

Homology Modeling, Molecular Dynamics Simulations and Site-directed Mutagenesis of Histamine H₂ Receptors

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der
Naturwissenschaftlichen Fakultät IV – Chemie und Pharmazie – der Universität Regensburg



vorgelegt von

Tobias Holzammer

aus Neumarkt i.d.Opf.

2013

Die vorliegende Arbeit entstand in der Zeit von Juli 2009 bis Juli 2013 unter der Leitung von Herrn Prof. Dr. S. Dove und Herrn Prof. Dr. A. Buschauer am Institut für Pharmazie der Naturwissenschaftlichen Fakultät IV – Chemie und Pharmazie – der Universität Regensburg.

Das Promotionsgesuch wurde eingereicht im Juli 2013.

Tag der mündlichen Prüfung: 02. August 2013

Prüfungsausschuss:

Prof. Dr. F.-M. Matysik	(Vorsitzender)
Prof. Dr. S. Dove	(Erstgutachter)
Prof. Dr. A. Buschauer	(Zweitgutachter)
Prof. Dr. S. Elz	(Drittprüfer)

Danksagungen

An dieser Stelle möchte ich mich bedanken bei:

Herrn Prof. Dr. Stefan Dove dass er mir die Möglichkeit gegeben hat, an diesem spannenden und vielschichtigen Projekt arbeiten zu dürfen, für seine fachliche Anleitung, die wissenschaftlichen Diskussionen sowie für seine Hilfsbereitschaft und konstruktive Kritik beim Verfassen dieser Arbeit,

Herrn Prof. Dr. Armin Buschauer für seine Unterstützung und sein Vertrauen in den vier Jahren, seine wissenschaftlichen Anregungen sowie seine konstruktive Kritik bei der Durchsicht dieser Arbeit,

Frau PD Dr. Andrea Straßer für ihre Hilfsbereitschaft in der Molekulardynamik Simulation sowie für die Durchsicht der Kapitel 4, 5 und 9,

Herrn Prof. Dr. Günther Bernhardt für seine stete Hilfsbereitschaft und fachliche Unterstützung,

Frau Dr. Irena Brunskole und Herrn Johannes Felixberger für ihre Hilfe bei molokularbiologischen und pharmakologischen Fragestellungen,

Frau M. Beer-Krön für die Einführung in den GTPase Assay, die $[\gamma\text{-}^{33}\text{P}]\text{GTP}$ -Herstellung und ihre ständige Hilfsbereitschaft im Labor,

Herrn Stefan Huber und Herrn Dr. Max Keller für die Einführung in die HPLC Analytik,

Frau Nicole Kagermeier, Herrn Dr. Mirosław Lopuch und Herrn Dr. Uwe Nordemann für die ausgezeichnete Zusammenarbeit im Labor und die zahlreichen Diskussionen,

Frau Dr. Janina Hamberger für die Einführung in die Western Blots und Frau Gertraud Wilberg für ihre Unterstützung auf dem Gebiet der Sf9-Zellkultur,

Herrn Dr. Tobias Birnkammer für die Bereitstellung der Acylguanidine und Herrn Paul Baumeister für die Bereitstellung von $[\text{35S}]\text{GTP}\gamma\text{S}$,

Herrn Prof. Dr. Roland Seifert (Medizinische Hochschule Hannover) für die Bereitstellung des pGEM-3Z-SF-hH₂R-His₆-G_{saS} Vektors,

Herrn Peter Richthammer für seine Hilfsbereitschaft und Kompetenz bei allen technischen Problemen sowie für die gute Zusammenarbeit bei der Durchführung der verschiedenen Praktika,

Frau Uta Hasselmann und Frau Karin Reindl für ihre freundliche Unterstützung bei allen organisatorischen Dingen,

dem Graduiertenkolleg 760 der DFG für die finanzielle Unterstützung und wissenschaftliche Förderung,

meinen aktuellen und ehemaligen Bürokollegen Frau Dr. Miriam Ertel, Herrn Dr. Peter Höcherl, Frau Dr. Agnieszka Kaczor, Herrn Dr. Max Keller, Frau Karoline Löffel, Herrn Dr. Johannes Mosandl und Herrn David Wifling für die angenehme Atmosphäre und die gute Zusammenarbeit,

allen Kollegen und studentischen Hilfskräften für die Durchführung verschiedener Praktika,

allen Mitgliedern des Lehrstuhls für ihre Kollegialität, Hilfsbereitschaft und das gute Arbeitsklima,

meinen Freunden Franz, Johannes und Patrick für die schöne Zeit in Regensburg,

meinen Eltern, meiner Schwester, und ganz besonders Christina und Karoline für ihre Unterstützung und Hilfe in jeglicher Art und Weise.

Contents

1	Introduction	1
1.1	G-Protein coupled receptors	1
1.1.1	Classification and relevance of GPCRs.....	1
1.1.2	GPCR structure.....	2
1.1.3	Ligand classification	5
1.1.4	G-protein cycle.....	6
1.1.5	Alternative signaling pathways and functional selectivity	7
1.1.6	GPCR activation	8
1.2	The histamine H₂R and the histamine receptor family.....	10
1.2.1	Histamine receptor subtypes.....	10
1.2.2	Characterization of the H ₂ R.....	14
1.2.3	H ₂ R ligands	15
1.2.4	Bivalent ligands for the H ₂ receptor	18
1.3	References.....	20
2	Scope and Objectives.....	31
3	Homology Models of Inactive and Active Human Histamine H₂ Receptor States.....	35
3.1	Introduction	35
3.2	Materials and methods.....	36
3.2.1	Sequence alignment	36
3.2.2	Generation of 3D structures	36
3.2.2.1	Generation of the inactive state hH ₂ R model.....	36
3.2.2.2	Generation of the active state hH ₂ R models.....	38
3.2.2.3	Structural refinement of the homology models	39
3.2.2.4	Insertion of the C-terminal part of the G _{sα} -protein and docking of histamine.....	40
3.2.3	Structure validation	40

3.3 Results and discussion	41
3.3.1 Template selection	41
3.3.2 Stereochemical quality of the models	48
3.3.2.1 Omega backbone angles	48
3.3.2.2 Ramachandran plot.....	49
3.3.2.3 Side chain dihedral angles	51
3.3.2.4 Planarity of aromatic groups and delocalized π -electron systems in amino acid side chains	52
3.3.2.5 Main chain bond lengths, bond angles and bad contacts	52
3.3.3 Comparison of inactive and active hH ₂ R receptor states	53
3.3.3.1 Overall topology of the protein backbone	53
3.3.3.2 Activation of the hH ₂ R	58
3.3.3.3 Analysis of inter-TM domain contacts.....	66
3.3.3.4 Contacts of extra- and intracellular loops.....	70
3.3.3.5 Interactions of helix 8	71
3.3.4 Ligand and G α CT interactions with the active hH ₂ R states.....	72
3.3.4.1 Histamine binding.....	72
3.3.4.2 Receptor – G-Protein interaction	73
3.4 Summary and conclusion	75
3.4.1 Template selection and quality of the models.....	75
3.4.2 Comparison of hH ₂ R states	76
3.4.3 Conclusion	77
3.5 References	77
 4 Molecular Dynamics Simulations of Inactive and Active Human Histamine H₂ Receptor States.....	 83
4.1 Introduction	83
4.2 Materials and methods.....	84
4.2.1 Materials	84
4.2.2 Parameters of the MD simulations	85
4.2.3 Construction of the hH ₂ R-DPPC-water systems	85
4.3 Results	87
4.3.1 Size and composition of the simulation systems	87
4.3.2 Equilibration of the hH ₂ R-DPPC-water systems	87
4.3.2.1 System parameters	88

4.3.2.2	Lipid bilayer	90
4.3.3	Analysis of system parameters.....	95
4.3.4	Structural analysis of the DPPC bilayer.....	98
4.3.5	Protein structure validation.....	100
4.3.5.1	Stereochemistry and planarity	101
4.3.5.2	Ramachandran analysis.....	104
4.3.5.3	Side chain rotamers	106
4.3.5.4	Saturation of heteroatoms with H-bonds.....	106
4.3.6	Secondary structure of hH ₂ R states	108
4.3.7	Flexibility of the proteins and comparison of the backbone positions of hH ₂ R states.....	113
4.3.8	Internal water molecules	118
4.3.8.1	Stability of putatively conserved water molecules inserted into the initial hH ₂ Rs.....	118
4.3.8.2	Solvation of polar amino acid side chains within TMs.....	121
4.3.8.3	Solvation of polar side chains near the H-bond network around TM7.....	123
4.3.9	Molecular differences between the inactive and active hH ₂ R state.....	125
4.3.9.1	Detailed analysis of molecular switches and amino acid interactions.....	125
4.3.9.2	Analysis of TM-TM contacts	137
4.3.9.3	Loops and helix 8.....	142
4.3.10	Ligand and GαCT interactions with the active hH ₂ R state	147
4.3.10.1	Interactions of histamine with the hH ₂ R binding pocket	147
4.3.10.2	Receptor – GαCT interaction.....	148
4.4	Discussion	151
4.4.1	Quality of the MD simulations.....	151
4.4.2	Activation of the hH ₂ R	152
4.4.3	Conclusion	156
4.5	References.....	156
5	Computational Tools for Analyzing Molecular Dynamics Simulations	163
5.1	Introduction	163
5.2	Materials and methods.....	165

5.3 Systematic calculation of direct and water mediated H-bonds:

<i>gro_hbonds</i>	166
5.3.1 Skills of the program	166
5.3.1.1 Systematic calculation of hydrogen bonds.....	166
5.3.1.2 Analysis of helical structures	166
5.3.1.3 Structure validation of H-bonds	167
5.3.1.4 Detailed analysis of specific hydrogen bonds	167
5.3.2 Structure of the program	167
5.3.3 Adjusting the program	168
5.3.4 Output files.....	171
5.3.4.1 Types of output	171
5.3.4.2 Direct and water mediated H-bonds	173
5.3.4.3 Helix analysis	173
5.3.4.4 Structure validation of H-bonds	175
5.3.4.5 Detailed output for a specific interaction	176
5.3.5 Performance	176

5.4 Analysis of hydrophobic interactions: ***gro_contacts***177

5.4.1 Skills of the program	177
5.4.2 Structure of the program	178
5.4.3 Output files.....	179
5.4.4 Performance	179

5.5 Structure validation of molecular dynamics simulations: ***gro_validation***....180

5.5.1 Skills of the program	180
5.5.1.1 Chirality check.....	180
5.5.1.2 Planarity check.....	181
5.5.1.3 Peptide bond analysis	182
5.5.1.4 Ramachandran analysis.....	183
5.5.1.5 Side chain rotamers	184
5.5.2 Structure of the program	184
5.5.2.1 Parameter file <i>gro_validation-para.txt</i>	187
5.5.2.2 Shell script <i>gro_validation.sh</i>	187
5.5.2.3 C-Program <i>gro_validation-calc</i>	188
5.5.2.4 Reference data.....	189
5.5.3 Output files.....	191
5.5.3.1 Chirality check.....	191
5.5.3.2 Planarity check.....	192

5.5.3.3	Peptide bond analysis	192
5.5.3.4	Ramachandran analysis	194
5.5.3.5	Side chain rotamers	196
5.5.4	Performance	199
5.6	Summary and conclusion	199
5.7	References	200
6	Point Mutation in the Orthosteric Binding Site of the Human Histamine H₂ Receptor: the Role of Tyr182 in TM5	203
6.1	Introduction	203
6.2	Materials and methods	205
6.2.1	Materials	205
6.2.2	Construction of the cDNA encoding the hH ₂ R-Y182F-G _{saS} fusion protein	206
6.2.3	Cell culture, generation of recombinant baculoviruses and membrane preparation	206
6.2.4	Immunoblot analysis	207
6.2.5	Steady-state GTPase activity assay	207
6.2.6	[³⁵ S]GTPγS binding assay	208
6.2.7	Miscellaneous	208
6.3	Results	209
6.3.1	Immunological detection of hH ₂ R-Y182F-G _{saS} in Sf9 cell membranes	209
6.3.2	Agonistic activities at hH ₂ R-G _{saS} and hH ₂ R-Y182F-G _{saS} in the GTPase assay	210
6.3.3	Potencies and intrinsic activities at hH ₂ R-G _{saS} and hH ₂ R-Y182F-G _{saS} in the [³⁵ S]GTPγS binding assay	211
6.3.4	Comparison of data obtained in the GTPase and [³⁵ S]GTPγS assay	212
6.4	Discussion	214
6.5	References	215
7	The Role of Acidic Amino Acids in the Third Extracellular Loop of the Guinea Pig Histamine H₂ Receptor	217
7.1	Introduction	217

7.2	Materials and methods.....	219
7.2.1	Materials	219
7.2.2	Construction of the cDNA for gpH ₂ R-D262S-D263S-E267S-E270S-G _{ssS}	220
7.2.3	Construction of the cDNA for gpH ₂ R-D262S-D263S-E267S-G _{ssS} and gpH ₂ R-E270S-G _{ssS}	221
7.2.4	Cell culture, generation of recombinant baculoviruses and membrane preparation.....	221
7.2.5	Immunoblot analysis	221
7.2.6	Steady-state GTPase activity assay	221
7.2.7	[³⁵ S]GTPγS binding assay.....	222
7.2.8	Multiple sequence alignment.....	222
7.2.9	Homology model of the gpH ₂ R	222
7.2.10	Miscellaneous	222
7.3	Results	222
7.3.1	Selection of a putative accessory binding site for bivalent H ₂ R agonists	222
7.3.2	Multiple sequence alignment analysis of the third extracellular loop of aminergic GPCRs	228
7.3.3	Immunological detection of recombinant proteins in Sf9 cell membranes.....	231
7.3.4	Intrinsic activities and potencies at the wild-type and mutant gpH ₂ Rs.....	232
7.4	Discussion	235
7.4.1	Second binding site for bivalent acylguanidine-type agonists and the role of Glu270 in the gpH ₂ R.....	235
7.4.2	Integrity of the orthosteric binding site	236
7.4.3	Path of the ligand into the binding site.....	237
7.4.4	Summary and conclusion	238
7.5	References.....	239
8	Summary.....	243
9	Appendix.....	247
9.1	Parameters for MD simulations in GROMACS	247
9.2	GROMOS96 53a6 force field parameters for histamine	250

9.3	Systematic calculation of hydrogen bonds: <i>gro_hbonds</i>	252
9.3.1	Parameter file <i>gro_hbonds-para.txt</i>	252
9.3.2	Shell script <i>gro_hbonds.sh</i>	252
9.3.3	C program <i>gro_hbonds-calc.c</i>	256
9.4	Calculation of hydrophobic contacts: <i>gro_contacts</i>	274
9.4.1	Parameter file <i>gro_contacts-para.txt</i>	274
9.4.2	Reference file <i>contacts-atoms.txt</i>	274
9.4.3	Shell script <i>gro_contacts.sh</i>	276
9.4.4	C program <i>gro_contacts-calc.c</i>	278
9.5	Structure validation of MD simulations: <i>gro_validation</i>	280
9.5.1	Parameter file <i>gro_validation-para.txt</i>	280
9.5.2	Shell script <i>gro_validation.sh</i>	281
9.5.3	C program <i>gro_validation-calc.c</i>	287
9.6	Time resolved Ramachandran analysis: <i>rama_time.sh</i>	301
9.7	References	302

Abbreviations

β AR	β -adrenergic receptor
3D	three-dimensional
AC	adenylyl cyclase
AMT	amthamine
ATP	adenosine triphosphate
bp	base pair(s)
BSA	bovine serum albumine
cAMP	cyclic 3',5'-adenosine monophosphate
cH ₂ R	canine histamine H ₂ receptor
CIM	cimetidine
DIM	dimaprit
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
EC ₅₀	agonist concentration which induces 50% of the maximum effect
ECL1, ECL2, ECL3	1 st , 2 nd and 3 rd extracellular loop of a G-protein coupled receptor
EDTA	ethylenediaminetetraacetic acid
E _{max}	maximal response, i. e. intrinsic activity
FAM	famotidine
FLAG	octapeptide epitope for the labeling of proteins
G-protein	guanine nucleotide-binding protein
G α CT	C-terminal part of the G _{saS} -subunit
GDP	guanosine diphosphate
GPCR	G-protein coupled receptor
gpH ₂ R	guinea pig histamine H ₂ receptor
gpH ₂ R-ECL3-4Ser-G _{saS}	gpH ₂ R-D262S-D263S-E267S-E270S-G _{saS}
gpH ₂ R-ECL3-3Ser-G _{saS}	gpH ₂ R-D262S-D263S-E267S-G _{saS}
GROMACS	GRONingen Machine for Chemical Simulations
G _{saS}	short splice variant of the G α protein G _{sa}
GTP	guanosine triphosphate
GTP γ S	guanosine 5'-O-[gamma-thio]triphosphate
GTPase	guanosine 5'-triphosphate hydrolase

H-bond	hydrogen bond
H ₁ R, H ₂ R, H ₃ R, H ₄ R	histamine receptor subtypes
H8	helix 8
HAH	histamine
hH ₂ R	human histamine H ₂ receptor
ICL1, ICL2, ICL3	1 st , 2 nd and 3 rd intracellular loop of a G- protein coupled receptor
IMP	impromidine
K _B	dissociation constant (functional assay)
MAPK	mitogen-activated protein kinase
MD	molecular dynamics
MSA	multiple sequence alignment
NPT ensemble	constant number of atoms (N), pressure (P) and temperature (T)
P _i	inorganic phosphate
PAGE	polyacrylamide gel electrophoresis
PAM250	Point Accepted Mutation matrix
PCR	polymerase chain reaction
pEC ₅₀	negative decadic logarithm of the EC ₅₀ value
PLC	phospholipase C
pK _B	negative decadic logarithm of the K _B value
PKC	protein kinase C
RAN	ranitidine
RMSD	root mean square deviation
RMSF	root mean square fluctuation
rpm	revolutions per minute
SASA	solvent-accessible surface area
S _{CD}	deuterium order parameter
SD	standard deviation
SDS	sodiumdodecylsulfate
SEM	standard error of the mean
SF	signal peptide and FLAG N-terminal tags
Sf9	<i>Spodoptera frugiperda</i> insect cell line
sn	stereospecific numbering (lipid nomenclature)
TIO	tiotidine
TM	transmembrane domain of a G-protein coupled receptor

TM1-TM7	numbering of transmembrane domains of a G-protein coupled receptor
Tris	tris(hydroxymethyl)aminomethan
vdW	van der Waals
wt	wild-type
X-ray	röntgen radiation (wavelength in the range of 0.01 to 10 nanometers)

Chapter 1

Introduction

1.1 G-Protein coupled receptors

1.1.1 Classification and relevance of GPCRs

G-protein coupled receptors (GPCRs), forming one of the largest protein family in the human proteome (Lander *et al.*, 2001; Venter *et al.*, 2001) and accounting for about 2% of the human genome (Fredriksson *et al.*, 2003), are embedded in the plasma membrane and are the most important proteins for transferring signals from the extracellular medium to the cytoplasm. A huge variety of external stimuli are able to activate GPCRs, including photons, neurotransmitters, peptides, proteases, glycoprotein hormones, purine ligands and chemokines (Liapakis *et al.*, 2012). In humans these ligands address more than 800 GPCRs with about 400 functional non-olfactory receptors (updated figures on the homepage of the International Union of Basic and Clinical Pharmacology, IUPHAR, <http://www.iuphar-db.org/index.jsp>; Sharman *et al.*, 2013). By phylogenetic analysis they were classified in five groups: glutamate, rhodopsin, adhesion, frizzled/taste2 and secretin (GRAFS classification; Fredriksson *et al.*, 2003). The rhodopsin family, also referred to as class A GPCRs according to the A-F clan system, which covers all GPCRs in both vertebrates and invertebrates (Attwood and Findlay, 1994; Kolakowski, 1994), is by far the largest and most diverse family with about 700 receptor proteins (Lagerstrom and Schioth, 2008). It is subdivided into four groups (α , β , γ and δ) with the histamine receptor family located in the amine receptor cluster of the α -subgroup (Fredriksson *et al.*, 2003). Class B (secretin and adhesion receptor families with 15 and 24 members, respectively) and class C (glutamate receptor family with 15 human proteins) receptors contain only a few GPCRs. Emphasizing their importance for drug discovery, GPCRs are associated with many physiological processes and diseases, such as asthma, cancer, inflammation, obesity, pain as well as cardiovascular, metabolic, gastrointestinal and CNS diseases (Pierce *et al.*, 2002). Furthermore, genetic variations in GPCRs are responsible for more than 30 different human

diseases (Schöneberg *et al.*, 2004). To date, about 30% of all drugs on the market address GPCRs (Overington *et al.*, 2006). This includes top-selling drugs targeting α - and β -adrenergic receptors, 5-HT receptors, dopamine receptors, histamine receptors and angiotensin receptors. However, only a small number of the superfamily of GPCRs is addressed by drugs at all (De los Frailes and Diez, 2009; Lappano and Maggiolini, 2011). Additionally, more than 100 GPCRs belong to the group of orphan receptors. This means that the corresponding endogenous ligands are still unknown (Chung *et al.*, 2008). Thus, there is a great potential for the discovery of 'new' disease-related GPCRs. Taking into consideration the broad experience with established valuable GPCR addressing pharmacotherapeutics and the advances in structural biology in the past 13 years, GPCRs will continue to be in the focus of drug discovery (Salon *et al.*, 2011).

1.1.2 GPCR structure

The sequence of GPCRs is characterized by seven stretches with mainly hydrophobic residues, forming the common architecture of seven α -helical segments (seven transmembrane domains, 7TM) permeating the lipid core of the plasma membrane (Figure 1.1). At the intracellular surface GPCRs are able to interact with G-proteins (guanine nucleotide-binding proteins), transmitting the external stimuli to second messenger systems. However, the coupling of GPCRs to heterotrimeric G-proteins is just one possible signaling pathway (section 1.1.5). The notation 7TM receptor seems therefore more applicable (Kobilka, 2007). The seven TM domains are connected by three extra- and intracellular loops, respectively, being the most variable structures in GPCRs concerning length, sequence identity and flexibility (Kobilka and Schertler, 2008; Mirzadegan *et al.*, 2003; Wheatley *et al.*, 2012). In TMs several highly conserved motifs were recognized which are associated with specific functions in GPCRs, e.g. the D/ERY motif at the cytoplasmic end of TM3 which is part of the ionic lock, restraining the TM6 position close to TM3 by interactions with Glu^{6.30} (1) in the inactive receptor state of some GPCRs (Rovati *et al.*, 2007; Vogel *et al.*, 2008), the CWxP motif in TM6 composed of Trp^{6.48} and Pro^{6.50}, responsible for the rotamer toggle switch and the kink in TM6 (Shi *et al.*, 2002), or the NPxxY(x)_{5,6}F motif at the cytoplasmic part of TM7 which performs conformational changes during GPCR activation (Fritze *et al.*, 2003; Scheerer *et al.*, 2008).

(1) Residues within TM domains are named according to the Ballesteros/Weinstein nomenclature; the most conserved residue in each TM is numbered as X.50 where X is the number of the respective TM domain (Ballesteros and Weinstein, 1995)

These conserved structural features suggested a common mechanism of activation (Ahuja and Smith, 2009; Karnik *et al.*, 2003; Nygaard *et al.*, 2009; Schwartz *et al.*, 2006). However, for a long time the activation of GPCRs remained unclear due to the lack of structural information.

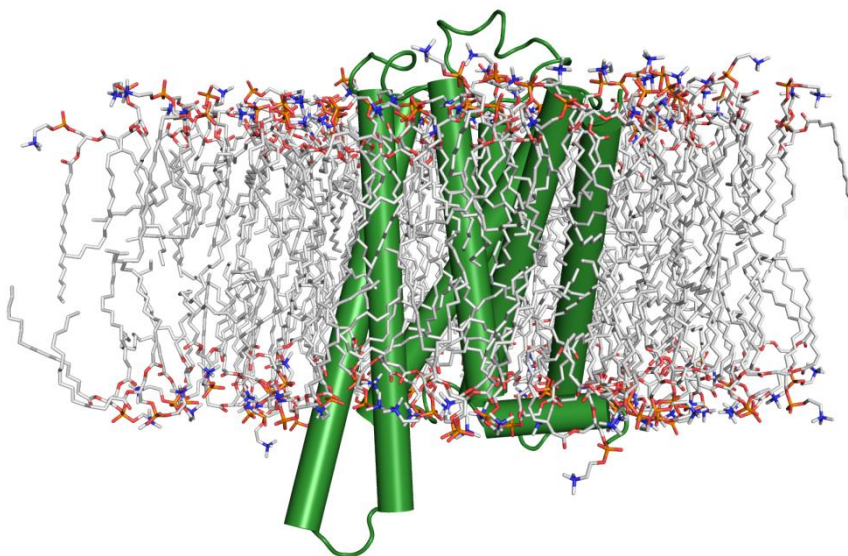


Figure 1.1: Model of a 7TM receptor embedded in a lipid bilayer

Large green tubes represent the seven TM domains extending from the extracellular medium (top) to the cytoplasmic side of the membrane (lipid core with grey carbons). TMs are connected by extra- and intracellular loops (green lines). At the intracellular side helix 8 (short green tube) is positioned parallel to the membrane within the polar lipid head groups.

First evidence for the 7TM structure of GPCRs was provided by electron microscopy studies of bacteriorhodopsin (Henderson and Unwin, 1975) and confirmed by a low resolution two-dimensional projection map of rhodopsin (Schertler *et al.*, 1993). A first 3D structure was published in 1995 with a resolution of about 9.5 Å, disclosing the overall shape of the rhodopsin molecule (Unger and Schertler, 1995). In the year 2000, the first high resolution X-ray structure of a GPCR, bovine rhodopsin (Palczewski *et al.*, 2000), was a breakthrough in structural biology. In the following years several rhodopsin crystal structures were resolved (Topiol and Sabio, 2009), which served as templates for molecular modeling studies of other GPCRs (Barton *et al.*, 2007). However, rhodopsin is unique among the class A receptor family. In contrast to GPCRs such as aminergic receptors the ligand 11-cis-retinal is covalently bound to rhodopsin by a Schiff base with Lys296^{7.43} (Li *et al.*, 2004). Furthermore,

the sequence is rather distant to that of other class A GPCRs. In later years, the X-ray structures of GPCRs (e.g. β ARs and adenosine A_{2A} receptor) proved that rhodopsin based homology models of GPCRs are too imprecise for drug design (Rosenbaum *et al.*, 2009). It was not until 2007 that the first non-rhodopsin GPCR structure was resolved, the human β_2 AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2007). The breakthrough was enabled by advances in GPCR crystallography (Venkatakrisnan *et al.*, 2013). For instance, the stabilization of the highly flexible third cytoplasmic loop by binding a F_{ab} antibody fragment or by generating a fusion protein (where T4 lysozyme was fused into ICL3 of the GPCR) resulted in more hydrophilic crystal contacts. This was as important as the insertion of point mutations in the protein to increase the thermostability of the receptor or the development of suitable crystallization matrices. Furthermore, using high affinity ligands with a slow rate of dissociation, receptors were locked in a single conformation and therefore stabilized during the crystallization (Tate, 2012). In the year 2008 even the first active receptor conformation of a GPCR, ligand free rhodopsin (opsin), was obtained (Park *et al.*, 2008; Scheerer *et al.*, 2008). Attaching a G-protein fragment to opsin helped to restrain the receptor in an active conformation. The comparison of inactive and active state GPCR conformations allowed invaluable insights into activating switches of GPCRs (Trzaskowski *et al.*, 2012). In 2011 Kobilka and coworkers released the ternary complex of an agonist bound β_2 AR with the nucleotide-free G_s heterotrimer, capturing the moment the receptor is activated by a ligand and sending a signal into the cell – a further milestone in GPCR research (Rasmussen *et al.*, 2011). Recently, the three-dimensional structure of the human CXCR1 in liquid crystalline phospholipid bilayers under physiological conditions and without stabilizing mutations was determined using NMR spectroscopy for the first time (Park *et al.*, 2012). All in all, since the year 2000, 20 different class A GPCRs have been determined (Venkatakrisnan *et al.*, 2013; Wacker *et al.*, 2013). Unfortunately, structural information about class B and class C GPCRs is limited, although the N-terminal ligand binding domain was already resolved for representative receptors of both classes. It is expected that whole protein structures of those GPCRs will be available within the next two years (Stevens *et al.*, 2013); metabotropic glutamate receptors for example are already in the pipeline (GPCR Network, <http://gpcr.scripps.edu>). The crystallization of a frizzled (class F) G-protein coupled receptor, the smoothened (SMO) receptor, was reported lately (Wang *et al.*, 2013). Also just recently the structure of active β -arrestin-1 bound to a carboxy-terminal G-protein coupled receptor phosphopeptide was published (Shukla *et al.*, 2013). Besides G-proteins β -arrestins are alternative signal transducing molecule interacting with 7TM receptors (see 1.1.5; Lefkowitz and Shenoy, 2005).

1.1.3 Ligand classification

Ligands able to activate and/or stabilize active GPCR conformations and, thus, to mobilize intracellular downstream effectors like G-proteins are termed agonists (Figure 1.2 A). However, ligand free receptors are also able to adopt active conformations and to interact with, e.g. G-proteins. Constitutive activity is frequently observed in wild-type and mutated GPCRs. The isomerization of inactive receptor conformations to active ones occur spontaneously, thus increasing the basal G-protein activity (Seifert and Wenzel-Seifert, 2002). In case of GPCRs exhibiting agonist-independent activity, inverse agonists can suppress basal activity. Such ligands stabilize receptor conformations which are not able to interact with intracellular signal transducers. According to the classical two state model (Figure 1.2 B), which assumes an inactive and active GPCR state (R and R^* , respectively), partial agonists and partial inverse agonists are able to shift the equilibrium to a certain degree, resulting in a submaximal effect compared to full agonists and full inverse agonists, respectively. However, based on the theory of multiple receptor states with certain energies this concept could be rendered more precisely (see 1.1.6). Antagonists are bound to receptors without changing the equilibrium between inactive and active states, i.e. without changing the basal activity of the respective receptor.

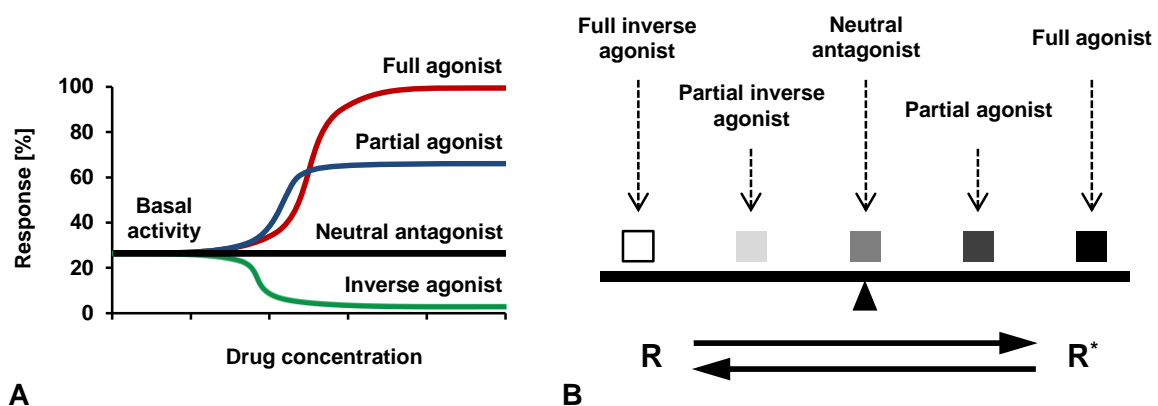


Figure 1.2: Classification of GPCR ligands and two state model of GPCR activation

A, Classification of GPCR ligands, adapted from Tate *et al.* (2012); B, Two state model of GPCR activation, adapted from Seifert and Wenzel-Seifert (2002). R^* and R , active and inactive receptor conformation, able to interact with signaling effectors (R^*) or not (R).

