C)=O)c1C)C(C)(C)C	NAM	mGluR1	7.30	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	6.80	R	+
O(C(=O)c1[nH]c(C)c(C(O)=O)c1C)CC	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c([nH]c1C)C(O)=O)CC	NAM	mGluR1	5.00	R	+
OC(=O)c1c(C)c([nH]c1C)C(O)=O	NAM	mGluR1	5.00	R	+
O(C(=O)c1[nH]c(C)c(C(OC(C)(C)C)=O)c1C)Cc1ccccc1	NAM	mGluR1	5.72	R	+
O(C(=O)c1c(C)c([nH]c1C)C(OCC=C)=O)C(C)(C)C	NAM	mGluR1	6.41	R	+
O(C(=O)c1[nH]c(C)c(C(OCC=C)=O)c1C)Cc1ccccc1	NAM	mGluR1	5.00	R	+
O=C(Nc1n(ncc1-c1ccccc1)-c1ccccc1)c1cc(ccc1)C#N	NAM	mGluR5	5.92	R	
O(C(C)(C)C)C(=O)N1CCC(=CC1)c1nnn(c1)-c1ccccc1	NAM	mGluR1	8.01	H	+
O=C(N1CCC(CC1)C)Cn1cc([N+](=O)[O-])nc1	NAM	mGluR5	5.00	R	-
O(C(=O)c1c(C)c(n(CC(=O)N(C)C)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c(n(CC(OCC=C)=O)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	4.90	R	+
O(C(=O)c1c(C)c(n(CN2CCCCC2)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c(n(CC(O)=O)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+

O(C(=O)c1c(C)c(n(CCN2CCN(CC2)CC=C)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.00	R	+
O1CCN(CC1)CCn1c(C(OCCC)=O)c(C)c(C(OC(C)(C)C)=O)c1C	NAM	mGluR1	5.00	R	+
O(C(=O)c1c(C)c(n(C(OC(C)(C)C)=O)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.64	R	+
[Si](OCCCCn1c(C(OCCC)=O)c(C)c(C(OC(C)(C)C)=O)c1C)(C(C)(C)C)(c1ccccc1)c1ccccc1	NAM	mGluR1	5.00	R	+
[Si](OCCn1c(C(OCCC)=O)c(C)c(C(OC(C)(C)C)=O)c1C)(C(C)(C)C)(c1ccccc1)c1ccccc1	NAM	mGluR1	5.00	R	+
n1n(ccc1C)CCN1CCCCC1	NAM	mGluR5	5.00	R	-
Clc1cc(NC(=O)Nc2nccn2C)ccc1	NAM	mGluR5	5.00	R / H	
Clc1ccccc1C#Cc1nc(ccc1)C	NAM	mGluR5	7.82		+
s1cc(nc1-c1cc(F)cc(F)c1)C#Cc1cc(F)ccc1	NAM	mGluR5	5.00	R	
n1ccccc1-c1nn(nn1)-c1cc(cc(c1)C)C#N	NAM	mGluR5	7.96	R	
O(C)c1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	7.96	R	
n1ccccc1-c1nn(nn1)-c1cc(cc(N)c1)C#N	NAM	mGluR5	6.51	R	
Oc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	6.25	R	
O(CC1CC1)c1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	6.01	R	
O(Cc1ccccc1)c1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	5.00	R	
O(c1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N)c1cccnc1	NAM	mGluR5	7.80	R	
Clc1ccc(cc1-n1nc(nn1)-c1ncccc1)C#N	NAM	mGluR5	5.00	R	
Clc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	7.59	R	
Clc1ccc(-n2nc(nn2)-c2ncccc2)cc1C#N	NAM	mGluR5	5.00	R	
Clc1c(-n2nc(nn2)-c2ncccc2)cccc1Cl	NAM	mGluR5	5.00	R	
n1ccccc1-c1nn(nn1)-c1cc(cc(c1)C#N)C#N	NAM	mGluR5	7.47	R	
O=[N+]([O-])c1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	7.38	R	
Fc1ccc(cc1)\C=N\N=C\c1ccc(F)cc1	NAM	mGluR5	5.00	H	
s1cc(nc1C)C#Cc1cc(cnc1)C	NAM	mGluR5	7.89	H	
s1cc(nc1C)C#Cc1cc(cnc1)-c1ccncc1	NAM	mGluR5	7.74	H	
s1cc(nc1C)C#Cc1cc(cnc1)COC	NAM	mGluR5	7.89	H	
s1cc(nc1C)C#Cc1cc(cc(c1)C)C#N	NAM	mGluR5	7.89	H	
s1cc(nc1C)C#Cc1cc(cc(F)c1)C#N	NAM	mGluR5	7.89	H	
s1cc(nc1C)C#Cc1cc(OC)cnc1	NAM	mGluR5	7.89	H	
s1c(-c2ccccc2)c(nc1C)C#Cc1cccnc1	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1cc2c(nc1)cccc2	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1cc2nccnc2cc1	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1cc2cc[nH]c2nc1	NAM	mGluR5	5.00	R	
s1cc(nc1NCc1ccccc1)C#Cc1cccnc1	NAM	mGluR5	5.00	R	
s1cc(nc1NC(=O)Nc1ccc(F)cc1F)C#Cc1cccnc1	NAM	mGluR5	5.00	R	
s1cc(nc1NC(OC)=O)C#Cc1cccnc1	NAM	mGluR5	5.00	R	
s1cc(nc1C#CCO)C#Cc1cc(F)ccc1	NAM	mGluR5	5.00	R	
s1cc(nc1C#C)C#Cc1cc(F)ccc1	NAM	mGluR5	5.53	R	
Clc1ncc(cc1)C#Cc1nc(sc1)C	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1ccc(nc1)C=C	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1cccnc1OS(=O)(=O)C	NAM	mGluR5	5.00	R	
s1cc(nc1C)C#Cc1cccnc1C#C	NAM	mGluR5	6.72	R	
s1cc(nc1C)\C=C\c1cc(cnc1)C	NAM	mGluR5	5.51	R	
s1cc(nc1C)C(=O)Nc1cc(F)ccc1	NAM	mGluR5	5.00	R	
s1cc(nc1C)C(=O)Nc1cc(cnc1)C	NAM	mGluR5	5.00	R	
Fc1cc(ccc1)C#Cc1nc(oc1)C	NAM	mGluR5	5.00	R	
o1cc(nc1C)C#Cc1cccnc1	NAM	mGluR5	5.00	R	
O=C1N\C(=N/C(=O)Nc2nc(ccc2)C)\N(C1)C	NAM	mGluR5	5.72	R / H	
O=C1N\C(=N/C(=O)Nc2cc(ccc2)C)\N(C)[C@@H]1c1ccccc1	NAM	mGluR5	5.00	R / H	
Clc1ccc(NC(=O)\N=C\2/NC(=O)[C@@H]3N/2CCC3)cc1	NAM	mGluR5	5.00	H	
Clc1cc(NC(=O)\N=C\2/NC(=O)[C@@H]3N/2CCC3)ccc1C	NAM	mGluR5	5.00	H	
O=C1N\C(\N2[C@@H]1CCC2)=N/C(=O)Nc1ccccc1C	NAM	mGluR5	5.00	H	
O=C1N\C(\N2[C@@H]1CCC2)=N/C(=O)Nc1cc(ccc1C)C	NAM	mGluR5	5.00	H	
O=C1N\C(\N2[C@@H]1CCC2)=N/C(=O)Nc1cccc(C)c1C	NAM	mGluR5	5.00	H	
Clc1cc(NC(=O)\N=C\2/NC(=O)CN/2C)cc(Cl)c1	NAM	mGluR5	5.00	H	
Clc1cc(NC(=O)\N=C\2/NCC(=O)N/2C)ccc1	NAM	mGluR5	5.00	R / H	
Clc1ccccc1NC(=O)\N=C\1/NC(=O)CN/1C	NAM	mGluR5	5.17		+
Clc1ccc(NC(=O)\N=C\2/NC(=O)CN/2C)cc1	NAM	mGluR5	5.00		+
O=C1N\C(=N/C(=O)Nc2ncccc2)\N(C1)C	NAM	mGluR5	5.96		+
O=C1N\C(=N/C(=O)Nc2cccnc2)\N(C1)C	NAM	mGluR5	5.00		+

10. Appendix

O=C1N\C(=N/C(=O)Nc2ccncc2)\N(C1)C	NAM	mGluR5	5.00		+
Clc1nc(NC(=O)\N=C\2/NC(=O)CN/2C)ccc1	NAM	mGluR5	6.78		+
Clc1nccc(NC(=O)\N=C\2/NC(=O)CN/2C)c1	NAM	mGluR5	6.18		+
O=C1N\C(=N/C(=O)Nc2cc(ncc2)C)\N(C1)C	NAM	mGluR5	5.57		+
s1cc(NC(=O)\N=C\2/NC(=O)CN/2C)cc1	NAM	mGluR5	5.40		+
Clc1cscc1NC(=O)\N=C\1/NC(=O)CN/1C	NAM	mGluR5	6.36		+
O=C1N\C(=N/C(=O)Nc2cc(ccc2)C(OC)=O)\N(C1)C	NAM	mGluR5	5.00	H	
O=C1N\C(=N/C(=O)NCCCCCC)\N(C1)C	NAM	mGluR5	5.00	H	
s1cccc1NC(=O)\N=C\1/NC(=O)CN/1C	NAM	mGluR5	5.31	H	
Fc1ncccc1-n1nnc(C=2CCN(CC=2)C)C(=O)N(C(C)C)C)c1C	NAM	mGluR5	7.85	H	
Fc1ncccc1-n1nnc(C=2CCN(CC=2)C)C(OC(C)(C)C)=O)c1C	NAM	mGluR5	7.85	H	
S(F)(=O)(=O)CCN1C(=O)[C@@H](N(c2ccccc2)C1=O)C	NAM	mGluR5	4.39	R	-
Fc1ccccc1OCC(=O)N1CCCC[C@H]1C	NAM	mGluR5	5.00	R	-
O1[C@@H](CN(C(C)(C)C)C1=O)COc1ccccc1	NAM	mGluR5	4.40	R	-
O1[C@@H](CN(CCC)C1=O)COc1ccccc1	NAM	mGluR5	5.00	R	-
O=C(NCCNc1ccccc1[N+](=O)[O-])C#CCCCC	NAM	mGluR5	7.77	H	
O=C1N(CC)C(=NC(N2C=Cc3c(C=C2)cccc3)=C1C#N)C	NAM	mGluR1	6.89	R	+
Clc1cccc(Cl)c1-c1nc(c(n1C)-c1ccc(Cl)cc1)-c1ccncc1	NAM	mGluR2	7.70		+
Clc1cccc(Cl)c1-c1nc(c(n1CC)-c1ccc(cc1)C)-c1ccncc1	NAM	mGluR2	7.70		
Clc1cccc(Cl)c1-c1nc(c(n1C1CC1)-c1ccc(Cl)cc1)-c1ccncc1	NAM	mGluR2	7.68		
Clc1ccc(cc1)-c1n(CC)c(nc1-c1ccncc1)C(O)(C1CCCCC1)C1CCCCC1	NAM	mGluR2	7.70		
O=C(N1CCN(CC1)c1ncccc1[N+](=O)[O-])C#Cc1ccccc1	NAM	mGluR5	7.66	H	
O=C1Nc2c(N=C(C1)c1cc(-n3ccnc3)ccc1)cc(cc2)C#Cc1ccccc1	NAM	mGluR2	7.66		
Ic1cc2N=C(CC(=O)Nc2cc1-n1cccc1)c1cc(ccc1)C#N	NAM	mGluR2	7.60		+
O=C1N(CCc2ccccc2)C(C)=C(C1)C(OC)=O	NAM	mGluR5	5.00	R	-
Fc1ccc(cc1)C#Cc1cc2N=C(CC(=O)Nc2cc1)c1cc(-n2ncnc2)ccc1	NAM	mGluR2	7.52		
Fc1c(cccc1F)-c1cc2N=C(CC(=O)Nc2cc1)c1cc(-n2nncc2)ccc1	NAM	mGluR2	7.52		
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c3[nH]ccc3cnc12	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5c([nH]cc5)c34)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
s1c2c(N=CN(C2=O)c2ccc(OC)cc2)c2c3[nH]ccc3cnc12	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5nc[nH]c5c34)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5ncn(c5c34)C)C2=O)cc1	NAM	mGluR1	6.99	R / H	+
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c3[nH]ncc3cnc12	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5c([nH]nc5)c34)C2=O)cc1	NAM	mGluR1	7.42	R / H	+
S1c2c(NC1=S)c1c3N=CN(C(=O)c3sc1nc2)c1ccc(cc1)C	NAM	mGluR1	6.00	R / H	+
s1c2c(nc1SC)c1c3N=CN(C(=O)c3sc1nc2)c1ccc(cc1)C	NAM	mGluR1	6.10	R / H	+
S1c2c(NC1=O)c1c3N=CN(C(=O)c3sc1nc2)c1ccc(cc1)C	NAM	mGluR1	6.00	R / H	+
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c3NCCOc3cnc12	NAM	mGluR1	6.15	R / H	
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c3NC[C@H](Oc3cnc12)C	NAM	mGluR1	6.22	R / H	
s1c2c(N=CN(C2=O)c2ccc(OC)cc2)c2c3NCCOc3cnc12	NAM	mGluR1	7.51	R / H	+
s1c2c(N=CN(C2=O)c2ccc(OC)cc2)c2c3NC[C@H](Oc3cnc12)C	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5OCCNc5c34)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5OC[C@@H](Nc5c34)C)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5OC[C@@H](Nc5c34)C)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5O[C@@H](CNc5c34)C)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
Clc1ccc(N2C=Nc3c(sc4ncc5O[C@@H](CNc5c34)C)C2=O)cc1	NAM	mGluR1	7.51	R / H	+
O1C[C@H]2CC(C[C@@]2(Cc2cc3c(cc2)cccc3)C1=O)=C	NAM	mGluR1	6.80	R	+
O(C)c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.03	R / H	+
Ic1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	6.79	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C#N	NAM	mGluR1	6.76	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)N	NAM	mGluR1	7.49	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(O)nc2cc1	NAM	mGluR1	7.07	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3N(CCCc3c2)C)cc1	NAM	mGluR1	7.49	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3CCCc3c2)cc1	NAM	mGluR1	7.07	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2c(nc(cc2)CCC)cc1	NAM	mGluR1	7.49	R / H	
S1CCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CCC)c(nc2cc1)C	NAM	mGluR1	7.49	R / H	
O1CCCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R / H	+
O(C)c1cc(ccc1OC)CC(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.00	R	+
O(C)C1CCC(CC1)C(=O)c1cc2cc(cnc2cc1)CC	NAM	mGluR1	7.29	R / H	
O(C(C)(C)C)C(=O)N1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.00	R	
s1cc(cc1)CC(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.13	R / H	