1.1.4 G-protein cycle

The first intracellular signal transducers identified to be activated by 7TM receptors were heterotrimeric G-proteins, composed of three subunits, α , β and γ (Oldham and Hamm, 2008). In total 21 $G\alpha$, 6 $G\beta$ and 12 $G\gamma$ subunits are found in humans (Downes and Gautam, 1999), whereas not all possible combinations of the three subunits are relevant for signal transduction (Denis *et al.*, 2012). Both the $G\alpha$ protein and the $G\beta\gamma$ -subunit are attached to the membrane (Chen and Manning, 2001; Dupre *et al.*, 2009). Compared to the number and diversity of ligands at GPCRs (peptides, biogenic amines, lipids and a magnitude of synthetic chemicals) the number of heterotrimers is rather low. This is in accord with the higher sequence similarity at the cytoplasmic side of the TM bundle (Mirzadegan *et al.*, 2003) which indicates a similar signal transduction mechanism of GPCRs. The heterotrimers are classified in four main classes based on the primary sequence of the $G\alpha$ subunit and, in part, on the selectivity of effectors ($G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$; Cabrera-Vera *et al.*, 2003; Milligan and Kostenis, 2006). Most important effectors for $G\alpha$ and $G\beta\gamma$ are shown in Figure 1.3. Some GPCRs couple to more than one G-protein subtype (Hermans, 2003; Lefkowitz *et al.*, 2002; Xiao, 2001).

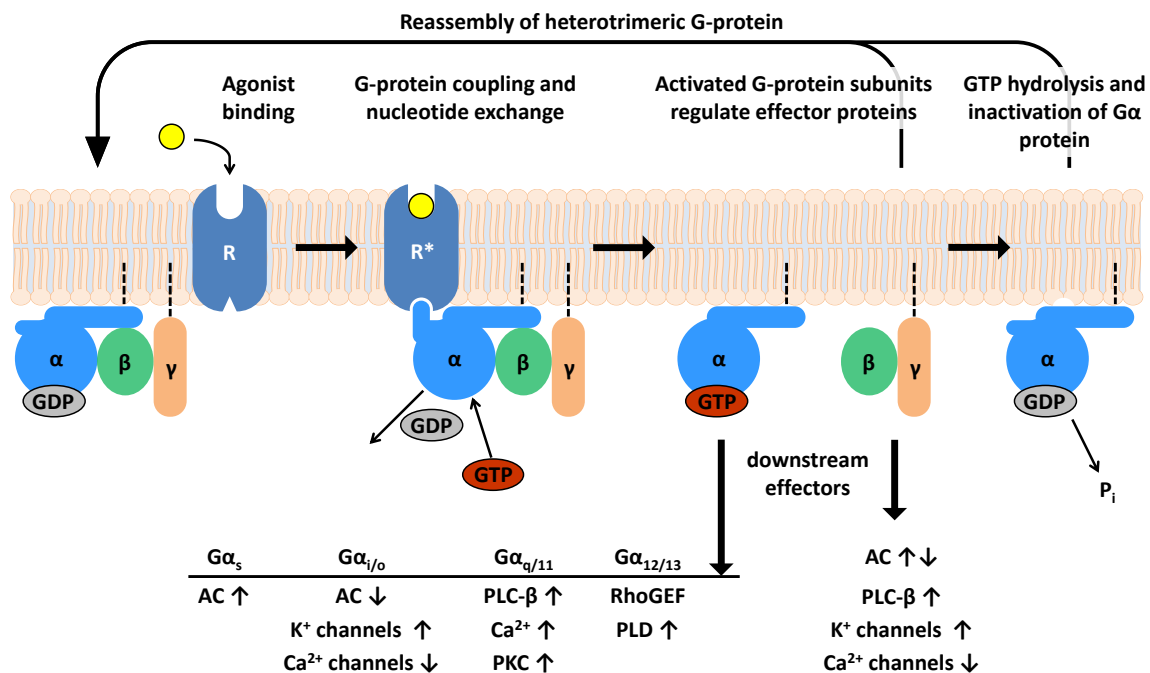


Figure 1.3: The G-protein cycle

R and R*, receptor conformation able to interact with signaling downstream effectors (R*) or not (R); AC, adenylyl cyclase; PLC- β , phospholipase C beta; PKC, protein kinase C; RhoGEF, structural domain of guanine nucleotide exchange factors for Rho/Rac/Cdc42-like GTPases; PLD, phospholipase D. Adapted from Rasmussen *et al.* (2011); downstream effectors were taken from Cabrera-Vera *et al.* (2003).

In the G-protein cycle the heterotrimer with GDP bound to the α -subunit interacts with an active GPCR conformation. The activated GPCR serves as catalytic activator for the GDP-GTP exchange of the α -subunit and thus conformational changes in $G\alpha$ results in a break of the heterotrimeric complex in $G\alpha$ and $G\beta\gamma$ and an adoption of subunit conformations capable of interacting with downstream effectors (Johnston and Siderovski, 2007). The GTPase activity of $G\alpha$ cleaves off the third phosphate group from GTP. The resulting GDP bound α -subunit recombines with the $G\beta\gamma$ subunit to return to the resting state (Figure 1.3). GTP hydrolysis can be accelerated by regulators of G-protein signaling (RGS) proteins of G_q and $G_{i/o}$ proteins (Magalhaes *et al.*, 2012).

1.1.5 Alternative signaling pathways and functional selectivity

Regardless of their role as ligand-regulated guanine nucleotide exchange factors for heterotrimeric G-proteins, GPCRs are capable of interacting with G-protein independent signal transduction pathways (Magalhaes *et al.*, 2012). Following receptor activation, G-protein coupled receptor kinases (GRKs) can phosphorylate serine and threonine residues in intracellular loops and C-terminus of GPCRs. As a consequence, β -arrestins bind to the modified receptor, preventing receptor–G-protein interactions and terminating GPCR signaling via heterotrimeric G-proteins (Kohout and Lefkowitz, 2003). For a long time, the recruitment of β -arrestin was considered necessary only for receptor desensitization, internalization and recycling of GPCRs (Lefkowitz, 1998; Lefkowitz and Shenoy, 2005). However, in the past decade evidence raised that GRKs and β -arrestins initiate G-protein independent pathways, such as the inhibition of nuclear factor κB (NF- κB)-targeted gene expression, scaffolding proteins of the mitogen-activated protein kinase (MAPK) signaling pathways and interactions with members of the c-Src family (Reiter *et al.*, 2012; Shukla *et al.*, 2011). The different G-protein and/or β -arrestin dependent pathways are preferentially or specifically activated by some ligands. These biased ligands are assumed to stabilize distinct receptor conformations which are capable of interacting exclusively or preferably with downstream signaling pathways (functional selectivity). Accordingly, the cubic ternary complex model, a model for 7TM receptor activation considering complexes of ligand-receptor, receptor-transducer and ligand-receptor-transducer for both, inactive and active GPCR conformations, has to be expanded to include receptor states interacting with different signal pathways (model of multiple signaling-component receptor conformations; Figure 1.4; Rajagopal *et al.*, 2010). Both models consider inactive (not able to interact with signaling downstream effectors) and active (able to initiate e.g. the G-protein cycle) GPCR conformations, explain constitutive activity (R_aT state: active receptor conformation without a ligand bound and linked to e.g. a signaling G-protein; cf. Figure 1.4), and the fact that also

inactive receptor states could couple to G-proteins without initiating the G-protein cycle (observed e.g. for opioid peptide receptors where antagonist bound inactive state conformations form GTP-sensitive, non-signaling ternary complexes, or for ligand free wild-type cannabinoid CB₁ receptors which sequester G-proteins – in the form of non-signaling ternary complexes – from other systems; Kenakin, 2004).

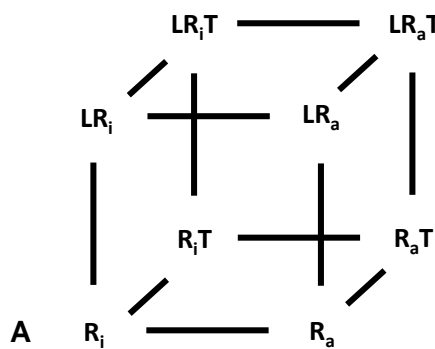
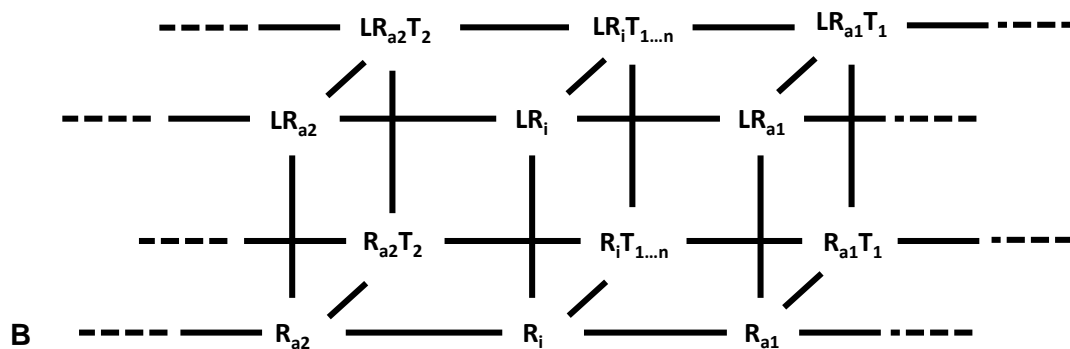


Figure 1.4: Receptor theory for GPCR activation

A, Cubic ternary complex model; B, A new model of GPCR activation considering multiple signaling-component receptor conformations (adapted from Rajagopal *et al.*, 2010). L, ligand; R_i, receptor in an inactive conformation which is not able to initiate the signaling of downstream effectors; R_a, active receptor conformation which could form signaling complexes with effectors; T, transducer, e.g. G-proteins or β -arrestins.



1.1.6 GPCR activation

GPCRs are dynamic, flexible molecules, capable of adopting specific receptor conformations, stabilized by functionally distinct ligands (Frauenfelder *et al.*, 1991; Kobilka and Deupi, 2007). To explain the different structural and functional receptor states, e.g. during the activation of rhodopsin where several conformations with different functionality are passed through (Hofmann *et al.*, 2009), it is not sufficient to consider a simple on-off switch model but to assume different conformations with distinct energies (Deupi and Kobilka, 2010). Stable conformations with low energy such as the structures obtained from X-ray

crystallography are favored. The transition from one local minimum energy conformation to another is determined by the differences between the respective energies of these states and the activation energy barrier (energy landscape of the receptor; Kobilka and Deupi, 2007). Conformational switches from one state to another occurring during receptor activation can happen in two different ways, by the induced-fit mechanism or by conformational selection. The major differences between these mechanisms consist in the first stages of ligand-receptor interaction and activation (Deupi and Kobilka, 2010). According to the induced-fit mechanism the ligand binds to the GPCR in the inactive conformation and thereby changes the energy landscape of the (ligand bound) receptor. Consequently, the energy difference between the inactive and active state is reduced. Furthermore, ligand binding provides the necessary energy for overcoming the transition barrier to achieve the active state, and the ligand stabilizes the receptor in its active conformation. A representative example of that mechanism is rhodopsin, which is characterized by low energy in the inactive conformation (and therefore locking the receptor in this state), as well as by a high energy barrier between inactive and active conformation (Figure 1.5 A). Hence, rhodopsin is devoid of constitutive activity (Govardhan and Oprian, 1994). The isomerization of covalently bound retinal upon light absorption involves energy transfer to the receptor, inducing conformational changes and the 'jump' over the activation barrier. In contrast, for the β_2 AR, representing an example of the conformational selection mechanism (Figure 1.5 B), the ligand-free receptor adopts multiple conformations (Ghanouni *et al.*, 2001; Peleg *et al.*, 2001) probably not separated by high energy barriers. This allows for a switch to an active receptor state, elevating the basal (constitutive) activity of the β_2 AR (Seifert *et al.*, 1998). In principle, the different conformational states can be stabilized by specific ligands, which in case of agonists lower the energy of the ligand-receptor complex LR* (ligand bound to an active receptor conformation). The increase in energy difference and barrier referred to R and LR hinders the transition to an inactive receptor state and therefore shifts the equilibrium of the receptor towards R*. In both mechanism G-protein binding further changes the energy landscape and shifts the equilibrium towards the active state of the receptor. To explain partial agonism two not-excluding theories are discussed. First, due to higher dissociation rates of partial agonists compared to full agonists not each ligand-receptor complex persists long enough to initiate the G-protein cycle. This results in a reduced effector response to ligand binding. Second, partial agonists are able to bind to receptor conformations which are not capable of interacting with downstream effectors in the same way as full agonists, but probably activate the G-protein only partially (Deupi and Kobilka, 2010).

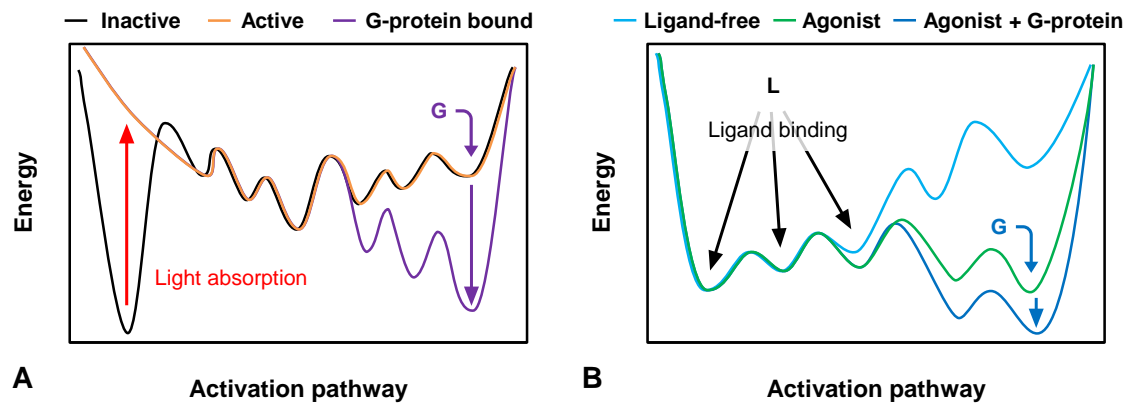


Figure 1.5: Energy landscape of the activation mechanism of rhodopsin and the β_2 AR

A, Energy landscape of rhodopsin (induced-fit mechanism); B, Energy landscape of the β_2 AR (conformational selection). G, G-protein; L, ligand. Adapted from Deupi and Kobilka (2010).

1.2 The histamine H₂R and the histamine receptor family

1.2.1 Histamine receptor subtypes

The biogenic amine histamine is synthesized in the cytosol via pyridoxal-5-phosphate dependent decarboxylation of histidine by L-histidine decarboxylase (HDC). The uptake of histamine in secretory granules is enabled by the vesicular monoamine transporter 2 (VMAT2; Kazumori *et al.*, 2004). High concentrations of histamine are found in mast cells, blood basophil and blood platelets, the skin, connective tissue, the lung and the gastrointestinal tract including enterochromaffin-like (ECL) cells in the stomach (Parsons and Ganellin, 2006). In the brain histaminergic neurons are involved in the sleep-wake cycle, energy and endocrine homeostasis, synaptic plasticity and learning (Haas and Panula, 2003). Histamine performs its actions as a local mediator and neurotransmitter via four histamine receptor subtypes (H₁R, H₂R, H₃R and H₄R), all belonging to aminergic class A GPCRs (Foord *et al.*, 2005; Seifert *et al.*, 2013). Despite binding the same endogenous ligand, a phylogenetic analysis (evolutionary ancestry based on sequence alignments) revealed that the histamine H₂ receptor clusters with adrenergic, dopamine and serotonin receptor subtypes (Figure 1.6; Vassilatis *et al.*, 2003). Within the histamine receptor family, H₃ and H₄ receptors are most closely related, sharing an overall and TM sequence identity of about 41% and 52%, respectively (Table 1.1).

TM sequence identity	Sequence identity			
	H ₁ R	H ₂ R	H ₃ R	H ₄ R
	H ₁ R	26	28	22
	H ₂ R	36	23	20
	H ₃ R	32	30	41
	H ₄ R	29	26	52

Table 1.1: Sequence identity of histamine receptor subtypes

Sequence identities (in %) are based on a multiple sequence alignment of the human H₁, H₂, H₃ and H₄ receptor. The definition of TM domains was taken from a crystal structure of the h β_2 AR (Cherezov *et al.*, 2007).

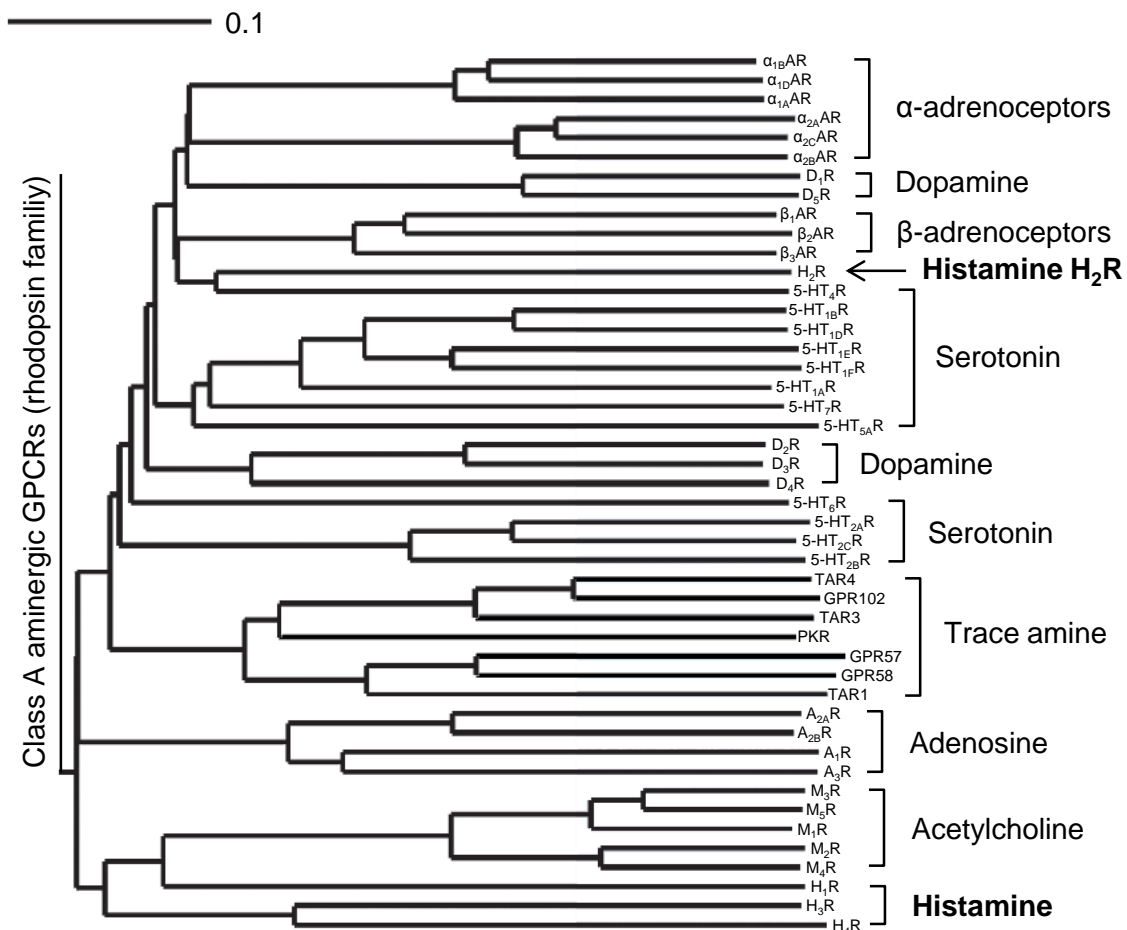


Figure 1.6: Phylogenetic tree of aminergic class A GPCRs

The ruler at the top indicates the horizontal distance equal to 10% sequence divergence. Adapted from Vassilatis *et al.* (2003).

The histamine H₁ receptor

The histamine H₁R, a 487 amino acids containing GPCR, is widely expressed in the body, e.g., in nerve cells, most smooth muscles especially in airways, gastrointestinal tract, endothelial and epithelial cells, neutrophils, genitourinary system and the cardiovascular system (Hill *et al.*, 1997). The H₁R couples predominantly to G_{q/11} proteins, initiating the phosphoinositide metabolism, resulting in inositol trisphosphate (IP₃) and diacylglycerol (DG). Consequently, Ca²⁺ is released from intracellular stores and protein kinase C is activated (Smit *et al.*, 1999). H₁R antagonists (antihistamines) have been used for decades for the treatment of allergic disorders (e.g. allergic rhinitis, chronic urticarial and atopic dermatitis), nausea and vomiting, and sedation (Du Buske, 1996; Simons and Simons, 2011; Simons, 2004). First generation antihistamines, like mepyramine or diphenhydramine, are highly lipophilic compounds which cross the blood brain barrier, block central H₁ receptors and cause sedation. More polar H₁R antagonists such as cetirizine were developed to reduce this undesired effect in the treatment of allergic diseases ('non-sedative' second generation of H₁R blockers). Besides, H₁R agonists such as the supra-/histaprodifens have been used as pharmacological tools to study H₁R functions in cellular systems. Betahistine (Aquamen®) is so far the only marketed H₁R agonist; the drug is therapeutically used in the treatment of Menière's disease (Barak, 2008; Seifert *et al.*, 2003).

A detailed description of the H₂R and its ligands is given in sections 1.2.2 to 1.2.4.

The histamine H₃ receptor

The histamine H₃ receptor (containing 445 amino acids) was first proposed in 1983. Experiments with rat cerebral cortical slices revealed that histamine inhibited its own synthesis and release via presynaptic feedback mechanisms which could not be attributed to H₁R or H₂R activity (Arrang *et al.*, 1983). The discovery of the agonist (R)- α -methylhistamine and the competitive antagonist thioperamide as well as cloning of the H₃R in 1999 enabled the investigation of its (patho)physiological roles and its intracellular mechanism (Arrang *et al.*, 1987; Lovenberg *et al.*, 1999). The H₃R is mostly expressed in the CNS and is important as a presynaptic auto- and heteroreceptor, controlling the release of histamine and other neurotransmitters such as dopamine, serotonin, noradrenalin and acetylcholine (Gemkow *et al.*, 2009; Hill *et al.*, 1997). The H₃R is involved in the regulation of several central functions like locomotor activity, wakefulness, food intake, thermoregulation and memory (Bakker *et al.*, 2004). In the periphery the H₃R was detected in the cardiovascular system, the gastrointestinal tract and the airways (Bertaccini and Coruzzi, 1995; Delaunois *et al.*, 1995;

Malinowska *et al.*, 1998). The activation of H₃Rs leads to a decrease in intracellular cAMP levels via coupling to G_{i/o} proteins and inhibition of the adenylyl cyclase. Besides, activation of phospholipase A₂ (PLA₂), MAPKs and phosphatidyl inositol-3 kinase, as well as inhibition of the Na⁺/H⁺ exchanger and modulation of intracellular calcium was demonstrated (Bongers *et al.*, 2007; Leurs *et al.*, 2005). Potential therapeutic applications for H₃R inverse agonists, antagonists or agonists include the treatment of migraine, asthma, allergic rhinitis, nociception, neuralgia, ischaemic arrhythmias, insomnia, cognitive disorders, tremor and obesity (Berlin *et al.*, 2011; Wijtmans *et al.*, 2007). Recently, the H₃R inverse agonist pitolisant was the first H₃R ligand to be introduced in the clinics for the treatment of narcolepsy (Schwartz, 2011).

The histamine H₄ receptor

The latest member of the histamine receptor family is the H₄R, comprising 390 amino acids. Although proposed 1994 (Raible *et al.*) it was not until the year 2000 that the H₄R was identified by several research groups due to its sequence homology with the H₃R (Liu *et al.*, 2001; Morse *et al.*, 2001; Nakamura *et al.*, 2000; Nguyen *et al.*, 2001; Oda *et al.*, 2000; Zhu *et al.*, 2001). The H₄R is mainly expressed in cells of hematopoietic origin like neutrophils, mast cells, eosinophils, basophils, dendritic cells, monocytes and T cells as well as the CNS (Connelly *et al.*, 2009; Leurs *et al.*, 2009). The location of H₄Rs in these cells suggested an important role in the modulation of immune and inflammatory responses, such as eosinophil chemotaxis, mast cell chemotaxis and chronic inflammation, dendritic cell activation and T cell differentiation, airway inflammation and allergy, chronic pruritus and autoimmune disorders (Thurmond *et al.*, 2008; Zampeli and Tiligada, 2009). The H₄R is coupled to pertussis toxin (PTX) sensitive G_{i/o} proteins and thus inhibits adenylyl cyclase (Leurs *et al.*, 2009). Additionally, the H₄R can activate the MAPK pathway via PTX sensitive mechanisms (Morse *et al.*, 2001) and induce calcium mobilization in mast cells and eosinophils, possibly initiated by the dissociated Gβγ subunit and PLC activation (de Esch *et al.*, 2005; Hofstra *et al.*, 2003). Besides coupling to G-proteins, recently, the activation of β-arrestin by several H₄R ligands was reported (Nijmeijer *et al.*, 2012; Seifert *et al.*, 2011). The H₄R represents an interesting target for the treatment of diseases like pruritus, atopic dermatitis, asthma and allergic rhinitis (Walter *et al.*, 2011). Some H₄R ligands already entered into clinical studies, e.g. UR-63325, the first H₄R antagonist from which clinical data has been reported (Lazewska and Kiec-Kononowicz, 2012), ZPL-38937887 (formerly PF-03893787) and JNJ-39758979 (Salcedo *et al.*, 2013).

1.2.2 Characterization of the H₂R

Effects of histamine at receptors different from those targeted by the classical antihistamines, were first detected at the stomach (acid secretion; 1941), the uterus (inhibition of the contraction in rats; 1946) and the heart (positive chronotrop; 1960). Ash and Schild (1966) noticed that these effects could not be blocked by histamine receptor antagonists such as mepyramine (later disclosed to be a H₁R selective antagonist) and therefore proposed an additional histamine receptor subtype. In 1972, this receptor was defined as the H₂R by pharmacological experiments using burimamide, the first antagonist at the H₂R (Black *et al.*, 1972). H₂Rs were found in gastric parietal cells, cardiac tissue, lung parenchyma, smooth muscles in airway, uterine and vascular smooth muscle and cells of the immune system (basophils, mast cells and lymphocytes) (Del Valle and Gantz, 1997; Hill *et al.*, 1997). Besides, the H₂R is widely distributed in the brain, e.g. the basal ganglia, hippocampus, amygdala and cerebral cortex (Traiffort *et al.*, 1992). An important physiological H₂R-mediated effect of histamine is the stimulation of parietal cells leading to acid secretion. Histamine is released from enterochromaffin-like cells upon stimulation of cholecystokinin CCK2 receptors by gastrin. Coupling of the H₂R to Gα_s proteins results in stimulation of cAMP production, initiating the fusion and activation of the H⁺/K⁺-ATPase (Schubert and Peura, 2008). Moreover, mobilisation of intracellular Ca²⁺ was reported (Delvalle *et al.*, 1992). Blocking the H₂R by antagonists such as cimetidine and ranitidine presented a breakthrough in the treatment of peptic ulcer disease (Malfertheiner *et al.*, 2009; Yeomans, 2002), offered new approaches to the pharmacotherapy of gastro-esophageal reflux disease (Katz and Tutuian, 2001) and Zollinger-Ellison syndrome (Wilcox and Hirschowitz, 2009) and, later on, proved to be useful in the triple therapy for helicobacter pylori eradication (Graham *et al.*, 2003). It should be noted, that nowadays H₂R antagonists are mostly replaced by proton pump inhibitors (PPI). The positive chronotropic and inotropic response on atrial and ventricular tissues and the vasodilatory effect via H₂R stimulation are mediated by an increase in cAMP production. In the heart the cAMP response results in both an increase in contractility and relaxation (positive lusitropic effect) (Levi and Alloatti, 1988). Besides, important roles of the H₂R in inflammation and modulation of the immune system were reported. H₂R activation leads to suppression of inflammatory functions by decreasing chemotaxis of eosinophils and neutrophils as well as inhibition of neutrophil activation, superoxide formation and degranulation. Suppression of immune response is mediated by reducing T cell proliferation and modulating cytokine production (Akdis and Simons, 2006; Schneider *et al.*, 2002). In the early 1990s the first H₂Rs were cloned, i.e. canine H₂R (Gantz *et al.*, 1991b), human H₂R (Gantz *et al.*, 1991a) rat H₂R (Ruat *et al.*, 1991) and guinea pig H₂R (Traiffort *et al.*, 1995). The sequence of the human H₂R is composed of 359 amino acids

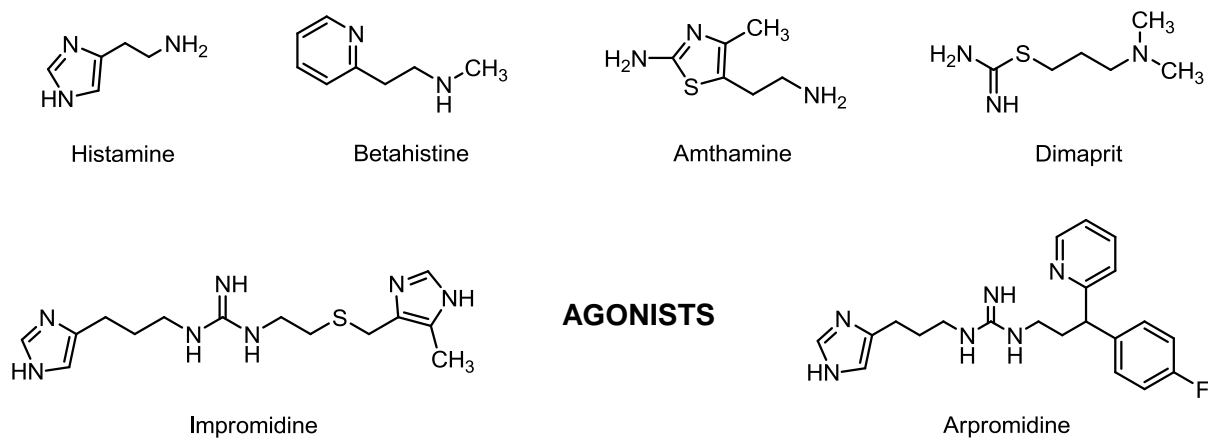
with seven stretches of mainly hydrophobic residues typical for 7TM receptors. Cloning of H₂R species and transfection in various cell lines and membranes enabled the detection of additional intracellular signaling mechanisms of the H₂R. Activation of the inositol 1,4,5-trisphosphate signal pathway via G α_q -proteins and phospholipase C (Kuhn *et al.*, 1996; Leopoldt *et al.*, 1997; Wang *et al.*, 1996) results in an increase in the intracellular Ca²⁺ concentration. This could contribute to the positive inotropic effects in cardiac myocytes (Wellner-Kienitz *et al.*, 2003). Agonistic stimulation of H₂R_s led to β -arrestin, dynamin and clathrin dependent desensitization and internalization (Fernandez *et al.*, 2008). H₂R desensitization was dependent on regulators of G-protein signaling (RGS) and G-protein coupled receptor kinase 2 (GRK2) without phosphorylation of the H₂R. However, GRK2 kinase activity was necessary for receptor internalization and the subsequent resensitization (Fernandez *et al.*, 2011).