182

O1CCCc2cc3cc(ccc3nc12)C(=O)Cc1ccc(cc1)C	NAM	mGluR1	6.80	R / H	
Fc1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	5.08	R	
Clc1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R / H	
s1cc(cc1)-c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C(OC(C)C)=O	NAM	mGluR1	6.79	R / H	
OC1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)NCCOC	NAM	mGluR1	6.98	H	
s1ccnc1-c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	5.00	R	
s1cccc1-c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	5.48	R	
Oc1nc2c(cc1CC)cc(cc2)C(=O)Cc1ccccc1	NAM	mGluR1	7.07	R / H	
O(C)c1nc2c(cc1CC)cc(cc2)/C(=N\O)/C1CCC(OC)CC1	NAM	mGluR1	7.10	R / H	
O(C)c1nc2c(cc1CC)cc(cc2)/C(=N\O)/C1CCC(OC)CC1	NAM	mGluR1	5.00	R	
O(C)c1nc2c(cc1CC)cc(cc2)/C(=N\N)/C1CCC(OC)CC1	NAM	mGluR1	5.30	R	
O(C)c1nc2c(cc1CC)cc(cc2)C(=C)C1CCC(OC)CC1	NAM	mGluR1	5.36	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3CCCCCc3c2)cc1	NAM	mGluR1	7.49	R / H	
O(C)C1CCC(CC1)C(=O)c1cc2C=C(CCC)[C@@H](Nc2cc1)C	NAM	mGluR1	7.49	R / H	
O=C(C1CCC(CC1)C)c1cc2c(nc3CCCCCc3c2)cc1	NAM	mGluR1	6.35	R / H	
O(C)c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	H	
Clc1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	5.63	R	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.84	R	
O1CCCc2cc3c(nc12)cc(cc3)C(=O)Cc1ccccc1	NAM	mGluR1	7.19	R	
Clc1cc(Cl)ccc1C(OC[C@H]1[C@@H]2N(CCC1)CCCC2)=O	NAM	mGluR5	5.00	R	-
O1C2=C(C=C([C@@H]1N(C)C)c1ccccc1)C(=O)CC(C2)(C)C	NAM	mGluR5	4.42	R	-
Fc1ccc(cc1)-c1cc2N=C(CC(=O)Nc2cc1)c1cc(-n2cc(nc2)C)ccc1	NAM	mGluR2	7.44		
s1cc(nc1-c1cc(ccc1)C1=Nc2cc(ccc2NC(=O)C1)-c1ccccc1F)CC	NAM	mGluR2	7.41		
Fc1cc(cc(-n2cc(cc2)-c2ncccc2)c1)C#N	NAM	mGluR5	7.40	R	
O=C(Nc1n(ncc1-c1ccccc1)-c1ccccc1)c1ccccc1	NAM	mGluR5	5.93	R	
n1c(cn(c1C)-c1nc(ccc1)C1CC1)C#Cc1cc(ncc1)C	NAM	mGluR5	7.37		
s1c2c(N=CN(C2=O)c2ccc(OC)cc2)c2c1nccc2N(C)C	NAM	mGluR1	7.35	R / H	+
O(C(=O)c1[nH]c(C)c(C(O[C@@H](C(C)(C)C)C)=O)c1C)CCC	NAM	mGluR1	7.32	R	+
S(C(=O)c1[nH]cc(C(OC(C)(C)C)=O)c1C)c1ccncc1	NAM	mGluR1	5.52	R	+
O(C(=O)c1[nH]c(C)c(C(O[C@H](C(C)(C)C)C)=O)c1C)CCC	NAM	mGluR1	7.32	R	+
O(C(=O)c1[nH]c(C)c(C(O[C@@H](C(C)(C)C)C)=O)c1C)CCC	NAM	mGluR1	7.32	R	+
o1ccccc1C(OC1C[C@@H]2N([C@H](C1)CC2)C)=O	NAM	mGluR5	5.00	R	-
O=C1c2c(-c3c1n(Cc1ccccc1)c(C)c3CN(C)C)cccc2	NAM	mGluR5	4.20	R	-
Clc1sc(S(=O)(=O)N)cc1NC(=O)c1c2n(nc1)C(=CC(=N2)c1ccc(c1)C(F)(F)F)C(F)(F)F	NAM	mGluR2	7.28		
O=C(Nc1n2c(nc1-c1ccccc1)C=CC=C2)c1ccccc1	NAM	mGluR5	5.41	R	
O=C(Nc1n2c(nc1-c1ccccc1)C=CC=C2)c1cc(ccc1)C#N	NAM	mGluR5	5.70	R	
Fc1ccccc1-c1cc2N=C(CC(=O)Nc2cc1)c1cc(-n2nncc2)ccc1	NAM	mGluR2	7.27		
s1cc(nc1-c1cc(ccc1)C1=Nc2cc(ccc2NC(=O)C1)-c1ccccc1F)C(O)=O	NAM	mGluR2	7.27		
O(C(=O)[C@@]12[C@@H](C1)\C(=N\O)\c1c(C2)cccc1)CC	NAM	mGluR1	5.82	H	+
O1c2c(cccc2)/C(=N\O)/[C@@H]2C[C@]12C(=O)Nc1ccccc1	NAM	mGluR3	7.11		77
o1ccccc1C(=O)N1CCN(CC1)C1CCC(CC1)c1ccccc1	NAM	mGluR5	5.00	R	-
Fc1cc(F)ccc1-c1cc2N=C(CC(=O)Nc2cc1)c1cc(-n2nncc2)ccc1	NAM	mGluR2	7.23		
o1c2c(cc1CN1CCCCC1)cccc2	NAM	mGluR5	5.00	R	-
n12c(ccc1)C(=NC(=C2)C)C#Cc1ccccc1	NAM	mGluR5	8.07	H	+
n12c(ccc1)C(=NC(=C2)C)C#Cc1cc(ccc1)C	NAM	mGluR5	8.03	H	+
Fc1ccc(cc1)C#CC1=NC(=Cn2c1ccc2)C	NAM	mGluR5	7.80	H	+
Clc1ccc(cc1)C#CC1=NC(=Cn2c1ccc2)C	NAM	mGluR5	6.20	H	+
Brc1ccc(cc1)C#CC1=NC(=Cn2c1ccc2)C	NAM	mGluR5	0.01	H	+
Oc1cc(ccc1)C#CC1=NC(=Cn2c1ccc2)C	NAM	mGluR5	6.50	H	+
n12c(ccc1)C(=NC(=C2)C)C#C[C@@H]1NC=CC=N1	NAM	mGluR5	5.70	H	+
n12c(ccc1)C(=NC(=C2)C)C#C[C@H]1NC=Cc2c1cccc2	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(c1)C=CN=C2C#Cc1ccccc1	NAM	mGluR5	6.80	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	7.40	H	+
n12c(cc(c1)C#N)C(=NC(=C2)C)C#Cc1ccccc1	NAM	mGluR5	8.05	H	+
O=C(N1CCN(CC1)C)c1cc2n(c=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	7.00	H	+
O=C(N1CCCC1)c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	7.10	H	+
O1CCN(CC1)C(=O)c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	7.80	H	+
n12c(cc(c1)C)C(=NC(=C2)C)C#Cc1ccccc1	NAM	mGluR5	7.30	H	+
O(C(=O)c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1)CC	NAM	mGluR5	8.00	H	+
o1nc(nc1-c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1)C	NAM	mGluR5	8.09	H	+

10. Appendix

O=C(NC)c1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	6.90	H	+
n12c(cc(c1)CN1CCCCC1)C(=NC(=C2)C)C#Cc1ccccc1	NAM	mGluR5	6.80	H	+
n12c(cc(c1)CN1CCN(CC1)C)C(=NC(=C2)C)C#Cc1ccccc1	NAM	mGluR5	6.30	H	+
O=C(N1CCN(CC1)Cc1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1)C	NAM	mGluR5	6.50	H	+
O=C1NCCN(C1)Cc1cc2n(C=C(N=C2C#Cc2ccccc2)C)c1	NAM	mGluR5	6.70	H	+
FC(F)(F)c1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C)C	NAM	mGluR5	0.01	H	+
Clc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C)C	NAM	mGluR5	0.01	H	+
Fc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C)C	NAM	mGluR5	6.40	H	+
n12c(cc(c1)C)C(=NC(=C2)C)C#Cc1ccc(cc1)C#N	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C)C	NAM	mGluR5	6.80	H	+
n12c(cc(c1)C)C(=NC(=C2)C)C#Cc1cc(ccc1)[C@@H]1NC=CC=C1	NAM	mGluR5	5.70	H	+
s1cc(cc1)-c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C)C	NAM	mGluR5	7.20	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2cc(OC)ccc2)C)c1	NAM	mGluR5	7.10	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	6.50	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	7.80	H	+
Clc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	7.00	H	+
Brc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	7.60	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2cc(ccc2)C#N)C)c1	NAM	mGluR5	7.90	H	+
S(=O)(=O)(Nc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C)C	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2cc(NC(=O)C)ccc2)C)c1	NAM	mGluR5	6.10	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2cc(ccc2)C(O)=O)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2cc(ccc2)C(OC(C)(C)C)=O)C)c1	NAM	mGluR5	0.01	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	6.90	H	+
Clc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	5.30	H	+
FC(F)(F)c1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	0.01	H	+
Fc1cc(F)ccc1C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	0.01	H	+
Fc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	7.00	H	+
Clc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	5.70	H	+
FC(F)(F)c1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2ccc(cc2)C#N)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2ccc(N(C)C)cc2)C)c1	NAM	mGluR5	0.01	H	+
S(=O)(=O)(Nc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C)C	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#CC=2C=CNCC=2)C)c1	NAM	mGluR5	7.10	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@@H]2NC=CC=C2)C)c1	NAM	mGluR5	7.30	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#CC=2CNC=C=2)C)c1	NAM	mGluR5	6.30	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2n(cnc2)C)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#Cc2ccoc2)C)c1	NAM	mGluR5	7.30	H	+
s1cccc1C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	7.20	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	8.00	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@H]2C=C(C=CC2)c2cn[nH]c2)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@H]2C=C(C=CC2)c2cnoc2)C)c1	NAM	mGluR5	6.10	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@H]2C=C(C=CC2)c2cn(nc2)C)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@H]2C=C(C=CC2)c2c(noc2C)C)C)c1	NAM	mGluR5	0.01	H	+
FC(F)(F)c1cc2n(C=C(N=C2C#C[C@H]2C=C(C=CC2)c2ccoc2)C)c1	NAM	mGluR5	5.00	H	+
s1cc(cc1)C=1C=CC[C@@H](C=1)C#CC1=NC(=Cn2c1cc(c2)C(F)(F)F)C	NAM	mGluR5	5.30	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C#N)C	NAM	mGluR5	8.07	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C#N)C	NAM	mGluR5	8.04	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C#N)C	NAM	mGluR5	7.80	H	+
s1cc(cc1)-c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C#N)C	NAM	mGluR5	8.03	H	+
Clc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	7.70	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	6.70	H	+
O(C(=O)c1cc2n(C=C(N=C2C#Cc2cc(ccc2)C#N)C)c1)CC	NAM	mGluR5	8.07	H	+
O(C(=O)c1cc2n(C=C(N=C2C#Cc2ccc(cc2)C#N)C)c1)CC	NAM	mGluR5	0.01	H	+
FC(F)(F)c1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	0.01	H	+
Clc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	0.01	H	+
Fc1ccc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	6.40	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	6.50	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(OCC)=O)C	NAM	mGluR5	8.09	H	+

O(C(=O)c1cc2n(C=C(N=C2C#C[C@@H]2NC=CC=C2)C)c1)CC	NAM	mGluR5	5.80	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(=O)N)C	NAM	mGluR5	6.30	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N)C	NAM	mGluR5	7.10	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N)C	NAM	mGluR5	7.30	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(=O)NC)C	NAM	mGluR5	5.80	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)NC)C	NAM	mGluR5	7.00	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)NC)C	NAM	mGluR5	7.30	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC1)C	NAM	mGluR5	6.60	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC1)C	NAM	mGluR5	7.70	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC1)C	NAM	mGluR5	6.80	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC1)C	NAM	mGluR5	7.70	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCOCC1)C	NAM	mGluR5	6.90	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCOCC1)C	NAM	mGluR5	7.70	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCOCC1)C	NAM	mGluR5	7.90	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCOCC1)C	NAM	mGluR5	8.00	H	+
Fc1ccccc1C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCN(CC1)C)C	NAM	mGluR5	5.90	H	+
Fc1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCN(CC1)C)C	NAM	mGluR5	6.70	H	+
FC(F)(F)c1cc(ccc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCN(CC1)C)C	NAM	mGluR5	6.60	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCN(CC1)C)C	NAM	mGluR5	6.80	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC1(C)C)C	NAM	mGluR5	5.90	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CC(F)(F)CC1)C	NAM	mGluR5	6.90	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1[C@@H](CC[C@H]1C)C)C	NAM	mGluR5	6.40	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1C[C@H](O[C@@H](C1)C)C)C	NAM	mGluR5	6.50	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCC[C@H]1C)C	NAM	mGluR5	7.40	H	+
s1cc(cc1)C#CC1=NC(=Cn2c1cc(c2)C(=O)N1CCCC[C@H]1C)C	NAM	mGluR5	7.00	H	+
O(C(=O)CCc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.05		+
Ic1cc(ccc1)C(Oc1cc(ccc1)C#Cc1nc(ccc1)C)=O	NAM	mGluR5	5.96		
O(C(=O)CCCc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.93		
S(Oc1cc(ccc1)C#Cc1nc(ccc1)C)(=O)(=O)C	NAM	mGluR5	6.74		
O(CCc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.80		
Fc1ccc(cc1)CCOc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.69		
O(CCCc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.92		
O(CCCCc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.94		
O(CC1CC1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.38		
O(Cc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.74		
Ic1cc(ccc1)COc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.70		
O(Cc1cc(OC)ccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.89		
O(CC(OC)=O)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.62		
n1c(cccc1C)C#Cc1cc(ccc1)-c1ccccc1	NAM	mGluR5	6.00		
O(C)c1cc(ccc1)-c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.63		
o1cccc1-c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.43		
s1cccc1-c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.18		
n1c(cccc1C)C#Cc1cc(cnc1)-c1ccccc1	NAM	mGluR5	7.70		
O(C)c1cc(ccc1)-c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.55		
o1cccc1-c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.69		
s1cccc1-c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.57		
n1c(cccc1C)-c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.59		
n1c(cccc1C)C#Cc1cc(N)ccc1	NAM	mGluR5	6.73		
O(C(=O)c1cc(ccc1)C#Cc1nc(ccc1)C)C	NAM	mGluR5	6.68		
O=C(C)c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.84		
O(C(=O)c1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.82		
BrCCCC(Oc1cc(ccc1)C#Cc1nc(ccc1)C)=O	NAM	mGluR5	5.47		
O(C(=O)c1cc(OC)ccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	5.76		
o1cccc1C(Oc1cc(ccc1)C#Cc1nc(ccc1)C)=O	NAM	mGluR5	6.00		
O(C(=O)Cc1ccccc1)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.13		
Clc1cc(ccc1)-c1onc(n1)C[C@H]1CCn2c1nnc2-c1ccncc1	NAM	mGluR5	7.31	H	
Fc1cc(cc(-n2nc(nn2)-c2ncc(cc2)C#N)c1)C#N	NAM	mGluR5	7.96	H	
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)C#N	NAM	mGluR5	7.85		
Clc1ccc(nc1)-c1nn(nn1)-c1cc(cc(F)c1)C#N	NAM	mGluR5	7.74	H	
Brc1ccc(nc1)-c1nn(nn1)-c1cc(cc(F)c1)C#N	NAM	mGluR5	7.21	H	

10. Appendix

Clc1cc(ccc1)-c1onc(n1)CCN1CCCCn2c1nnc2-c1ccncc1	NAM	mGluR5	7.09	H	
Clc1cc(ccc1)-c1nc(on1)[C@@H]1N(CCOC1)c1nnc(n1C)-c1ccc(OC(F)F)cc1	NAM	mGluR5	6.70	H	
Clc1cc(-n2nc(cn2)[C@H](Oc2nnc(n2C)-c2ccncc2)C)ccc1	NAM	mGluR5	7.39	H	
Clc1cc(-n2nnc(c2)CSc2nnc(n2C2CC2)-c2ccncc2)ccc1	NAM	mGluR5	6.58	H	
Clc1cc(-n2nc(nn2)[C@H](Sc2nnc(n2C)-c2ccncc2)C)ccc1	NAM	mGluR5	7.26	H	
Clc1cc(-n2nc(nn2)[C@@H](N2CCN(CC2)C(OCC)=O)C)ccc1	NAM	mGluR5	6.88	H	
Fc1ccc(cc(-n2nc(nn2)-c2ncc(F)cc2)c1)C#N	NAM	mGluR5	7.96	H	
Clc1nc(C(=O)Nc2ncccc2)c(nc1N(C)C)N	NAM	mGluR5	6.87		
Clc1nc(cnc1N(C)C)C(=O)Nc1ncccc1	NAM	mGluR5	6.80		
Clc1nc(C(=O)Nc2ncccc2)c(nc1N1CCCCC1)N	NAM	mGluR5	6.99		
Clc1nc(C(=O)Nc2ncccc2)c(nc1NC(C)C)N	NAM	mGluR5	7.44		
Clc1nc(C(=O)Nc2ncccc2)c(nc1N)N	NAM	mGluR5	5.42		
Clc1nc(C(=O)Nc2ncccc2)c(nc1)N	NAM	mGluR5	7.44		
Clc1nc(cnc1)C(=O)Nc1ncccc1	NAM	mGluR5	6.60		
O=C(Nc1ncccc1)c1nc(cnc1)C	NAM	mGluR5	6.78		
O=C(Nc1ncccc1)c1nccnc1N	NAM	mGluR5	5.45		
Brc1nc(C(=O)Nc2ncccc2)c(nc1)N	NAM	mGluR5	7.44		
O=C(Nc1ncccc1)c1nc(cnc1N)C	NAM	mGluR5	7.55		
Clc1nc(C(=O)Nc2ncccc2)c(nc1)NC	NAM	mGluR5	6.84		
Clc1nc(C(=O)Nc2nc(ccc2)C)c(nc1)N	NAM	mGluR5	7.44		
O=C(Nc1nc(ccc1)C)c1nc(cnc1N)C	NAM	mGluR5	7.44		
O=C(Nc1nc(ccc1)C)c1nc(ccc1)C	NAM	mGluR5	7.44		
Oc1ccc(nc1C(=O)Nc1nc(ccc1)C)C	NAM	mGluR5	7.44		
Brc1nc(ccc1)C(=O)Nc1nc(ccc1)C	NAM	mGluR5	7.44		
O=C(Nc1ncccc1)c1ncc(nc1N)N(C)C	NAM	mGluR5	5.29		
O=C(Nc1nc(ccc1)C)c1nc(ccc1N)C	NAM	mGluR5	7.44		
n1c(cccc1C)C#Cc1ccccc1	NAM	mGluR5	7.52	R/H	
s1ccnc1C#Cc1ccccc1	NAM	mGluR5	7.09		
s1cc(nc1)C#Cc1ccccc1	NAM	mGluR5	7.01		
s1cc(nc1C#Cc1ccccc1)C	NAM	mGluR5	7.89		
s1cc(nc1C)C#Cc1ccccc1	NAM	mGluR5	7.01	R/H	+
s1cc(nc1C)C#Cc1ncccc1	NAM	mGluR5	7.28		+
s1cc(nc1C)C#Cc1cccnc1	NAM	mGluR5	7.61	R/H	+
s1cc(nc1C)C#Cc1ccncc1	NAM	mGluR5	6.91		
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccccc1	NAM	mGluR5	6.48		
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccccc1C	NAM	mGluR5	6.95		
n1ccccc1-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	6.72		
O(c1cc(-n2nc(nn2)-c2ncccc2)ccc1)c1ccccc1	NAM	mGluR5	5.84		
O(c1cc(-n2nc(nn2)-c2ncccc2)ccc1)c1ncccc1	NAM	mGluR5	5.54		
O(c1cc(-n2nc(nn2)-c2ncccc2)ccc1)c1cccnc1	NAM	mGluR5	7.23		
O(c1ccc(-n2nc(nn2)-c2ncccc2)cc1)c1cccnc1	NAM	mGluR5	6.13		
Fc1cc(-n2nc(nn2)-c2ncccc2)cc(Oc2cccnc2)c1	NAM	mGluR5	7.92		
S(c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1)c1cccnc1	NAM	mGluR5	7.80		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)Cc1cccnc1	NAM	mGluR5	7.42		
Fc1cc(Nc2cccnc2)cc(-n2nc(nn2)-c2ncccc2)c1	NAM	mGluR5	6.91		
Fc1cc(N(C)c2cccnc2)cc(-n2nc(nn2)-c2ncccc2)c1	NAM	mGluR5	7.52		
Fc1cc(-n2c3c(cc2)ccnc3)cc(-n2nc(nn2)-c2ncccc2)c1	NAM	mGluR5	7.60		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1	NAM	mGluR5	6.61		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1C	NAM	mGluR5	6.92		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cc(ccc1)C	NAM	mGluR5	5.83		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1C#N	NAM	mGluR5	7.96		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1N	NAM	mGluR5	7.96		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ncccc1	NAM	mGluR5	7.02		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cccnc1	NAM	mGluR5	7.43		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1nccc1C	NAM	mGluR5	7.51		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccncc1C	NAM	mGluR5	7.20		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cccnc1C	NAM	mGluR5	7.35		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cnccc1C	NAM	mGluR5	7.96		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccc(nc1)C	NAM	mGluR5	7.24		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cc(cnc1)C	NAM	mGluR5	6.30		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cnccc1N	NAM	mGluR5	6.97		
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ncccc1N	NAM	mGluR5	7.51		

186

Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccncc1	NAM	mGluR5	6.02		
n1ccccc1-c1ccn(c1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
n1ccccc1-n1nc(cc1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
n1ccccc1-c1nn(cc1)-c1cc(ccc1)C#N	NAM	mGluR5	6.28		
n1ccccc1-n1ncc(c1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
n1ccccc1-c1cn(nc1)-c1cc(ccc1)C#N	NAM	mGluR5	6.60		
n1ccccc1-c1ncn(c1)-c1cc(ccc1)C#N	NAM	mGluR5	5.94		
n1ccccc1-n1nc(cn1)-c1cc(ccc1)C#N	NAM	mGluR5	6.68		
n1ccccc1-c1nnn(c1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
n1ccccc1-c1nn(nc1)-c1cc(ccc1)C#N	NAM	mGluR5	6.52		
n1cn(nc1-c1cc(ccc1)C#N)-c1ncccc1	NAM	mGluR5	5.00		
n1ccccc1-c1ncn(n1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
n1ccccc1-n1nc(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	6.35		
o1nc(nc1-c1cc(ccc1)C#N)-c1ncccc1	NAM	mGluR5	6.31		
n1ccccc1-n1cc(cc1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00		
Fc1cc(cc(-n2nc(nn2)-c2[nH]c(cn2)C)c1)C#N	NAM	mGluR5	7.11		
Fc1cc(-n2nc(nn2)-c2[nH]ccn2)cc(Oc2cccnc2)c1	NAM	mGluR5	7.96		
Fc1cc(cc(-n2nc(nn2)-c2[nH]ccn2)c1)-c1cnccc1C	NAM	mGluR5	7.70		
s1ccnc1-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	6.74		
s1cc(nc1)-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	6.80		
[nH]1ccnc1-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	7.11		
Fc1cc(cc(-n2nc(nn2)-c2[nH]ccn2)c1)C#N	NAM	mGluR5	7.33		
Clc1cc(NC(=O)\N=C\2/NC(=O)CN/2C)ccc1	NAM	mGluR5	6.48		
Clc1cc(ccc1CC#N)-c1oc2c(n1)cccc2	NAM	mGluR5	7.52		
o1c2c(nc1-c1cc(OC)c(cc1)CC#N)cccc2	NAM	mGluR5	7.38		
o1c2c(nc1-c1cc(OC)c(cc1)-c1cccnc1)cccc2	NAM	mGluR5	6.38		
O(C)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	8.26	R/H	
n1c(cccc1C)C#Cc1cc(ccc1)C#N	NAM	mGluR5	8.52	R/H	
n1c(cccc1C)C#Cc1cccnc1	NAM	mGluR5	7.30	R	+
o1c2c(nc1-c1cc(OC)c(cc1)-c1ncccc1)cccc2	NAM	mGluR5	6.80	R/H	
s1c2c(N=CN(N3CCCCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR1	5.97	H	-
n1c(cccc1Nc1ncc(c2c1nccc2)-c1cccnc1)C	NAM	mGluR5	6.91	R/H	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49	R	
O(C)c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49		
O(C)C1CCC(CC1)C(=O)c1cc2cc(cnc2cc1)CC	NAM	mGluR1	7.49	R	
Clc1nc2cc(C)c(cc2cc1CC)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R	
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)N(C)C	NAM	mGluR1	7.49		
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)COC(=O)C	NAM	mGluR1	7.04		
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(O)nc2cc1	NAM	mGluR1	7.49		
O=C(Cc1ccccc1)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49	R	
O=C([C@H]1[C@H]2CC[C@@H](C1)C2)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR6	7.49	R	
O(C)C1CCC(CC1)CC(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49		
O=C(C1CCC(CC1)C)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49		
FC1(CCC(OC)CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.49		
O=C(C1Cc2c(C1)cccc2)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.25		
O=C(C[C@@H]1[C@@H]2CC[C@H](C1)C2)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.22		
O(C)c1ccccc1CC(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	7.17		
O1c2c(OC[C@@H]1C(=O)c1cc3cc(CC)c(nc3cc1)C)cccc2	NAM	mGluR1	7.08		
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)C#N	NAM	mGluR1	7.49		
S1CCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R	
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3N(CCCc3c2)C)cc1	NAM	mGluR1	7.49	R	
O(C)C1CCC(CC1)C(=O)c1cc2C=C(CC)C(=O)Nc2cc1	NAM	mGluR1	7.49	R	
O(C)C1CCC(CC1)C(=O)c1cc2nc(C)c(nc2cc1)C	NAM	mGluR1	7.44		
O(C)C1CCC(CC1)C(=O)c1cc2C=C(c3n(nnn3)-c2cc1)CC	NAM	mGluR1	7.36		
s1cc(cc1)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.49		
Brc1ccc(cc1)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.49		
Fc1ccc(cc1)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.49		
Fc1c(cccc1F)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.13		
O1CCCc2cc3cc(ccc3nc12)Cc1ccccc1	NAM	mGluR1	7.07		
Fc1ccccc1C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.38		
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(OCC)=O	NAM	mGluR1	5.47		
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@]12C(OCC)=O	NAM	mGluR1	5.52		

10. Appendix

O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(OCC)=O	NAM	mGluR1	5.82
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(=O)N[C@H](Cc1cccc1)C(OC)=O	NAM	mGluR1	6.37
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(=O)N[C@@H](Cc1ccccc1)C(OC)=O	NAM	mGluR1	5.85
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(=O)N[C@@H](Cc1ccccc1)C(OC)=O	NAM	mGluR1	6.03
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(=O)N[C@H](Cc1cccc1)C(OC)=O	NAM	mGluR1	5.85
S(CCO)c1nc(Nc2ccc(OC)cc2)c2cc(OC)ccc2n1	NAM	mGluR1	7.33
Clc1cccc(Cl)c1CSCCNc1nc(nc2c1CCCC2)C	NAM	mGluR1	7.57
Clc1cc2c(nc(SCCO)nc2NC2C3CCC2CC3)cc1	NAM	mGluR1	7.36
Clc1cccc(Cl)c1CSCCNc1ncnc2c1cccc2	NAM	mGluR1	7.34
O(C)c1ccc(Nc2ncnc3c2cc(OC)cc3)cc1	NAM	mGluR1	7.02
Clc1cc2c(ncnc2NC2C3CCC2CC3)cc1	NAM	mGluR1	6.40
Clc1c2ncnc(NC3C4CCC3CC4)c2ccc1	NAM	mGluR1	5.72
O(CCOC)c1nc(N2CCc3c(CC2)cccc3)c(nn1)C#N	NAM	mGluR1	5.52
n1c(N2CCc3c(CC2)cccc3)c(nnc1N)C#N	NAM	mGluR1	7.28
n1c(N2CCc3c(CC2)cccc3)c(nnc1N(C)C)C#N	NAM	mGluR1	5.86
n1c(N2CCc3c(CC2)cccc3)c(nnc1NCC1CC1)C#N	NAM	mGluR1	7.28
OCCNc1nc(N2CCc3c(CC2)cccc3)c(nn1)C#N	NAM	mGluR1	7.28
O[C@@H](CNc1nc(N2CCc3c(CC2)cccc3)c(nn1)C#N)C	NAM	mGluR1	7.28
n1c(N2CCc3c(CC2)cccc3)c(nnc1NN)C#N	NAM	mGluR1	6.43
O(C(C)(C)C)C(=O)NCCNc1nc(N2CCc3c(CC2)cccc3)c(nn1)C#N	NAM	mGluR1	7.28
n1c(N2CCc3c(CC2)cccc3)c(nnc1NCCc1cccnc1)C#N	NAM	mGluR1	7.28
n1c(CC)c(nc(N2CCc3c(CC2)cccc3)c1C#N)C	NAM	mGluR1	7.28
n1c(C)c(nc(N2CCc3c(CC2)cccc3)c1C#N)CC	NAM	mGluR1	6.99
n1ccnc(N2CCc3c(CC2)cccc3)c1C#N	NAM	mGluR1	6.33
n1c(-c2ccccc2)c(nc(N2CCc3c(CC2)cccc3)c1C#N)C	NAM	mGluR1	7.28
OCCNc1nc(N2CCc3c(CC2)cccc3)c(nc1)C#N	NAM	mGluR1	6.30
s1c2CCN(CCc2nc1C)C=1N=C(NC(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	4.52
s1c2CCN(CCc2nc1C)c1nc(nc(OCC)c1[N+](=O)[O-])C	NAM	mGluR1	5.38
s1c2CCN(CCc2nc1C)C=1N=C(N(CC)C(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	5.68
s1c2CCN(CCc2nc1N)C=1N=C(NC(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	4.31
s1c2CCN(CCc2nc1N)C=1N=C(N(CC)C(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	5.22
s1c2CCN(CCc2nc1)C=1N=C(NC(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	4.37
s1c2CCN(CCc2cc1)C=1N=C(NC(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	5.72
s1c2CCN(CCc2cc1)c1nc(nc(OCC)c1[N+](=O)[O-])C	NAM	mGluR1	6.36
s1c2CCN(CCc2cc1)C=1N=C(N(CC)C(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	7.16
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)N1CCCCC1	NAM	mGluR1	5.44
o1cccc1-c1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	6.59
O(C)C1CCC(CC1)C(=O)Nc1cc2cc(CC)c(O)nc2cc1	NAM	mGluR1	5.00
O1c2c(cccc2)/C(=N/O)/[C@H]2C[C@@]12C(=O)N(C)c1ccccc1	NAM	mGluR1	5.70
O1CCCc2cc3c(nc12)cc(cc3)C(=O)c1ccccc1	NAM	mGluR1	5.08
O(C)C1CCC(CC1)C(=O)c1ccc(NC(=O)C)cc1	NAM	mGluR1	5.12
O(C)C1CCC(CC1)C(=O)c1cc2[nH]c(nc2cc1)C	NAM	mGluR1	6.24
O=C(CCc1ccccc1)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.64
O=C(Cc1ccc(N(C)C)cc1)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.19
O=C([C@@H]1CCc2c1cccc2)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.61
O=C(C1CCNCC1)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.00
O=C(C1CCC(N(C)C)CC1)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.00
O(C)c1cc(ccc1OC)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	5.00
O1C[C@@H](Cc2c1cccc2)C(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.94
O=C(C12CC3CC(C1)CC(C2)C3)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.90
O=C(CC(C)C)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.79
O(C)c1cc(ccc1)CC(=O)c1cc2cc(CC)c(nc2cc1)C	NAM	mGluR1	6.87
O(C)C1CCC(NC(=O)c2cc3cc(CC)c(O)nc3cc2)CC1	NAM	mGluR1	5.00
O(C)C1CCC(CC1)\C(=N\O)\c1cc2cc(CC)c(O)nc2cc1	NAM	mGluR1	5.00
Fc1cc(ccc1)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.32
n1c(cccc1C)C#Cc1ccc(cc1)C#N	NAM	mGluR5	6.00
Oc1ccccc1C#Cc1nc(ccc1)C	NAM	mGluR5	8.06
n1c(cccc1C)C#Cc1ccccc1C	NAM	mGluR5	8.70
n1c(cccc1C)C#Cc1cc(ccc1)C	NAM	mGluR5	7.48
O=C(Nc1ccccc1)c1nc(ccc1)C	NAM	mGluR5	7.44