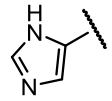
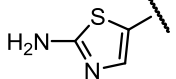
1.2.3 H₂R ligands

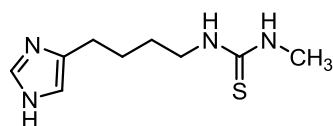
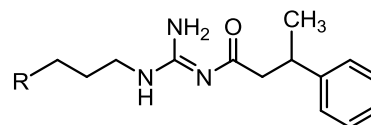
Metiamide (Black *et al.*, 1973), chemically derived from burimamide, was the first orally available H₂R antagonist which entered controlled trials in humans (Pounder *et al.*, 1975). Due to cases of agranulocytosis, a severe adverse effect, metiamide had to be withdrawn (Burland *et al.*, 1975), whereas cimetidine (trademark Tagamet™), the cyanoguanidine analogue of the thiourea metiamide, was developed and launched onto the market in 1976 (Brimblecombe *et al.*, 1975; Molinder, 1994). Cimetidine revolutionized the treatment of peptic ulcer and became one of the first blockbuster drugs in history (sales of \$1 billion annually). Stimulated by the proof of principle, additional H₂R antagonists were marketed as drugs, such as ranitidine or famotidine, and other compounds were developed as pharmacological tools to characterize the H₂R, e.g. tiotidine (Figure 1.7; Ganellin, 1992). Among a huge number of H₂R antagonists described in the literature, zolantidine is unique, as it was designed as a CNS-penetrating pharmacological tool for the investigation of H₂R_s in the brain (Calcutt *et al.*, 1988; Young *et al.*, 1988). Later on, most of the classical H₂R antagonists were characterized as inverse agonists (Monczor *et al.*, 1998; Smit *et al.*, 1996). With the discovery of the proton pump inhibitors, which irreversibly block the H⁺/K⁺-ATPase located in parietal cells of the gastric mucosa, the relevance of H₂R antagonists as drugs for the treatment of gastric acid related diseases has substantially declined (Fellenius *et al.*, 1981; Sachs *et al.*, 2007). The treatment of chronic heart failure has been explored as a potential indication for H₂R antagonists (Kim *et al.*, 2006; Takahama *et al.*, 2010).

In contrast to H₂R antagonists, agonists at the H₂R have not been routinely used in therapy. The potential therapeutic value of H₂R agonists as inotropic vasodilators in severe

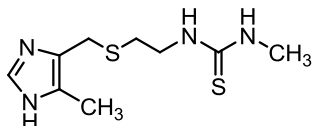
congestive heart failure, demonstrated in patients using impromidine, stimulated the search for 'cardiohistaminergics' with improved pharmacological properties (Baumann *et al.*, 1984; Buschauer, 1989; Buschauer and Baumann, 1991; Felix *et al.*, 1991). Moreover, one may speculate about H₂R agonists as anti-inflammatory agents (Burde *et al.*, 1990; Burde *et al.*, 1989). In 2005 histamine dihydrochloride (Ceplene[®]) was approved by the EMA (European Medicines Agency) in combination with interleukin-2 for maintenance treatment in adults with acute myeloid leukemia, AML (Yang and Perry, 2011). Besides, H₂R agonists are important pharmacological tools to study the physiological and pathophysiological role of this histamine receptor (Birnkammer *et al.*, 2012; Coruzzi *et al.*, 1993). 4-Methylhistamine (5-methylhistamine according to the IUPAC rules of chemical nomenclature) was used by Black *et al.* (1972) as a selective agonist to define the H₂R; it should be noted that 30 years later the compound turned out to be much more potent as a H₄R agonist. Impromidine was described as the first highly potent H₂R agonist, exhibiting 50-fold higher potency than histamine regarding the positive chronotropic response at the isolated, spontaneously beating guinea pig right atrium (Durant *et al.*, 1978). Numerous impromidine analogues and amine-type compounds such as amthamine and dimaprit have been synthesized and evaluated for H₂ agonism (for a review, cf. Dove *et al.*, 2004). In terms of drug-like properties, the major disadvantage of the basic N-[3-(1H-imidazol-4-yl)propyl]guanidines (pK_a ~ 12) is the lack of oral bioavailability (and CNS penetration). The insertion of a carbonyl next to the guanidine moiety decrease the basicity by 4-5 orders of magnitude and improved the pharmacokinetic properties of these acylguanidine-type agonists so that they were absorbed from the gut of mice and detectable in the brain (Ghorai *et al.*, 2008). After the discovery of the H₃R and the H₄R, the imidazolylpropylguanidines, regardless of being acylated or not, turned out to possess considerably high affinity to histamine receptors other than H₂R. This problem was solved by a bioisosteric approach: replacement of the imidazolyl ring by a 2-amino-4-methylthiazol-5-yl moiety (cf. structure of amthamine) resulted in highly potent and selective guanidine-type H₂R agonists (Figure 1.7, cf. compounds AK24 and AK470; Ghorai *et al.*, 2008; Kraus *et al.*, 2009). Remarkably, both guanidines and acylguanidines show higher potencies at guinea pig H₂Rs than at the human orthologue. An aspartate in the gpH₂R in position 7.36 (alanine in the hH₂R) was shown to contribute to the distinct interaction of H₂R species isoforms with acyl-/guanidines (Birnkammer *et al.*, 2012; Burde *et al.*, 1990; Buschauer, 1989; Kelley *et al.*, 2001; Preuss *et al.*, 2007). Recently synthesized bivalent compound exhibited a similar preference for the guinea pig compared to the human H₂R orthologue.

**AGONISTS**

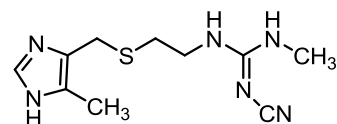
Compound	R	H _x R selectivity
AK24		H ₃ > H ₄ > H ₂ >> H ₁
AK470		H ₂ >> H ₁ , H ₃ , H ₄



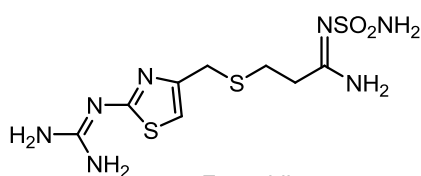
Burimamide



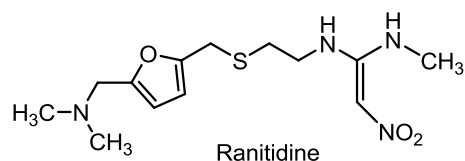
Metiamide



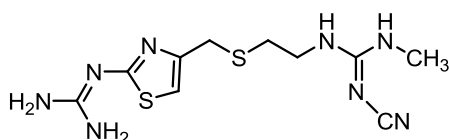
Cimetidine



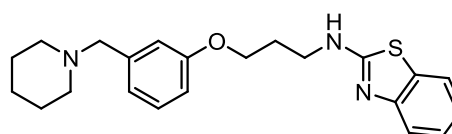
Famotidine

ANTAGONISTS

Ranitidine



Tiotidine



Zolantidine

Figure 1.7: Structures of H₂R ligands

Selectivity data (pEC₅₀) for AK24 and AK470 were determined in the GTPase activity assay (Ghorai *et al.*, 2008; Kraus *et al.*, 2009).

1.2.4 Bivalent ligands for the H₂ receptor

Bivalent ligands consist of two pharmacophores, separated by an appropriate spacer to define their distance and, if necessary, an additional linker to connect the pharmacophoric units and the spacer (Shonberg *et al.*, 2011). At first developed by Portoghese *et al.* in the 1980s for opioid receptors (Portoghese *et al.*, 1986), bivalent ligands are now widely used in targeting GPCRs and have been reported, e.g., for the CXCR4 (Tanaka *et al.*, 2010), dopamine D₂ receptor (Kühhorn *et al.*, 2011), β_2 -adrenergic and adenosine A₁ receptors (Karellas *et al.*, 2008), serotonin receptors (Halazy *et al.*, 1996) and histamine receptors (Birnkammer *et al.*, 2012).

Compared to the corresponding monovalent parent compound, bivalent ligands can have increased potency, intrinsic activity and receptor subtype selectivity, and the pharmacokinetic properties can be improved (Halazy, 1999). With regards to the study of ligand-GPCR interactions, bivalent ligands have been of special interest as pharmacological tools to explore putative receptor homo- and heterodimers, for instance, in case of adrenergic receptors (Angers *et al.*, 2000; Xu *et al.*, 2003), opioid receptors (Portoghese, 2001; Ramsay *et al.*, 2002), muscarinic receptors (Hern *et al.*, 2010; Zeng and Wess, 2000), dopamine receptors (Lukasiewicz *et al.*, 2010; Scarselli *et al.*, 2001) and histamine receptors (Fukushima *et al.*, 1997; van Rijn *et al.*, 2006). However, it is extremely challenging to determine the ligand-receptor stoichiometry and to get insight into the molecular binding mode as a crystal structure of such a GPCR dimer in complex with a bivalent ligand is not yet available. As a conceivable alternative to simultaneous binding to the orthosteric binding sites of dimerizing receptor protomers, interactions with two binding sites at the same protomer should be taken into account, i.e., one pharmacophoric moiety could bind to the orthosteric site and the second pharmacophoric moiety to an additional (allosteric) site (Halazy, 1999; May *et al.*, 2007; Messer, 2004; Perez *et al.*, 1998; Valant *et al.*, 2009). For 'bridging' of dimerizing receptor protomers, the distance between the pharmacophoric units is a critical issue. Although the ideal spacer length varies between 18 and 25 atoms for several GPCRs studied, the optimal length has to be elucidated empirically (Berque-Bestel *et al.*, 2008). Molecular modeling of opioid receptors suggested a distance of ~ 27 Å between two recognition sites of neighboring receptors with a TM5/TM6 interface (Portoghese, 2001). This corresponds to an extended alkyl chain with 20 methylene groups.

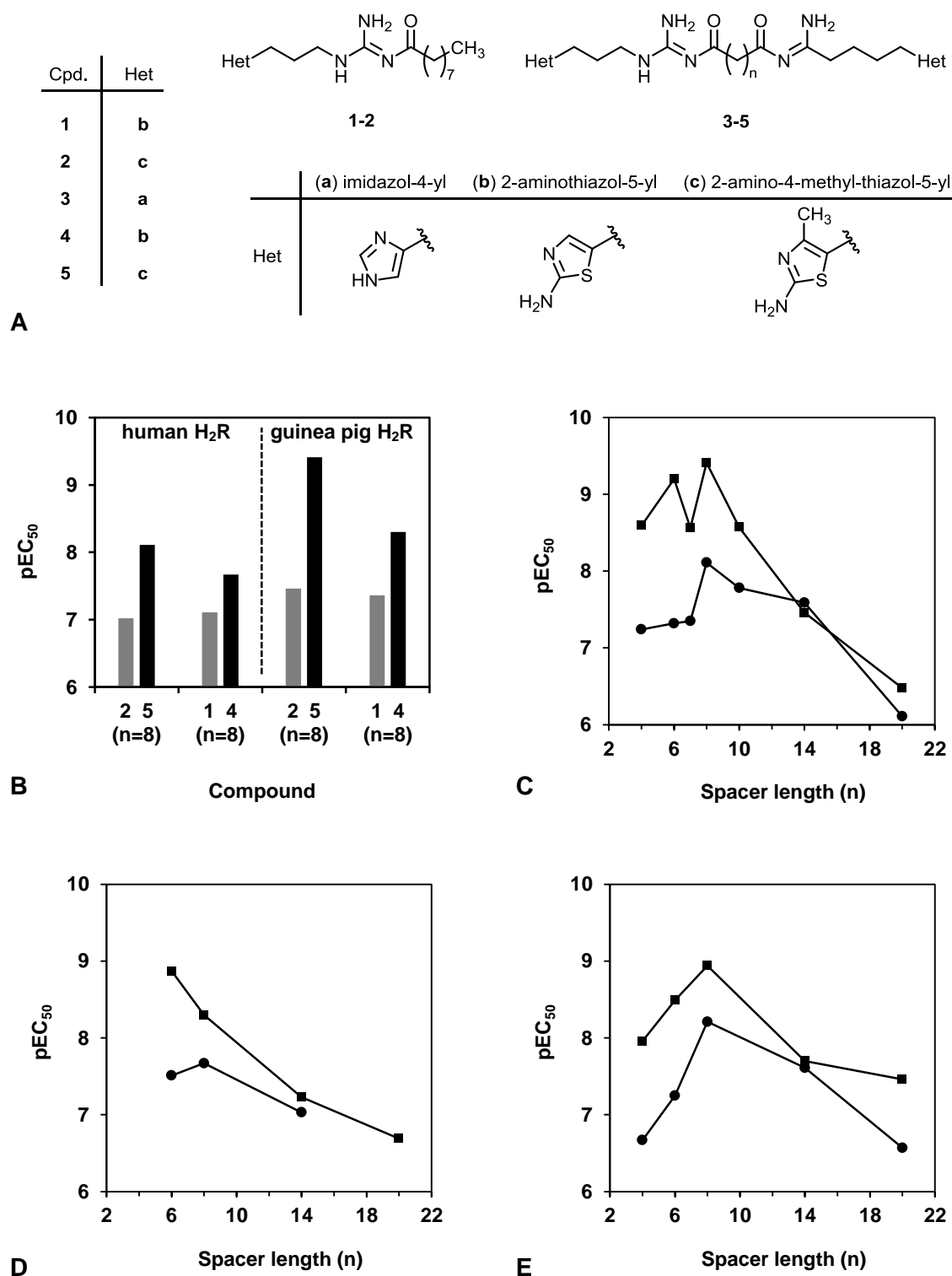


Figure 1.8: Potencies of bivalent acylguanidine-type H₂R agonists compared to their monovalent congeners and dependence on their spacer length

A, Structures of selected mono- and bivalent acylguanidine-type agonists. B, pEC₅₀ values of monovalent (grey) and bivalent (black) acylguanidine-type agonists at the human (left) and guinea pig (right) H₂R determined in the GTPase assay (Birnkammer, 2011; Birnkammer *et al.*, 2012). C - E, Potencies of bivalent acylguanidine-type agonists 5 (C), 4 (D) and 3 (E) with increasing spacer length n at the human (●) and guinea pig (■) H₂R in the GTPase assay (Birnkammer *et al.*, 2012). Cpd. 4 with 20 methylene groups in the spacer showed no agonistic activity at the hH₂R (D).

Recently reported bivalent H₂R agonists, hetarylpropylguanidines connected by dialkanoyl spacers attached to the guanidine moiety (Figure 1.8 A; Birnkammer, 2011; Birnkammer et al., 2012), were more potent than their monovalent parent compounds (Figure 1.8 B) and revealed highest potencies at spacer lengths of 6 to 8 methylene groups (~ 9 Å to 11.5 Å; Figure 1.8 C-E). Obviously, the distance of < 12 Å between the pharmacophores is insufficient to bind simultaneously to the orthosteric sites of dimerizing GPCRs. Instead, the increase in potency might be caused by interaction of the second acylguanidine moiety with an additional recognition site at the same receptor protomer. This hypothesis was investigated by site-directed mutagenesis experiments at the gpH₂R presented in chapter 7 of this thesis.

1.3 References

- Ahuja S and Smith SO. Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol Sci* **2009**, 30, 494-502.
- Akdis CA and Simons FE. Histamine receptors are hot in immunopharmacology. *Eur J Pharmacol* **2006**, 533, 69-76.
- Angers S, Salahpour A, Joly E, Hilairret S, Chelsky D, Dennis M and Bouvier M. Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET). *Proc Natl Acad Sci U S A* **2000**, 97, 3684-3689.
- Arrang JM, Garbarg M, Lancelot JC, Lecomte JM, Pollard H, Robba M, Schunack W and Schwartz JC. Highly potent and selective ligands for histamine H₃-receptors. *Nature* **1987**, 327, 117-123.
- Arrang JM, Garbarg M and Schwartz JC. Auto-inhibition of brain histamine release mediated by a novel class (H₃) of histamine receptor. *Nature* **1983**, 302, 832-837.
- Ash AS and Schild HO. Receptors mediating some actions of histamine. *Br J Pharmacol Chemother* **1966**, 27, 427-439.
- Attwood TK and Findlay JB. Fingerprinting G-protein-coupled receptors. *Protein Eng* **1994**, 7, 195-203.
- Bakker RA, Dees G, Carrillo JJ, Booth RG, Lopez-Gimenez JF, Milligan G, Strange PG and Leurs R. Domain swapping in the human histamine H₁ receptor. *J Pharmacol Exp Ther* **2004**, 311, 131-138.
- Ballesteros JA and Weinstein H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In *Methods Neurosci*, Stuart CS, Ed. Academic Press: **1995**; Vol. Volume 25, pp 366-428.
- Barak N. Betahistidine: what's new on the agenda? *Expert Opin Investig Drugs* **2008**, 17, 795-804.
- Barton N, Blaney FE, Garland S, Tehan B and Wall I. 4.26 - Seven Transmembrane G Protein-Coupled Receptors: Insights for Drug Design from Structure and Modeling. In *Comprehensive Medicinal Chemistry II*, Editors-in-Chief: John BT and David JT, Eds., Elsevier: Oxford, **2007**; pp 669-701.
- Baumann G, Permanetter B and Wirtzfeld A. Possible value of H₂-receptor agonists for treatment of catecholamine-insensitive congestive heart failure. *Pharmacol Ther* **1984**, 24, 165-177.
- Berlin M, Boyce CW and Ruiz Mde L. Histamine H₃ receptor as a drug discovery target. *J Med Chem* **2011**, 54, 26-53.

- Berque-Bestel I, Lezoualc'h F and Jockers R. Bivalent ligands as specific pharmacological tools for G protein-coupled receptor dimers. *Curr Drug Discov Technol* **2008**, 5, 312-318.
- Bertaccini G and Coruzzi G. An update on histamine H3 receptors and gastrointestinal functions. *Dig Dis Sci* **1995**, 40, 2052-2063.
- Birnkammer T. Highly potent and selective acylguanidine-type histamine H2 receptor agonists: synthesis and structure-activity relationships of mono- and bivalent ligands. Doctoral thesis, Regensburg, Regensburg, **2011**. <http://epub.uni-regensburg.de/22237/>
- Birnkammer T, Spickenreither A, Brunskole I, Lopuch M, Kagermeier N, Bernhardt G, Dove S, Seifert R, Elz S and Buschauer A. The Bivalent Ligand Approach Leads to Highly Potent and Selective Acylguanidine-Type Histamine H-2 Receptor Agonists. *J Med Chem* **2012**, 55, 1147-1160.
- Black JW, Duncan WA, Durant CJ, Ganellin CR and Parsons EM. Definition and antagonism of histamine H2-receptors. *Nature* **1972**, 236, 385-390.
- Black JW, Duncan WA, Emmett JC, Ganellin CR, Hesselbo T, Parsons ME and Wyllie JH. Metiamide--an orally active histamine H2-receptor antagonist. *Agents Actions* **1973**, 3, 133-137.
- Bongers G, Bakker RA and Leurs R. Molecular aspects of the histamine H3 receptor. *Biochem Pharmacol* **2007**, 73, 1195-1204.
- Brimblecombe RW, Duncan WA, Durant GJ, Ganellin CR, Parsons ME and Black JW. The pharmacology of cimetidine, a new histamine H2-receptor antagonist. *Br J Pharmacol* **1975**, 53, 435P-436P.
- Burde R, Buschauer A and Seifert R. Characterization of histamine H2-receptors in human neutrophils with a series of guanidine analogues of impromidine. Are cell type-specific H2-receptors involved in the regulation of NADPH oxidase? *Naunyn Schmiedeberg's Arch Pharmacol* **1990**, 341, 455-461.
- Burde R, Seifert R, Buschauer A and Schultz G. Histamine inhibits activation of human neutrophils and HL-60 leukemic cells via H2-receptors. *Naunyn Schmiedeberg's Arch Pharmacol* **1989**, 340, 671-678.
- Burland WL, Duncan WA, Hesselbo T, Mills JG, Sharpe PC, Haggie SJ and Wyllie JH. Pharmacological evaluation of cimetidine, a new histamine H2-receptor antagonist, in healthy man. *Br J Clin Pharmacol* **1975**, 2, 481-486.
- Buschauer A. Synthesis and in vitro pharmacology of arpromidine and related phenyl(pyridylalkyl)guanidines, a potential new class of positive inotropic drugs. *J Med Chem* **1989**, 32, 1963-1970.
- Buschauer A and Baumann G. Structure-activity relationships of histamine H2-agonists, a new class of positive inotropic drugs. *Agents Actions Suppl* **1991**, 33, 231-256.
- Cabrera-Vera TM, Vanhauwe J, Thomas TO, Medkova M, Preininger A, Mazzoni MR and Hamm HE. Insights into G protein structure, function, and regulation. *Endocr Rev* **2003**, 24, 765-781.
- Calcutt CR, Ganellin CR, Griffiths R, Leigh BK, Maguire JP, Mitchell RC, Mylek ME, Parsons ME, Smith IR and Young RC. Zolantidine (SK&F 95282) is a potent selective brain-penetrating histamine H2-receptor antagonist. *Br J Pharmacol* **1988**, 93, 69-78.
- Chen CA and Manning DR. Regulation of G proteins by covalent modification. *Oncogene* **2001**, 20, 1643-1652.
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK and Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, 318, 1258-1265.
- Chung S, Funakoshi T and Civelli O. Orphan GPCR research. *Br J Pharmacol* **2008**, 153, 10.
- Connelly WM, Shenton FC, Lethbridge N, Leurs R, Waldvogel HJ, Faull RL, Lees G and Chazot PL. The histamine H4 receptor is functionally expressed on neurons in the mammalian CNS. *Br J Pharmacol* **2009**, 157, 55-63.

- Coruzzi G, Timmerman H, Adami M and Bertaccini G. The new potent and selective histamine H₂ receptor agonist amthamine as a tool to study gastric secretion. *Naunyn Schmiedebergs Arch Pharmacol* **1993**, 348, 77-81.
- de Esch IJ, Thurmond RL, Jongejan A and Leurs R. The histamine H₄ receptor as a new therapeutic target for inflammation. *Trends Pharmacol Sci* **2005**, 26, 462-469.
- De los Frailes M and Diez E. Screening technologies for G protein-coupled receptors: from HTS to uHTS. *Methods Mol Biol* **2009**, 552, 15-37.
- Del Valle J and Gantz I. Novel insights into histamine H₂ receptor biology. *Am J Physiol* **1997**, 273, G987-996.
- Delaunois A, Gustin P, Garbarg M and Ansay M. Modulation of acetylcholine, capsaicin and substance P effects by histamine H₃ receptors in isolated perfused rabbit lungs. *Eur J Pharmacol* **1995**, 277, 243-250.
- Delvalle J, Wang L, Gantz I and Yamada T. Characterization of H₂ histamine receptor: linkage to both adenylate cyclase and [Ca²⁺]_i signaling systems. *Am J Physiol* **1992**, 263, G967-972.
- Denis C, Sauliere A, Galandrin S, Senard JM and Gales C. Probing heterotrimeric G protein activation: applications to biased ligands. *Curr Pharm Des* **2012**, 18, 128-144.
- Deupi X and Kobilka BK. Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology* **2010**, 25, 293-303.
- Dove S, Elz S, Seifert R and Buschauer A. Structure-activity relationships of histamine H₂ receptor ligands. *Mini Rev Med Chem* **2004**, 4, 941-954.
- Downes GB and Gautam N. The G protein subunit gene families. *Genomics* **1999**, 62, 544-552.
- Du Buske LM. Clinical comparison of histamine H₁-receptor antagonist drugs. *J Allergy Clin Immunol* **1996**, 98, S307-318.
- Dupre DJ, Robitaille M, Rebois RV and Hebert TE. The role of Gbetagamma subunits in the organization, assembly, and function of GPCR signaling complexes. *Annu Rev Pharmacol Toxicol* **2009**, 49, 31-56.
- Durant GJ, Duncan WA, Ganellin CR, Parsons ME, Blakemore RC and Rasmussen AC. Impromidine (SK&F 92676) is a very potent and specific agonist for histamine H₂ receptors. *Nature* **1978**, 276, 403-405.
- Felix SB, Buschauer A and Baumann G. Therapeutic value of H₂-receptor stimulation in congestive heart failure. Hemodynamic effects of BU-E-76, BU-E-75 and arpromidine (BU-E-50) in comparison to impromidine. *Agents Actions Suppl* **1991**, 33, 257-269.
- Fellenius E, Berglinth T, Sachs G, Olbe L, Elander B, Sjostrand SE and Wallmark B. Substituted benzimidazoles inhibit gastric acid secretion by blocking (H⁺ + K⁺)ATPase. *Nature* **1981**, 290, 159-161.
- Fernandez N, Gottardo FL, Alonso MN, Monczor F, Shayo C and Davio C. Roles of phosphorylation-dependent and -independent mechanisms in the regulation of histamine H₂ receptor by G protein-coupled receptor kinase 2. *J Biol Chem* **2011**, 286, 28697-28706.
- Fernandez N, Monczor F, Baldi A, Davio C and Shayo C. Histamine H₂ receptor trafficking: role of arrestin, dynamin, and clathrin in histamine H₂ receptor internalization. *Mol Pharmacol* **2008**, 74, 1109-1118.
- Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, Spedding M and Harmar AJ. International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacol Rev* **2005**, 57, 279-288.
- Frauenfelder H, Sligar SG and Wolynes PG. The energy landscapes and motions of proteins. *Science* **1991**, 254, 1598-1603.
- Fredriksson R, Lagerstrom MC, Lundin LG and Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**, 63, 1256-1272.

- Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP and Ernst OP. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. *Proc Natl Acad Sci U S A* **2003**, 100, 2290-2295.
- Fukushima Y, Asano T, Saitoh T, Anai M, Funaki M, Ogihara T, Katagiri H, Matsushashi N, Yazaki Y and Sugano K. Oligomer formation of histamine H2 receptors expressed in Sf9 and COS7 cells. *FEBS Lett* **1997**, 409, 283-286.
- Ganellin CR. Pharmacochemistry of H1 and H2 receptors. In *The Histamine Receptor*, Schwartz JC and Haas H, Eds., Wiley-Liss: New York, **1992**; pp 1-56.
- Gantz I, Munzert G, Tashiro T, Schaffer M, Wang L, DelValle J and Yamada T. Molecular cloning of the human histamine H2 receptor. *Biochem Biophys Res Commun* **1991a**, 178, 1386-1392.
- Gantz I, Schaffer M, DelValle J, Logsdon C, Campbell V, Uhler M and Yamada T. Molecular cloning of a gene encoding the histamine H2 receptor. *Proc Natl Acad Sci U S A* **1991b**, 88, 429-433.
- Gemkow MJ, Davenport AJ, Harich S, Ellenbroek BA, Cesura A and Hallett D. The histamine H3 receptor as a therapeutic drug target for CNS disorders. *Drug Discov Today* **2009**, 14, 509-515.
- Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR and Kobilka BK. Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. *J Biol Chem* **2001**, 276, 24433-24436.
- Ghorai P, Kraus A, Keller M, Götte C, Igel P, Schneider E, Schnell D, Bernhardt Gn, Dove S, Zabel M, Elz S, Seifert R, et al. Acylguanidines as Bioisosteres of Guanidines: NG-Acylated Imidazolylpropylguanidines, a New Class of Histamine H2 Receptor Agonists. *J Med Chem* **2008**, 51, 7193-7204.
- Govardhan CP and Oprian DD. Active site-directed inactivation of constitutively active mutants of rhodopsin. *J Biol Chem* **1994**, 269, 6524-6527.
- Graham DY, Hammoud F, El-Zimaity HM, Kim JG, Osato MS and El-Serag HB. Meta-analysis: proton pump inhibitor or H2-receptor antagonist for *Helicobacter pylori* eradication. *Aliment Pharmacol Ther* **2003**, 17, 1229-1236.
- Haas H and Panula P. The role of histamine and the tuberomammillary nucleus in the nervous system. *Nat Rev Neurosci* **2003**, 4, 121-130.
- Halazy S. G-protein coupled receptors bivalent ligands and drug design. *Expert Opin Ther Pat* **1999**, 9, 431-446.
- Halazy S, Perez M, Fourrier C, Pallard I, Pauwels PJ, Palmier C, John GW, Valentin JP, Bonnafeous R and Martinez J. Serotonin dimers: application of the bivalent ligand approach to the design of new potent and selective 5-HT(1B/1D) agonists. *J Med Chem* **1996**, 39, 4920-4927.
- Henderson R and Unwin PN. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **1975**, 257, 28-32.
- Hermans E. Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacol Ther* **2003**, 99, 25-44.
- Hern JA, Baig AH, Mashanov GI, Birdsall B, Corrie JE, Lazareno S, Molloy JE and Birdsall NJ. Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules. *Proc Natl Acad Sci U S A* **2010**, 107, 2693-2698.
- Hill SJ, Ganellin CR, Timmerman H, Schwartz JC, Shankley NP, Young JM, Schunack W, Levi R and Haas HL. International union of pharmacology .13. Classification of histamine receptors. *Pharmacol Rev* **1997**, 49, 253-278.
- Hofmann KP, Scheerer P, Hildebrand PW, Choe HW, Park JH, Heck M and Ernst OP. A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem Sci* **2009**, 34, 540-552.
- Hofstra CL, Desai PJ, Thurmond RL and Fung-Leung WP. Histamine H4 receptor mediates chemotaxis and calcium mobilization of mast cells. *J Pharmacol Exp Ther* **2003**, 305, 1212-1221.

- Johnston CA and Siderovski DP. Receptor-Mediated Activation of Heterotrimeric G-Proteins: Current Structural Insights. *Mol Pharmacol* **2007**, 72, 219-230.
- Karellas P, McNaughton M, Baker SP and Scammells PJ. Synthesis of bivalent beta2-adrenergic and adenosine A1 receptor ligands. *J Med Chem* **2008**, 51, 6128-6137.
- Karnik SS, Gogonea C, Patil S, Saad Y and Takezako T. Activation of G-protein-coupled receptors: a common molecular mechanism. *Trends Endocrinol Metab* **2003**, 14, 431-437.
- Katz PO and Tutuian R. Histamine receptor antagonists, proton pump inhibitors and their combination in the treatment of gastro-oesophageal reflux disease. *Best Pract Res Clin Gastroenterol* **2001**, 15, 371-384.
- Kazumori H, Ishihara S, Rumi MA, Ortega-Cava CF, Kadowaki Y and Kinoshita Y. Transforming growth factor-alpha directly augments histidine decarboxylase and vesicular monoamine transporter 2 production in rat enterochromaffin-like cells. *Am J Physiol Gastrointest Liver Physiol* **2004**, 286, 16.
- Kelley MT, Burckstummer T, Wenzel-Seifert K, Dove S, Buschauer A and Seifert R. Distinct interaction of human and guinea pig histamine H2-receptor with guanidine-type agonists. *Mol Pharmacol* **2001**, 60, 1210-1225.
- Kenakin T. Principles: receptor theory in pharmacology. *Trends Pharmacol Sci* **2004**, 25, 186-192.
- Kim J, Ogai A, Nakatani S, Hashimura K, Kanzaki H, Komamura K, Asakura M, Asanuma H, Kitamura S, Tomoike H and Kitakaze M. Impact of blockade of histamine H2 receptors on chronic heart failure revealed by retrospective and prospective randomized studies. *J Am Coll Cardiol* **2006**, 48, 1378-1384.
- Kobilka B and Schertler GF. New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol Sci* **2008**, 29, 79-83.
- Kobilka BK. G protein coupled receptor structure and activation. *Biochim Biophys Acta* **2007**, 4, 794-807.
- Kobilka BK and Deupi X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol Sci* **2007**, 28, 397-406.
- Kohout TA and Lefkowitz RJ. Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization. *Mol Pharmacol* **2003**, 63, 9-18.
- Kolakowski LF, Jr. GCRDb: a G-protein-coupled receptor database. *Receptors Channels* **1994**, 2, 1-7.
- Kraus A, Ghorai P, Birnkammer T, Schnell D, Elz S, Seifert R, Dove S, Bernhardt G and Buschauer A. N-G-Acylated Aminothiazolylpropylguanidines as Potent and Selective Histamine H2 Receptor Agonists. *ChemMedChem* **2009**, 4, 232-240.
- Kühhorn J, Hubner H and Gmeiner P. Bivalent dopamine D2 receptor ligands: synthesis and binding properties. *J Med Chem* **2011**, 54, 4896-4903.
- Kuhn B, Schmid A, Harteneck C, Gudermann T and Schultz G. G proteins of the Gq family couple the H2 histamine receptor to phospholipase C. *Mol Endocrinol* **1996**, 10, 1697-1707.
- Lagerstrom MC and Schioth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **2008**, 7, 339-357.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, 409, 860-921.
- Lappano R and Maggiolini M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat Rev Drug Discov* **2011**, 10, 47-60.
- Lazewska D and Kiec-Kononowicz K. Azines as histamine H4 receptor antagonists. *Front Biosci* **2012**, 4, 967-987.
- Lefkowitz RJ. G protein-coupled receptors. III. New roles for receptor kinases and beta-arrestins in receptor signaling and desensitization. *J Biol Chem* **1998**, 273, 18677-18680.

- Lefkowitz RJ, Pierce KL and Luttrell LM. Dancing with different partners: protein kinase a phosphorylation of seven membrane-spanning receptors regulates their G protein-coupling specificity. *Mol Pharmacol* **2002**, 62, 971-974.
- Lefkowitz RJ and Shenoy SK. Transduction of receptor signals by beta-arrestins. *Science* **2005**, 308, 512-517.
- Leopoldt D, Harteneck C and Nurnberg B. G proteins endogenously expressed in Sf 9 cells: interactions with mammalian histamine receptors. *Naunyn Schmiedebergs Arch Pharmacol* **1997**, 356, 216-224.
- Leurs R, Bakker RA, Timmerman H and de Esch IJ. The histamine H3 receptor: from gene cloning to H3 receptor drugs. *Nat Rev Drug Discov* **2005**, 4, 107-120.
- Leurs R, Chazot PL, Shenton FC, Lim HD and de Esch IJ. Molecular and biochemical pharmacology of the histamine H4 receptor. *Br J Pharmacol* **2009**, 157, 14-23.
- Levi RC and Alloatti G. Histamine modulates calcium current in guinea pig ventricular myocytes. *J Pharmacol Exp Ther* **1988**, 246, 377-383.
- Li J, Edwards PC, Burghammer M, Villa C and Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **2004**, 343, 1409-1438.
- Liapakis G, Cordomi A and Pardo L. The G-protein coupled receptor family: actors with many faces. *Curr Pharm Des* **2012**, 18, 175-185.
- Liu C, Ma X, Jiang X, Wilson SJ, Hofstra CL, Blevitt J, Pyati J, Li X, Chai W, Carruthers N and Lovenberg TW. Cloning and pharmacological characterization of a fourth histamine receptor (H(4)) expressed in bone marrow. *Mol Pharmacol* **2001**, 59, 420-426.
- Lovenberg TW, Roland BL, Wilson SJ, Jiang X, Pyati J, Huvar A, Jackson MR and Erlander MG. Cloning and functional expression of the human histamine H3 receptor. *Mol Pharmacol* **1999**, 55, 1101-1107.
- Lukasiewicz S, Polit A, Kedracka-Krok S, Wedzony K, Mackowiak M and Dziejzicka-Wasylewska M. Hetero-dimerization of serotonin 5-HT(2A) and dopamine D(2) receptors. *Biochim Biophys Acta* **2010**, 12, 8.
- Magalhaes AC, Dunn H and Ferguson SS. Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins. *Br J Pharmacol* **2012**, 165, 1717-1736.
- Malfertheiner P, Chan FKL and McColl KEL. Peptic ulcer disease. *Lancet* **2009**, 374, 1449-1461.
- Malinowska B, Godlewski G and Schlicker E. Histamine H3 receptors--general characterization and their function in the cardiovascular system. *J Physiol Pharmacol* **1998**, 49, 191-211.
- May LT, Leach K, Sexton PM and Christopoulos A. Allosteric Modulation of G Protein-Coupled Receptors. *Annual Review of Pharmacology and Toxicology* **2007**, 47, 1-51.
- Messer WS, Jr. Bivalent ligands for G protein-coupled receptors. *Curr Pharm Des* **2004**, 10, 2015-2020.
- Milligan G and Kostenis E. Heterotrimeric G-proteins: a short history. *British Journal of Pharmacology* **2006**, 147, S46-S55.
- Mirzadegan T, Benko G, Filipek S and Palczewski K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* **2003**, 42, 2759-2767.
- Molinder HK. The development of cimetidine: 1964-1976. A human story. *J Clin Gastroenterol* **1994**, 19, 248-254.
- Monczor F, Legnazzi BL, Rivera E and Davio C. Tiotidine, a classical H2-antagonist, presents characteristics of an inverse agonist in U937 cell line. *Inflamm Res* **1998**, 47, S42-43.
- Morse KL, Behan J, Laz TM, West RE, Jr., Greenfeder SA, Anthes JC, Umland S, Wan Y, Hipkin RW, Gonsiorek W, Shin N, Gustafson EL, et al. Cloning and characterization of a novel human histamine receptor. *J Pharmacol Exp Ther* **2001**, 296, 1058-1066.
- Nakamura T, Itadani H, Hidaka Y, Ohta M and Tanaka K. Molecular cloning and characterization of a new human histamine receptor, HH4R. *Biochem Biophys Res Commun* **2000**, 279, 615-620.

- Nguyen T, Shapiro DA, George SR, Setola V, Lee DK, Cheng R, Rauser L, Lee SP, Lynch KR, Roth BL and O'Dowd BF. Discovery of a novel member of the histamine receptor family. *Mol Pharmacol* **2001**, 59, 427-433.
- Nijmeijer S, Vischer HF, Rosethorne EM, Charlton SJ and Leurs R. Analysis of multiple histamine H(4) receptor compound classes uncovers Galphai protein- and beta-arrestin2-biased ligands. *Mol Pharmacol* **2012**, 82, 1174-1182.
- Nygaard R, Frimurer TM, Holst B, Rosenkilde MM and Schwartz TW. Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol Sci* **2009**, 30, 249-259.
- Oda T, Morikawa N, Saito Y, Masuho Y and Matsumoto S. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J Biol Chem* **2000**, 275, 36781-36786.
- Oldham WM and Hamm HE. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* **2008**, 9, 60-71.
- Overington JP, Al-Lazikani B and Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* **2006**, 5, 993-996.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M and Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, 289, 739-745.
- Park JH, Scheerer P, Hofmann KP, Choe HW and Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **2008**, 454, 183-187.
- Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM and Opella SJ. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **2012**, 491, 779-783.
- Parsons ME and Ganellin CR. Histamine and its receptors. *Br J Pharmacol* **2006**, 147, S127-135.
- Peleg G, Ghanouni P, Kobilka BK and Zare RN. Single-molecule spectroscopy of the beta(2) adrenergic receptor: observation of conformational substates in a membrane protein. *Proc Natl Acad Sci U S A* **2001**, 98, 8469-8474.
- Perez M, Pauwels PJ, Fourier C, Chopin P, Valentin JP, John GW, Marien M and Halazy S. Dimerization of sumatriptan as an efficient way to design a potent, centrally and orally active 5-HT_{1B} agonist. *Bioorg Med Chem Lett* **1998**, 8, 675-680.
- Pierce KL, Premont RT and Lefkowitz RJ. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* **2002**, 3, 639-650.
- Portoghese PS. 2000 Alfred Burger Award Address in Medicinal Chemistry. From Models to Molecules: Opioid Receptor Dimers, Bivalent Ligands, and Selective Opioid Receptor Probes. *J Med Chem* **2001**, 44, 3758-3758.
- Portoghese PS, Larson DL, Sayre LM, Yim CB, Ronsisvalle G, Tam SW and Takemori AE. Opioid agonist and antagonist bivalent ligands. The relationship between spacer length and selectivity at multiple opioid receptors. *J Med Chem* **1986**, 29, 1855-1861.
- Pounder RE, Williams JG, Milton-Thompson GJ and Misiewicz JJ. Relief of duodenal ulcer symptoms by oral metiamide. *Br Med J* **1975**, 2, 307-309.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Mutations of Cys-17 and Ala-271 in the human histamine H-2 receptor determine the species selectivity of guanidine-type agonists and increase constitutive activity. *J Pharmacol Exp Ther* **2007**, 321, 975-982.
- Raible DG, Lenahan T, Fayvilevich Y, Kosinski R and Schulman ES. Pharmacologic characterization of a novel histamine receptor on human eosinophils. *Am J Respir Crit Care Med* **1994**, 149, 1506-1511.
- Rajagopal S, Rajagopal K and Lefkowitz RJ. Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nat Rev Drug Discov* **2010**, 9, 373-386.
- Ramsay D, Kellett E, McVey M, Rees S and Milligan G. Homo- and hetero-oligomeric interactions between G-protein-coupled receptors in living cells monitored by two variants of bioluminescence resonance energy transfer (BRET): hetero-oligomers

- between receptor subtypes form more efficiently than between less closely related sequences. *Biochem J* **2002**, 365, 429-440.
- Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **2007**, 450, 383-387.
- Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **2011**, 477, 549-555.
- Reiter E, Ahn S, Shukla AK and Lefkowitz RJ. Molecular mechanism of beta-arrestin-biased agonism at seven-transmembrane receptors. *Annu Rev Pharmacol Toxicol* **2012**, 52, 179-197.
- Rosenbaum DM, Rasmussen SGF and Kobilka BK. The structure and function of G-protein-coupled receptors. *Nature* **2009**, 459, 356-363.
- Rovati GE, Capra V and Neubig RR. The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. *Mol Pharmacol* **2007**, 71, 959-964.
- Ruat M, Traiffort E, Arrang JM, Leurs R and Schwartz JC. Cloning and tissue expression of a rat histamine H2-receptor gene. *Biochem Biophys Res Commun* **1991**, 179, 1470-1478.
- Sachs G, Shin JM, Vagin O, Lambrecht N, Yakubov I and Munson K. The gastric H,K ATPase as a drug target: past, present, and future. *J Clin Gastroenterol* **2007**, 41, S226-242.
- Salcedo C, Pontes C and Merlos M. Is the H4 receptor a new drug target for allergies and asthma? *Front Biosci* **2013**, 5, 178-187.
- Salon JA, Lodowski DT and Palczewski K. The Significance of G Protein-Coupled Receptor Crystallography for Drug Discovery. *Pharmacol Rev* **2011**, 63, 901-937.
- Scarselli M, Novi F, Schallmach E, Lin R, Baragli A, Colzi A, Griffon N, Corsini GU, Sokoloff P, Levenson R, Vogel Z and Maggio R. D2/D3 dopamine receptor heterodimers exhibit unique functional properties. *J Biol Chem* **2001**, 276, 30308-30314.
- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP and Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, 455, 497-U430.
- Schertler GF, Villa C and Henderson R. Projection structure of rhodopsin. *Nature* **1993**, 362, 770-772.
- Schneider E, Rolli-Derkinderen M, Arock M and Dy M. Trends in histamine research: new functions during immune responses and hematopoiesis. *Trends Immunol* **2002**, 23, 255-263.
- Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Rompler H and Sangkuhl K. Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther* **2004**, 104, 173-206.
- Schubert ML and Peura DA. Control of gastric acid secretion in health and disease. *Gastroenterology* **2008**, 134, 1842-1860.
- Schwartz JC. The histamine H3 receptor: from discovery to clinical trials with pitolisant. *Br J Pharmacol* **2011**, 163, 713-721.
- Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM and Elling CE. Molecular mechanism of 7TM receptor activation--a global toggle switch model. *Annu Rev Pharmacol Toxicol* **2006**, 46, 481-519.
- Seifert R, Schneider EH, Dove S, Brunskole I, Neumann D, Strasser A and Buschauer A. Paradoxical stimulatory effects of the "standard" histamine H4-receptor antagonist JNJ7777120: the H4 receptor joins the club of 7 transmembrane domain receptors exhibiting functional selectivity. *Mol Pharmacol* **2011**, 79, 631-638.
- Seifert R, Strasser A, Schneider EH, Neumann D, Dove S and Buschauer A. Molecular and cellular analysis of human histamine receptor subtypes. *Trends Pharmacol Sci* **2013**, 34, 33-58.

- Seifert R and Wenzel-Seifert K. Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn Schmiedeberg's Arch Pharmacol* **2002**, 366, 381-416.
- Seifert R, Wenzel-Seifert K, Burckstummer T, Pertz HH, Schunack W, Dove S, Buschauer A and Elz S. Multiple differences in agonist and antagonist pharmacology between human and guinea pig histamine H1-receptor. *J Pharmacol Exp Ther* **2003**, 305, 1104-1115.
- Seifert R, Wenzel-Seifert K, Lee TW, Gether U, Sanders-Bush E and Kobilka BK. Different effects of Gsalpha splice variants on beta2-adrenoreceptor-mediated signaling. The Beta2-adrenoreceptor coupled to the long splice variant of Gsalpha has properties of a constitutively active receptor. *J Biol Chem* **1998**, 273, 5109-5116.
- Sharman JL, Benson HE, Pawson AJ, Lukito V, Mpamhanga CP, Bombail V, Davenport AP, Peters JA, Spedding M and Harmar AJ. IUPHAR-DB: updated database content and new features. *Nucleic Acids Res* **2013**, 41, 18.
- Shi L, Liapakis G, Xu R, Guarneri F, Ballesteros JA and Javitch JA. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J Biol Chem* **2002**, 277, 40989-40996.
- Shonberg J, Scammells PJ and Capuano B. Design Strategies for Bivalent Ligands Targeting GPCRs. *ChemMedChem* **2011**, 6, 963-974.
- Shukla AK, Manglik A, Kruse AC, Xiao K, Reis RI, Tseng WC, Staus DP, Hilger D, Uysal S, Huang LY, Paduch M, Tripathi-Shukla P, et al. Structure of active beta-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* **2013**, 497, 137-141.
- Shukla AK, Xiao K and Lefkowitz RJ. Emerging paradigms of beta-arrestin-dependent seven transmembrane receptor signaling. *Trends Biochem Sci* **2011**, 36, 457-469.
- Simons FE and Simons KJ. Histamine and H1-antihistamines: celebrating a century of progress. *J Allergy Clin Immunol* **2011**, 128, 1139-1150.
- Simons FER. Advances in H1-Antihistamines. *New Engl J Med* **2004**, 351, 2203-2217.
- Smit MJ, Hoffmann M, Timmerman H and Leurs R. Molecular properties and signalling pathways of the histamine H1 receptor. *Clin Exp Allergy* **1999**, 3, 19-28.
- Smit MJ, Leurs R, Alewijnse AE, Blauw J, Van Nieuw Amerongen GP, Van De Vrede Y, Roovers E and Timmerman H. Inverse agonism of histamine H2 antagonist accounts for upregulation of spontaneously active histamine H2 receptors. *Proc Natl Acad Sci U S A* **1996**, 93, 6802-6807.
- Stevens RC, Cherezov V, Katritch V, Abagyan R, Kuhn P, Rosen H and Wuthrich K. The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat Rev Drug Discov* **2013**, 12, 25-34.
- Takahama H, Asanuma H, Sanada S, Fujita M, Sasaki H, Wakeno M, Kim J, Asakura M, Takashima S, Minamino T, Komamura K, Sugimachi M, et al. A histamine H(2) receptor blocker ameliorates development of heart failure in dogs independently of beta-adrenergic receptor blockade. *Basic Res Cardiol* **2010**, 105, 787-794.
- Tanaka T, Nomura W, Narumi T, Masuda A and Tamamura H. Bivalent ligands of CXCR4 with rigid linkers for elucidation of the dimerization state in cells. *J Am Chem Soc* **2010**, 132, 15899-15901.
- Tate CG. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem Sci* **2012**, 37, 343-352.
- Thurmond RL, Gelfand EW and Dunford PJ. The role of histamine H1 and H4 receptors in allergic inflammation: the search for new antihistamines. *Nat Rev Drug Discov* **2008**, 7, 41-53.
- Topiol S and Sabio M. X-ray structure breakthroughs in the GPCR transmembrane region. *Biochem Pharmacol* **2009**, 78, 11-20.
- Traiffort E, Pollard H, Moreau J, Ruat M, Schwartz JC, Martinez-Mir MI and Palacios JM. Pharmacological characterization and autoradiographic localization of histamine H2 receptors in human brain identified with [125I]iodoaminopotentidine. *J Neurochem* **1992**, 59, 290-299.

- Traiffort E, Vizuite ML, Tardivel-Lacombe J, Souil E, Schwartz JC and Ruat M. The guinea pig histamine H₂ receptor: gene cloning, tissue expression and chromosomal localization of its human counterpart. *Biochem Biophys Res Commun* **1995**, 211, 570-577.
- Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A and Filipek S. Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr Med Chem* **2012**, 19, 1090-1109.
- Unger VM and Schertler GF. Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. *Biophys J* **1995**, 68, 1776-1786.
- Valant C, Sexton PM and Christopoulos A. Orthosteric/allosteric bitopic ligands: going hybrid at GPCRs. *Mol Interv* **2009**, 9, 125-135.
- van Rijn RM, Chazot PL, Shenton FC, Sansuk K, Bakker RA and Leurs R. Oligomerization of recombinant and endogenously expressed human histamine H(4) receptors. *Mol Pharmacol* **2006**, 70, 604-615.
- Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE and Gaitanaris GA. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci U S A* **2003**, 100, 4903-4908.
- Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF and Babu MM. Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, 494, 185-194.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, et al. The sequence of the human genome. *Science* **2001**, 291, 1304-1351.
- Vogel R, Mahalingam M, Ludeke S, Huber T, Siebert F and Sakmar TP. Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. *J Mol Biol* **2008**, 380, 648-655.
- Wacker D, Wang C, Katritch V, Han GW, Huang XP, Vardy E, McCorvy JD, Jiang Y, Chu M, Siu FY, Liu W, Xu HE, et al. Structural features for functional selectivity at serotonin receptors. *Science* **2013**, 340, 615-619.
- Walter M, Kottke T and Stark H. The histamine H(4) receptor: targeting inflammatory disorders. *Eur J Pharmacol* **2011**, 668, 1-5.
- Wang C, Wu H, Katritch V, Han GW, Huang XP, Liu W, Siu FY, Roth BL, Cherezov V and Stevens RC. Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **2013**, 497, 338-343.
- Wang L, Gantz I and DelValle J. Histamine H₂ receptor activates adenylate cyclase and PLC via separate GTP-dependent pathways. *Am J Physiol* **1996**, 271, G613-620.
- Wellner-Kienitz MC, Bender K, Meyer T and Pott L. Coupling to G_s and G_(q/11) of histamine H₂ receptors heterologously expressed in adult rat atrial myocytes. *Biochim Biophys Acta* **2003**, 23, 1-2.
- Wheatley M, Wootten D, Conner MT, Simms J, Kendrick R, Logan RT, Poyner DR and Barwell J. Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **2012**, 165, 1688-1703.
- Wijtmans M, Leurs R and de Esch I. Histamine H₃ receptor ligands break ground in a remarkable plethora of therapeutic areas. *Expert Opin Investig Drugs* **2007**, 16, 967-985.
- Wilcox CM and Hirschowitz BI. Treatment strategies for Zollinger-Ellison syndrome. *Expert Opin Pharmacother* **2009**, 10, 1145-1157.
- Xiao RP. Beta-adrenergic signaling in the heart: dual coupling of the beta₂-adrenergic receptor to G_(s) and G_(i) proteins. *Sci STKE* **2001**, 16.
- Xu J, He J, Castleberry AM, Balasubramanian S, Lau AG and Hall RA. Heterodimerization of alpha 2A- and beta 1-adrenergic receptors. *J Biol Chem* **2003**, 278, 10770-10777.
- Yang LP and Perry CM. Histamine dihydrochloride: in the management of acute myeloid leukaemia. *Drugs* **2011**, 71, 109-122.

- Yeomans ND. Management of peptic ulcer disease not related to *Helicobacter*. *J Gastroenterol Hepatol* **2002**, 17, 488-494.
- Young RC, Mitchell RC, Brown TH, Ganellin CR, Griffiths R, Jones M, Rana KK, Saunders D, Smith IR, Sore NE and et al. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H₂ receptor histamine antagonists. *J Med Chem* **1988**, 31, 656-671.
- Zampeli E and Tiligada E. The role of histamine H₄ receptor in immune and inflammatory disorders. *Br J Pharmacol* **2009**, 157, 24-33.
- Zeng F and Wess J. Molecular aspects of muscarinic receptor dimerization. *Neuropsychopharmacology* **2000**, 23, S19-31.
- Zhu Y, Michalovich D, Wu H, Tan KB, Dytko GM, Mannan IJ, Boyce R, Alston J, Tierney LA, Li X, Herrity NC, Vawter L, et al. Cloning, expression, and pharmacological characterization of a novel human histamine receptor. *Mol Pharmacol* **2001**, 59, 434-441.

Chapter 2

Scope and Objectives

G-protein coupled receptors, membrane spanning proteins transferring extracellular signals across the cell membrane to the intracellular side, are fundamental for most physiological processes and are the targets of about 30% of all drugs on the market (Katritch *et al.*, 2012; Stevens *et al.*, 2013). Ligand binding to GPCRs stabilizes different, active or inactive states of the receptor. Active states couple to effectors of cytosolic signaling pathways, like heterotrimeric G-proteins and arrestins. However, the molecular basis for GPCR activation remained unclear for a long time. In 2000 the first crystal structure of a GPCR was obtained, bovine rhodopsin (Palczewski *et al.*, 2000), and since 2007 several other GPCRs in their inactive state were crystallized. The recent resolution of active state GPCRs, namely activated opsin, the β_2 AR and the adenosine A_{2A} receptor (Lebon *et al.*, 2011; Park *et al.*, 2008; Rasmussen *et al.*, 2011b) has been the basis for more detailed investigations on GPCR activation mechanisms. Relatedness of the β ARs to the histamine H_2 receptor (Vassilatis *et al.*, 2003) make the latter an interesting GPCR for structural analysis. Results of such experimental and theoretical studies may advance the development of potent and selective H_2 R ligands in our group, including bivalent hetarylpropylguanidines which showed an increased potency compared to their monovalent congeners (Birnkammer *et al.*, 2012; Kraus *et al.*, 2009).

In the first part of this thesis, the aim was to investigate the structural differences between inactive and active hH_2 R states. Homology models had to be constructed based on the inactive $t\beta_1$ AR state (Warne *et al.*, 2008) and on the active $h\beta_2$ AR state (Rasmussen *et al.*, 2011a). Additionally, in a multiple template approach another active hH_2 R state variant was to be generated using the structures of the $t\beta_1$ AR and opsin (Scheerer *et al.*, 2008). The intention was to compare different models of active hH_2 R states with respect to reliability and to select the 'best' model (chapter 3).

Subsequently, it was intended to embed hH_2 R models of both states into a natural environment consisting of a 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) bilayer and

water molecules as solvent, and to subject the models to 80 ns molecular dynamics (MD) simulations with the MD package GROMACS (Van Der Spoel *et al.*, 2005). This should further validate the hH₂R models, prove their relatively stable existence in a natural, dynamic environment and allow the investigation of the impact of conformational flexibility on the process of receptor activation (chapter 4).

The analysis of the results of the MD simulations should be performed with two novel routines extending the capabilities of GROMACS (chapter 5). They had to systematically calculate direct and water mediated H-bonds as well as van der Waals contacts in MD simulation systems. Major goals were a user-friendly setup of the analysis, a rapid calculation and a suitable output format. Additionally, to assure the quality of the simulated proteins, it was planned to write a program that is able to analyze the stereochemistry at C_α atoms as well as the planarity of aromatic side chains, delocalized π -electron systems and peptide bonds. The program should also measure backbone Φ/Ψ dihedral angles and side chain torsions and compare them to experimental reference values (Lovell *et al.*, 2000).

In the second part of this thesis, site-directed mutagenesis studies were to study the influence of amino acids in the hH₂R and the gpH₂R on receptor activation. In the first project, Tyr182^{5.38} suggested to contribute to ligand binding (Nederkoorn *et al.*, 1996) should be mutated in the hH₂R to Phe (chapter 6). After expression in Sf9 cells using baculoviruses, membranes containing the recombinant protein should be obtained, and several ligands were planned to be investigated at the wild-type and mutant hH₂Rs in the GTPase and GTP γ S assay.

In a further project a putative second binding site for bivalent hetarylpropylguanidine-type agonists should be identified in order to explore their higher potency at the H₂R compared to monovalent compounds (chapter 7). Receptor mutants of the gpH₂R should be generated and, after expression in Sf9 cells, ligands should be tested in the GPT γ S assay at the wild-type and mutated gpH₂Rs.

References

- Birnkammer T, Spickenreither A, Brunskole I, Lopuch M, Kagermeier N, Bernhardt G, Dove S, Seifert R, Elz S and Buschauer A. The Bivalent Ligand Approach Leads to Highly Potent and Selective Acylguanidine-Type Histamine H-2 Receptor Agonists. *J Med Chem* **2012**, 55, 1147-1160.
- Katritch V, Cherezov V and Stevens RC. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2012**, 8, 8.
- Kraus A, Ghorai P, Birnkammer T, Schnell D, Elz S, Seifert R, Dove S, Bernhardt G and Buschauer A. N-G-Acylated Amino-thiazolylpropylguanidines as Potent and Selective Histamine H-2 Receptor Agonists. *ChemMedChem* **2009**, 4, 232-240.
- Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AG and Tate CG. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, 474, 521-525.
- Lovell SC, Word JM, Richardson JS and Richardson DC. The penultimate rotamer library. *Proteins* **2000**, 40, 389-408.
- Nederkoorn PH, van Lenthe JH, van der Goot H, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H2 receptor. I. Theoretical investigations of histamine binding to an oligopeptide mimicking a part of the fifth transmembrane alpha-helix. *J Comput Aided Mol Des* **1996**, 10, 461-478.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M and Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, 289, 739-745.
- Park JH, Scheerer P, Hofmann KP, Choe HW and Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **2008**, 454, 183-187.
- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011a**, 469, 175-180.
- Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **2011b**, 477, 549-555.
- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP and Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, 455, 497-U430.
- Stevens RC, Cherezov V, Katritch V, Abagyan R, Kuhn P, Rosen H and Wuthrich K. The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat Rev Drug Discov* **2013**, 12, 25-34.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, 26, 1701-1718.
- Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE and Gaitanaris GA. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci U S A* **2003**, 100, 4903-4908.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG and Schertler GFX. Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* **2008**, 454, 486-U482.

Chapter 3

Homology Models of Inactive and Active Human Histamine H₂ Receptor States

3.1 Introduction

Since the first crystallization of a G-protein coupled receptor in the year 2000, bovine rhodopsin (Palczewski *et al.*, 2000), several Class A GPCR structures have been resolved. Until 2007 all obtained structures adopted an inactive receptor conformation (Katritch *et al.*, 2012). Advances in protein engineering and crystallography allowed the resolution of the first crystal structure of an active state GPCR in 2008, activated opsin (Park *et al.*, 2008; Scheerer *et al.*, 2008). This enabled invaluable insights into GPCR structure and function. For the histamine H₂ receptor no structural data from crystallography or NMR spectroscopy exist so far. In such cases the technique of homology modeling may be used to better understand structure-function relationships of these receptors (Costanzi, 2012). Some inactive state models of the H₂R were already described, based on crystal structures of rhodopsin (Kelley *et al.*, 2001; Kim *et al.*, 2006; Preuss, 2007), the β_2 -adrenoceptor (Ghorai *et al.*, 2008) and the adenosine A_{2A} receptor (Zhang *et al.*, 2012). However, to date no receptor conformations of the H₂R in its putative active state were constructed and compared to the inactive state.

In order to unveil the structural determinants which keep the hH₂R in its inactive and active conformation, respectively, and to analyze possible key structural elements for the transition from the inactive to the active state, homology models of both receptor states were constructed and examined. Because of differences in the template structures, two comparative models of the active receptor conformation were created. The coupling of the activated hH₂R with a protein fragment of the C-terminal part of the G_{sα}-protein further enabled the exploration of a part of the receptor–G-protein interface at an atomic level.

3.2 Materials and methods

3.2.1 Sequence alignment

All sequences were received from the UniProt Knowledgebase (UniProtKB; Consortium, 2012). The multiple sequence alignment (MSA) of the hH₂R and the GPCRs resolved so far was performed with ClustalW2 (Goujon *et al.*, 2010; Larkin *et al.*, 2007). To achieve a correct alignment especially for the sequences following the intracellular loop 3 (ICL3; very variable in length), the MSA was performed with about 330 other GPCR sequences. The alignment was modified to attain a gap-free alignment in the region of the transmembrane domains (TMs). The lengths of the α -helices were calculated after assigning the secondary structure to the proteins with the DSSP (Define Secondary Structure of Proteins) algorithm (Kabsch and Sander, 1983) included in Sybyl-X 1.3 (Tripos, St. Louis, MO, USA), using the inactive state crystal structures listed in Table 3.1. All 15 GPCR structures were superimposed by aligning their common alpha helical regions (Figure 3.1). This structural superimposition was used to refine and control the initial multiple sequence alignment. The conservation score of amino acids at identical positions in the hH₂R and the respective template sequence was calculated with ClustalX 2.1 (Larkin *et al.*, 2007; Thompson *et al.*, 1997), using the Gonnet PAM250 matrix (Gonnet *et al.*, 1992) and a score plot scale of 5. The conservation score, ranging from 0 to 100, is an indication for the quality of the alignment, in which a high score indicates a well-conserved amino acid pair.