188

Oc1ccc(cc1)C#Cc1nc(ccc1)C	NAM	mGluR5	8.70	
O(C)c1ccccc1C#Cc1nc(ccc1)C	NAM	mGluR5	8.70	
n1c(cccc1C)C#Cc1ccc(cc1)C	NAM	mGluR5	8.70	
O(C(=O)c1c(C)c([nH]c1C)C(OCCCC)=O)C(C)(C)C	NAM	mGluR1	6.80	
O1CCCc2cc3cc(ccc3nc12)C(=O)c1ccccc1	NAM	mGluR1	7.49	
O1CCCc2cc3cc(ccc3nc12)C(=O)Cc1ccccc1	NAM	mGluR1	7.49	R
Brc1cc2SC3=C(CCC(C3)(C)C)C(=O)c2cc1	NAM	mGluR1	8.05	
S1C2=C(CCC(C2)(C)C)C(=O)c2c1cc(cc2)-c1cccnc1	NAM	mGluR1	7.51	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(cc2)-c1cc(ccc1)C#N	NAM	mGluR1	7.64	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(N1CCCCC1)cc2	NAM	mGluR1	7.89	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2cc(F)c(N3CCCCC3)cc12	NAM	mGluR5	7.55	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(N(COCC)C)cc2	NAM	mGluR1	7.66	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(cc2)-c1cncnc1	NAM	mGluR1	7.46	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(cc2)C=1NC(ON=1)=O	NAM	mGluR1	7.09	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(NS(=O)(=O)C1CC1)cc2	NAM	mGluR1	7.60	
o1cc(nc1-c1cc(ccc1)C#N)-c1nccccc1	NAM	mGluR5	7.35	
S(CC(OCC)=O)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	6.85	
Brc1ccc(cc1)Cc1cnc(SCC(OCC)=O)nc1N	NAM	mGluR5	6.74	
s1cc(cc1)Cc1cnc(SCC(OCC)=O)nc1N	NAM	mGluR5	6.74	
S(CC(OCC)=O)c1nc(N)c(cn1)Cc1ccoc1	NAM	mGluR5	6.80	
S(CC(OCC=C)=O)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	6.92	
s1cccc1Cc1cnc(SCC(OCC)=O)nc1NCC(C)C	NAM	mGluR5	6.80	
S1C2=C(CC[C@@H](C2)CC)C(=O)c2c1cc(cc2)-c1[nH]nnn1	NAM	mGluR1	8.05	
FC(F)(F)COc1nc(nc(N2CCC(CC2)c2ccccc2)c1C#N)NCCO	NAM	mGluR1	7.44	
Clc1nc(nc(N2CCN(CC2)c2ccccc2)c1C#N)NCC1CC1	NAM	mGluR1	7.60	
Fc1ccc(N2CCN(CC2)C=2N=C(N(CCO)C(=O)C=2[N+](=O)[O-])C)cc1	NAM	mGluR1	7.38	
Fc1ccc(N2CCN(CC2)c2nc(nc(OCCO)c2[N+](=O)[O-])C)cc1	NAM	mGluR1	7.24	
Fc1ccc(N2CCN(CC2)C=2N=C(N(CC)C(=O)C=2[N+](=O)[O-])C)cc1	NAM	mGluR1	7.31	
n1c(N2CCN(CC2)c2ccccc2)c(C#N)c(nc1NC1CC1)NC1CC1	NAM	mGluR1	7.19	
OCCNc1nc(nc(N2CCN(CC2)c2ccccc2)c1C#N)NC1CC1	NAM	mGluR1	7.48	
OCCNc1nc(nc(N2CCC(CC2)c2ccccc2)c1C#N)NCc1cccnc1	NAM	mGluR1	7.52	
ClC=1C=Cn2cc(nc2C=1)-c1cc(C)c(cc1)C	NAM	mGluR5	8.09	
O(C)C=1C=Cn2cc(nc2C=1)-c1cc(C)c(cc1)C	NAM	mGluR5	8.10	
n1c2n(cc1-c1cc(C)c(cc1)C)C=CC=2	NAM	mGluR5	8.10	
n1c2n(cc1-c1cc(C)c(cc1)C)C=CC(=C2)C	NAM	mGluR5	8.10	
Brc1cc(ccc1F)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.10	
s1c(C)c(cc1C)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.10	
O(C(=O)c1[nH]cc(c1)C(O[C@@H](C(C)(C)C)C)=O)CCC	NAM	mGluR1	7.32	
Oc1ccc(nc1N=Nc1ccccc1)C	NAM	mGluR5	5.43	
O(C)c1cc(ccc1OC)-c1nc2n(c1)C=CC(=C2)C	NAM	mGluR5	8.06	
O(C)c1cc(ccc1OC)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
O(C)c1cc(ccc1OC)-c1nc2n(c1)C=CC(OC)=C2	NAM	mGluR5	8.06	
O(Cc1ccccc1)c1ccc(cc1OC)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
Brc1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
Ic1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
Clc1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
n1c2n(cc1-c1cc(ccc1)C)C=CC=2	NAM	mGluR5	8.06	
FC(F)(F)c1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
Fc1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
n1c2n(cc1-c1cc(C)c(cc1)C)C=CC(=C2)CC	NAM	mGluR5	8.06	
s1c(ccc1C)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
n1c2n(cc1-c1ccccc1)C=CC(=C2)C	NAM	mGluR5	8.09	
n1c2n(cc1-c1ccc(cc1)C)C=CC=2	NAM	mGluR5	8.06	
O(C)c1cc(ccc1)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	
O(C)c1cc(ccc1OC)-c1nc2n(C=C(C=C2)C)c1	NAM	mGluR5	8.06	
O1c2c(OC[C@H]1c1nc3n(c1)C=CC=C3)cccc2	NAM	mGluR5	8.06	
n1c2n(C=C(C=C2)C)cc1-c1ccc(cc1)C	NAM	mGluR5	8.09	
n1c2n(cc1-c1cc3CCCc3cc1)C=CC=2	NAM	mGluR5	8.06	
o1c2c(cc1-c1nc3n(c1)C=CC=C3)cccc2	NAM	mGluR5	8.06	
s1cc(c2c1cccc2)-c1nc2n(c1)C=CC=2	NAM	mGluR5	8.06	

10. Appendix

O1CCc2cc(ccc12)-c1nc2n(c1)C=CC=C2	NAM	mGluR5	8.06
S(CC(OCC)=O)c1nc(N)c(cn1)Cc1n(ccc1)C	NAM	mGluR5	6.42
Clc1cc(Cl)ccc1-c1cnc(SCC(OCC)=O)nc1N	NAM	mGluR5	5.41
S(CC(OCC)=O)c1nc(N)c(cn1)C(OCC)=O	NAM	mGluR5	6.57
s1cccc1Cc1cnc(SCC(OCC)=O)nc1NCC	NAM	mGluR5	6.22
s1cccc1Cc1cnc(SCC#C)nc1N	NAM	mGluR5	5.86
S(CC(OC[C@@H]1C[C@H]1C)=O)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	6.20
S(CC(OCC1CCC1)=O)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	5.84
S(CC(OCC1CC1)=O)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	6.70
S(Cc1onc(n1)C1CC1)c1nc(N)c(cn1)Cc1ccccc1	NAM	mGluR5	6.35
s1cccc1Cc1cnc(SCC=C)nc1N	NAM	mGluR5	5.55
s1cccc1Cc1cnc(SCc2ncon2)nc1N	NAM	mGluR5	6.40
Fc1ccc(N2CCN(CC2)c2nc(nc(OCC)c2[N+](=O)[O-])C)cc1	NAM	mGluR1	6.74
O=C1NC(=NC(N2CCC(CC2)c2ccccc2)=C1[N+](=O)[O-])C	NAM	mGluR1	7.20
Fc1ccc(N2CCN(CC2)C=2N=C(NC(=O)C=2[N+](=O)[O-])C)cc1	NAM	mGluR1	6.09
S(C)c1ccccc1N1CCN(CC1)c1nc(nc(NCCO)c1C#N)NCCO	NAM	mGluR1	6.46
Fc1ccc(cc1)C1CCN(CC1)C=1N=C(N(CCCO)C(=O)C=1[N+](=O)[O-])C	NAM	mGluR1	6.55
Fc1ccccc1N1CCN(CC1)c1nc(nc(NCCO)c1C#N)NCCO	NAM	mGluR1	6.54
OCCNc1nc(nc(N2CCN(CC2)c2ccccc2[N+](=O)[O-])c1C#N)NCCO	NAM	mGluR1	6.12
Fc1ccc(cc1)C1CCN(CC1)c1nc(nc(NCCO)c1C#N)NCCO	NAM	mGluR1	6.15
Fc1ccc(cc1)C=1CCN(CC=1)c1nc(nc(NCCO)c1C#N)NCCO	NAM	mGluR1	6.41
OCCNc1nc(nc(N2CCC(CC2)c2ccc(cc2)C#N)c1C#N)NCCO	NAM	mGluR1	5.74
n1c(N2CCC(CC2)c2ccccc2)c(C#N)c(nc1NCc1ncccc1)NCc1nccc1	NAM	mGluR1	5.82
OCCNc1nc(nc(N2CCN(CC2)c2ccccc2)c1C#N)NCc1cccnc1	NAM	mGluR1	6.82
Clc1nc(nc(N2CCC(CC2)c2ccc(F)cc2)c1C#N)NCCO	NAM	mGluR1	6.85
S(C)c1nc(nc(N2CCN(CC2)c2ccc(F)cc2)c1C#N)N	NAM	mGluR1	6.68
S(C)c1nc(nc(N2CCC(CC2)c2ccc(F)cc2)c1C#N)N	NAM	mGluR1	6.80
n1c(CC)c(nc(N2CCN(CC2)c2ccccc2)c1C#N)C	NAM	mGluR1	7.28
n1c(C)c(nc(N2CCN(CC2)c2ccccc2)c1C#N)CC	NAM	mGluR1	7.28
Fc1ccc(N2CCN(CC2)c2nc(NCCO)cnc2C#N)cc1	NAM	mGluR1	6.00
OCCNc1nc(N2CCC(=CC2)c2ccccc2)c(nn1)C#N	NAM	mGluR1	6.18
Fc1ccc(N2CCN(CC2)c2nc(C)c([n+]([O-])c2C#N)CC)cc1	NAM	mGluR1	7.48
O(C(=O)c1n(C)c(nc1C)C#Cc1ccccc1)CC	NAM	mGluR5	8.00
O(C(=O)c1[nH]c(nc1C)C#Cc1ccccc1)CC	NAM	mGluR5	7.66
O(C)c1cc(ccc1)C#Cc1nc(C)c(n1C)C(OCC)=O	NAM	mGluR5	7.25
n1ccn(C)c1C#Cc1ccccc1	NAM	mGluR5	8.00
[nH]1ccnc1C#Cc1ccccc1	NAM	mGluR5	7.66
OCCn1c(ncc1[N+](=O)[O-])C#Cc1ccccc1	NAM	mGluR5	7.66
Clc1cccc(Cl)c1C#Cc1nc(C)c(n1C)C(OCC)=O	NAM	mGluR5	8.00
O(C(=O)c1nc(n(c1C)-c1ccccc1)C#Cc1ccccc1)CC	NAM	mGluR5	7.66
O(C(=O)c1n(C)c(nc1C)C#Cc1cc(ccc1)C)CC	NAM	mGluR5	8.00
O(C(=O)c1n(C)c(nc1C)C#Cc1cc(NC(=O)C)ccc1)CC	NAM	mGluR5	7.57
O(C(=O)c1n(C)c(nc1C)C#Cc1cc(-n2c(ccc2C)C)ccc1)CC	NAM	mGluR5	7.66
o1nc(nc1-c1n(C)c(nc1C)C#Cc1ccccc1)C	NAM	mGluR5	7.66
Clc1ccc(cc1)C#Cc1nc(C)c(n1C)C(OCC)=O	NAM	mGluR5	8.00
Fc1ccc(cc1)C#Cc1nc(C)c(n1C)C(OCC)=O	NAM	mGluR5	8.00
O(C(=O)c1n(C)c(nc1C)C#Cc1ccc(cc1)-c1ccccc1)CC	NAM	mGluR5	8.00
Fc1ccccc1C#Cc1nc(C)c(n1C)C(OCC)=O	NAM	mGluR5	8.00
Fc1ccccc1C#Cc1nccn1C	NAM	mGluR5	8.00
O(C(=O)c1n(C)c(nc1C)C#Cc1ccc(N)cc1)CC	NAM	mGluR5	7.89
Clc1ccccc1C#Cc1nccn1C	NAM	mGluR5	8.00
Clc1n(CC(OCC)=O)c(nc1Cl)C#Cc1ccccc1	NAM	mGluR5	8.00
n1cc(n(c1)C)C#Cc1ccccc1	NAM	mGluR5	7.66
O(C)c1cc2c([nH]c(C#Cc3ccccc3)c2CCNC(=O)C)cc1	NAM	mGluR5	7.47
S1Cc2n(-c3cccnc13)cnc2C#Cc1ccccc1	NAM	mGluR5	7.66
O1Cc2n(-c3c1cccc3)cnc2C#Cc1ccccc1	NAM	mGluR5	7.66
ClC[C@@H](O)Cn1c([N+](=O)[O-])c(nc1C)C#Cc1ccccc1	NAM	mGluR5	7.66
O=Cc1n(cnc1C#Cc1ccccc1)C	NAM	mGluR5	7.66
[nH]1cc(nc1)C#Cc1ccccc1	NAM	mGluR5	7.66
n1cn(cc1C#Cc1ccccc1)C	NAM	mGluR5	7.66
O=[N+]([O-])c1n(C)c(nc1C#Cc1ccccc1)C	NAM	mGluR5	7.25

n1n(C)c(cc1C)C#Cc1ccccc1	NAM	mGluR5	7.66	
n1c(C(C)C)c(n(C)c1\C=C\c1ccccc1)C(C)C	NAM	mGluR5	7.25	
Fc1ccc(cc1)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
Clc1ccc(cc1)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
O(CCCC)c1ccc(cc1)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
O(C)c1cc(C)c(\C=C\c2nc(C(C)C)c(n2C)C(C)C)c(C)c1C	NAM	mGluR5	7.25	
O(C)c1ccc(cc1)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
Clc1ccc(cc1F)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
O(CC)c1ccc(cc1)\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
O(C)c1c(OC)c(OC)ccc1\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
Clc1cc(Cl)ccc1\C=C\c1nc(C(C)C)c(n1C)C(C)C	NAM	mGluR5	7.25	
Brc1ncn(C)c1\C=C\c1ccccc1	NAM	mGluR5	8.00	
n1cc(n(c1)C)\C=C\c1ccccc1	NAM	mGluR5	7.28	
O=C1CC(=C2N=CC(N(C)C)=C(N=C2C1)C#Cc1ccccc1)c1cc(cc1)C#N	NAM	mGluR2	7.52	
Fc1ccccc1C1=CNC2=C(N=C1OCC(F)(F)F)C=C(CC2=O)c1cc(ccc1)-c1onc(c1)C	NAM	mGluR2	7.52	
FC(F)(F)C1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-c1onc(c1)CN1CCOCC1	NAM	mGluR2	6.90	
FC(F)(F)C=1N=C2CC(=O)CC(=C2N=CC=1N(C)C1CC1)c1cc(ncc1)C#N	NAM	mGluR2	7.60	
FC(F)(F)C=1N=C2CC(=O)CC(=C2N=CC=1OCC(F)(F)F)c1cc(ncc1)C#N	NAM	mGluR2	7.60	
O=C1CC(=C2N=CC(N(CCC)C)=C(N=C2C1)C)c1cc(ccc1)C#N	NAM	mGluR2	7.60	
o1nc(cc1-c1cc(ccc1)C=1CC(=O)CC2=NC(C#N)=C(N(CC)CC)C=NC=12)C	NAM	mGluR2	7.77	
O1CCCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R
Clc1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R
Fc1nc2c(cc1CC)cc(cc2)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R
O1CCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R
S1CCCc2cc3cc(C(=O)Cc4ccccc4)c(cc3nc12)C	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3CCCCc3c2)cc1	NAM	mGluR1	7.49	R
S1CCCc2cc3cc(ccc3nc12)C(=O)C1CCC(OC)CC1	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2c(nc(cc2)CCC)cc1	NAM	mGluR1	7.49	R
FC1(CCC(OC)CC1)C(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.49	R
O1CCCc2cc3cc(C(=O)Cc4ccccc4)c(cc3nc12)C	NAM	mGluR1	7.49	R
Fc1ccccc1CC(=O)c1cc2c(nc3OCCCc3c2)cc1	NAM	mGluR1	7.49	R
S1CCCc2cc3cc(ccc3nc12)C(=O)Cc1ccccc1	NAM	mGluR1	7.49	R
O1CCCc2cc3cc(ccc3nc12)C(=O)C[C@@H]1[C@@H]2CC[C@H](C1)C2	NAM	mGluR1	7.49	R
S1CCCc2cc3cc(ccc3nc12)C(=O)Cc1ccccc1F	NAM	mGluR1	7.49	R
s1cc(cc1)CC(=O)c1cc2c(cc1)cc1OCCCc1c2	NAM	mGluR1	7.49	R
O1c2nc3c(cc(cc3)C(=O)Cc3ccccc3)cc2C[C@H]1C	NAM	mGluR1	7.49	R
S1CCc2cc3cc(ccc3nc12)C(=O)C1CCCCC1	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1)N	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2cc(CCC)c(nc2cc1)C	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2cc(CC)c(nc2cc1C)C	NAM	mGluR1	7.49	R
O(C)C1CCC(CC1)C(=O)c1cc2c(nc3CCCCc3c2)cc1	NAM	mGluR1	7.49	R
FC(F)(F)C=1N=C2CC(=O)CC(=C2N=CC=1N1CCOCC1)c1cc(ncc1)C#N	NAM	mGluR2	7.60	
O=C1CC(=C2N=CC=C(n3cccc3)N=C2C1)c1cc(ccc1)C#N	NAM	mGluR2	7.46	
IC=1C=NC=2C(=NC=1n1ccccc1)CC(=O)CC=2c1cc(ncc1)C#N	NAM	mGluR2	7.12	
O=C1CC(=C2N=CC=C(n3cc(cc3)-c3ccccc3)N=C2C1)c1cc(ccc1)C#N	NAM	mGluR2	7.12	
O(C)C=1C=NC=2C(=NC=1n1cccc1)CC(=O)CC=2c1cc(ccc1)C#N	NAM	mGluR2	7.36	
O(C)C=1C=NC=2C(=NC=1n1cccc1C(C)(C)C)CC(=O)CC=2c1cc(ccc1)C#N	NAM	mGluR2	7.10	
o1cnc(C(OCC)=O)c1-c1cc(ccc1)C=1CC(=O)CC2=NC(n3cccc3)=CC=NC=12	NAM	mGluR2	7.54	
o1cc(nc1-c1cc(ccc1)C=1CC(=O)CC2=NC(n3cccc3)=CC=NC=12)C(=O)N	NAM	mGluR2	7.21	
o1cc(nc1-c1cc(ccc1)C=1CC(=O)CC2=NC(n3cccc3)=CC=NC=12)C(=O)NCCO	NAM	mGluR2	7.04	
FC(F)(F)C=1N=C2CC(=O)CC(=C2N=CC=1OCC)c1cc(ncc1)C#N	NAM	mGluR2	7.60	
FC(F)(F)C=1N=C2CC(=O)CC(=C2N=CC=1C)c1cc(ncc1)C#N	NAM	mGluR2	7.60	