3.2.2 Generation of 3D structures

3.2.2.1 Generation of the inactive state hH₂R model

For the generation of the homology model of the hH₂R in the inactive receptor conformation (hH₂R_i) a crystal structure of the turkey β_1 -adrenoceptor (t β_1 AR) was selected as template (PDB ID 2VT4, chain B; Warne *et al.*, 2008b). The coordinates for the hH₂R from TM1 to the N-terminus of ECL2 (Ala16^{1.30} to Asn162), from the C-terminal part of ECL2 to TM5 (Ser165 to Ala213^{5.69}), from TM6 (Arg228^{6.29} to Leu259^{6.60}), TM7 (Glu267^{7.32} to Tyr288^{7.53}) and helix 8 with a few C-terminal amino acids (Leu291 to Cys305) were taken from the template. In the t β_1 AR this corresponds to Gln39^{1.30} to Pro187, Leu190 to Ile238^{5.69}, Arg284^{6.29} to Phe315^{6.60}, Asp322^{7.32} to Tyr343^{7.53}, and Ser346 to Phe359, respectively. The α -helix present in ECL2 of the t β_1 AR was preserved in the model. The e2 loop of the t β_1 AR is two amino acids longer than that of the hH₂R. Thus, Asp186 to Ala189 of the t β_1 AR were deleted and Glu163 and Thr164 of the hH₂R were inserted using the loop search application included in Sybyl-X 1.3.

Table 3.1: GPCR crystal structures of inactive states considered on homology modeling of the hH₂R

GPCR	UniprotKB entry	PDB ID	Resolution [Å]	Publication date
Turkey β_1-adrenergic receptor (t β_1 AR) ¹	P07700	2VT4	2.7	2008
Human β_2-adrenergic receptor (h β_2 AR) ²	P07550	2RH1	2.4	2007
Human adenosine A_{2A} receptor (hA _{2A} AR) ³	P29274	3EML	2.6	2008
Bovine rhodopsin (bovRhod) ⁴	P02699	1U19	2.2	2004
Squid rhodopsin (squidRhod) ⁵	P31356	2Z73	2.5	2007
Human histamine H₁ receptor (hH ₁ R) ⁶	P35367	3RZE	3.1	2010
Human D₃ dopamine receptor (hD ₃ R) ⁷	P35462	3PBL	2.89	2010
Human muscarinic acetylcholine receptor M₂ (hmAChRM ₂) ⁸	P08172	3UON	3.0	2011
Rat muscarinic acetylcholine receptor M₃ (rmAChRM ₃) ⁹	P08483	4DAJ	3.4	2012
Human C-X-C chemokine receptor type 4 (hCXCR4) ¹⁰	P61073	3ODU	2.5	2010
Human sphingosine 1-phosphate receptor 1 (hS1PR1) ¹¹	P21453	3V2W	3.35	2011
Mouse delta-type opioid receptor (mOPRD) ¹²	P32300	4EL4	3.4	2012
Human kappa-type opioid receptor (hOPRK) ¹³	P41145	4DJH	2.9	2012
Mouse mu-type opioid receptor (mOPRM) ¹⁴	P42866	4DKL	2.8	2012
Human nociceptin receptor (hNOPR) ¹⁵	P41146	4EA3	3.01	2012

¹ Warne *et al.*, 2008b, chain B; ² Cherezov *et al.*, 2007; ³ Jaakola *et al.*, 2008; ⁴ Okada *et al.*, 2004, chain A; ⁵ Murakami and Kouyama, 2008, chain A; ⁶ Shimamura *et al.*, 2011; ⁷ Chien *et al.*, 2010, chain B; ⁸ Haga *et al.*, 2012; ⁹ Kruse *et al.*, 2012, chain D; ¹⁰ Wu *et al.*, 2010, chain A; ¹¹ Hanson *et al.*, 2012; ¹² Granier *et al.*, 2012; ¹³ Wu *et al.*, 2012, chain B; ¹⁴ Manglik *et al.*, 2012; ¹⁵ Thompson *et al.*, 2012, chain A.

This approach scans for loop fragments contained in a protein database (a subset of the Protein Data Bank, PDB; Bernstein *et al.*, 1977). Loops of appropriate residue lengths and with terminal residues well fitting the anchor regions of the modeled protein are then chosen (Rossi *et al.*, 2007). ICL3 (Lys214 to Ile227) which is not resolved in the template structure and ECL3 (Arg260 to Asn266) were also constructed with loop searches. The junction of TM7 to H8 is one amino acid longer in the hH₂R sequence than in the tβ₁AR. Therefore, Cys344 and Arg345 were deleted in the template and a loop search was performed with Ala-Ala-Leu (Ala289 to Leu291) to close the gap. N- and C-terminal amino acids were not present in the template structure. Predictions about these peptide sequences in the hH₂R would be highly speculative and they were therefore not modeled.

3.2.2.2 Generation of the active state hH₂R models

The first homology model of the hH₂R in the active receptor conformation (hH₂R_{act}) was constructed by a multiple template approach using the crystal structure of the tβ₁AR (PDB ID 2VT4, chain B; Warne *et al.*, 2008b) and the crystal structure of opsin in its G-protein-interacting conformation (PDB ID 3DQB; Scheerer *et al.*, 2008). To locate appropriate alignments of hH₂R domains with the templates, the structures of the tβ₁AR and opsin were superimposed based on their common TM domains and helix 8. TM1 of the hH₂R (Ala16^{1.30} to Leu45^{1.59}) was constructed by mutating the corresponding TM1 of the tβ₁AR (Gln39^{1.30} to Ser68^{1.59}). The result was aligned to 20 N-terminal amino acids of TM1 of the tβ₁AR (Gln39^{1.30} to Gly58^{1.49}) and 2 C-terminal amino acids of TM1 of opsin (Val63^{1.58} and Gln64^{1.59}). Accordingly, TM2 of the hH₂R (Leu52^{2.38} to Ser81^{2.67}) was based on mutation of tβ₁AR-TM2 (Leu75^{2.38} to Arg104^{2.67}) and superimposed with 4 N-terminal amino acids of TM2 of opsin (Leu72^{2.39} to Ile75^{2.42}) and 4 C-terminal amino acids of TM2 of the tβ₁AR (Leu101^{2.64} to Arg104^{2.67}). The coordinates from ECL1 to TM4 (Cys82^{2.68} to His155^{4.62}) were directly adopted from the tβ₁AR (Gly105^{2.68} to Met178^{4.62}). TM5 of the hH₂R (Glu180^{5.36} to Ala213^{5.69}) was generated by mutation of tβ₁AR-TM5 (Arg205^{5.36} to Ile238^{5.69}) and fitted to 13 N-terminal amino acids of TM5 of the tβ₁AR (Arg205^{5.36} to Tyr217^{5.48}) and 8 C-terminal amino acids of TM5 of opsin (Leu226^{5.61} to Ala233^{5.68}). TM6 of the hH₂R (Glu229^{6.30} to Gly258^{6.59}) was also based on the tβ₁AR (Glu285^{6.30} to Val314^{6.59}). It was superimposed to TM6 of opsin considering the common α-helical domains (tβ₁AR, Glu285^{6.30} to Asn313^{6.58}; opsin, Glu247^{6.30} to Ile275^{6.58}). TM7 and helix 8 of the hH₂R (Glu270^{7.35} to Ala308) were constructed by mutating opsin (Met288^{7.35} to Asn326). ICL1 (Asn46 to Asn51) was adopted from opsin (His65 to Thr70), and ECL2 (Leu156 to Asn179) from the inactive hH₂R state model described above. Atoms at the junction of the inserted ICL1 and the adjacent TMs 1 and 2, as well as at ECL2 and its surrounding residues were carefully modified to obtain an

appropriate geometry of the connecting bonds (e.g. planarity of peptide omega angles). ICL3 (Lys214 to Arg228) and ECL3 (Leu259 to Leu269) were modeled with loop searches. N- and C-terminal amino acids were not included in the homology model due to the low sequence identity to the template opsin (N-terminus) and because they were not resolved in the template structures (C-terminus), respectively.

For the construction of the second homology model of the hH₂R in the active receptor conformation (hH₂R_{a2}) the crystal structure of the nanobody-stabilized active state of the β_2 AR (PDB ID 3P0G; Rasmussen *et al.*, 2011a) was used as template. The coordinates of a few amino acids preceding TM1 till the N-terminus of ECL2 (Ser8 to Asn162), of most of ECL2 till TM5 (Ser165 to Arg210^{5.66}), of TM6 (Ile227^{6.28} to Arg260^{6.61}), of TM7 (Glu267^{7.32} to Ala289^{7.54}) and of helix 8 with some C-terminal amino acids (Asn292 to Leu307) were taken from the template. In the template β_2 AR this corresponds to Asp23 to Thr177, Ile182 to Lys227^{5.66}, Leu266^{6.28} to Gln299^{6.61}, Lys305^{7.32} to Cys327^{7.54}, and Ser329 to Arg344, respectively. Due to a two amino acids longer ECL2 in the h β_2 AR, Glu163 and Thr164 of the hH₂R replacing His178 to Ala181 in the h β_2 AR were inserted with a loop search. ICL3 which is not resolved in the template structure was also constructed with a loop search (Asp211 to Thr226). The junction between TM7 and H8 of the hH₂R is one amino acid longer than in the h β_2 AR. Therefore Arg328 was deleted in the template, and a loop search was performed with Ala290 and Leu291 to close the gap. Most of the N- and C-terminal amino acids are not present in the template structure and were therefore not modeled for the hH₂R.

3.2.2.3 Structural refinement of the homology models

Side chain conformations of identical amino acids in corresponding positions of the hH₂R and the template were not changed. All other amino acids were mutated into the appropriate hH₂R residue. Conformations were adjusted with respect to corresponding side chain torsion angles in the template in order to reproduce contacts between adjacent amino acids. Additionally, the agreement with frequently occurring rotamer states provided by Lovell *et al.* (2000) was ensured, and the structures were checked for bumps between neighboring amino acids. From 15 conserved water molecules identified by Angel *et al.* (2009), 10 were inserted in the homology models of the hH₂R. If necessary, the positions were manually adjusted to enable similar hydrogen bonds as suggested from the analysis of crystal structures (Angel *et al.*, 2009). Amber-FF99 charges were assigned to the protein and to the water molecules. Finally, all three homology models were energy-minimized (100 iterations) with the steepest descent method, using the Amber-FF99 force field and a dielectric constant of 4.

3.2.2.4 Insertion of the C-terminal part of the G_{sa} -protein and docking of histamine

Both active state models were completed by insertion of the carboxy terminus of the α -subunit of the G_s -protein ($G\alpha CT$). It was constructed as described in the supplementary information of Scheerer *et al.* (2008). The 11 amino acid synthetic peptide (ILENLKDCGLF) derived from the C-terminus of the transducin G_{at} subunit which was co-crystallized in the structure of opsin in its G-protein-interacting conformation (PDB ID 3DQB; Scheerer *et al.*, 2008), was mutated to the sequence of the short isoform of the G_α subunit of the guanine nucleotide-binding protein G_s , G_{saS} (384-QRMHLRQYELL-394). The backbone of the 4 N-terminal amino acids (QRMH) was rebuilt to obtain the geometry of an ideal α -helix ($\Phi = -57^\circ$, $\Psi = -47^\circ$; CBN, 1970). Thr369 to Ile383 of G_{saS} were likewise constructed, and finally both peptide chains were connected, resulting in a 26 amino acid containing C-terminal fragment of G_{saS} ($G\alpha CT$). The active state model based on the crystal structures of the $t\beta_1AR$ and opsin (hH_2R_{a1}) was superimposed with opsin, regarding all common TM domains. The active state model based on the crystal structure of the active $h\beta_2AR$ state (hH_2R_{a2}) was superimposed with opsin, considering only TM5 and TM6. The newly constructed $G\alpha CT$ peptide was aligned to G_{at} of the opsin- G_{at} -complex and inserted in each of the two active state models. If necessary, side chain torsions were modified at the interface between the receptor and $G\alpha CT$ to avoid too close van der Waals contacts.

Histamine was manually docked into the active state models in an energy-minimum conformation, interacting with Asp98^{3,32} and Asp186^{5,42} (Gantz *et al.*, 1992). Gasteiger-Hückel charges were assigned to the ligand, and the active state homology models were energy-minimized again as described above.

3.2.3 Structure validation

For validation of the hH_2R models, the planarity of peptide bonds, aromatic side chains and polar side chain groups containing a planar π -electron system were checked. Furthermore, the distribution of the peptide backbone dihedral angles phi (Φ) and psi (Ψ) (Ramachandran plot) and the side chain rotamers were measured and compared to experimental values (Lovell *et al.*, 2000). All parameters were analyzed with the program *gro_validation* described in chapter 5. Main chain bond angles and lengths, as well as bad contacts within the protein were controlled using PROCHECK (Laskowski *et al.*, 1993).

3.3 Results and discussion

3.3.1 Template selection

The first step in homology modeling is the selection of an appropriate template. In the last years several GPCRs were successfully crystallized (Katritch *et al.*, 2012). GPCRs available to date for homology modeling of inactive state receptors are given in Table 3.1 and shown in Figure 3.1. Sequence identity, conservation of amino acids, and structural features like the distribution of prolines in transmembrane domains and cysteins forming disulfide bridges are crucial for the selection of an appropriate template (Worth *et al.*, 2009). As some structural features of the target GPCR are probably distributed over different crystal structures, the application of multiple templates may often improve the quality of the final model (Mobarec *et al.*, 2009; Yarnitzky *et al.*, 2010).

The t β ₁AR (33%), the h β ₂AR (32%) and the hD₃R (29%) display the highest sequence identity with the hH₂R. However, since a GPCR homology model is mainly based on TM domains of the template protein, TM sequence identity is of special importance (Table 3.2). For TM1 to 7, again the t β ₁AR (44%), the hD₃R (42%) and the h β ₂AR (38%) share the highest sequence identity with the hH₂R. Mostly this is also the case when analyzing single TMs. Possible alternative templates are for TM1 the histamine H₁ receptor, for TM2 the adenosine A_{2A} receptor, for TM3 and TM7 the κ -type opioid receptor, for TM4 the sphingosine 1-phosphate receptor 1, and for TM5 the muscarinic acetylcholine receptor M₃.

In contrast to sequence identity which can only be true (1) for identical or false (0) for different amino acids at the same position, the conservation score distinguishes between the types of amino acids compared. Higher values are assigned for conservative exchanges as valine \rightarrow isoleucine (65) and lower for non-conservative exchanges like valine \rightarrow threonine (28). Thus, conservation scores enable a refined comparison of protein sequences. In case of the hH₂R, sequence identity and similarity analyses lead to comparable results. The t β ₁AR, the h β ₂AR and the hD₃R show the highest conservation scores for the sequence as a whole (53, 54 and 49, respectively), as well as the TM domains 1 to 7 (64, 61 and 64, respectively) (Table 3.3). Regarding conservation, alternative possible templates to model the hH₂R are for TM1, TM2, TM5 and TM7 the histamine H₁ receptor, for TM2 the adenosine A_{2A} receptor, for TM4 the κ -type opioid receptor, and for TM5 the muscarinic acetylcholine receptor M₃.

Concerning sequence identity and conservation, a multiple template approach using TM coordinates from different templates is possible. However, this approach is only appropriate if the target receptor shares low sequence identity with the templates (Yarnitzky *et al.*, 2010).

Since the $t\beta_1AR$, the $h\beta_2AR$ and the hD_3R show adequate sequence identity and conservation with the hH_2R , the use of only one template for the modeling process is suitable.

Proline kinks in the α -helices of the 7 transmembrane receptors appear in the $t\beta_1AR$ (Warne *et al.*, 2008b), the $h\beta_2AR$ (Cherezov *et al.*, 2007) and the hD_3R (Chien *et al.*, 2010) next to conserved prolines in all TMs except 1 and 3. Apart from TM4, these prolines are also present in the hH_2R (Figure 3.2), indicating that helical kinks exist in its 3D structure, too. A proline in position 4.59 or 4.60 is present in 14 of the 15 crystallized GPCRs. The kinks in TM4 of the $t\beta_1AR$, the $h\beta_2AR$ and the hD_3R are rather weak and located at the extracellular end.

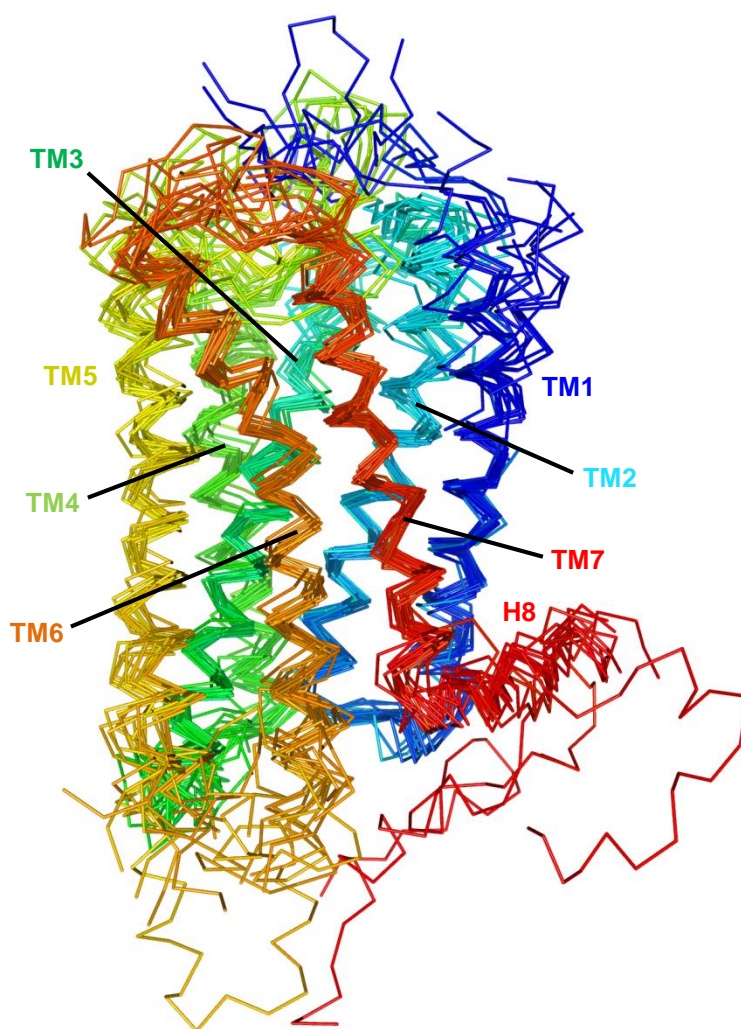


Figure 3.1: Alignment of inactive state GPCR structures considered on homology modeling of the hH_2R

Side view of the C_α trace of the 15 GPCR crystal structures given in Table 3.1. All structures were superimposed with the β_2AR as reference structure using their shared α -helical regions. Stabilizing proteins (e.g. the T4 lysozyme protein) and co-crystallized ligands are omitted.

However, the κ -type opioid receptor without one of these prolines contains an even more pronounced kink in TM4. Recent studies revealed that only 33% to 38% of the kinks in TM domains of GPCRs are associated with proline (Hall *et al.*, 2009; Huang and Chen, 2012; Langelaan *et al.*, 2010). According to an evolutionary hypothesis, a mutation to proline initially induces a kink in a helix, and the resulting packing defects are subsequently repaired by further mutations, thereby locking the kink in the structure (Yohannan *et al.*, 2004). Therefore, the introduction of a helical kink in TM4 of the hH₂R despite the missing proline is reasonable.

Table 3.2: Sequence identity between the hH₂R and GPCRs with available crystal structure

The numbers are the percentage sequence identity with the hH₂R for the complete receptor sequence (all), each single TM and TM1 to 7. The values were calculated as the amount of identical amino acids at the same position divided by the number of positions analyzed. The three highest identity scores are highlighted in decreasing grey-scales (dark grey, highest score).

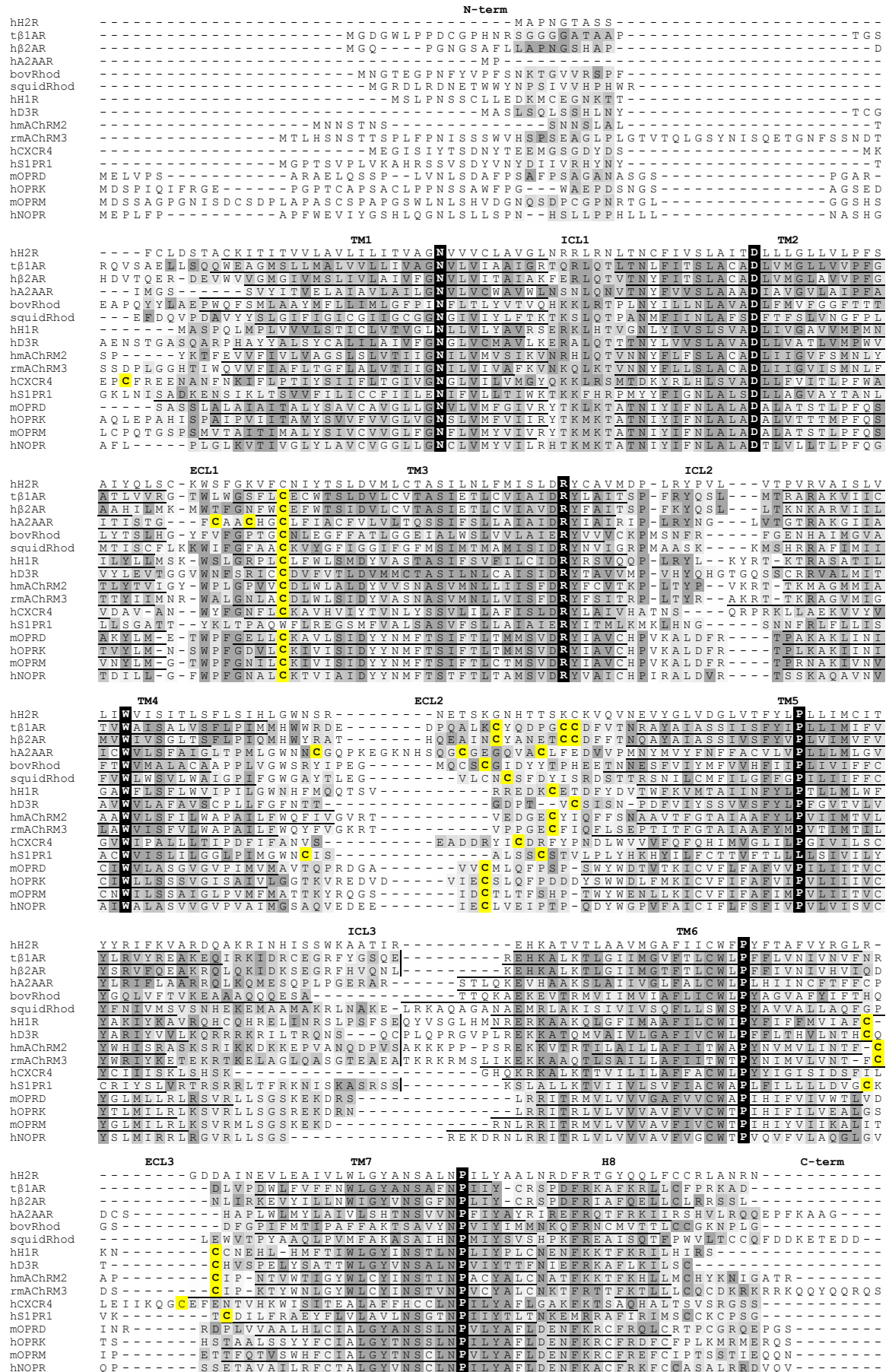
GPCR	all	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1-7
t β ₁ AR	33	34	57	47	33	35	42	57	44
h β ₂ AR	32	27	53	46	25	39	41	33	38
hA _{2A} AR	23	33	46	30	21	29	25	32	31
bovRhod	15	10	20	21	11	29	28	17	19
squidRhod	16	16	25	33	21	18	12	16	20
hH ₁ R	24	38	42	37	13	31	40	45	35
hD ₃ R	29	43	41	59	30	27	45	48	42
hmAChRM ₂	24	29	31	41	17	27	37	40	32
rmAChRM ₃	22	28	30	36	27	36	38	36	33
hCXCR4	18	27	17	34	10	18	25	25	22
hS1PR1	17	21	32	18	29	23	26	38	27
mOPRD	22	26	33	38	16	21	31	46	30
hOPRK	20	27	30	46	28	13	24	52	31
mOPRM	18	20	30	39	21	15	24	46	28
hNOPR	18	20	30	29	12	17	19	46	25

Table 3.3: Conservation scores between the hH₂R and GPCRs with available crystal structures

The numbers are the conservation scores between the hH₂R and the respective template GPCR for the complete receptor sequence (all), each single TM and TM1 to 7, calculated as described in section 3.2. The three highest scores are highlighted in decreasing grey-scales (dark grey, highest score).

GPCR	all	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1-7
tβ ₁ AR	53	63	72	66	55	63	60	69	64
hβ ₂ AR	54	55	68	65	56	65	60	60	61
hA _{2A} AR	43	55	68	54	51	47	44	53	53
bovRhod	38	36	46	49	46	53	44	47	46
squidRhod	38	42	45	55	49	46	33	45	45
hH ₁ R	45	60	69	60	42	56	56	66	58
hD ₃ R	49	61	63	78	62	53	62	68	64
hmAChRM ₂	45	57	61	63	37	53	54	58	55
rmAChRM ₃	41	53	60	62	43	56	55	56	55
hCXCR ₄	38	54	45	55	37	39	43	42	45
hS1PR1	37	48	50	47	52	38	45	55	48
mOPRD	42	50	51	58	47	45	52	65	52
hOPRK	39	52	53	62	55	39	47	65	53
mOPRM	38	47	49	58	47	42	45	64	50
hNOPR	36	43	49	52	40	44	38	59	46

A conserved disulfide bond between Cys^{3.25} of TM3 and Cys^{5.30} of ECL2 is present in all 15 GPCR structures except the sphingosine 1-phosphate receptor 1. It fixes the second half of ECL2 to TM3 and thereby limits its mobility. This restraint probably occurs also in the hH₂R between the corresponding Cys91^{3.25} and Cys174^{ECL2}. Further disulfide bonds observed in GPCR crystal structures (Figure 3.2) are absent in the hH₂R structure because of missing cysteine residues in the respective positions.

Figure 3.2: Multiple sequence alignment of the hH₂R and 15 GPCRs

The multiple sequence alignment was performed as described in section 3.2. The most conserved residue of each TM is shown in a white character on black background. Cysteines forming disulfide bridges are highlighted in yellow. The conservation score of amino acids of the respective GPCR with the hH₂R was calculated as described in section 3.2 and is shown in grey scales (dark grey, high conservation; bright grey, low conservation). The α -helical domains are indicated with black lines above the sequence. For reasons of clarity, in some sequences (t β ₁AR, h β ₂AR, hH₁R, hD₃R, hmAChRM₂, rmAChRM₃ and hS1PR1) amino acids of ICL3 are omitted (indicated by |). The sequences following helix 8 are truncated.

The t β ₁AR shows the highest sequence identity and conservation score with the hH₂R. Furthermore, its structural features, i.e. kinks and disulfide bonds, are also expected in the hH₂R. Consequently, the t β ₁AR was considered to be an appropriate template for an inactive state model of the hH₂R.

Recently, structures of some active state GPCRs were published, namely opsin (Park *et al.*, 2008; Scheerer *et al.*, 2008), rhodopsin (Choe *et al.*, 2011; Deupi *et al.*, 2012a; Standfuss *et al.*, 2011), the adenosine A_{2A} receptor (Lebon *et al.*, 2011; Xu *et al.*, 2011) and the β ₂-adrenergic receptor bound to an agonist and stabilized by a heterotrimeric G_s-protein (Rasmussen *et al.*, 2011b) or by a nanobody (Rasmussen *et al.*, 2011a).

The first model of the hH₂R in the active state (hH₂R_{a1}) was created at a point of time when only 3D structures of opsin (PDB IDs 3DQB and 3CAP) were available as active state templates. However, opsin has a low overall and TM sequence identity with the hH₂R (15% and 19%, respectively; Table 3.2). Therefore, a multiple template approach using opsin (PDB ID 3DQB) and the template for the inactive state hH₂R, the t β ₁AR, was applied, resulting in a chimeric protein (Figure 3.3). This approach is not only suitable to combine structural properties which are distributed over more than one template, but also to insert characteristics of active GPCR states (Schneider *et al.*, 2010).

Since the β ₂AR is one of the GPCRs exhibiting highest sequence identity and conservation with the hH₂R, a second active state model of the hH₂R (hH₂R_{a2}) was created using the nanobody bound h β ₂AR as template (Rasmussen *et al.*, 2011a). Its overall structure is very similar to the G_s-protein bound β ₂AR (Rasmussen *et al.*, 2011b) released afterwards (Figure 3.4). Thus, the use of either structure as template will not lead to significantly different target models.

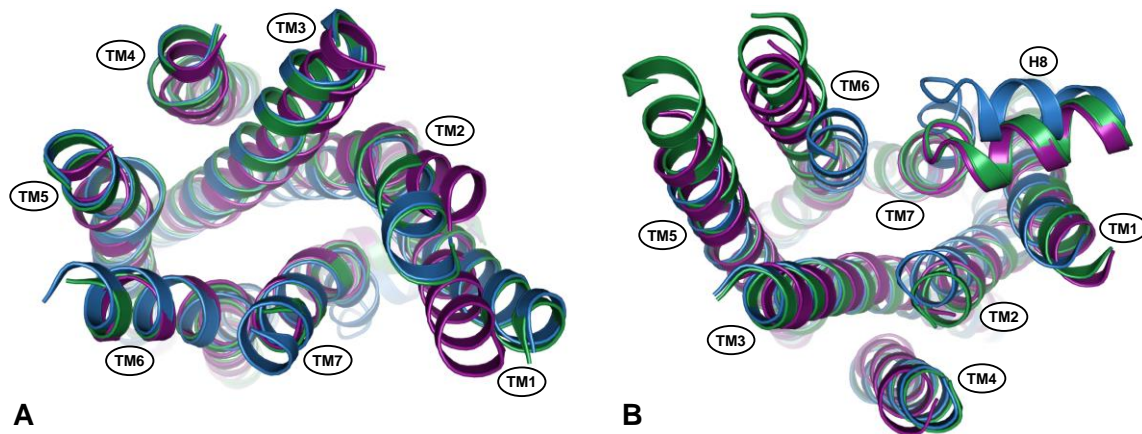


Figure 3.3: Comparison of the TM domains of the t β_1 AR, opsin and the hH₂R_{a1} model

Extracellular (A) and intracellular (B) view into the t β_1 AR (PDB ID 2VT4, chain B; ■), opsin (PDB ID 3DQB; ■) and the hH₂R_{a1} model (■) after superposition of their common TM domains. TMs are shown as ribbons. Extra- and intracellular loops are omitted for reasons of clarity.