10. Appendix

o1nc(cc1-c1ccc(cc1)C=1CC(=O)CC2=NC(C#N)=C(N3CCOCC3)C=NC=12)C	NAM	mGluR2	7.49	
O[C@@]1([C@@H]2[C@H](N(CC2)C(OC)=O)CCC1)C#Cc1cc(ccc1)C	NAM	mGluR5	7.60	
O(C)c1cc2c(ncnc2N[C@H]2[C@@H]3CC[C@H](C2)C3)cc1	NAM	mGluR1	5.61	
O(C)c1cc2c(ncnc2NC2CCCCC2)cc1	NAM	mGluR1	6.48	
Clc1nc(NC2Cc3c(C2)cccc3)c2cc(OC)ccc2n1	NAM	mGluR1	7.33	
Clc1ccccc1CCNc1ncnc2c1cc(OC)cc2	NAM	mGluR1	6.52	
Clc1nc(Nc2ccc(OC)cc2)c2cc(OC)ccc2n1	NAM	mGluR1	7.33	
Clc1cc2c(nc(Cl)nc2NC2Cc3c(C2)cccc3)cc1	NAM	mGluR1	7.52	
Clc1ccccc1CCNc1nc(Cl)nc2c1cc(OC)cc2	NAM	mGluR1	7.33	
Clc1cc2c(ncnc2N[C@H]2[C@H]3CC[C@@H](C2)C3)cc1	NAM	mGluR1	5.95	
S(CC)c1nc(NC[C@H](F)c2ccccc2)c2CCCCc2n1	NAM	mGluR1	7.57	
S(CC)c1nc(NOC)c2CCCCc2n1	NAM	mGluR1	7.49	
Clc1ccccc1[C@@H](O)CNc1nc(SCC)nc2c1CCCC2	NAM	mGluR1	7.57	
S(CC)c1nc(NCC(F)(F)c2ccccc2)c2CCCCc2n1	NAM	mGluR1	7.57	
Clc1ccccc1CCNc1nc(SCC)nc2c1CCCC2	NAM	mGluR1	7.57	
S(CC)c1nc(Nc2ccc(F)cc2)c2CCCCc2n1	NAM	mGluR1	7.49	
S(CC)c1nc(NC2Cc3c(C2)cccc3)c2CCCCc2n1	NAM	mGluR1	7.41	
Clc1cccc(Cl)c1CSCCNc1nc(SCC)nc2c1CCCC2	NAM	mGluR1	6.61	
S(CC)c1nc(NN2Cc3c(C2)cccc3)c2CCCCc2n1	NAM	mGluR1	6.26	
Clc1ccccc1[C@@H](OC)CNc1nc(SCC)nc2c1CCCC2	NAM	mGluR1	6.21	
Clc1ccccc1OCCNc1nc(SCC)nc2c1CCCC2	NAM	mGluR1	6.00	
S(CC)c1nc(N[C@H]2[C@H]3CC[C@@H](C2)C3)c2CCCCc2n1	NAM	mGluR1	6.49	
O(C(=O)c1c[nH]c(C(OCC)=O)c1C)C(C)(C)C	NAM	mGluR1	6.47	+
O(C(=O)c1c[nH]c(C(OCCC)=O)c1C)C(C)(C)C	NAM	mGluR1	7.32	+
O(C(=O)c1c(C)c([nH]c1COC(=O)C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	8.30	+
O(C(=O)c1c(C)c([nH]c1C=O)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.70	+
O(C(=O)c1c(C)c([nH]c1C(OC)=O)C(OCCC)=O)C(C)(C)C	NAM	mGluR1		R	+
O(C(=O)c1c(C)c([nH]c1C(=O)NCC1CC1)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	5.60	+
O(C(=O)c1c[nH]c(C(OC(C)(C)C)=O)c1C)C(C)(C)C	NAM	mGluR1	7.00	+
o1cccc1COC(=O)c1[nH]cc(C(OC(C)(C)C)=O)c1C	NAM	mGluR1	5.92	+
O(C(=O)c1[nH]cc(C(OC(C)(C)C)=O)c1C)CC1CC1	NAM	mGluR1	7.32	+
O(C(=O)c1[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c1C)CCC	NAM	mGluR1	7.32	+
O(C(=O)c1[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c1C)C(C)(C)C	NAM	mGluR1	7.32	+
Fc1c(OC(=O)c2[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c2C)c(F)c(F)c(F)c1F	NAM	mGluR1	6.80	+
O([C@@H](C)(C)(C)C)C(=O)c1c[nH]c(C(O[C@@H](CN(C)C)C)=O)c1C	NAM	mGluR1	6.41	+
Clc1[nH]c(C(OCCC)=O)c(C)c1C(O[C@@H](C)(C)(C)C)C)=O	NAM	mGluR1	6.80	+
O(C(=O)c1[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c1C)c1cccnc1	NAM	mGluR1	6.14	+
O(C(=O)c1[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c1C)[C@@H]1CCN(C1)CC	NAM	mGluR1	6.59	+
O(C(=O)c1[nH]cc(C(O[C@@H](C)(C)(C)C)=O)c1C)c1ncncc1	NAM	mGluR1	5.84	+
O(C(=O)c1c(C)c(n(C)c1C)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	8.05	+
O(C(=O)c1n(CCCO)c(C)c(C(O[C@@H](C)(C)(C)C)=O)c1C)CCC	NAM	mGluR1	8.05	+
O(C(=O)c1n(nc(c1)C(O[C@@H](C)(C)(C)C)=O)C)CCC	NAM	mGluR1	7.59	
O(C(=O)c1nn(C)c(c1)C(O[C@@H](C)(C)(C)C)=O)CCC	NAM	mGluR1	7.59	
O([C@@H](C)(C)(C)C)C(=O)c1n(nc(c1)C(O[C@@H](C)(C)(C)C)=O)C	NAM	mGluR1	7.59	
O(C(=O)c1n(nc(c1)C(O[C@@H](C)(C)(C)C)=O)C)CC	NAM	mGluR1	7.59	
O(C(=O)c1n[nH]c(c1)C(O[C@@H](C)(C)(C)C)=O)CCCC	NAM	mGluR1	7.59	
O(C(=O)c1n[nH]c(c1)C(O[C@@H](C)(C)(C)C)=O)CCC	NAM	mGluR1	7.59	
s1c(C(OC)=O)c(C)c(C(OCC)=O)c1N	NAM	mGluR1	7.29	
s1c(C)c(C(OCC)=O)c(N)c1C(OCC)=O	NAM	mGluR1	7.29	
s1cc(C(OCC)=O)c(C)c1C(OCC)=O	NAM	mGluR1	7.29	
s1cc(C(OC(C)(C)C)=O)c(C)c1C(OCCC)=O	NAM	mGluR1	7.29	
s1cc(C(O[C@@H](C)(C)(C)C)=O)c(C)c1C(OCCC)=O	NAM	mGluR1	7.29	
s1c(C(OCCC)=O)c(C)c(C(O[C@@H](C)(C)(C)C)=O)c1N	NAM	mGluR1	7.32	
O(C(=O)c1ccc(cc1)C(OCCC)=O)CCC	NAM	mGluR1	7.59	
O(C(=O)c1ccc(cc1)C(OCCC)=O)C(C)(C)C	NAM	mGluR1	7.59	
O(C(=O)c1ccc(cc1)C(O[C@@H](C)(C)(C)C)=O)CCC	NAM	mGluR1	7.59	
O(C(=O)c1ccc(cc1)C(O[C@@H](C)(C)(C)C)=O)CC1CC1	NAM	mGluR1	7.59	

O(C)c1cc(ccc1-c1cccnc1)-c1nc2n(c1)C=CC=C2	NAM	mGluR5	7.04
S1C=Cn2cc(nc12)-c1cc(OC)c(cc1)-c1cccnc1	NAM	mGluR5	7.03
S1CCn2cc(nc12)-c1cc(OC)c(cc1)-c1cccnc1	NAM	mGluR5	6.49
Clc1cc2C3=C(CCCC3)C(Oc2cc1OC(C)C)=O	NAM	mGluR1	9.00
Fc1cc(-n2c(C)c(cc2C)C=C2C(=O)c3c(cccc3)C2=O)ccc1	NAM	mGluR1	7.40
S\1c2c(cccc2)C(=O)/C/1=C/c1ccc(OC)c(CC)c1OC	NAM	mGluR1	7.24
OC=1n2nc(cc2N=C(C=1)c1ccccc1)-c1cc(ccc1)C	NAM	mGluR1	7.28
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccc(cc1)C	NAM	mGluR5	7.96
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cc(ccc1)C#N	NAM	mGluR5	7.92
O(c1ccccc1-n1nc(nn1)-c1ncccc1)c1ncccc1	NAM	mGluR5	7.96
[nH]1cc(nc1)-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	5.70
[nH]1nccc1-c1nn(nn1)-c1cc(ccc1)C#N	NAM	mGluR5	5.70
O=[N+]([O-])c1cc(ccc1)-c1nc2n(c1)C=CC=C2	NAM	mGluR5	8.06
n1cn(cc1-c1cc(ccc1)C#N)-c1ncccc1	NAM	mGluR5	5.00
n1ccccc1-n1nnc(c1)-c1cc(ccc1)C#N	NAM	mGluR5	5.00
Clc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.82
Fc1cccc(F)c1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.82
O(C(=O)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1)C	NAM	mGluR5	6.65
FC(F)(F)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.64
Oc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.54
O(C)c1ccc(cc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1)C	NAM	mGluR5	6.51
O(CC)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.50
Fc1ccc(F)cc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.47
FC(F)(F)Oc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.42
O=C(C)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.38
O(c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1)c1ccccc1	NAM	mGluR5	6.35
O(Cc1ccccc1)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.35
O(C)c1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.34
O=Cc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.30
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccccc1CC	NAM	mGluR5	6.18
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1c(cccc1C)C	NAM	mGluR5	6.02
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccccc1-c1ccccc1	NAM	mGluR5	6.01
O(C(=O)c1cc(ccc1)-c1cc(-n2nc(nn2)-c2ncccc2)ccc1)C	NAM	mGluR5	6.47
Clc1cc(ccc1)-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.02
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1cc(ccc1)C	NAM	mGluR5	5.93
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccc(cc1)C	NAM	mGluR5	5.81
Clc1ccc(cc1)-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	5.52
Clc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1	NAM	mGluR5	6.97
Fc1cccc(F)c1-c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1	NAM	mGluR5	6.97
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1O	NAM	mGluR5	6.70
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1OC	NAM	mGluR5	6.57
Fc1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1	NAM	mGluR5	6.54
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1[N+](=O)[O-]	NAM	mGluR5	6.51
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1OC(F)(F)F	NAM	mGluR5	6.50
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1C=O	NAM	mGluR5	6.44
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1C(=O)N	NAM	mGluR5	6.41
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1Oc1ccccc1	NAM	mGluR5	6.38
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1OCC	NAM	mGluR5	6.35
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1c(cccc1C)C	NAM	mGluR5	6.06
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1-c1ccccc1	NAM	mGluR5	6.00
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccccc1CC	NAM	mGluR5	5.52
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cc(F)ccc1	NAM	mGluR5	6.66
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cc(ccc1)C(OC)=O	NAM	mGluR5	6.50
Clc1cc(ccc1)-c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1	NAM	mGluR5	6.08
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1ccc(F)cc1	NAM	mGluR5	6.47
Clc1ccc(cc1)-c1cc(-n2nc(nn2)-c2ncccc2)cc(F)c1	NAM	mGluR5	5.69
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1cccnc1	NAM	mGluR5	7.16
o1nc(C)c(-c2cc(-n3nc(nn3)-c3ncccc3)ccc2)c1C	NAM	mGluR5	6.60
O(C)c1ncc(cn1)-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	6.51
o1nc(cc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1)-c1ccccc1	NAM	mGluR5	6.05
o1ccccc1-c1cc(-n2nc(nn2)-c2ncccc2)ccc1	NAM	mGluR5	5.52
n1ccccc1-c1nn(nn1)-c1cc(ccc1)-c1ccncc1	NAM	mGluR5	5.68
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1c(noc1C)C	NAM	mGluR5	7.55

193

10. Appendix

Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1onc(c1)-c1ccccc1	NAM	mGluR5	7.12	
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cnc(OC)nc1	NAM	mGluR5	7.03	
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1cncnc1	NAM	mGluR5	6.70	
Fc1cc(cc(-n2nc(nn2)-c2ncccc2)c1)-c1occc1	NAM	mGluR5	6.24	
c1c(cccc1C)C#Cc1ccccc1	NAM	mGluR5	5.96	
n1ccccc1C#Cc1ccccc1	NAM	mGluR5	7.04	
n1cccc(C)c1C#Cc1ccccc1	NAM	mGluR5	7.20	
n1c(cc(cc1C)C)C#Cc1ccccc1	NaM	mGluR5	7.21	
Brc1nc(ccc1)C#Cc1ccccc1	NAM	mGluR5	6.18	
S(C)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.68	
O(CC=C)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.53	
n1c(cccc1C)C#Cc1ncccc1	NAM	mGluR5	6.74	+
n1c(cccc1C)C#Cc1ccncc1	NAM	mGluR5	7.03	+
O(C)c1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	7.66	
Brc1cc(cnc1)C#Cc1nc(ccc1)C	NAM	mGluR5	7.92	
O(C)c1nc(ccc1)C#Cc1ccccc1	NAM	mGluR5	5.71	
O(C)c1cc(OC)ccc1C#Cc1nc(ccc1)C	NAM	mGluR5	6.56	
O(C)c1ccc(OC)cc1C#Cc1nc(ccc1)C	NAM	mGluR5	7.09	
O(C)c1cc(ccc1OC)C#Cc1nc(ccc1)C	NAM	mGluR5	6.09	
O1c2cc(ccc2OC1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.20	
O=[N+]([O-])c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	7.87	
FC(F)(F)c1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.45	
Fc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.45	
OCc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	6.62	
n1c(cccc1C)C#Cc1cc2c(nc1)cccc2	NAM	mGluR5	6.63	
n1c(cccc1C)C#Cc1c2c(cccc2)cnc1	NAM	mGluR5	7.68	
n1c(cccc1C)C#Cc1cc2n(ccc2cc1)C	NAM	mGluR5	5.74	
n1c(cccc1C)C#Cc1cc2c(n(cc2)C)cc1	NAM	mGluR5	6.26	
n1c(cccc1C)C#Cc1cncnc1	NAM	mGluR5	6.76	
s1cc(nc1C)C#Cc1cc(OC)ccc1C	NAM	mGluR5	7.25	
s1cc(nc1C)C#Cc1cc(ccc1)CO	NAM	mGluR5	6.37	
s1cc(nc1C)C#Cc1c2c(cccc2)cnc1	NAM	mGluR5	7.70	
s1cc(nc1C)C#Cc1cc([N+](=O)[O-])ccc1	NAM	mGluR5	7.82	
Brc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	8.70	
Clc1c(cc(Cl)cc1Cl)\C=C\c1nc(ccc1)C	NAM	mGluR5	8.06	
O(C)c1ccc(nc1C#Cc1cc(ccc1)C)C	NAM	mGluR5	7.66	
Clc1c(cc(Cl)cc1Cl)C#Cc1nc(ccc1)C	NAM	mGluR5	8.06	
Clc1cc(cc(Cl)c1)C#Cc1nc(ccc1)C	NAM	mGluR5	8.70	
FC(F)(F)Oc1cc(ccc1)C#Cc1nc(ccc1)C	NAM	mGluR5	8.06	
Fc1cc(ccc1)C#Cc1nc(cc(c1)C)C	NAM	mGluR5	8.00	
n1c(cccc1C)C#Cc1cc(ccc1)C#C	NAM	mGluR5	8.70	
Fc1ccc(-n2cc(nc2C)C#Cc2cc(F)ncc2)cc1	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1nc(ccc1)C(F)(F)F)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1cc(F)cnc1)C	NAM	mGluR5	7.20	
Fc1ccc(-n2cc(nc2C)C#Cc2cc(ncc2)C#N)cc1	NAM	mGluR5	7.36	
Clc1cc(-n2cc(nc2C)C#Cc2cc(Cl)ncc2)ccc1Cl	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1cc(C)c(F)cc1)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1ccc(cc1)C)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1ccc(OC)cc1)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1nc(ccn1)C)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1nc(OC)ccn1)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1nc(ncc1)C(F)(F)F)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1ncc(cc1)C)C	NAM	mGluR5	7.37	
Clc1ncc(-n2cc(nc2C)C#Cc2cc(ncc2)C#N)cc1	NAM	mGluR5	7.16	
Clc1nccc(c1)C#Cc1nc(n(c1)-c1ccc(nc1)C1CC1)C	NAM	mGluR5	7.37	
Clc1nccc(c1)C#Cc1nc(n(c1C)-c1ccc(F)cc1)C	NAM	mGluR5	7.37	
o1c2c(nc1-c1cc(OC)c(cc1)-c1ccncc1)cccc2	NAM	mGluR5	7.38	
O=C(Nc1ncccc1)c1nccnc1	NAM	mGluR5	7.44	
O=C(Nc1nc(ccc1)C)c1nc(ccc1Nc1cccnc1)C	NAM	mGluR5	7.44	
Brc1cc(cnc1)C(=O)Nc1ncccc1	NAM	mGluR5	7.44	
Brc1cc(ccc1)C(=O)Nc1ncccc1	NAM	mGluR5	7.27	
S(Oc1cc(ccc1)C#Cc1nc(ccc1)C)(=O)(=O)c1ccc(cc1)C	NAM	mGluR5	8.70	
O=C1N(c2ccc(cc2)C(OCC(=O)c2ccc(cc2)C)=O)C(=O)[C@H]2[	NAM	mGlur5	5.02	

194

C@H]1CC=CC2

SMILES				
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	6.76	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc3CCCc3cc2)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	6.40	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc3CCCCc3cc2)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	6.77	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc3c(cc2)cccc3)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	6.52	
O=C1N(c2ccc(cc2)C(OCC(=O)c2ccc(cc2)C2CCCCC2)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	4.74	
O(C)c1cc(ccc1OC)C(=O)COC(=O)c1ccc(N2C(=O)[C@H]3[C@@H](CC=CC3)C2=O)cc1	NAM	mGlur5	5.79	
O1CCOc2c1cc(cc2)C(=O)COC(=O)c1ccc(N2C(=O)[C@H]3[C@@H](CC=CC3)C2=O)cc1	NAM	mGlur5	5.69	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc([N+](=O)[O-])c(cc2)C)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	5.42	
O(C)c1ccc(cc1[N+](=O)[O-])C(=O)COC(=O)c1ccc(N2C(=O)[C@H]3[C@@H](CC=CC3)C2=O)cc1	NAM	mGlur5	5.77	
O=C1N(c2ccc(cc2)C(OCC(=O)c2ccc(N3CCCC3)cc2)=O)C(=O)[C@H]2[C@H]1CC=CC2	NAM	mGlur5	5.96	
O1CCN(CC1)c1ccc(cc1)C(=O)COC(=O)c1ccc(N2C(=O)[C@H]3[C@@H](CC=CC3)C2=O)cc1	NAM	mGlur5	5.47	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@@H]2[C@H]1CCCC2	NAM	mGlur5	6.72	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@H]2[C@H]1CCCC2	NAM	mGlur5	6.82	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@H]2[C@H]1C[C@@H](CC2)C	NAM	mGlur5	6.95	
Br[C@@H]1C[C@@H]2[C@@H](C[C@H]1Br)C(=O)N(c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O)C2=O	NAM	mGlur5	7.01	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@H]2[C@H]1CC(=CC2)C	NAM	mGlur5	6.67	
O=C1N(c2ccc(cc2)C(OCC(=O)c2cc(C)c(cc2)C)=O)C(=O)[C@@H]2[C@H]1[C@H]1C=C[C@@H]2C1	NAM	mGlur5	5.99	
O=C1N(C(=O)CC1)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	6.30	
O=C1N(C(=O)C[C@H]1C)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	6.39	
O=C1N(C(=O)C=C1C)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	5.32	
O=C1N(C(=O)C(C)=C1C)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	5.19	
O=C1N(C(=O)CCC1)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	5.49	
O=C1N(C(=O)CC(C1)C)c1ccc(cc1)C(OCC(=O)c1cc(C)c(cc1)C)=O	NAM	mGlur5	5.74	
Clc1cc(-n2nc(nn2)[C@@H]2N(CCCC2)c2nnc(n2C)-c2ccncc2)ccc1	NAM	mGluR5	7.39	H
s1c2c(N=CN(C2=O)c2ccccc2)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c1nccc2N(C)C	NAM	mGluR5	6.05	
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2N(C)C	NAM	mGluR5	6.35	
s1c2c(N=CN(C2=O)c2ccc(cc2)CCC)c2c1nccc2N(C)C	NAM	mGluR5	5.69	
s1c2c(N=CN(C2=O)c2ccc(cc2)C(C)C)c2c1nccc2N(C)C	NAM	mGluR5	5.45	
s1c2c(N=CN(C2=O)c2ccc(cc2)C(C)(C)C)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(C2=O)c2ccc(cc2)C2CC2)c2c1nccc2N(C)C	NAM	mGluR5	5.34	
s1c2c(N=CN(C2=O)c2ccc(cc2)C2CCCCC2)c2c1nccc2N(C)C	NAM	mGluR5	3.00	
s1c2c(N=CN(C2=O)c2ccc(cc2)C2CCC2)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(C2=O)c2ccc(cc2)C2CCCC2)c2c1nccc2N(C)C	NAM	mGluR5	5.94	
s1c2c(N=CN(C3CCCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.19	
s1c2c(N=CN([C@@H]3CCCC[C@H]3C)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN([C@H]3C[C@@H](CCC3)C)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.67	
s1c2c(N=CN(C3CCC(CC3)C)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.23	
s1c2c(N=CN(C3CCCCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.83	
s1c2c(N=CN(C34CC5CC(C3)CC(C4)C5)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.81	
s1c2c(N=CN(C(CC)CC)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(CC3CCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(CC3CCCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=CN(C3CC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	5.00	
s1c2c(N=C(N(C2=O)c2ccc(cc2)CC)C)c2c1nccc2N(C)C	NAM	mGluR5	6.35	

10. Appendix

s1c2c(N=C(N(C2=O)c2ccc(cc2)CC)CC)c2c1nccc2N(C)C	NAM	mGluR5	5.34		
s1c2c(N=C(N(C2=O)c2ccc(cc2)CC)CCCC)c2c1nccc2N(C)C	NAM	mGluR5	6.05		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1cccc2N(C)C	NAM	mGluR5	5.00		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1cncc2N(C)C	NAM	mGluR5	5.67		
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c1cncc2N(C)C	NAM	mGluR5	5.39		
s1c2c(N=CN(C2=O)c2ccc(cc2)C)c2c1nccc2N(CC)C	NAM	mGluR5	5.59		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2N(CCOC)C	NAM	mGluR5	5.00		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2N1CCCC1	NAM	mGluR5	5.00		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2NC1CC1	NAM	mGluR5	5.00		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2NCC#N	NAM	mGluR5	5.68		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2NCC(F)(F)F	NAM	mGluR5	5.00		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2NC	NAM	mGluR5	5.22		
s1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2[N+]([O-])(C)C	NAM	mGluR5	5.00		
s1c2c(N=CN(N3CCCCCC3)C2=O)c2c1nccc2[N+]([O-])(C)C	NAM	mGluR5	5.00		
o1c2c(N=CN(C2=O)c2ccc(cc2)CC)c2c1nccc2N(C)C	NAM	mGluR5	5.00		
o1c2c(N=CN(C3CCC(CC3)C)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.05		
s1c2c(N=NN(C3CCCCCC3)C2=O)c2c1nccc2N(C)C	NAM	mGluR5	6.09		
s1ccnc1N1CCN(CC1)C(=O)C#Cc1ccccc1	NAM	mGluR5	7.52		+
O=C1N(CNc2cc(ccc2)C)C(=O)[C@@H]2[C@H]1[C@H]1C=C[C@@H]2C1	NAM	mGluR5	7.72		
O1c2c(C[C@@H]1CN(C(OC)=O)C)cccc2	NAM	mGluR5	5.00	R	-
O1CCN(CC1)C(=O)CCCOc1ccc(cc1)C	NAM	mGluR5	5.46	R	-
O1CCN(CC1)C(=O)Cc1ccc(OCCCC)cc1	NAM	mGluR5	5.00	R	-
O=C(N1CCCC1)\C=C\c1ccccc1	NAM	mGluR5	5.00	R	-
Clc1ccc(cc1)\C=C\C(=O)N1CCCC1	NAM	mGluR5	5.00	R	-
O=C1N(CCC1)CC(=O)N(Cc1ccccc1)C	NAM	mGluR5	5.00	R	-
O1c2cc(ccc2OC1)\C=C\C(=O)N1CCC(CC1)C	NAM	mGluR5	4.48	R	-
S\1c2c(N(C)/C/1=C/C(=O)C(C)C)cccc2	NAM	mGluR5	4.62	R	-
S\1c2c(N(C)/C/1=C\C(=O)C1CCC1)cccc2	NAM	mGluR5	5.00	R	-
O1c2c(C[C@@H]1C(=O)N1CCCCC1)cccc2	NAM	mGluR5	5.00	R	-
O=C1N(C)C(=O)N(c2c1cn(c2)-c1cc(ccc1)C)C	NAM	mGluR5	5.00	R	-
O=C(\C=C\c1ccccc1)c1nc(ccc1)C	NAM	mGluR5	7.55		
n1c(C(C)C)c(n(C)c1\C=C\c1ccc(cc1)C)C(C)C	NAM	mGluR5	7.25		
Fc1c(cccc1F)C1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(-n2nncc2)ccc1	NAM	mGluR2	7.15		
ClC1=CNC2=C(N=C1N(CCOC)C)C=C(CC2=O)c1cc(ccc1)-c1onc(c1)C	NAM	mGluR2	7.60		
ClC1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-c1n(ncc1)C	NAM	mGluR2	7.64		
ClC1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-c1scc(n1)CO	NAM	mGluR2	7.52		
ClC1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-c1occ(n1)CO	NAM	mGluR2	7.41		
ClC1=CNC2=C(N=C1N(CCC)C)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.52		
ClC1=CNC2=C(N=C1N(CC)CC)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.36		
ClC1=CNC2=C(N=C1N1CCCC1)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.72		
ClC1=CNC2=C(N=C1N(C)C1CC1)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	6.80		
ClC1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(-n2nccc2)ccc1	NAM	mGluR2	6.96		
S(C)c1nccn1-c1noc(c1)-c1cc(ccc1)C=1CC(=O)C=2NC=C(C(F)(F)F)C(=NC=2C=1)N(C)C	NAM	mGluR2	7.72		
FC(F)(F)C1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(-n2nncc2CNC2CC2)ccc1	NAM	mGluR2	7.31		
ClC1=CNC2=C(N=C1N([C@@H](CC)C)C)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.70		
ClC1=CNC2=C(N=C1N(CCC)C)C=C(CC2=O)c1cc(-n2nncc2CN2CCC2)ccc1	NAM	mGluR2	7.77		
ClC1=CNC2=C(N=C1N([C@@H](CC)C)C)C=C(CC2=O)c1cc(-n2ncnc2CO)ccc1	NAM	mGluR2	7.05		
FC(F)(F)C1=CNC2=C(N=C1N(CCC)C)C=C(CC2=O)c1cc(-n2ncnc2)ccc1	NAM	mGluR2	7.57		
FC(F)(F)C1=CNC2=C(N=C1N([C@@H](CC)C)C)C=C(CC2=O)c1cc(-n2ncnc2)ccc1	NAM	mGluR2	7.77		
ClC1=CNC2=C(N=C1N([C@@H](CC)C)C)C=C(CC2=O)c1cc(ccc1)-c1scc(n1)CO	NAM	mGluR2	7.77		
ClC1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-	NAM	mGluR2	6.32		

196

c1nc(sc1)NCC			
s1c(nnc1CO)-c1cc(ccc1)C=1CC(=O)C=2NC=C(C(F)(F)F)C(=NC=2C=1)N(C)C	NAM	mGluR2	7.77
FC(F)(F)C1=CNC2=C(N=C1N(C)C)C=C(CC2=O)c1cc(ccc1)-c1nc(oc1CCC)C	NAM	mGluR2	7.34
s1c(nnc1CO)-c1cc(ccc1)C=1CC(=O)C=2NC=C(C(F)(F)F)C(=NC=2C=1)N(CC)C	NAM	mGluR2	7.77
FC(F)(F)C1=CNC2=C(N=C1N1CCOCC1)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.13
O=C1CC(=CC=2N=CC(n3cccc3)=CNC1=2)c1cc(-n2ccnc2)ccc1	NAM	mGluR2	7.60
s1cc(nc1-c1cc(ccc1)C=1CC(=O)C=2NC=C(n3cccc3)C=NC=2C=1)CO	NAM	mGluR2	7.55
O=C1CC(=CC=2N=CC(n3cccc3)=CNC1=2)c1cc(-n2nncc2)ccc1	NAM	mGluR2	7.77
o1ccnc1-c1cc(ccc1)C=1CC(=O)C=2NC=C(n3cccc3)C=NC=2C=1	NAM	mGluR2	7.64
O1CCN(CC1)C1=NC2=C(NC=C1)C(=O)CC(=C2)c1cc(ccc1)-c1n(ncc1)C	NAM	mGluR2	7.77
FC(F)(F)C1=CNC2=C(N=C1OCC)C=C(CC2=O)c1cc(-n2nncc2CO)ccc1	NAM	mGluR2	7.08
ClC1=NC2=C(NC=C1C(F)(F)F)C(=O)CC(=C2)c1cc(-n2nncc2CNC2CC2)ccc1	NAM	mGluR2	7.77
FC(F)(F)C1=CNC2=C(N=C1C)C=C(CC2=O)c1cc(-n2nncc2CNC2CC2)ccc1	NAM	mGluR2	7.59
s1cc(nc1CO)-c1cc(ccc1)C=1CC(=O)C=2NC=C(C(F)(F)F)C(=NC=2C=1)C	NAM	mGluR2	7.77
FC(F)(F)C1=CNC2=C(N=C1C)C=C(CC2=O)c1cc(-n2nccc2)ccc1	NAM	mGluR2	7.15
Clc1cc(Cl)ccc1/C(/OC(C)C)=C\n1ncnc1	NAM	mGluR2	6.19
Brc1ccc(cc1)/C(/OCCCC)=C\n1ncnc1	NAM	mGluR2	6.18
Clc1ccc(cc1)/C(/OCCCC)=C\n1ncnc1	NAM	mGluR2	6.22
Clc1cccc(Cl)c1/C(/OCCCC)=C\n1ncnc1	NAM	mGluR2	6.22
Clc1cc(Cl)ccc1/C(/OCc1ccccc1)=C\n1ncnc1	NAM	mGluR2	6.37
Clc1cc(Cl)ccc1/C(/OC1CCCCC1)=C\n1ncnc1	NAM	mGluR2	7.00
Clc1cc(Cl)ccc1/C(/OCc1ccccc1)=C\n1nnnc1	NAM	mGluR2	6.57
Clc1cc(Cl)ccc1/C(/OCc1ccc(OC)cc1)=C\n1ncnn1	NAM	mGluR2	6.00
Clc1cc(Cl)ccc1[C@@H](OC(=O)c1ccccc1)Cn1nnnc1	NAM	mGluR2	6.05
O\1\C(=C\2/N(c3c(cc(cc3)C)C=C/2)CC)\C(=O)N(CC)/C/1=N/c1ccc(Oc2ccccc2)cc1	NAM	mGluR1?	7.34
S(=O)(=O)(CCc1oc(cc1)\C=C\1/C(C(OC)=O)=C(N(C/1=O)c1ccccc1)C)c1ccc(cc1)C	NAM	mGluR1?	7.49
O(C)c1ccc(cc1)\C=C\C(=O)N1CCCCCC1	NAM	mGluR5	5.92
Oc1ccc(cc1C(=O)Nc1nc(ccc1)C)C	NAM	mGluR5	7.44
O=C(N(C)c1nc(ccc1)C)c1nc(cnc1N)C	NAM	mGluR5	7.44

10.5 Vendor Ligand Data Bases

Table A 2: Vendor ligand data collections sources and versions.