The active state structures of the adenosine A_{2A} receptor diverge from those of rhodopsin and the β_2 AR (Lebon *et al.*, 2011; Xu *et al.*, 2011). The relocation of TM6 is much smaller (3-4 Å vs. 8 Å), partially occluding the G-protein binding site. The adenosine A_{2A} receptor structures rather represent an intermediate conformation between the inactive and active state, and thus were disqualified as appropriate templates for the construction of an active state hH₂R homology model.

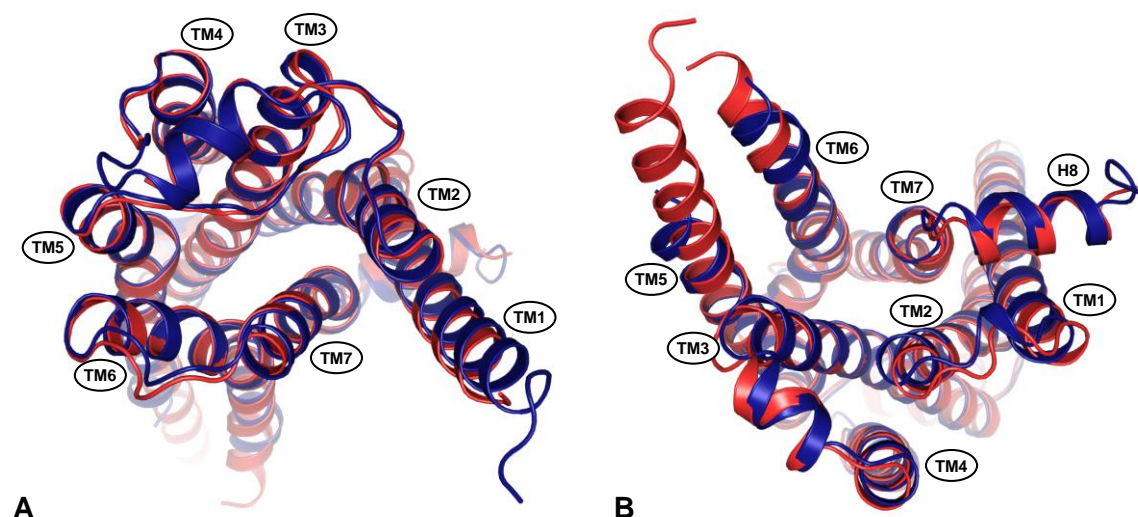


Figure 3.4: Comparison of the active state structures of the h β_2 AR

Structures of the nanobody-stabilized active state of the h β_2 AR (PDB ID 3P0G; ■) and the h β_2 AR-G_s complex (PDB ID 3SN6; ■) after superimposing their common TM domains. View from the extracellular (A) and the intracellular (B) side.

3.3.2 Stereochemical quality of the models

3.3.2.1 Omega backbone angles

The planarity of the peptide bond is one of the major constraints forcing protein configurations (Edison, 2001). However, in experimentally derived peptide and protein structures, peptide bonds have significant deviations from planarity (MacArthur and Thornton, 1996). The mean ω (\pm SD) of the hH₂R_i, hH₂R_{a1} and hH₂R_{a2} were 178.7° (\pm 5.8°), 176.7° (\pm 6.2°) and 177.0° (\pm 6.1°), respectively. This is in good agreement with the mean ω values reported in literature (MacArthur and Thornton, 1996; Morris *et al.*, 1992). Most of the coordinates of the backbone atoms defining the amide bonds were directly taken from the templates or the protein fragments derived from loop searches, and they were only marginally changed by energy minimization. In the hH₂R_i, hH₂R_{a1} and hH₂R_{a2} models, 27 out of 289 (9.3%), 40 out of 314 (12.7%), and 38 out of 324 (11.7%) ω angles deviated more than 10° from planarity (180°), respectively (Figure 3.5). Only two (hH₂R_i and hH₂R_{a2}) and three (hH₂R_{a1}) of these angles were introduced by connecting TMs with protein fragments received from loop searches, or by connecting parts of the receptor obtained from different templates.

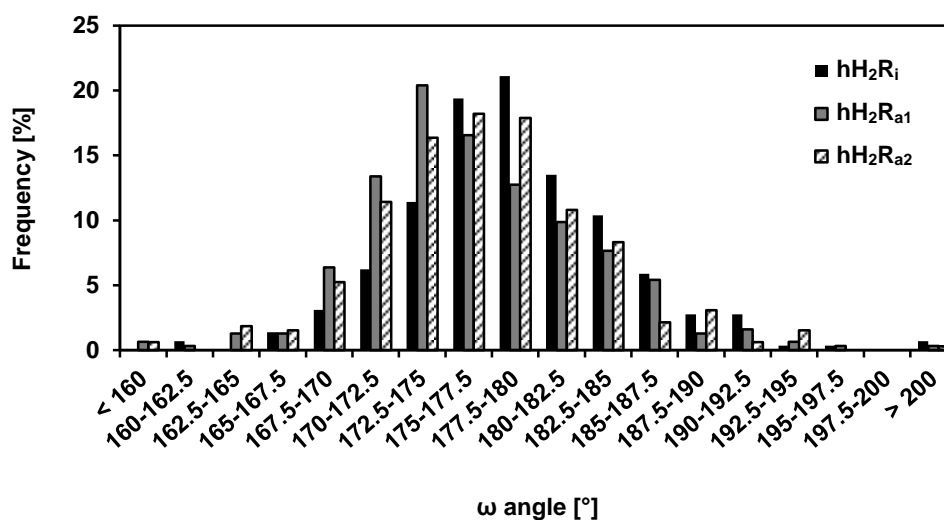


Figure 3.5: Distribution of the backbone dihedral angle ω in the hH₂R models

3.3.2.2 Ramachandran plot

The backbone dihedral angles Φ (phi) and Ψ (psi) were generally found in the favored conformational regions (Figure 3.6, Table 3.4). In total, for the hH₂R_i, hH₂R_{a1} and hH₂R_{a2} models, only two, four and one amino acid, respectively, were located in disallowed regions. 5 out of these 7 amino acids were obtained from protein crystal structures, either from the templates t β_1 AR and opsin, or from a protein fragment of the Protein Data Bank received by a loop search. The two remaining backbone angles of Asn266 in the hH₂R_i and Leu269 in the hH₂R_{a1} model are located at the junction of ECL3 (obtained by a loop search) and TM7 (taken from a template GPCR).

Table 3.4: Ramachandran analysis

Residue	Region ³	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}
General¹	favored	95.4% (249/261)	88.8% (253/285)	91.9% (271/295)
	allowed	3.8% (10/261)	9.8% (28/285)	7.8% (23/295)
	outlier	0.8% (2/261)	1.4% (4/285)	0.3% (1/295)
Glycine	favored	100% (13/13)	100% (13/13)	84.6% (11/13)
	allowed	0% (0/13)	0% (0/13)	15.4% (2/13)
	outlier	0% (0/13)	0% (0/13)	0% (0/13)
Proline	favored	100% (7/7)	85.7% (6/7)	85.7% (6/7)
	allowed	0% (0/7)	14.3% (1/7)	14.3% (1/7)
	outlier	0% (0/7)	0% (0/7)	0% (0/7)
Pre-proline²	favored	100% (7/7)	85.7% (6/7)	100% (7/7)
	allowed	0% (0/7)	14.3% (1/7)	0% (0/7)
	outlier	0% (0/7)	0% (0/7)	0% (0/7)
Summary	favored	95.8% (276/288)	89.1% (278/312)	91.6% (295/322)
	allowed	3.5% (10/288)	9.6% (30/312)	8.1% (26/322)
	outlier	0.7% (2/288)	1.3% (4/312)	0.3% (1/322)

¹ All amino acids except glycine, proline and residues preceding proline.

² Amino acids preceding proline.

³ For the definition of the regions favored, allowed and outlier see chapter 5.

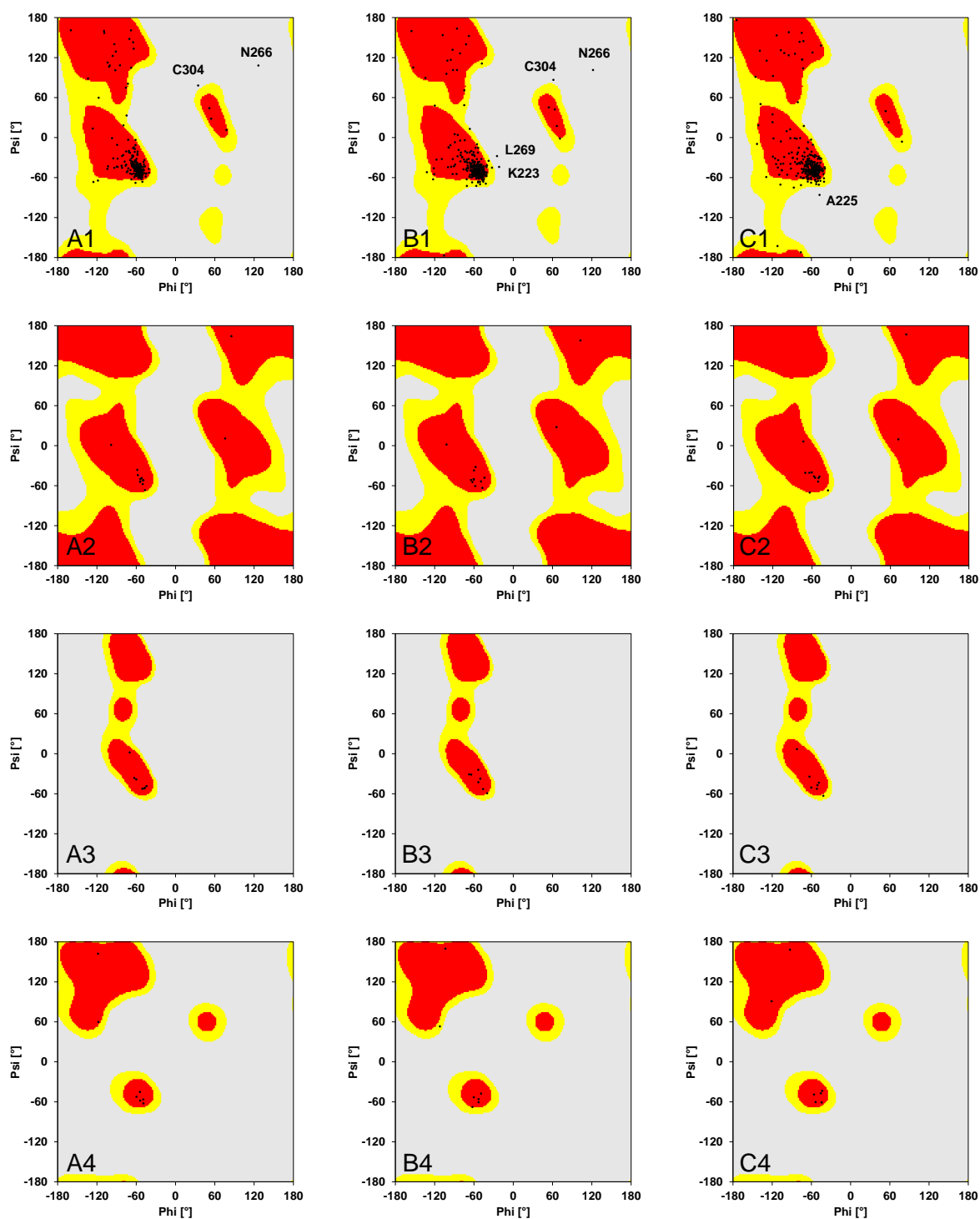


Figure 3.6: Ramachandran plots of hH₂R models

Ramachandran plots of the hH₂R_i (A), hH₂R_{a1} (B) and hH₂R_{a2} (C) model. Plots are shown for (1) all amino acids except glycine, proline and residues preceding proline, (2) glycine, (3) proline and (4) pre-proline (amino acid before proline). Red fields, most favored regions; yellow fields, additionally allowed regions; grey fields, disallowed regions. Outliers are labeled. Ramachandran analysis was performed as described in chapter 5.

3.3.2.3 Side chain dihedral angles

Since the TM sequence identity between the templates and the hH₂R is below 50%, more than every second dihedral angle had to be manually adjusted. As the dihedral angles affect the positions of side chains and the contacts to amino acids of neighboring TMs, accurately checking for reasonable torsions is feasible. The majority of amino acid side chains adopt dihedral angles frequently occurring in protein crystal structures. Only 2.0% (hH₂R_i), 2.2% (hH₂R_{a1}) and 3.2% (hH₂R_{a2}) of these torsions are found in unfavorable regions (Table 3.5). These uncommon side chains were either directly adopted from a template structure, adjusted to point in the same direction as the respective side chain in the template, or generated to avoid clashes with neighboring amino acids.

Table 3.5: Side chain dihedral angles

Region ¹	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}
Favored	94.3% (232/246)	92.6% (252/272)	94.0% (265/282)
Allowed	3.7% (9/246)	5.1% (14/272)	2.8% (8/282)
Disallowed	2.0% (5/246)	2.2% (6/272)	3.2% (9/282)
	⇓	⇓	⇓
Outlier reason			
Same side chain dihedrals angle as the respective amino acid in the template	Trp158 ^{ECL2}	Ser150 ^{4.57} Trp158 ^{ECL2} Asn179 ^{ECL2}	Leu30 ^{1.44} Asn217 ^{ICL3} Leu129 ^{ICL2} Thr131 ^{ICL2} Leu152 ^{4.59} Trp158 ^{ECL2} Arg296 ^{H8}
Adjusted to occupy the same area as in the template and/ or to interact with neighboring amino acids	Tyr94 ^{3.28} Asn159 ^{ECL2} Asn162 ^{ECL2}	Arg50 ^{ICL1} Leu52 ^{2.38} Asn159 ^{ECL2}	Phe110 ^{3.44} Asn159 ^{ECL2}
Adjusted to avoid an overlap with proximate amino acids	Tyr202 ^{5.58}		

¹ The definition of the regions for amino acid side chains (favored, allowed and disallowed) is given in chapter 5.

3.3.2.4 Planarity of aromatic groups and delocalized π -electron systems in amino acid side chains

The dihedral angles of one (hH_2R_i) and ten (hH_2R_{a1} and hH_2R_{a2} , respectively) amino acids with aromatic groups or delocalized π -electron system deviate between 10° and 30° from planarity (Figure 3.7). An unnatural deformation of originally planar moieties of amino acids could therefore be excluded.

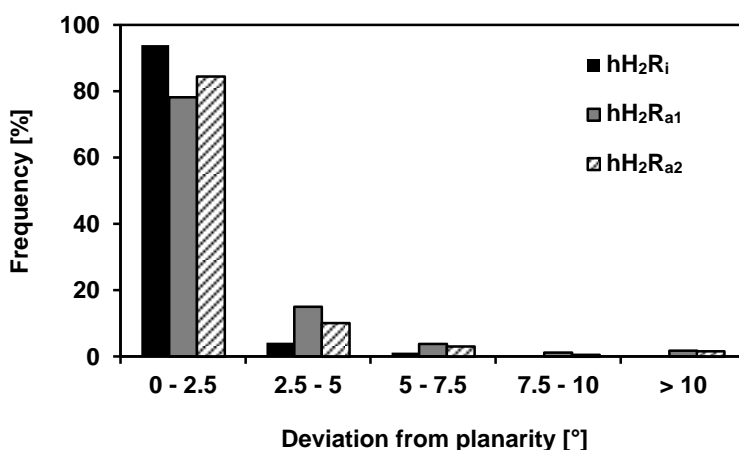


Figure 3.7: Planarity check of aromatic groups and delocalized π -electron systems in amino acid side chains

3.3.2.5 Main chain bond lengths, bond angles and bad contacts

Main chain bond lengths and angles are in acceptable limits (Table 3.6). A bad contact in the hH_2R_{a2} was detected with PROCHECK for Ala224^{6.25} near ICL3. However, a bump was neither observed visually nor was the energy of that residue conspicuously higher than for other amino acids.

Table 3.6: PROCHECK analysis of main chain parameters and bad contacts

Parameter	hH_2R_i	hH_2R_{a1}	hH_2R_{a2}
Main chain bond lengths within limits	100%	100%	100%
Main chain bond angles within limits	96.1%	93.6%	91.6%
Bad contacts	0	0	1

3.3.3 Comparison of inactive and active hH₂R receptor states

3.3.3.1 Overall topology of the protein backbone

After superimposing the backbone atoms of the shared TM domains 1 to 7 of the hH₂R_i, hH₂R_{a1} and hH₂R_{a2} models, root mean square deviations (RMSD) of 1.81 Å for hH₂R_i vs. hH₂R_{a1}, 2.66 Å for hH₂R_i vs. hH₂R_{a2}, and 1.74 Å for hH₂R_{a1} vs. hH₂R_{a2} were calculated. The RMSD values for all backbone atoms, i.e. TMs and loops, were only slightly higher (1.93 Å, 2.91 Å and 2.07 Å, respectively). Figure 3.8 depicts the RMSD values of backbone and side chain atoms as a function of the residue number.

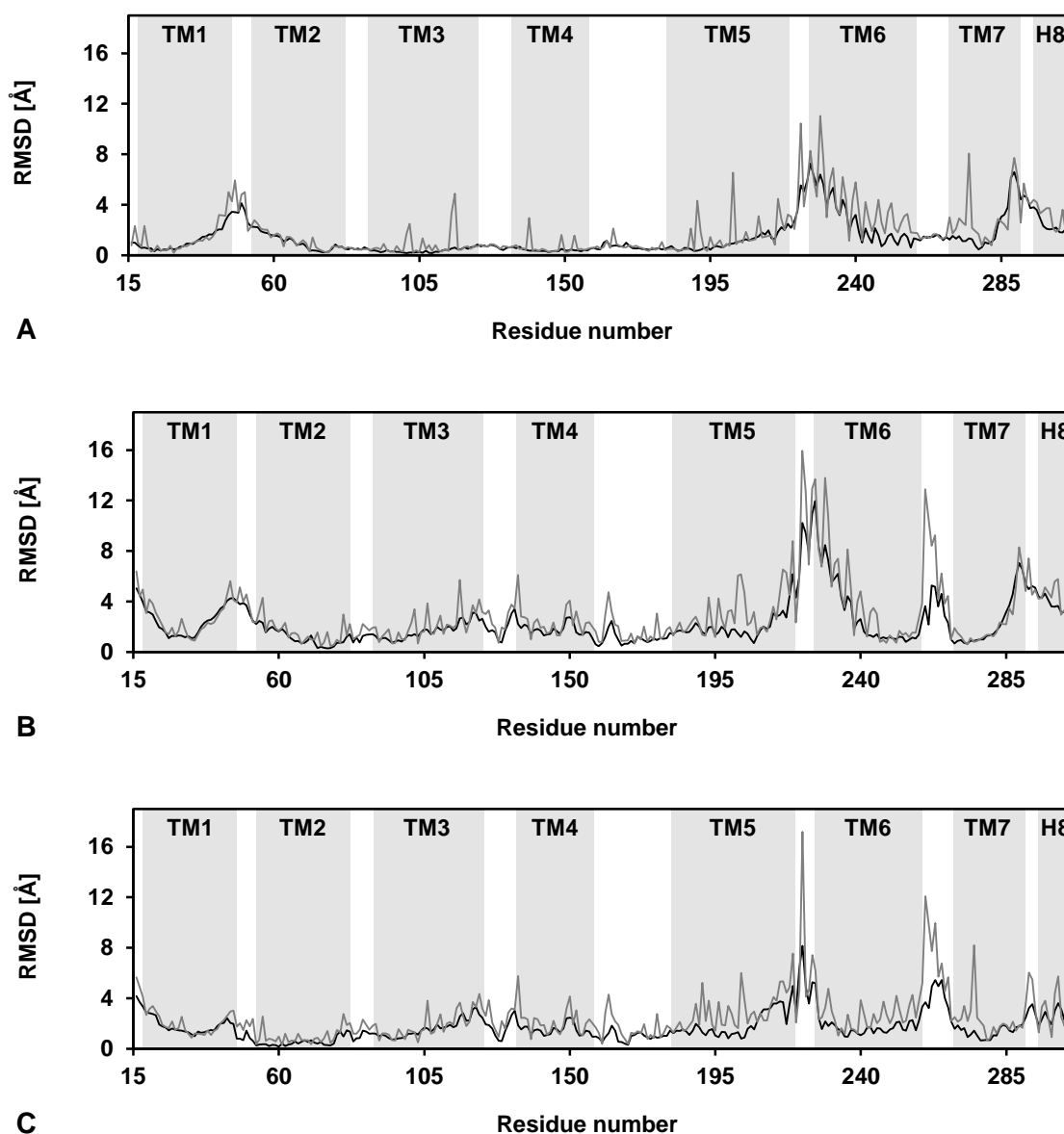
Comparing hH₂R_i and hH₂R_{a1}, high backbone RMSD values (ca. > 4 Å) were only observed in intracellular parts, namely around ICL1, from the end of TM5 via ICL3 to the intracellular part of TM6, as well as the end of TM7 and H8 (Figure 3.8 A and Figure 3.9 D). In relation to the hH₂R in the inactive state, the hH₂R_{a1} model is characterized by outward movements of TM6 and of the cytoplasmic ends of TM1 and TM2, including ICL1, as well as by the inward move of TM7 and helix 8. In hH₂R_{a1} this results in an increase of the distance between TM6 and TM2 (by 4.7 Å), TM3 (by 2.1 Å) and TM7 (by 1.2 Å), respectively (Table 3.7). The inward position of TM7 in the hH₂R_{a1} model also leads to a decreased cytoplasmic distance between TM2 and TM7 (by 1.0 Å) as well as TM3 and TM7 (by 2.9 Å). Remaining parts differ only slightly between both models, since their backbone coordinates were either directly taken from the tβ₁AR in each case (the protein fragment from ECL1 till the beginning of ECL2), since domains of hH₂R_{a1} were adapted to the template tβ₁AR of the hH₂R_i model (extracellular parts of TM1, TM2 and TM5), or since fragments were generated the same way in both models (ECL2 and ECL3).

Although the hH₂R_{a2} model was derived from a different template than hH₂R_i and the hH₂R_{a1}, the extracellular parts of the three models fit very well. Exceptions occur at the N-terminus of TM1, which is facing more outwards in the hH₂R_{a2} model (Figure 3.10 A), and loop ECL3 which was generated by a special loop search (individual anchor regions at the extracellular ends of TM6 and TM7 in the hH₂R_{a2} model). Compared to hH₂R_i and hH₂R_{a1}, the distance between TM1 and TM7 decreased in hH₂R_{a2} by 1.9 Å and 2.2 Å, respectively. Major changes occurred at the intracellular part. TM1 of the hH₂R_{a2} model was even further away from helix 8 of the inactive state than TM1 of hH₂R_{a1}, because helix 8 of hH₂R_{a2} moved further inwards compared to the other models (Figure 3.9 D). However, ICL1, the cytoplasmic part of TM2 and the end of TM7 overlap very well in both active state hH₂Rs (Figure 3.8 C). In the hH₂R_{a2} model, the backbone coordinates at the cytoplasmic ends of TM3 and TM4 deviate slightly from those of the other two models. TM3 is moderately shifted to the center of the receptor, and TM4 moves outwards in the hH₂R_{a2} model (Figure 3.10 C). Whereas the intracellular

distances between TM2 and TM3 as well as TM3 and TM7 are the smallest in the $\text{hH}_2\text{R}_{\text{a}2}$ model (9.9 Å and 13.3 Å, respectively), the distance between TM3 and TM5 is increased by 1.2 Å and 1.6 Å compared to $\text{hH}_2\text{R}_{\text{i}}$ and $\text{hH}_2\text{R}_{\text{a}1}$, respectively. This increase is caused by the intracellular parts of TM5 and TM6 of $\text{hH}_2\text{R}_{\text{a}2}$, located more outwards than in the other models (Figure 3.9 D). In the $\text{hH}_2\text{R}_{\text{a}2}$ model, the distances of TM6 to TM2, TM3 and TM7 are 4.8 Å, 2.6 Å and 1.2 Å greater, respectively, than in the $\text{hH}_2\text{R}_{\text{i}}$ model.

Figure 3.8: Root mean square deviations between inactive and active hH_2R states

Shown is the RMSD of each residue for backbone atoms (N, C $_{\alpha}$, C, O; black line) and side chain heavy atoms (all side chain atoms except hydrogens; grey line) for $\text{hH}_2\text{R}_{\text{i}}$ vs. $\text{hH}_2\text{R}_{\text{a}1}$ (A), $\text{hH}_2\text{R}_{\text{i}}$ vs. $\text{hH}_2\text{R}_{\text{a}2}$ (B), and $\text{hH}_2\text{R}_{\text{a}1}$ vs. $\text{hH}_2\text{R}_{\text{a}2}$ (C). The RMSD was calculated after superimposition of the shared α -helical region of the $\text{hH}_2\text{R}_{\text{i}}$, $\text{hH}_2\text{R}_{\text{a}1}$ and $\text{hH}_2\text{R}_{\text{a}2}$.



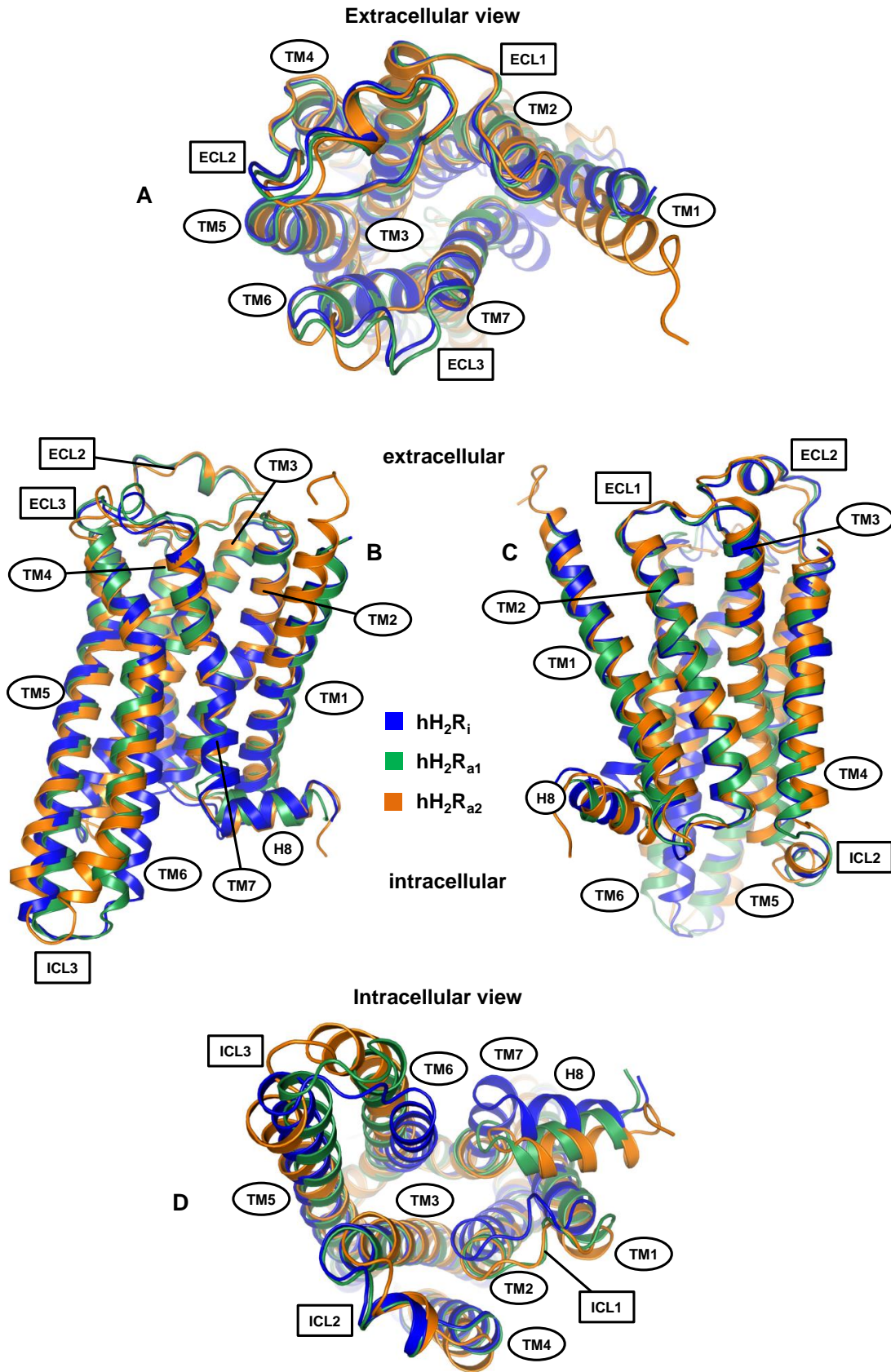


Figure 3.9: Cartoon representation of the hH₂R homology models

Comparison of the hH₂R_i, hH₂R_{a1} and hH₂R_{a2} after superimposing their shared TM domains. A, view from the extracellular side; B, C, side view; D, view from the intracellular side.

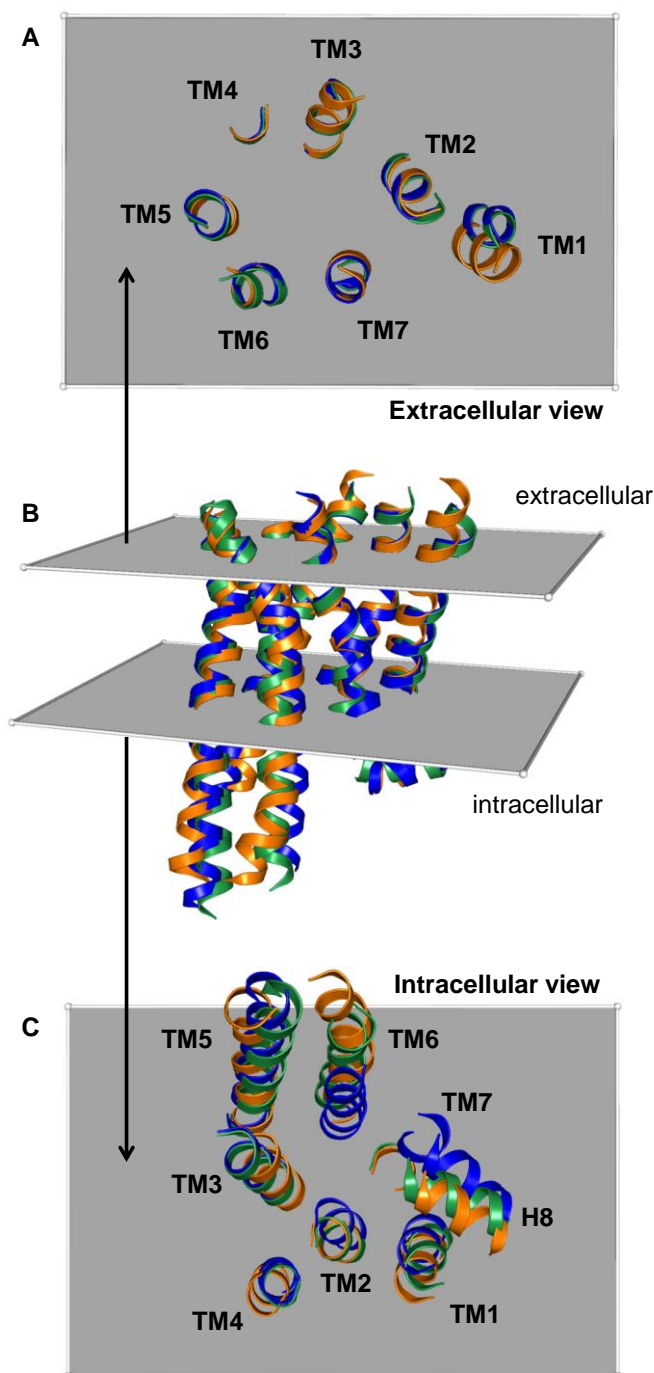


Figure 3.10: Planes parallel to the membrane at the extra- and intracellular side of the hH₂R models used for the calculation of TM distances

After superimposition of the shared TM domains of the hH₂R_i (■), hH₂R_{a1} (■) and hH₂R_{a2} (■) (B, side view), the least extending α -helix at the extracellular and intracellular side, respectively, was determined. Geometric centers of the heavy backbone atoms of the terminal four amino acids were defined for both TMs. Planes parallel to the plasma membrane were constructed through these centers. The intersections of the planes with the other TMs were used to define four amino acids in the other six α -helices at their extra- and intracellular ends, respectively. The geometric centers of the heavy backbone atoms of these amino acids were used for the calculation of pairwise distances (Table 3.7). For each receptor model, the calculations were based on the same amino acids. A, View from the extracellular side; C, View from the intracellular side. Loop regions and termini were not shown for reasons of clarity.