Data source URL	Version
www.acbblocks.com	2007.3
www.acros.com	2006.4
www.akosgmbh.de	2008.7
www.alfa.com	2008.2
www.amriglobal.com	2007.9
www.analyticon.com	2007.2
www.apolloscientific.co.uk	2006.6
www.asinex.com	2006.11
www.bachem.com	2005.6
www.biofocus.com	2007.4
www.chemdiv.com	2007.8

10. Appendix

www.chemblock.com	2007.3
www.chemstar.ru	2008
www.combi-blocks.com	2007.4
www.enamine.net	2008.1
www.frinton.com	2006.12
www.ibscreen.com	2008.2
www.keyorganics.ltd.uk	2008.2
www.lifechemicals.com	2008.8
www.maybridge.com	2008.5
www.mdpi.org	2005.6
www.msdiscovery.com	2008.1
http://mlsmr.glpg.com/MLSMR_HomePage/	2005.9
www.nanosyn.com	2008.1
www.nchlab.com	2008.4
www.oakwoodchemical.com	2007.4
www.otavachemicals.com	2007.12
www.peakdale.com	2006
www.pharmeks.com	2007.4
www.prestwickchemical.fr	2008.1
http://pubchem.ncbi.nlm.nih.gov/	2003.11
www.sigmaaldrich.com	2006.5
www.specs.net	2008.1
www.synphabase.ch	2006.12
www.timtec.net	2008.1
www.ubichem.com	2008.1
www.toslab.com	2008.1
www.vitasmlab.com	2008.2

10.6 PyMol script for mutation highlighting

```
Load mGluR5_structure.pdb
color gray
show cartoon
hide lines
cmd.hide("(mGluR5_structure and hydro)")
hide all

select rest, resi 607-609+640-644+674-686+718-733+764-769+796-801+830-
850
color gray, rest
show cartoon, rest

select ec2, resi 674-686
#color yellow, ec2

select TM_region, resi 577-606+610-639+644-673+687-717+734-763+769-
795+802-829
cmd.spectrum("b",selection=("TM_region"),quiet=0)
show cartoon, TM_region
#r = cmd.create ( "entropy", "TM_region")
#show cartoon,"entropy"
#cmd.spectrum("b",selection=("entropy"),quiet=0)

######family C mutations from literature translated using WS numbering
on proposed alignment######
```

```
select lit_mutations, resid
635+646+647+648+651+652+655+656+658+659+702+703+740+743+744+745+747+78
1+785+786+787+788+789+790+791+792+802+803+805+806+807+810+813


##################negative modulators###############################
#CPCCOEt human mGluR1 antagonist T815+A818
select CPCCOET, resi 802+805


#EM_TBPC rat mGluR1 antagonist Y672+V757+798+F801+Y805+T815
select EM_TBPC, resi 659+744+788+785+792+802


#M-MPEP human mGluR5 antagonist P655+S658+A810
select M_MPEP, resi 655+658+810


#MPEP rat mGluR5 antagonist P654+Y658+L743+T780+W784+F787+Y791+A809
select MPEP_rat, resi 655+659+744+781+785+788+792+810


#MPEP human mGluR5 antagonist P655+S658+A810
select MPEP_human, resi 655+658+810


#NPS_2143 Antagonist of CaSR resi F668+F684+R680+688+E837+I841
#on mGluR5 full length select NPS_2143, resi 635+651+646+655+803+807
select NPS_2143, resi 635+651+646+655+803+807


#Calhex_231 CaSR antagonist F684+F688+E837+I841
#on mGluR5 full length select Calhex_231, resi 651+655+803+807
select Calhex_231, resi 651+655+803+807


#Fenobam rmGR5 NAM resi
R647A+P654S+S657C+Y658V+L743V+T780A+W784+F787A+V788M+Y791A+809
#on mGluR5 full length select 648+655+658+659+744+781+785+788+792+810
select fenobam, resid 648+655+658+659+744+781+785+788+792+810


#Lactisole Nam for hT1R3
640(3.32)+641(3.33)+733(5.42)+735(5.39)+738(5.42)+739(5.43)+778+782+79
8
#on mGluR5 full length select 649+650+740+743+745+748+749+788+792+806
select lactisole, resid 649+650+740+743+745+748+749+788+792+806


select NAM_mGR1,resi 659+744+788+785+792+802+805
select NAM_mGR5,resi 648+655+658+659+744+781+785+788+792+810
select NAM_HUM_mGR5, resi 655+658+810
select NAM_HUM_mGR1, resi 802+805
select NAM_RAT_mGR5,resi 648+655+658+659+744+781+785+788+792+810
select NAM_RAT_mGR1,resi 659+744+788+785+792+802
select NAM_CASR,resi 635+651+646+803+655+807
select NAM_T1R3, resi 649+650+740+743+745+748+749+788+792+806
select mGR_NAM,resi 659+744+788+785+792+802+805+648+655+658+781+810
select NAM, resi
635+646+648+651+655+658+659+744+781+785+788+792+802+803+805+806+807+81
0


##################positive modulators###############################
#DFB agonist of rat mGluR5 S657+L743+T780+W784+F787+M801
#on mGluR5 human full length DFB S658+L744+T781+W785+F788+M802
select DFB, resi 658+744+781+785+788+802


#LY 487379 PAM for hmGluR2 S688L+G689V+N735D+A733T
#on mGluR5 full length select 702+703+747+745
select ly_487379, resid 702+703+747+745


#Ro67_7476 rat mGluR1 agonist S668+C671+V757
select RO67_7476, resi 655+658+792
```

10. Appendix

```
#cyclamate HT1R3 pam Q636+F730+L782+S640
#on mGluR5 human full length 647+740+788+651
select cyclamate, resi 647+740+788+651

#d-tryptophan orthosteric agonist on hT1R3 L644+T645+Y771+Q794+I805
#on mGluR5 human full length 655+656+745+802+813
#select d_tryptophan, resid 655+656+745+802+813

select PAM, DFB or ly_487379 or RO67_7476 or cyclamate or d-tryptophan
select mGR_PAM, DFB + ly_487379 + RO67_7476
select T1R3_PAM, cyclamate or d-tryptophan

select mGR_PAM_NAM, (NAM_mGR1+NAM_mGR5) and (mGR_PAM)
select both_NAM_PAM, NAM and PAM

##########positions_of_mutation_with_no_effect#############

#3.32+6.52+6.55+5.43+45.53+5.39+5.44+5.48+7.38+7.4+7.45+7.46+6.48+6.51
+3.27+3.36+5.42
select no_effect, resi
651+789+792+744+734+740+745+749+808+810+815+817+785+788+646+655+743
```

10.7 PyMol script for structural features and comparison

```
load mGluR5_structure.pdb
color gray
show cartoon
hide lines
cmd.hide("(mGluR5_structure and hydro)")
hide all

select rest, resi 607-609+640-644+674-686+718-733+764-769+796-801+830-
850
color gray, rest
show cartoon, rest

select ec2, resi 674-686
#color yellow, ec2

select TM_region, resi 577-606+610-639+644-673+687-717+734-763+769-
795+802-829
cmd.spectrum("b",selection=("TM_region"),quiet=0)
show cartoon, TM_region
#r = cmd.create ( "entropy", "TM_region")
#show cartoon,"entropy"
#cmd.spectrum("b",selection=("entropy"),quiet=0)

########positions_with_identical_property_in familyC_GPCRs#########

select ident_prop_to_famC, resi
580+583+586+588+590+593+598+599+600+602+603+604+605+610+617+618+622+62
3+626+629+630+632+637+644+645+646+649+650+653+654+657+661+662+666+672+
700+701+703+708+709+716+717+737+740+741+743+749+750+752+756+757+759+77
1+772+774+778+780+783+785+786+788+795+804+805+807+809+811+817+818+820+
822+823+824+826+828

##########positions_with_low_H_using_9function_coding############
select selow_9f, resi
590+594+620+622+644+662+665+666+669+703+705+750+758+774+777+778+786+80
6+820+825
```

```
###########positions_with_high_H_using_9function_coding#############
select se_high_9f, resi
606+616+645+689+690+691+692+734+735+737+739+747+763+773+794

################dimerization###############################
select dimeFromFamA, resi
584+588+591+598+634+638+698+701+704+705+708+712+715+717+735+737+738+74
2+779+705+712+715
select expDimFromFamA, resi 584+588+591+634+638+704+717+779

################g-protein coupling###############################
select gprotein, resi
670+672+673+759+762+773+774+823+827+610+611+616+668+669+670+672
select gproteinProp, resi 606+687+690+693+694+770+829

################helix helix contacts###############################
# weistein positions
1.5+2.45+2.5+3.35+3.37+3.42+4.49+4.50+4.52+5.46+6.44+6.47+7.46+7.48
select hhcontacts, resi
594+616+621+654+656+661+706+707+709+747+781+784+816+818

################DRY motiv###############################
select DRYmotiv, resi 668+669+670

################NPXXY motiv###############################
select NPXXYmotiv, resi 819+820+821+822+823

################FXXXW motiv###############################
select FXXXWmotiv, resi 781+782+783+784+785

################FamA ligand binding###############################
select FamA_BP, resi
623+624+625+629+644+652+654+655+665+666+740+751+785+822+647+813

################constantly active###############################
select CA, resi
594+614+616+624+644+654+662+665+666+668+669+740+758+774+777+787

################carazolol binding
pocket###############################
select carazolol_bp, resi
647+651+652+655+656+739+743+747+785+788+789+792+805+809+813
```

10. Appendix

10.8 Mutation data collection

Table A 3: Mutation data collection

Receptor	Mutation	BW pos	TM or loop	Modulator	Effect	Reference
hCasR	F668A	**2.64**	2	NPS R-568	reduced NPS R-568 binding	Miedlich *et al.* 2004
hCasR	F668A	**2.64**	2	NSP 2143	Reduced NSP 2143 binding and expression on cell membrane	Miedlich *et al.* 2004
hCasR	R680A	**3.27**	3	NPS 2143, Calhex 233	decreased inhibition	Petrel *et al.* 2004
hCasR	R680A	**3.27**	3	NPS R-568, Calindol	**no effect on Ca2+ response or PAM effects**	Petrel *et al.* 2004
hCasR	R680A	**3.27**	3	NSP 2143	Reduced NSP 2143 binding	Miedlich *et al.* 2004
hT1R3	Q636	**3.28**	3	Cyclamate	reduced sensitivity to Cyclamate	Jiang *et al.* 2005 b
rmGluR5	R647A	**3.29**	3	Fenobam	Decreased binding	Malherbe *et al.* 2006
hCasR	F684A	**3.32**	3	Calhex 231	No binding of Calhex 231	Petrel *et al.* 2003
hCasR	F684A	**3.32**	3	NPS 2143, Calhex 231	no block of Ca2+	Petrel *et al.* 2004
hCasR	F684A	**3.32**	3	NPS R-568, Calindol	**reduced Ca2+ potentiation, no effect on PAM**	Petrel *et al.* 2004
hT1R3	S640A	**3.32**	3	Lactisole	Increased sensitivity to lactisole	Jiang *et al.* 2005 a
hT1R3	S640A	**3.32**	3	Cyclamate, D-Tryptophan	**no effect on ligand effect**	Jiang *et al.* 2005 b
hCasR	F684A	**3.32**	3	NPS R-568	reduced NPS R-568 binding	Miedlich *et al.* 2004
hCasR	F684A	**3.32**	3	NSP 2143	Reduced NSP 2143 binding	Miedlich *et al.* 2004
hT1R3	H641A	**3.33**	3	Lactisole	No lactisole binding	Jiang *et al.* 2005 a
hCasR	F688A	**3.36**	3	Calhex 231	Reduced binding of Calhex 231	Petrel *et al.* 2003
hCasR	F688A	**3.36**	3	NPS 2143, Calhex 231	no block of Ca2+	Petrel *et al.* 2004

hCasR	F688A	**3.36**	3	NPS R-568, Calindol	**reduced Ca2+ potentiation, no effect on PAM**	Petrel *et al.* 2004
rmGluR1a	S668P	**3.36**	3	RO 67-7476	loss of PAM effect	Knoflach *et al* 2001
rmGluR5	P654S	**3.36**	3	Fenobam	no binding	Malherbe *et al.* 2006
hT1R3	L644A	**3.36**	3	D-Tryptophan	No binding	Jiang *et al.* 2005 a
hT1R3	L644A	**3.36**	3	Cyclamate, D-Tryptophan	no respose to cyclamate or d-tryptophan	Jiang *et al.* 2005 b
rmGluR5a	P654S	**3.36**	3	[H3]MPEP	Decrease in MPEP binding affinity	Malherbe *et al.* 2003 b
hmGluR5a	P655S	**3.36**	3	[H3]M-MPEP	Decrease in M-MPEP binding affinity	Pagano *et al.* 2000
hT1R3	T645A	**3.37**	3	D-Tryptophan	No binding	Jiang *et al.* 2005 a
hT1R3	T645A	**3.37**	3	Cyclamate, D-Tryptophan	no respose to cyclamate or d-tryptophan	Jiang *et al.* 2005 b
rmGluR5a	S657C	**3.39**	3	DFB	Abolished DFB enhancement	Mühlemann *et al.* 2006
rmGluR5	S657C	**3.39**	3	Fenobam	no binding	Malherbe *et al.* 2006
hmGluR5a	S658C	**3.39**	3	[H3]M-MPEP	Decrease in M-MPEP binding affinity	Pagano *et al.* 2000
rmGluR1a	Y672V	**3.4**	3	EM-TBPC	Decrease of EM-TBPC-binding	Malherbe *et al.* 2003 a
rmGluR5	Y658V	**3.4**	3	Fenobam	no binding	Malherbe *et al.* 2006
rmGluR5	Y658F	**3.4**	3	Fenobam	slightly increased binding	Malherbe *et al.* 2006
rmGluR5a	Y658V	**3.4**	3	[H3]MPEP	Complete loss of MPEP binding and inhibition	Malherbe *et al.* 2003 b
hT1R3	F730A	**5.39**	5	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
hT1R3	F730A	**5.39**	5	Cyclamate	Diminished response to cyclamate	Jiang *et al.* 2005 b
hT1R3	F730Y	**5.39**	5	Cyclamate	slightly enhanced cyclamate activity	Jiang *et al.* 2005 b
rT1R3	L735F	**5.39**	5	lactisole	gain of function, lactisole activity like in humans	Winning *et al.* 2005
hCasR	L776A	**5.42**	5	Calhex 231	Increased affinity for Calhex 231	Petrel *et al.* 2003