Some slight anomalies are also observed in TM4 and TM5. The α -helix of TM4 has a marginally different course compared to the other models (Figure 3.9 C), and in TM5 a bulge is present near the orthosteric binding site (Figure 3.9 A). Taken together, the rearrangements of TM domains in the active hH₂R state models are very similar to evident changes when comparing the inactive and active states of rhodopsin/opsin (Okada *et al.*, 2004; Scheerer *et al.*, 2008) and of the β_2 AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011a).

Table 3.7: Distances of TM domains at the extra- and intracellular side of the hH₂R models

The distances (Å) were calculated between the center of the backbone atoms (N, C α , C, O) of 4 consecutive residues of each TM at the extra- and intracellular side as explained in Figure 3.10. Shown are TM couples where at least one interaction occurs as described in Figure 3.16 and Table 3.9. Numbers in brackets indicate that there is no interaction between the respective TMs.

TM contact	Extracellular side			Intracellular side		
	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}
TM1 – TM2	10.8	11.0	10.6	9.4	10.1	9.5
TM1 – TM7	16.0	(16.3)	14.1	10.8	11.7	11.3
TM2 – TM3	11.8	11.7	12.3	10.9	11.6	9.9
TM2 – TM4	(21.4)	(21.3)	(21.4)	11.5	10.6	11.3
TM2 – TM6	(21.6)	(22.3)	(21.8)	13.9	(18.6)	(18.7)
TM2 – TM7	12.7	13.2	12.5	12.3	11.3	10.8
TM3 – TM4	9.8	9.8	9.4	11.7	11.7	11.9
TM3 – TM5	17.2	17.4	17.2	9.6	9.2	10.8
TM3 – TM6	19.6	20.2	20.1	11.2	13.3	13.8
TM3 – TM7	(16.9)	17.0	(17.3)	(17.5)	14.6	13.3
TM4 – TM5	13.9	14.2	14.8	(20.5)	(20.2)	(21.2)
TM5 – TM6	10.2	10.2	9.8	11.9	11.0	11.7
TM6 – TM7	11.3	11.8	11.5	10.2	11.4	11.4

Notably, TM domains with highly conserved prolines introducing kinks in the α -helical structure differ from the inactive state in the hH₂R_{a2} model (TM2 and TM6) or in both active state models (TM7). The latter could be a general characteristic and be part of the rearrangement of the cytoplasmic end of TM7 (Table 3.8).

Table 3.8: Proline kinks (in degree) in transmembrane domains of inactive and active hH₂R states

TM	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}
2	29.8	30.1	26.8
5	12.0	12.2	13.4
6	33.3	32.3	29.1
7	40.2	32.5	36.1

3.3.3.2 Activation of the hH₂R

The alterations of the relative positions of TM domains during receptor activation are accompanied by conformational changes (molecular switches) which were proposed to be responsible for activation and signal transduction from the ligand binding site to the G-protein interacting surface at the intracellular part of the receptor (Trzaskowski *et al.*, 2012).

The so-called rotamer toggle switch was previously suggested to be part of the signal transduction in GPCRs (Shi *et al.*, 2002). According to this hypothesis, conformational changes of amino acid side chains located close to the orthosteric binding site (Cys/Ser/Thr^{6.47}, Trp^{6.48} and Phe^{6.52}) modulate the proline kink in TM6 (Kobilka and Deupi, 2007), thereby leading to the outward move of its cytoplasmic part. In this way, the rotamer toggle switch was supposed to transfer the signal from the ligand binding site to the intracellular side. However, in the active state structures of rhodopsin (Park *et al.*, 2008), the adenosine A_{2A} receptor (Xu *et al.*, 2011) and the β_2 AR (Rasmussen *et al.*, 2011a), Trp^{6.48} does not show the predicted rotamer transition. Relocation of Trp^{6.48}, as part of the so-called transmission switch (Deupi and Standfuss, 2011), is only indicated in the crystal structure of opsin (Scheerer *et al.*, 2008) and therefore also present in the hH₂R_{a1} model. Summarizing data from active state structures, a rearrangement of amino acids in TM3, TM5 and TM6 (Ile^{3.40}, Pro^{5.50}, Leu^{5.51}, Phe^{6.44} and Trp^{6.48}) may be responsible for signal transmission including TM6 movement.

In the hH₂R_{a1} model, changes of amino acids touching the orthosteric binding site are observed for Tyr250^{6.51} and Phe254^{6.55}. Both side chains are located deeper in the binding pocket compared to the inactive state. Consequently, Trp247^{6.48} turns down, stacks with Phe251^{6.52} and forces Phe243^{6.44} in a new position closer to Leu195^{5.51} (Figure 3.11 A). These changes resemble those of the respective amino acids in the transition from rhodopsin to opsin. The modification of the side chains of Trp247^{6.48} and Tyr250^{6.51}, which were reported to contribute to the stabilization of inactive state conformations (Deupi *et al.*, 2012b),

results in a break of interactions with amino acids of TM7 (Figure 3.16 B1), partly impairing the hydrogen bond network around TM7 (see below). To avoid an overlap of Phe243^{6.44} with Leu195^{5.51} and other clashes with TM5, TM6 rotates horizontally outwards. Furthermore, the new position of the Phe254^{6.55} side chain results in a movement of Phe251^{6.52}, breaking the π - π -interactions with Phe191^{5.47} and forming a new aromatic interaction with Tyr192^{5.48}. Since the position of Phe191^{5.47} in the inactive state partly overlaps with Phe251^{6.52} in the hH₂R_{a1} model, Phe191^{5.47} is relocated, enabling the Trp247^{6.48} movement towards the position of Phe243^{6.44} in the inactive state. Hence, the crucial move of Phe243^{6.44} is initiated by Tyr250^{6.51} and Phe254^{6.55} near the binding pocket. TM3 is only slightly affected by this rearrangement. Solely the χ 2 dihedral angle of Ile106^{3.40} is changed to enable the new Phe243^{6.44} position. Caused by the construction of the hH₂R_{a1} model, where TM6 was adapted to the structure of opsin, the mechanism is comparable to the rhodopsin activation (Figure 3.11 A).

The activation mechanism suggested from the hH₂R_{a2} model is similar to that described for the β_2 AR (Deupi and Standfuss, 2011; Rasmussen *et al.*, 2011a). The hH₂R_{a2} binding site is contracted, leading to a 2.1 Å inward shift of the C $_{\alpha}$ atom of Thr190^{5.46} compared to the inactive state. The inward movement of TM5 is relayed up to the highly conserved Pro194^{5.50}.

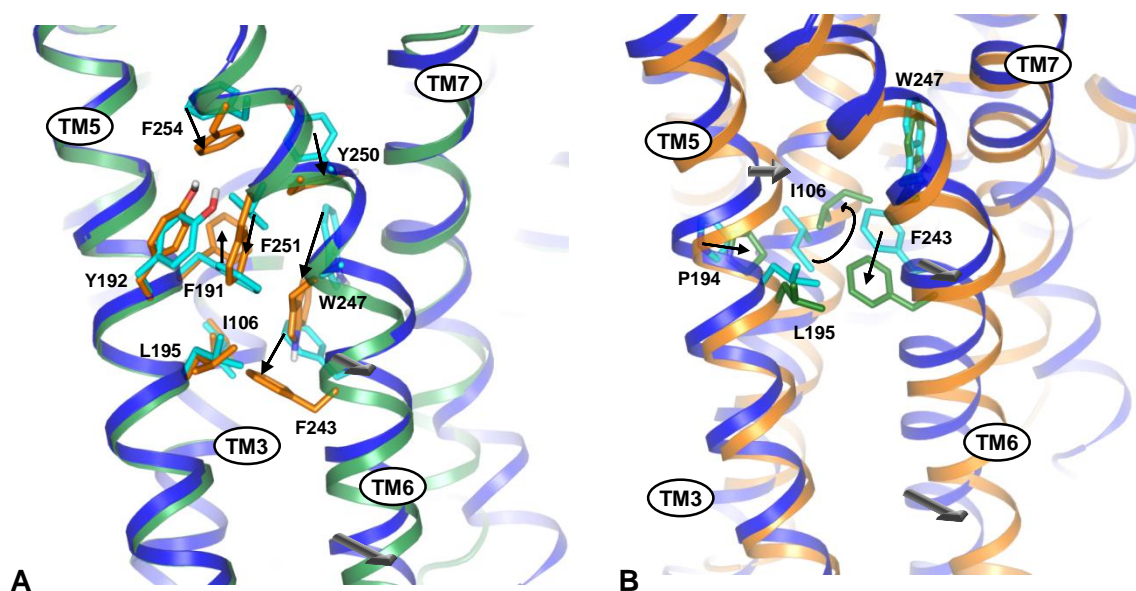


Figure 3.11: Signal transduction from the ligand binding site to the motion of TM6 in the active hH₂R states

Superimposed structures of hH₂R_i (blue ribbons for TMs, carbon atoms in cyan) with hH₂R_{a1} (A, green ribbons for TMs, carbon atoms in orange) and hH₂R_{a2} (B, orange ribbons for TMs, carbon atoms in green). The movements of amino acid side chains by receptor activation are indicated with thin black arrays. The moves of transmembrane domains are shown with grey arrays.

To avoid an overlap of TM5 with Ile106^{3.40}, this residue changes its conformation, pointing towards TM7 in its new position and enabling the side chain movement of Phe243^{6.44} into the former Ile106^{3.40} position. As in the hH₂R_{a1} model and in the active state crystal structures, the new position of Phe243^{6.44} requires the outward motion of TM6. Thus, Phe^{6.44} seems to be a key amino acid in receptor activation (Figure 3.11 B).

The motion of TM6 affects a further highly conserved motif. Concluded from biochemical studies, Arg^{3.50} of the D/ERY motif and Glu^{6.30} are in close contact forming a salt bridge. This so-called ionic lock was discussed to be a key structural element for restraining GPCRs in an inactive receptor conformation (Angelova *et al.*, 2002; Ballesteros *et al.*, 2001; Greasley *et al.*, 2002; Shapiro *et al.*, 2002). In the crystal structures of rhodopsin (Palczewski *et al.*, 2000), the dopamine D₃ receptor (Chien *et al.*, 2010) and some antagonist bound adenosine A_{2A} receptors (Dore *et al.*, 2011) the ionic lock is observed, but all other crystallized inactive states of GPCRs do not contain this restraint. For the β_2 AR an equilibrium of conformations with and without the lock was suggested (Dror *et al.*, 2009), and the relatively high basal activity and structural instability of the β_2 AR was discussed as reason (Rasmussen *et al.*, 2007). Furthermore, high crystallographic B-values in the structure of rhodopsin (Palczewski *et al.*, 2000) indicated that the side chains of Arg^{3.50} and Glu^{6.30} may adopt different conformations. The salt bridge between these amino acids is not present in the hH₂R_i model (Figure 3.12 A). The distance between the closest heteroatoms of Arg116^{3.50} and Glu229^{6.30} is 5.9 Å. Similar to the t β_1 AR (Warne *et al.*, 2008b), Glu229^{6.30} points towards the cytoplasm. The close contact of Asp115^{3.49} and Arg116^{3.50} enables a salt bridge between them. Asp115^{3.49} is furthermore hydrogen bonded to Tyr126^{ICL2}, possibly restraining the α -helical conformation of ICL2. Both the tyrosine residue and the α -helix of ICL2 were discussed to be important for G-protein coupling (Warne *et al.*, 2008b). The side chain of Arg116^{3.50} is additionally retained in its position by interacting with the hydrophobic amino acids Ile112^{3.46}, Ile205^{5.61} and Leu236^{6.37}. However, by torsion of the side chain of Glu229^{6.30}, a closer contact to Arg116^{3.50} is possible (distance between the closest heteroatoms ~ 3.5 Å). Therefore, an equilibrium as suggested for the β_2 AR cannot be ruled out. In the hH₂R_{a1} model Glu229^{6.30} moves towards TM5 and interacts with Arg210^{5.66} (salt bridge) and Thr233^{6.34} (hydrogen bond), restraining the cytoplasmic ends of TM5 and TM6 in their close contact (Figure 3.12 B). The ionic interaction between Asp115^{3.49} and Arg116^{3.50} is broken. The aspartate moves towards TM4 and is close to Tyr391 of G α CT. Arg116^{3.50} is flanked by the hydrophobic side chains of Ile112^{3.46}, Tyr202^{5.58}, Ile205^{5.61} and Leu236^{6.37}, filling a part of the newly created gap between TM3 and TM6. Additionally, Arg116^{3.50} interacts with Tyr391 and Leu393 of G α CT. The side chain positions are in agreement with the structure of opsin (Scheerer *et al.*, 2008). In the hH₂R_{a2} model Glu229^{6.30} is also linked to Arg210^{5.66} (salt bridge). Similar to the

inactive state, Asp115^{3.49} is hydrogen bonded to Tyr126^{ICL2}, and additionally to Thr53^{2.39}. Arg116^{3.50} is surrounded by the hydrophobic side chains of Thr53^{2.39} and Ile112^{3.46}. It further contacts Tyr391 and Leu393 (van der Waals interactions) as well as Glu392 (salt bridge) of GaCT (Figure 3.12 C). The ionic interaction with the G-protein is enabled by the more inward position of TM3 in the hH₂R_{a2} model compared to hH₂R_{a1} (Figure 3.9 D).

The differences of the D/ERY motif in the hH₂R models are in agreement with a previous study on the rat histamine H₂ receptor, where the aspartate and arginine residues were separately replaced by Ala or Asn (Alewijns *et al.*, 2000). Mutation of the aspartate resulted in receptors with high constitutive activity, increased agonist affinity and increased signaling properties. On the atomic level, the substitution of Asp^{3.49} releases Arg^{3.50} from its constrained 'inactive' position.

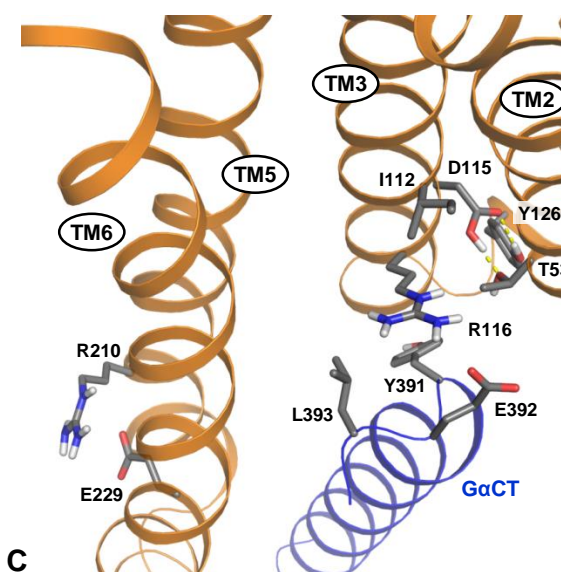
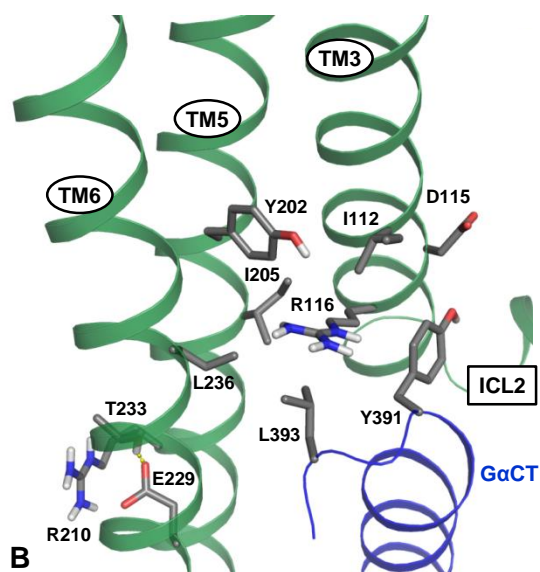
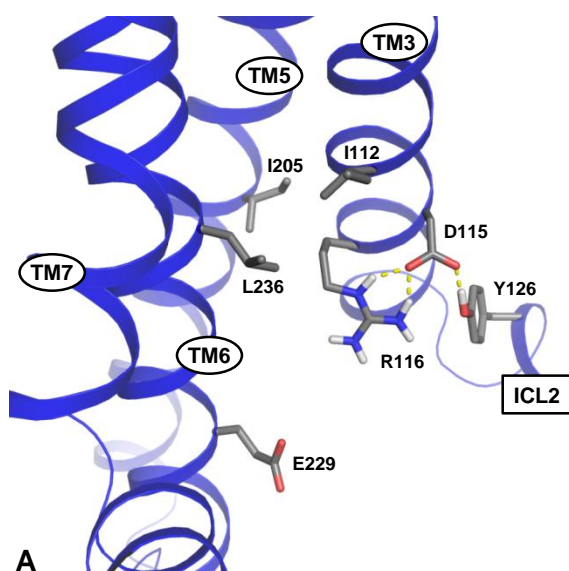


Figure 3.12: Rearrangement of the DRY motif in the hH₂R

Side view of hH₂R_i (A, blue ribbons), hH₂R_{a1} (B, green ribbons) and hH₂R_{a2} (C, orange ribbons). Shown are amino acids (sticks) close to Asp115^{3.49}, Arg116^{3.50} and Glu229^{6.30}. Hydrogen bonds are indicated with yellow dashed lines. Not all TMs are shown for reasons of clarity.

The putative ionic lock with Glu^{6.30} is then impossible, since Arg^{3.50} is not fixed in the appropriate position. Furthermore, Arg^{3.50} is free to fill the gap between TM3 and TM6, to interact with the G-protein and therefore to stabilize an active receptor conformation with increased affinity for agonists (Lefkowitz *et al.*, 1993; Rasmussen *et al.*, 2011a). The mutation of the arginine also led to an increased agonist affinity, but decreased signal transduction properties. Again, the potential disruption of the (possibly temporary) ionic lock could cause a shift towards an active receptor conformation, characterized by higher agonist affinity. Because Arg^{3.50} contributes to the receptor-G-protein interaction, replacing this residue could result in decreased signaling properties of the receptor as identified by the rH₂R mutant.

Both active state models indicate that the hydrophobic interface of TM2/TM6 and TM3/TM6 is broken up on receptor activation. In the inactive state, amino acids of TM2, TM3 and TM6 are packed in the middle of the TM bundle, resulting in smaller distances in the intracellular part of the inactive state receptor (TM2-TM6, 13.9 Å; TM3-TM6, 11.2 Å; Table 3.7) compared to hH₂R_{a1} (TM2-TM6, 18.6 Å; TM3-TM6, 13.3 Å) and hH₂R_{a2} (TM2-TM6, 18.7 Å; TM3-TM6, 13.8 Å). Only in the inactive conformation van der Waals interactions between TM2 (Thr53^{2.39}, Ile57^{2.43}, Leu60^{2.46}) and TM6 (Leu236^{6.37}, Val239^{6.40}), as well as TM3 (Ile112^{3.46}, Ser113^{3.47}) and TM6 (Leu236^{6.37}, Val239^{6.40}, Met240^{6.41}) are present (Figure 3.13 A). These amino acids correspond partly with those of rhodopsin (Leu76^{2.43}, Leu79^{2.46}, Leu128^{3.43}, Leu131^{3.46}, Met253^{6.36}, Met257^{6.40}), forming the so called hydrophobic barrier. This barrier separates the water-mediated hydrogen bond network around TM2 and TM7 from the D/ERY motif (Li *et al.*, 2004; Standfuss *et al.*, 2011). In the active receptor conformation this hydrophobic cluster is broken by the rotation of TM6 (Figure 3.13 B). Instead, in the core of the receptor amino acids of TM3, TM5 and TM7 approach each other. Tyr288^{7.53} is linked to Leu109^{3.43} and Ile112^{3.46}, as well as Ser105^{3.39} to Asn284^{7.49} in the hH₂R_{a1} model (water-mediated hydrogen bond) and Leu109^{3.43} to Asn284^{7.49} in the hH₂R_{a2} model (van der Waals interaction). Furthermore, Tyr202^{5.58} is extended by a conformational switch into the gap between TM3 and TM6, being in close contact to Tyr288^{7.53} of the NPxxY(x)_{5,6}F motif.

The motion of TM6 at the cytoplasmic part of the receptor enables the rearrangement of the highly conserved NPxxY(x)_{5,6}F motif at the end of TM7 and helix 8. In the inactive state Tyr288^{7.53} points towards TM2, similar as in the structures of the β₁AR (Warne *et al.*, 2008b), the β₂AR (Cherezov *et al.*, 2007), squid rhodopsin (Murakami and Kouyama, 2008), the adenosine A_{2A} receptor (Jaakola *et al.*, 2008) and the dopamine D₃ receptor (Chien *et al.*, 2010).

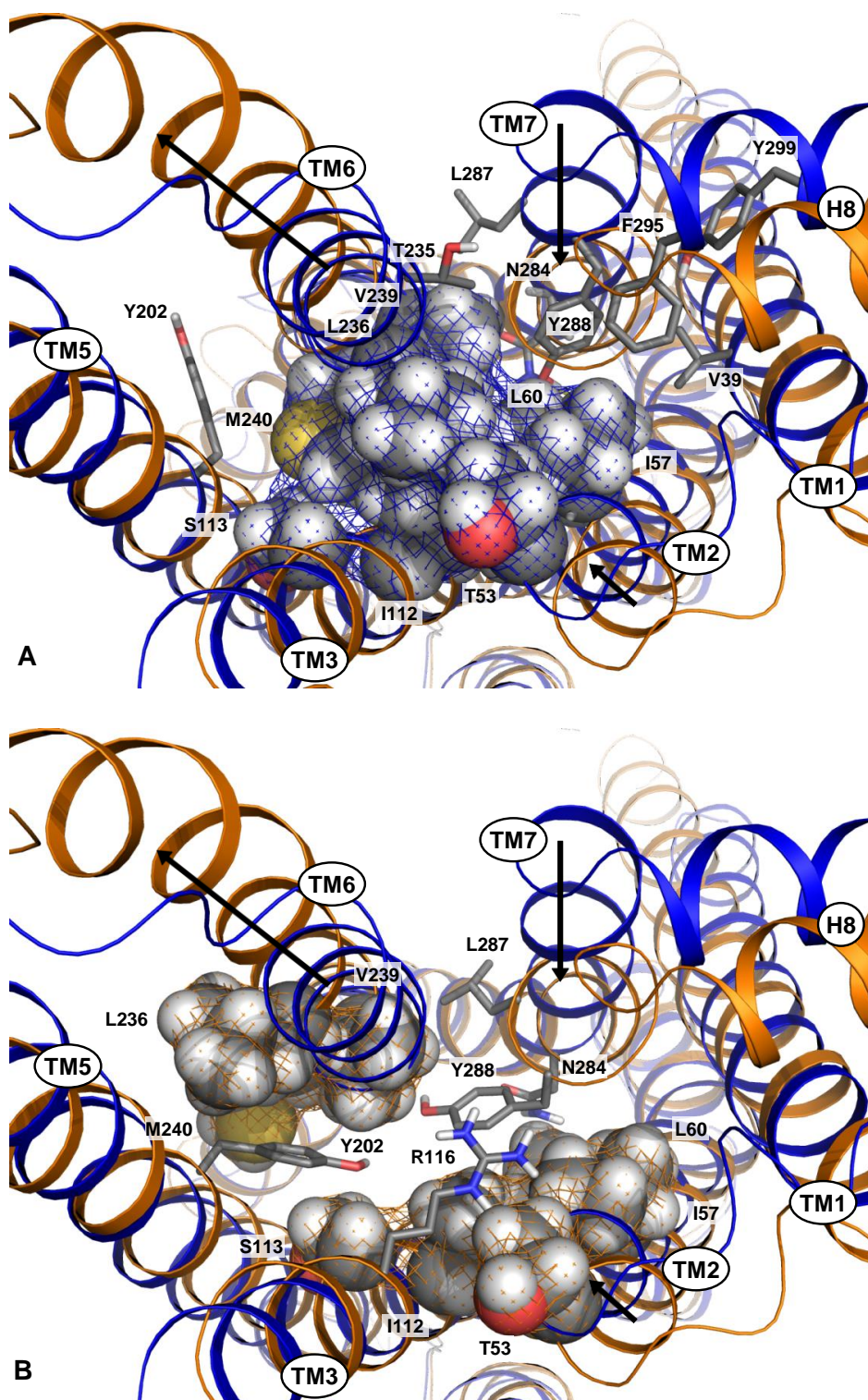


Figure 3.13: Disruption of the hydrophobic barrier by activation of the hH₂R

Cytoplasmic view of the superimposed structures of hH₂R_i (blue ribbons) and hH₂R_{a2} (orange ribbons). Black arrays indicate the movement of TMs 2, 6 and 7 on receptor activation. Amino acids of the hydrophobic barrier are shown in spheres surrounded by a mesh. Additionally drawn residues (sticks) are either in close contact to Tyr288^{7,53} or fill the gap between TM3 and TM6. Amino acids are from hH₂R_i (A) and hH₂R_{a2} (B), respectively (hH₂R_{a1} is not shown since, in the analyzed region, the positions of TM domains and side chains do not significantly differ in both active state models).

Whereas van der Waals interactions with Ile57^{2.43}, Val239^{6.40}, Asn284^{7.49} and Leu287^{7.52} do also occur in both active hH₂R state models, the 'inactive' position of Tyr288^{7.53} is stabilized by hydrophobic contacts with Val39^{1.53} and Thr235^{6.36} as well as with Phe295 and Tyr299 of helix 8 (Figure 3.13 A). In the hH₂R_i model, a conserved water molecule placed between TM2 and TM7 connects the side chain hydroxyl group of Tyr288^{7.53} to the backbone oxygen of Ile57^{2.43} and the side chain of Asn284^{7.49} via hydrogen bonds (Figure 3.14 A). In contrast, in the active state models Tyr288^{7.53} is switched into the newly created gap between TM3 and TM6 (Figure 3.13 B). This orientation is also present in the active state structures of the β_2 AR (Rasmussen *et al.*, 2011a; Rasmussen *et al.*, 2011b) and opsin (Scheerer *et al.*, 2008). Furthermore, Tyr288^{7.53} approaches Tyr202^{5.58}, which changes its place with Met240^{6.41} on receptor activation and extends into the gap between TM3 and TM6. The water mediated hydrogen bonds of Tyr288^{7.53} are broken, and new interactions are formed with Leu60^{2.46}, Leu109^{3.43} and Ile112^{3.46} (Figure 3.13 B). In rhodopsin the disruption of the Tyr-Phe interaction within the NPxxY(x)_{5,6}F motif (alanine replacement mutation) results in a partly activated receptor (metarhodopsin II). However, replacement of Tyr^{7.53} led to decreased activation rate of the G-protein compared to the wild-type (determined from GTP γ S uptake by G_i), suggesting that Tyr^{7.53} is not only important for stabilizing an inactive receptor state but also for efficient G-protein coupling (Fritze *et al.*, 2003). Indeed, the position of Tyr^{7.53} in active GPCR states seems to keep TM6 away from moving inwards and to maintain the crevice at the intracellular surface (Scheerer *et al.*, 2008).

The rearrangement of TM7 around Tyr288^{7.53} and the motion of TM6 change the hydrogen bond network at the cytoplasmic part of the receptor, affecting TM1, TM2, TM3, TM6 and TM7. In the hH₂R_i model, a water molecule connects the backbone oxygen of Val239^{6.40} to the side chains of Asn280^{7.45} and Asn284^{7.49} (Figure 3.14 A). This water molecule was derived from the structure of the h β_2 AR (Cherezov *et al.*, 2007), connecting the respective amino acids Ile295^{6.40}, Asn318^{7.45} and Asn322^{7.49}. In the active state models of the hH₂R these contacts are unclosed (Figure 3.14 B, C). The distances between the corresponding heteroatoms are raised by 1.5 Å to 2.7 Å. The outward move of TM6 and, as a consequence, the partial disintegration of the water network, could be a trigger for the modification of the NPxxY(x)_{5,6}F motif. Asn284^{7.49} and the subsequent part of TM 7 up to helix 8 are rearranged in the active state hH₂R models, and the proline kink in TM7 is modified, too. Direct hydrogen bonds to other amino acids are formed by the side chains of Asn280^{7.45} (hH₂R_{a1}) and Asp64^{2.50} (hH₂R_{a2}).