10. Appendix

hCasR	L776A	**5.42**	5	NPS 2143, Calhex 231	**decreased inhibition for Calhex, no effect on NAM activity for NPS**	Petrel *et al.* 2004
hCasR	L776A	**5.42**	5	NPS R-568, Calindol	**reduced Ca2+ potentiation, no effect on PAM**	Petrel *et al.* 2004
rT1R3	V738A	**5.42**	5	lactisole	increased lactisole sensitivity to half of hTaslr3 level	Winning *et al.* 2005
hT1R3	A733V	**5.42**	5	Lactisole	Decreased sensitivity to lactisole	Jiang *et al.* 2005 a
mT1R3	V738A	**5.42**	5	Lactisole	Increased sensitivity to lactisole	Jiang *et al.* 2005 a
rT1R3	V738A	**5.42**	5	lactisole	gain of function, lactisole activity like in humans	Winning *et al.* 2005
rmGluR5a	L743A	**5.43**	5	DFB	Increased DFB enhancement	Mühlemann *et al.* 2006
rmGluR5a	L743V	**5.43**	5	DFB	Increased DFB enhancemen	Mühlemann *et al.* 2006
rmGluR1a	V757L	**5.43**	5	RO 67-7476	Critical for RO-67-7476 and EM-TBPC-binding	Knoflach *et al* 2001
hmGluR1a	L757V	**5.43**	5	EM-TBPC	Mutation converts to rmGluR1a, high binding affinity to EM-TBPC	Malherbe *et al.* 2003 a
rmGluR1a	V757L	**5.43**	5	EM-TBPC	Mutation converts to hmGluR1a, reduction of EM-TBPC binding	Malherbe *et al.* 2003 a
rmGluR1a	V757A	**5.43**	5	EM-TBPC	reduction of EM-TBPC binding	Malherbe *et al.* 2003 a
rmGluR5	L743V	**5.43**	5	Fenobam	**no effect on binding**	Malherbe *et al.* 2006
rmGluR5	L743A	**5.43**	5	Fenobam	slightly increased binding	Malherbe *et al.* 2006
rmGluR5a	L743A	**5.43**	5	[H3]MPEP	Decrease in MPEP binding affinity	Malherbe *et al.* 2003 b
hT1R3	A735I	**5.44**	5	Lactisole	**no change on Lactisole effect**	Jiang *et al.* 2005 a
hGluR2	N735D	**5.46**	5	LY 487379	Reduced LY 487379 binding	Schaffhauser *et al.* 2003

hT1R3	T739M	**5.48**	5	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
rmGluR5a	T780A	**6.44**	6	DFB	Abolished DFB enhancement	Mühlemann *et al.* 2006
rmGluR5	T780A	**6.44**	6	Fenobam	no binding	Malherbe *et al.* 2006
hT1R3	Y771A	**6.44**	6	D-Tryptophan	No binding	Jiang *et al.* 2005 a
hT1R3	Y771A	**6.44**	6	Cyclamate, D-Tryptophan	no respose to cyclamate or d-tryptophan	Jiang *et al.* 2005 b
rmGluR5a	T780A	**6.44**	6	[H3]MPEP	Decrease in MPEP binding affinity	Malherbe *et al.* 2003 b
rmGluR5a	W784A	**6.48**	6	DFB	Increased DFB enhancemen	Mühlemann *et al.* 2006
rmGluR5a	W784F	**6.48**	6	DFB	Increased DFB enhancemen	Mühlemann *et al.* 2006
hCasR	W818A	**6.48**	6	Calhex 231	Reduced binding of Calhex 231	Petrel *et al.* 2003
hCasR	W818A	**6.48**	6	NPS 2143, Calhex 231	**no effect on NAM effect**	Petrel *et al.* 2004
hCasR	W818A	**6.48**	6	NPS R-568, Calindol	left-shift (stronger for NPS), Emax increase	Petrel *et al.* 2004
rmGluR1a	W798F	**6.48**	6	EM-TBPC	Increased binding affinity to EM-TBPC	Malherbe *et al.* 2003 a
rmGluR5	W784F	**6.48**	6	Fenobam	Decreased binding	Malherbe *et al.* 2006
rmGluR5	W784A	**6.48**	6	Fenobam	no binding	Malherbe *et al.* 2006
rmGluR5a	W784A	**6.48**	6	[H3]MPEP	Complete loss of MPEP binding and inhibition	Malherbe *et al.* 2003 b
hCasR	I819A	**6.49**	6	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	S820F	**6.5**	6	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
rmGluR5a	F787A	**6.51**	6	DFB	Converts DFB into a partial antagonist	Mühlemann *et al.* 2006
hCasR	F821A	**6.51**	6	Calhex 231	Increased affinity for Calhex 231	Petrel *et al.* 2003
hCasR	F821A	**6.51**	6	NPS 2143, Calhex 231	**no effect on NAM effect**	Petrel *et al.* 2004

hCasR	F821A	**6.51**	6	NPS R-568, Calindol	small left-shift, no Emax effect	Petrel *et al.* 2004
hCasR	F821L	**6.51**	6	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
rmGluR1a	F801A	**6.51**	6	EM-TBPC	Complete loss of EM-TBPC-binding	Malherbe *et al.* 2003 a
rmGluR5	F787A	**6.51**	6	Fenobam	no binding	Malherbe *et al.* 2006
hT1R3	F778A	**6.51**	5	Lactisole	Decreased sensitivity to lactisole	Jiang *et al.* 2005 a
hT1R3	L782A	**6.51**	6	Cyclamate	reduced sensitivity to Cyclomate	Jiang *et al.* 2005 b
rmGluR5a	F787A	**6.51**	6	[H3]MPEP	Complete loss of MPEP binding and inhibition	Malherbe *et al.* 2003 b
hCasR	I822A	**6.52**	6	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
rmGluR5	V788M	**6.52**	6	Fenobam	**no effect on binding**	Malherbe *et al.* 2006
hT1R3	V776A	**6.52**	6	Cyclamate, D-Tryptophan	**no effect on ligand effect**	Jiang *et al.* 2005 b
hCasR	P823A	**6.53**	6	Ca	Reduction of Ca response, slitly increased with NPS R-568	Hu *et al.* 2005
hCasR	A824S	**6.54**	6	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	Y825A	**6.55**	6	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
rmGluR1a	Y805A	**6.55**	6	EM-TBPC	Complete loss of EM-TBPC-binding	Malherbe *et al.* 2003 a
rmGluR5	Y791A	**6.55**	6	Fenobam	Decreased binding	Malherbe *et al.* 2006
hT1R3	L782A	**6.55**	6	Lactisole	Increased sensitivity to lactisole	Jiang *et al.* 2005 a
hT1R3	V779A	**6.55**	6	Cyclamate, D-Tryptophan	**no effect on ligand effect**	Jiang *et al.* 2005 b
rmGluR5a	Y791A	**6.55**	6	[H3]MPEP	Increased MPEP-potency but not binding affinity	Malherbe *et al.* 2003 b
hCasR	V836L	**7.31**	7	Ca	Increased Ca-sensitivity	Hu *et al.* 2005

hCasR	V836L	**7.31**	7	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
rmGluR5a	M801T	**7.32**	7	DFB	Abolished DFB enhancement	Mühlemann *et al.* 2006
hCasR	E837D	**7.32**	7	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	E837K	**7.32**	7	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	E837	**7.32**	7	Ca	Crutial salt bridge to NPS R-568 and NPS 2143	Hu *et al.* 2005
hmGluR1b	T815	**7.32**	7	(-)-CPCCOEt	Critical for CPCCOEt-binding	Litching *et al.* 1999
rmGluR1a	T815M	**7.32**	7	EM-TBPC	Complete loss of EM-TBPC-binding	Malherbe *et al.* 2003 a
hT1R3	Q794A	**7.32**	7	D-Tryptophan	No binding	Jiang *et al.* 2005 a
hT1R3	Q794A	**7.32**	7	Cyclamate, D-Tryptophan	no respose to cyclamate or d-tryptophan	Jiang *et al.* 2005 b
hCasR	E837A	**7.33**	7	Calhex 231	No binding of Calhex 231	Petrel *et al.* 2003
hCasR	E837A	**7.33**	7	NPS 2143, Calhex 233	no block of Ca2+	Petrel *et al.* 2004
hCasR	E837A	**7.33**	7	NPS R-568, Calindol	no left-shift, Emax decrease only for Calindol	Petrel *et al.* 2004
hCasR	E837A	**7.33**	EC3	NPS R-568	Critical for NPS R-568 binding	Hu *et al.* 2002
hCasR	E837I	**7.33**	7	NPS R-568	reduced NPS R-568 binding	Miedlich *et al.* 2004
hCasR	E837I	**7.33**	7	NSP 2143	Reduced NSP 2143 binding	Miedlich *et al.* 2004
hmGluR1b	A818S	**7.35**	7	(-)-CPCCOEt	Critical for CPCCOEt-binding	Litching *et al.* 1999
hT1R3	L798I	**7.36**	7	Lactisole	Decreased sensitivity to lactisole	Jiang *et al.* 2005 a
hCasR	I841A	**7.37**	7	Calhex 231	Reduced binding of Calhex 231	Petrel *et al.* 2003
hCasR	I841A	**7.37**	7	NPS 2143, Calhex 233	decreased inhibition	Petrel *et al.* 2004
hCasR	I841A	**7.37**	7	NPS R-568, Calindol	no left-shift, Emax decrease	Petrel *et al.* 2004
hT1R3	L800V	**7.38**	7	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a

rmGluR5	A809V	**7.4**	7	Fenobam	no binding	Malherbe *et al.* 2006
hT1R3	V802A	**7.4**	7	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
rmGluR5a	A809V	**7.4**	7	[H3]MPEP	Complete loss of MPEP binding and inhibition	Malherbe *et al.* 2003 b
hmGluR5a	A810V	**7.4**	7	[H3]M-MPEP	Loss of M-MPEP binding affinity	Pagano *et al.* 2000
hmGluR5a	A810G	**7.4**	7	[H3]M-MPEP	Loss of M-MPEP binding affinity	Pagano *et al.* 2000
hT1R3	I805A	**7.43**	7	D-Tryptophan	No binding	Jiang *et al.* 2005 a
hT1R3	I805A	**7.43**	7	Cyclamate, D-Tryptophan	no respose to cyclamate or d-tryptophan	Jiang *et al.* 2005 b
hT1R3	A807V	**7.45**	7	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
hT1R3	A808T	**7.46**	7	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
rmGluR1a	N747A	**45.53**	EC2	EM-TBPC	Complete loss of EM-TBPC-binding	Malherbe *et al.* 2003 a
rmGluR5	N733A	**45.53**	EC2	Fenobam	**no effect on binding**	Malherbe *et al.* 2006
rmGluR5	P654S/ S657C	**3.36, 3.39**	3	Fenobam	no binding	Malherbe *et al.* 2006
hmGluR5a	P655S/S658C	**3.36 / 3.39**	3	[H3]M-MPEP	Loss of M-MPEP binding affinity	Pagano *et al.* 2000
hGluR2	S688L, G689V	**4.45, 4.46**	4	LY 487379	Reduced LY 487379 binding	Schaffhauser *et al.* 2003
hGluR2	S688L, N735D	**4.45, 5.46**	4.5	LY 487379	Strongly reduced LY 487379 binding	Schaffhauser *et al.* 2003
hGluR2	G689V, N735D	**4.46, 5.46**	4.5	LY 487379	Strongly reduced LY 487379 binding	Schaffhauser *et al.* 2003
rT1R3	V738A, L735F	**5.39, 5.42**	5	lactisole	Increased lactisole sensitivity to hTas1r3 level	Winnig *et al.* 2005
hGluR2	A733T, N735D	**5.44, 5.46**	5	LY 487379	Strongly reduced LY487379 binding	Schaffhauser *et al.* 2003
hCasR	K831A	**7.26 / EC3**	7	Ca	Increased Ca-sensitivity	Hu *et al.* 2005

hCasR	F832S	**7.27 / EC3**	7	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	A835T	**7.30, EC3**	7	Ca	Naturaly occuring mutation,Increased Ca-sensitivity	Hu *et al.* 2005
hCasR	T764A		EC2	Calhex231	**no effect**	Petrel *et al.* 2003
hCasR	H766A		EC2	Calhex231	**no effect**	Petrel *et al.* 2003
hCasR	T764A		EC2	NPS 2143, Calhex 231	**no effect on NAM effect**	Petrel *et al.* 2004
hCasR	H766A		EC2	NPS 2143, Calhex 231	**no effect on NAM effect**	Petrel *et al.* 2004
hT1R2	S144A		EC			Xu *et al.* 2004
hT1R2	Y218A		EC		Abolished response to sweeteners, possibly not expressed on cell surface	Xu *et al.* 2004
hT1R2	E302A		EC		Selectively affected response to diff. sweeteners	Xu *et al.* 2004
hCasR	D758A		EC2	Ca	Conserved in all CaSR, increased Ca-sensitivity	Hu *et al.* 2002
hCasR	E759A		EC2	Ca	Conserved in all CaSR, increased Ca-sensitivity	Hu *et al.* 2002
hCasR	E767A		EC2	Ca	Conserved in all CaSR, increased Ca-sensitivity	Hu *et al.* 2002
hCasR	F788C		EC2	Ca	Impaired Ca-sesitivity	Hu *et al.* 2002
hCasR	E755A, E757A, D758A, E759A, E767A		EC2	Ca	Incorrect fold, poorly expressed at cell surface	Hu *et al.* 2002
hCasR	G830A		EC3	Ca	Increased Ca-sensitivity	Hu *et al.* 2005
hT1R3	A537T		CRD	Brazzein	Diminished response to brazzein	Jiang *et al.* 2004
hT1R3	F540P		CRD	Brazzein	Diminished response to brazzein	Jiang *et al.* 2004

10. Appendix

mT1R3	T542A	CRD	Brazzein	Increased response to brazzein	Jiang *et al.* 2004
hT1R3	R790Q	EC3, H7	Lactisole	Decreased sensitivity to lactisole	Jiang *et al.* 2005 a
hT1R3	V788A	EC3	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
hT1R3	L789Y	EC3	Lactisole	**no effect on Lactisole effect**	Jiang *et al.* 2005 a
hT1R3	R790	EC3	Cyclamate	Diminished response to cyclamate	Jiang *et al.* 2005 b
hT1R3	H721	EC2	Cyclamate	reduced sensitivity to Cyclamate	Jiang *et al.* 2005 b
hT1R3	R723	EC2	Cyclamate	reduced sensitivity to Cyclamate	Jiang *et al.* 2005 b
hT1R3	R790H	EC3	Cyclamate	no response to both, nonfunctional?	Jiang *et al.* 2005 b

Curriculum Vitae

Swetlana Derksen
Dipl. Bioinformatician

Personal Details

Born on 28 September 1980 in Orsk, Russia
Nationality: German and Russian

Current Position

Since 01.12.2009	Molecular Modeler at BASF Aktiengesellschaft, Computational Chemistry & Biology, Carl-Bosch-Straße, 67056 Ludwigshafen/Rhein, Germany.

Former positions

01.2006-11.2008	Ph.D. student in the group of Prof. Dr. G. Schneider (Johann-Wolfgang von Goethe University, Frankfurt am Main)
	Scholarship holder of Merz Pharmaceuticals GmbH & Co. KGaA, Eckenheimer Landstraße 100, 60318 Frankfurt am Main

Education

June 2000	High school graduation (German Abitur) from Goethe-Gymnasium in Kassel, Germany
October 2000	studies in Bioinformatics at Goethe-University in Frankfurt am Main, Germany
September 2005	Diploma examination; overall grade: *very good* Diploma thesis, "De novo design using combinatorial molecule libraries" advisor Prof. Dr. G. Schneider (University of Frankfurt)
01.2006 - 08.2009	Postgraduate studies at group of Prof. Dr. G. Schneider (University of Frankfurt) "Investigation of Structure and Allosteric Modulation of Family C GPCRs by Sequence-, Structure- and Ligand-based Approaches"

Internship

11.2005-12.2005	at Merz Pharmaceuticals, Medicinal Chemistry group (Dr. Tanja Weil), Task: Programming a tool for fingerprint calculation and clustering of ligand binding poses based on molecular interactions in Java

Scholarship

Scholarship from Merz Pharmaceuticals granted for 2006-2007, extended for 2008.

Publications

Journal Articles

Renner S., <u>Derksen S.</u>, Radestock S. and Mörchen F. (2008) Maximum common binding modes (MCBM): consensus docking scoring using multiple ligand information and interaction fingerprints. J. Chem. Inf. Model. 48, 319-332.

Proschak E., Rupp M., <u>Derksen S.</u> and Schneider G. (2007) Shapelets: Possibilities and limitations of shape-based virtual screening. *J. Comp. Chem.* 29, 108-114.

<u>Derksen S.</u>, Rau O., Schneider P., Schubert-Zsilavecz M. and Schneider G. (2006) Virtual screening for PPAR modulators using a probabilistic neural network, *ChemMedChem*, 1, 1346-1350.

Boecker A., <u>Derksen S.</u>, Schmidt E., Teckentrup A. and Schneider G. (2005) A Hierarchical Clustering Approach for Large Compound Libraries, *J. Chem. Inf. Model*, 45, 807-815.
Download at http://gecco.org.chemie.uni-frankfurt.de/hkmeans/index.html

Conference Posters

<u>Derksen S.</u>, Weil T. and Schneider G., (2007) Improving Models of Transmembrane Domains of Class C GPCR using Property Conservation Analysis, 4th Joint Sheffield Conference on Chemoinformatics.

2. German Conference on Chemoinformatics Goslar, Germany (2006)
I) Noeske T., <u>Derksen S.</u>, Weil T. and Schneider G., "Selectivity Profiles for Ligands of Family C GPCRs"
II) <u>Derksen S.</u>, Rupp M., Proschak E., Schneider G., "Kernel based classifier in virtual screening"

<u>Derksen S.</u>, Rau O., Syha Y., Popescu L., Schubert-Zsilavecz M., Schneider G. (2006) Quest for Novel Ligands of Peroxisome Proliferator-Activated Receptors (PPAR) Using Probabilistic Neural Networks, Annual Meeting "Frontiers in Medicinal Chemistry", Frankfurt am Main.

Jimenez T., <u>Derksen S.</u>, Schmidt E., Gohlke H. (2005) Binding pocket comparison based on knowledge-based potential fields, 19. Darmstädter Molecular Modelling Workshop, Erlangen