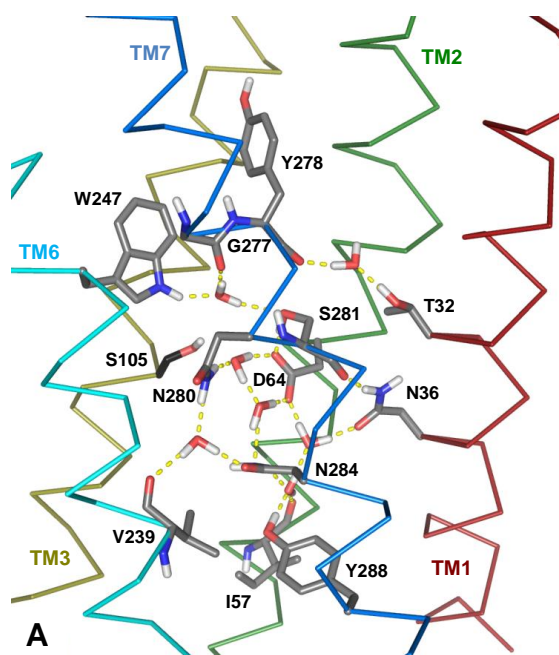
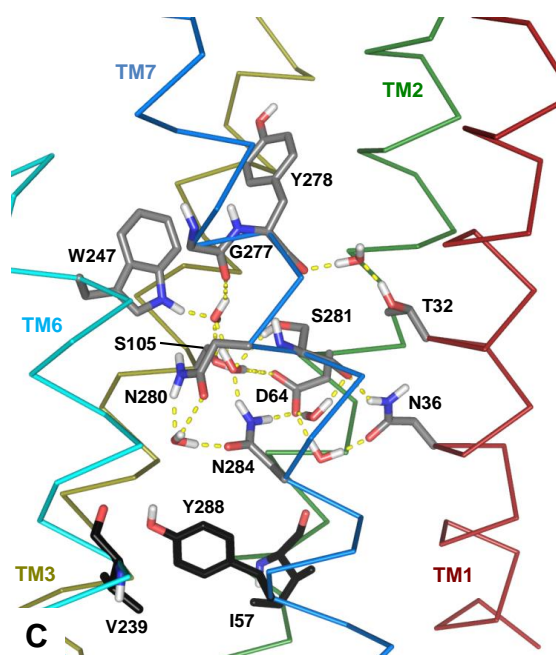
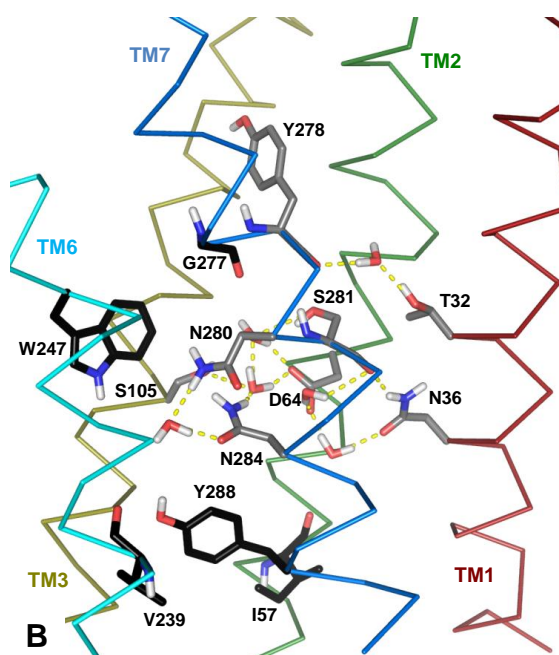


Figure 3.14: Hydrogen bond network at the intracellular part of the hH₂R

Side view of hH₂R_i (A), hH₂R_{a1} (B) and hH₂R_{a2} (C). All TM domains except TM4 and TM5 are shown as tubes. Amino acids (sticks) with grey carbon atoms contribute to the hydrogen bond network of the respective hH₂R model. Amino acids with black carbon atoms are not part of the network. The water molecules were inserted as described in section 3.2.2. Hydrogen bonds are indicated with yellow dashed lines (maximum distance of 3.5 Å between corresponding heteroatoms, hydrogen bond angle of at least 120°).



In the hH₂R_{a1} model the inward position of the intracellular end of TM7 enables a water mediated hydrogen bond between Ser105^{3,39} and the side chains of Asp64^{2,50} and Asn284^{7,49}. In the hH₂R_{a2} model Ser105^{3,39} is connected by direct hydrogen bonds to Thr63^{2,49} and Asp64^{2,50}, and additionally by a water molecule to Gly277^{7,42} and the side chain of Trp247^{6,48} which is also part of the hydrogen bond network in the inactive state. This connection is absent in hH₂R_{a1} due to the motion of Trp247^{6,48} towards Phe243^{6,44}, indicating that the interruption of the hydrogen bond network near Trp247^{6,48} could be a further link from

the rearrangements below the ligand binding site to the modifications around Tyr288^{7.49} of the NPxxY(x)_{5,6}F motif. Similar observations were reported for a constitutively active rhodopsin mutant. The movement of the side chain of Trp^{6.48}, following the TM6 motion, disrupts the water mediated hydrogen bond with Ser^{7.45}. This results in a reorganization of the hydrogen bond network, including highly conserved amino acids in TM1, TM2 and in the NPxxY(x)_{5,6}F motif (Standfuss *et al.*, 2011).

In rhodopsin the linkage of the amino acids in position 3.28 and 7.43 are of special importance for the activation mechanism. Concomitant with the metarhodopsin II formation, the Schiff base nitrogen of Lys296^{7.43} is deprotonated and the Glu113^{3.28}/Lys296^{7.43} salt bridge is broken. This so called 3-7 lock switch was discussed to be the fundamental trigger for receptor activation (Kim *et al.*, 2004). However, in the nanobody and G_s-protein stabilized active state structures of the β_2 AR (Rasmussen *et al.*, 2011a; Rasmussen *et al.*, 2011b), the respective amino acids in TM3 and TM7 are still in close contact to each other. Thus, the switch in TM3/TM7 seems not to be a common element required for GPCR activation. Accordingly, in the inactive and active state hH₂R models the respective heteroatoms of Tyr94^{3.28}, Asp98^{3.32} and Tyr278^{7.43} were in close contact (distances of heteroatoms below 3.9 Å). Only in the hH₂R_{a1} model hydrogen bonds from Tyr278^{7.43} to Tyr94^{3.28} and Asp98^{3.32} were detected (Figure 3.16 B1, Table 3.9).

3.3.3.3 Analysis of inter-TM domain contacts

Besides molecular switches of highly conserved motifs, the types and counts of interactions (salt bridges, direct and water mediated hydrogen bonds, van der Waals contacts) between neighboring TMs were systematically analyzed in order to identify structural features crucial for the inactive and active hH₂R states (Figure 3.16, Table 3.9). Most of the contacts are of the van der Waals type, taking into account that about 65% of all amino acids in the α -helical region of the hH₂R are hydrophobic. The majority of the differences of inter-TM contacts between inactive and active hH₂R states apply to molecular switches around highly conserved motifs described in the previous section.

The relatively slight differences of TM domains at the extracellular part (Figure 3.8) indicate analogous interactions in all three receptor models. Indeed, remarkable anomalies occur only in the hH₂R_{a1} model with respect to TM6 and TM7 (Figure 3.16 B1). The reason is the adjustment of TM6 and some amino acid side chains to the active structure of opsin. The reduced number of interactions in the TM1-TM7 and TM2-TM7 interfaces of hH₂R_{a1} is caused by the adaption of the side chain conformation of Trp275^{7.40} to the respective amino acid of opsin (Scheerer *et al.*, 2008). In contrast to rhodopsin, where the side chain of the corresponding Phe293^{7.40} is located in the interface between TM1-TM2-TM7 (Okada *et al.*,

2004), in opsin it points towards the plasma membrane. Therefore, van der Waals interactions from Trp275^{7.40} to TM1 and TM2 were broken. Receptor activation is not affected by this modification. The adaption of Trp247^{6.48} and Phe250^{6.51} to the respective positions of Trp^{6.48} and Phe^{6.51} in opsin, respectively, results in a break of contacts to TM7 (see above). The reduced number of interactions in the TM5-TM6 interface is caused by the alignment of TM6 of the t β ₁AR to the respective TM of opsin during model generation. Consequently, the suggested activation mechanism of hH₂R_{a1} differs from that of hH₂R_{a2} (cf. section 3.3.3.2).

Major changes occur at the intracellular side. The cytoplasmic outward movement of TM6 and the rearrangement and inward movement of TM7 have the greatest impact. In general, the contacts of TM6 to TM2, TM3, TM5 and TM7, respectively, and the contacts between TM3 and TM7 are modified by receptor activation (Figure 3.16 A2, B2, C2). The differences between TM2 and TM6, TM3 and TM6 as well as TM3 and TM7 were already described in context with the hydrophobic barrier. The rearrangement of TM7 including the change of the NPxxY(x)_{5,6}F motif and the water mediated hydrogen bond network affecting the TM6-TM7 interface was also discussed above. Although the TM5-TM6 distance is only slightly reduced from 11.9 Å in the hH₂R_i to 11.0 Å in the hH₂R_{a1} and 11.7 Å in the hH₂R_{a2} model, the number of interactions increases from 8 (inactive) to 16 in both active state models (Figure 3.15).

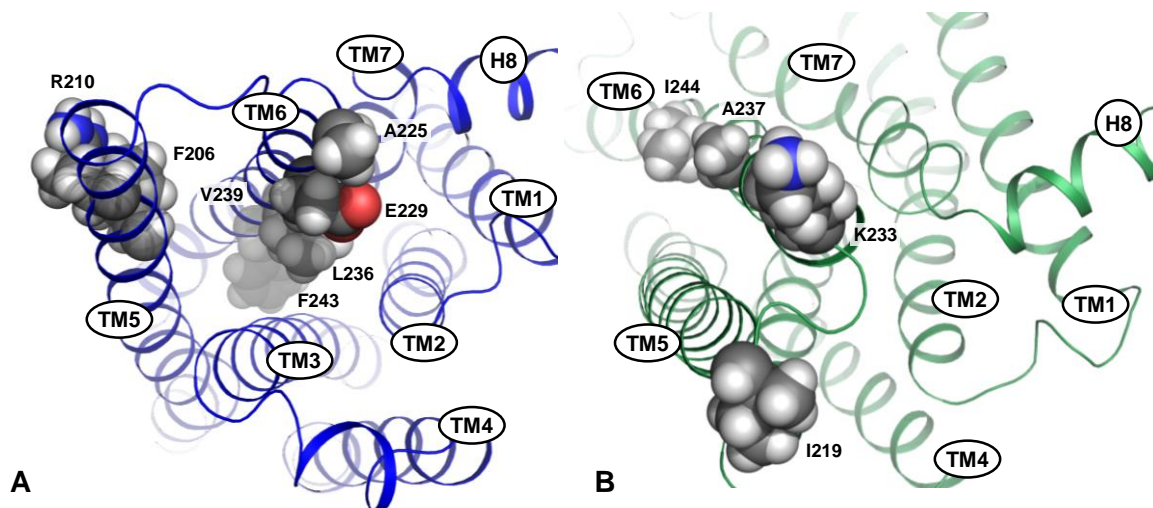


Figure 3.15: Interface between TM5 and TM6 at the intracellular side of the hH₂R

View from the intracellular side. In hH₂R_i (A, TMs as blue ribbons) Phe206^{5.62}, Arg210^{5.66}, Ala225^{6.26}, Glu229^{6.30}, Leu236^{6.37}, Val239^{6.40} and Phe243^{6.44} do not contribute to the TM5-TM6 interaction observed in both active hH₂R states. In hH₂R_{a1} (B, TMs as green ribbons) Ile219^{5.75}, Lys223^{6.24}, Ala237^{6.38} and Ile244^{6.45}, connecting TM5 and TM6 in the inactive state, do not point to the other TM domain, respectively.

Table 3.9: Counts and types of contacts between TM domains of the hH₂R states

Given are the counts of interactions at the extra- and intracellular part of the three receptor states for a specific TM-TM contact. In brackets: counts of ionic interactions (distance of charged heteroatoms below 5 Å), hydrogen bonds (distance of heteroatoms below 3.5 Å, angle of heteroatom 1, hydrogen and heteroatom 2 at least 120°), water mediated hydrogen bonds and van der Waals interactions (distance of hydrophobic heavy atoms below 5 Å). Repulsive ionic interactions between neighboring TMs were not observed.

	TM contact	hH ₂ R _i	hH ₂ R _{a1}	hH ₂ R _{a2}
Extracellular side	TM1 - TM2	9 (0, 0, 0, 9)	8 (0, 0, 0, 8)	9 (0, 0, 0, 9)
	TM1 - TM7	2 (0, 0, 0, 2)	0 (0, 0, 0, 0)	3 (0, 0, 0, 3)
	TM2 - TM3	9 (0, 0, 0, 9)	10 (0, 0, 0, 10)	8 (0, 1, 0, 7)
	TM2 - TM7	9 (0, 1, 1, 7)	3 (0, 0, 0, 3)	7 (0, 0, 1, 6)
	TM3 - TM4	10 (0, 1, 0, 9)	10 (0, 1, 0, 9)	11 (0, 0, 0, 11)
	TM3 - TM5	2 (0, 0, 0, 2)	3 (0, 0, 0, 3)	3 (0, 0, 0, 3)
	TM3 - TM6	2 (0, 0, 0, 2)	1 (0, 0, 0, 1)	1 (0, 0, 0, 1)
	TM3 - TM7	0 (0, 0, 0, 0)	2 (0, 2, 0, 0)	0 (0, 0, 0, 0)
	TM4 - TM5	8 (0, 0, 4, 4)	6 (0, 0, 2, 4)	7 (0, 0, 3, 4)
	TM5 - TM6	14 (0, 1, 0, 13)	9 (0, 1, 0, 8)	13 (0, 1, 0, 12)
	TM6 - TM7	19 (1, 1, 4, 13)	11 (1, 1, 0, 9)	18 (2, 0, 3, 13)
	total	84 (1, 4, 9, 70)	63 (1, 5, 2, 55)	80 (2, 2, 7, 69)
Intracellular side	TM1 - TM2	12 (0, 0, 1, 11)	9 (0, 0, 1, 8)	13 (0, 0, 1, 12)
	TM1 - TM7	9 (0, 1, 1, 7)	5 (0, 1, 0, 4)	7 (0, 1, 0, 6)
	TM2 - TM3	14 (0, 2, 0, 12)	11 (0, 1, 1, 9)	17 (0, 4, 0, 13)
	TM2 - TM4	12 (0, 1, 0, 11)	13 (0, 1, 0, 12)	13 (0, 1, 0, 12)
	TM2 - TM6	3 (0, 0, 0, 3)	0 (0, 0, 0, 0)	0 (0, 0, 0, 0)
	TM2 - TM7	7 (0, 1, 4, 2)	6 (0, 0, 2, 4)	8 (0, 1, 2, 5)
	TM3 - TM4	7 (0, 0, 0, 7)	8 (0, 0, 0, 8)	10 (0, 0, 0, 10)
	TM3 - TM5	19 (0, 0, 0, 19)	21 (0, 0, 0, 21)	17 (0, 1, 0, 16)
	TM3 - TM6	9 (0, 0, 0, 9)	3 (0, 0, 0, 3)	2 (0, 0, 0, 2)
	TM3 - TM7	0 (0, 0, 0, 0)	3 (0, 0, 1, 2)	3 (0, 0, 0, 3)
	TM5 - TM6	8 (0, 0, 0, 8)	16 (1, 1, 0, 14)	16 (1, 0, 0, 15)
	TM6 - TM7	9 (0, 0, 2, 7)	3 (0, 0, 0, 3)	6 (0, 0, 0, 6)
	total	96 (0, 5, 8, 96)	98 (1, 4, 5, 88)	112 (1, 8, 3, 100)

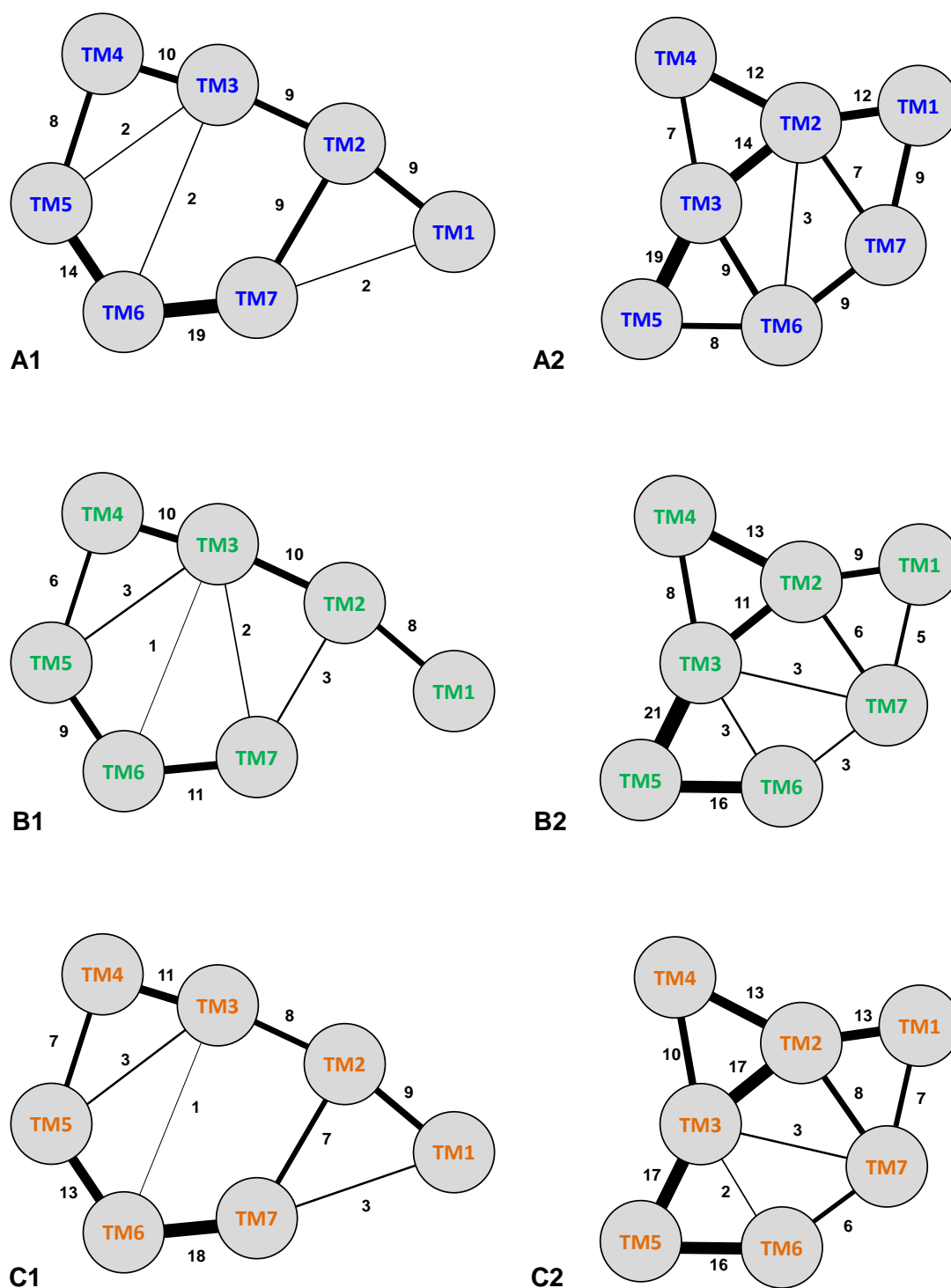


Figure 3.16: Contacts between TM domains at the extra- and intracellular part of the hH₂R states

The total number of interactions (salt bridge, hydrogen bond, water mediated hydrogen bond, van der Waals contacts; for definition cf. Table 3.9) between TM domains at the extracellular (1) and the intracellular (2) part of hH₂R_i (A), hH₂R_{a1} (B) and hH₂R_{a2} (C). The border between the extra- and intracellular side is the middle between the planes shown in Figure 3.10.

Moreover, since the TM5-TM6 interface changes upon the outward move of TM6, none of the 8 hydrophobic contacts in hH₂R_i is retained in hH₂R_{a1} and only one in hH₂R_{a2}. However, some of the amino acids connecting TM5 and TM6 in the hH₂R_i model form new contacts to the other TM. Ile219^{5.75}, Lys223^{6.24}, Ala237^{6.38} and Ile244^{6.45} only contribute to the interface in the hH₂R_i. In the active state models these amino acids do not point towards TM5 but face the lipid bilayer or the solvent (Figure 3.15 B). Phe206^{5.62}, Arg210^{5.66}, Ala225^{6.26}, Glu229^{6.30}, Leu236^{6.37}, Val239^{6.40} and Phe243^{6.44} are part of the TM5-TM6 interface only in both active state conformations. In the inactive state they are displayed towards the plasma membrane or the cytoplasm (amino acids of TM5), or towards the core of the receptor (amino acids of TM6).

3.3.3.4 Contacts of extra- and intracellular loops

In the active state models the backbones of ICL1, which fit very well among each other (Figure 3.8 C), are placed away from TM7 as well as H8 and located more outwards than in hH₂R_i (Figure 3.9 D). This allows the inward motion of the cytoplasmic part of TM7 and helix 8, also observed in opsin and the activated β_2 AR, followed by the rearrangement around the NPxxY(x)_{5,6}F motif (Cherezov *et al.*, 2007; Okada *et al.*, 2004; Rasmussen *et al.*, 2011b; Scheerer *et al.*, 2008). Interactions are formed with TM1, TM2 and H8, but they do not significantly differ between the inactive and active hH₂R states. This is also the case for ECL1, whose shape is very similar in the three models (Figure 3.9 A). Contacts are mainly formed with the adjacent TMs 2 and 3, as well as with the backbone of ECL2 (Figure 3.9 A). For the hH₂R_i and hH₂R_{a1} models, ICL2 and its neighbors TM3 and TM4 were taken from the same template, the t β_1 AR. Therefore, the backbone geometry, the position of the side chains and their interactions are almost identical (Figure 3.8 A). ICL2 of the hH₂R_{a2} model is slightly shifted due to the more inward position of TM3 and the more outward position of TM4 at their cytoplasmic ends, respectively (Figure 3.9 D). Besides intra-loop interactions, contacts are formed mainly with terminal residues of TM3 and TM4. Additionally two hydrophobic interactions with Leu52^{2.38} are obvious. In the active states, side chains in ICL2 contact residues of G α CT (section 3.3.4.2). ECL2 of the hH₂R_i and hH₂R_{a1} models was constructed in the same way (section 3.2.2). Therefore, backbones and side chains overlap well in both models. Despite using another template for ECL2 in the hH₂R_{a2} model, the deviation from the other two models is only marginal (Figure 3.9 A). Merely around the solvent exposed Glu163, ECL2 is positioned closer to ECL3 resulting in higher RMSD values for Asn162 and Glu163 (Figure 3.8 B, C). Interactions of ECL2 are similar in all hH₂R models, and the loops are mainly stabilized by intra-loop contacts. Contacts are formed with ECL1, ECL3 (hH₂R_i and hH₂R_{a1}), as well as with the termini of TMs 4 and 5. Furthermore, the close distance to TM3

enables in all three models a highly conserved disulfide bridge of Cys174 with Cys91^{3.25}, a hydrogen bond of Asn168 with Lys88^{3.22} and hydrophobic interactions of Thr171 with Gly87^{3.21} and Lys88^{3.22}, Cys174 with Tyr94^{3.28}, and Val176 with Thr95^{3.29}. In the hH₂R_{a1} model, Val176 contacts histamine (section 3.3.4.1). Since the α -helices of TM5 and TM6 extend far at their cytoplasmic ends, interactions of the modeled parts of ICL3 are rare. Only some contacts to TM5 and TM6 as well as to G α CT in the hH₂R_{a2} were detected (section 3.3.4.2). Like ECL2, ECL3 of the hH₂R_i and hH₂R_{a1} models was constructed in the same way, resulting in a similar course in both models (Figure 3.9 A) as well as in identical interactions to nearby domains. Since ECL3 of hH₂R_{a2} was generated by another loop search, the backbone and the side chains deviate from the other two models (Figure 3.8 B, C).

3.3.3.5 Interactions of helix 8

The region surrounding helix 8 is characterized by changes in the active state models, namely the inward movement of TM7 and H8 itself, and the outward shifts of TM1 and TM2 compared to the inactive state (Figure 3.9 D). Despite these changes, the interaction of H8 with the cytoplasmic part of TM1 is similar in the three models, consisting of about 10 hydrophobic interactions. Furthermore, ICL1 is connected to H8 in the same way in all models (Arg48 to Asp294, and Leu49 to Asp294 and Phe295). Besides TM1 and ICL1, also TM7 stabilizes H8 in its position by interactions of the aromatic side chains of Phe295 and Tyr299 with TM7. The contact of Tyr288^{7.53} to Phe295 and Tyr299 observed in hH₂R_i is broken in both active state models due to the rearrangement of the cytoplasmic part of TM7 (section 3.3.3.1). Instead of that, Phe295 interacts with Ala289^{7.54}, Ile57^{2.43} and Asn54^{2.40} (the latter only in hH₂R_{a1}). A hydrogen bond of the side chain of Arg296 to the backbone oxygens of Ala290^{7.55} and Leu291^{7.56} is present in the hH₂R_i and the hH₂R_{a1} models. This is also observed for the respective amino acids in the crystal structures of the t β ₁AR and opsin. In the hH₂R_{a2} model and its template, the h β ₂AR, this contact is missing, but Asp294 is hydrogen bonded to Asn292 at the junction of TM7 and H8. In the hH₂R_i model Asn292 is hydrogen bonded to Thr235^{6.36}, possibly contributing to the inward position of TM6 in the inactive state.

3.3.4 Ligand and GαCT interactions with the active hH₂R states

3.3.4.1 Histamine binding

The interactions of the endogenous agonist histamine with the orthosteric binding site of the hH₂R are similar in both active state models (Figure 3.17). Salt bridges and/or charge-assisted hydrogen bonds are formed with Asp98^{3.32} and Asp186^{5.42}. This docking is in agreement with mutational data received at the canine H₂R (Gantz *et al.*, 1992). Van der Waals contacts occur with Val99^{3.33}, Tyr250^{6.51} and Phe254^{6.55} in both models. These amino acids were reported to contribute to ligand binding in several GPCRs (Cherezov *et al.*, 2007; Shi and Javitch, 2002; Warne *et al.*, 2008b; Wieland *et al.*, 1999; Wieland *et al.*, 1996). Additionally, Tyr182^{5.38} is in close contact to the ligand, however, without being hydrogen bonded to the imidazole moiety of histamine as supposed formerly (Nederkoorn *et al.*, 1996a; Nederkoorn *et al.*, 1996b). To test this hypothesis a Tyr182Phe hH₂R mutant was generated and pharmacologically characterized (chapter 6).

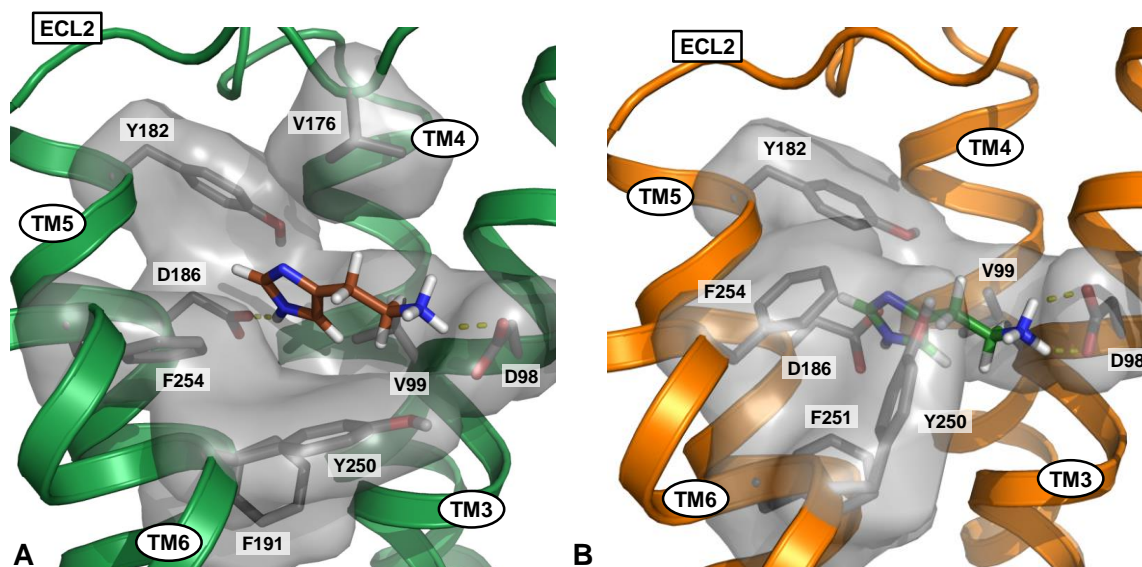


Figure 3.17: Histamine binding in the orthosteric binding site of the hH₂R

Side view of the hH₂R_{a1} (A) and the hH₂R_{a2} (B) model in complex with histamine. The orthosteric binding site is shown from TM7. ECL3 and TM7 are omitted for reasons of clarity. Histamine (carbon atoms in brown (A) and green (B), respectively) was manually docked into the putative binding pocket. Amino acid side chains in a sphere of 5 Å radius around histamine are displayed in sticks and are surrounded by a grey surface. Hydrogen bonds are indicated with yellow dashed lines.

In the hH₂R_{a2} model Phe251^{6.52} contacts histamine at the same position as Phe191^{5.47} in the hH₂R_{a1} model. In hH₂R_{a2} the ligand is located somewhat deeper in the binding pocket due to the more downward position of Tyr250^{6.51} and Phe254^{6.55}. Compared to hH₂R_{a1}, TM5 to TM7 are slightly shifted to the cytoplasm, whereas TM3 is moved to the extracellular side (Figure 3.9). Therefore, only in the hH₂R_{a1} model the ligand is able to contact Val176 in ECL2 at the top of the binding pocket. A further remarkable difference between both active state models is the position of TM5. The distance calculated between the corresponding C_α atoms of TM5 (Thr190^{5.46}) and TM3 (Thr103^{3.37}), TM6 (Trp247^{6.48}) and TM7 (Gly277^{7.42}) is decreased in the hH₂R_{a2} model by 0.52 Å, 0.47 Å, and 1.18 Å, respectively. Accordingly, in the template for the hH₂R_{a2} model, the hβ₂AR, a similar compression is observed (Rasmussen *et al.*, 2011a).

3.3.4.2 Receptor – G-Protein interaction

After superimposing the active state models, the docked C-terminal part of the G_{sa}-protein (GαCT) is apparently closer to ICL3 and TM6 in the hH₂R_{a2} model (Figure 3.18 A, B). This is due to a different docking procedure of GαCT in both models. Since the cytoplasmic parts of TM5 and TM6 of hH₂R_{a2} were more protruding and outwards shifted than in hH₂R_{a1} (Figure 3.9 D), GαCT was docked closer to ICL3 and TM6 in the hH₂R_{a2} model in order to enable adequate interactions with this region of the receptor. However, the C-terminal ends of GαCT fit well in both models. Several van der Waals contacts of GαCT are formed with the crevice between TM2, TM3, ICL2 and the junction TM7-H8. Additionally, the helices of TM5 and TM6, which both expand farthest into the cytoplasm, contact the G_{sa}-subunit. Most of the interactions are also present in the crystal structure of opsin in its G-protein-interacting conformation and in the structure of the β₂AR-G_s complex (Table 3.10). In general, more interactions are formed in the hH₂R_{a1} model (Figure 3.18 C, D).

Table 3.10: Van der Waals contacts between GαCT and active state hH₂R models

domain	hH ₂ R _{a1}	hH ₂ R _{a2}
TM2	2 (T53 ¹ , N54)	0
TM3	6 (D115, R116 ^{1,2} , A119 ^{1,2} , V120 ^{1,2})	6 (R116, A119, V120)
ICL2	4 (P123 ² , Y126 ²)	1 (P123)
TM5	3 (I205 ^{1,2} , Q212 ^{1,2} , I219)	1 (I216 ²)
TM6	5 (R228 ¹ , A232 ^{1,2} , T235 ² , L236 ²)	4 (A232, T235, V239)
TM7/H8	1 (N292 ¹)	1 (N292)
total	21	13

^{1,2} Hydrophobic interactions are also observed for the respective amino acid in the crystal structure of opsin (¹; Scheerer *et al.*, 2008) and in the β₂AR (²; Rasmussen *et al.*, 2011b), respectively.

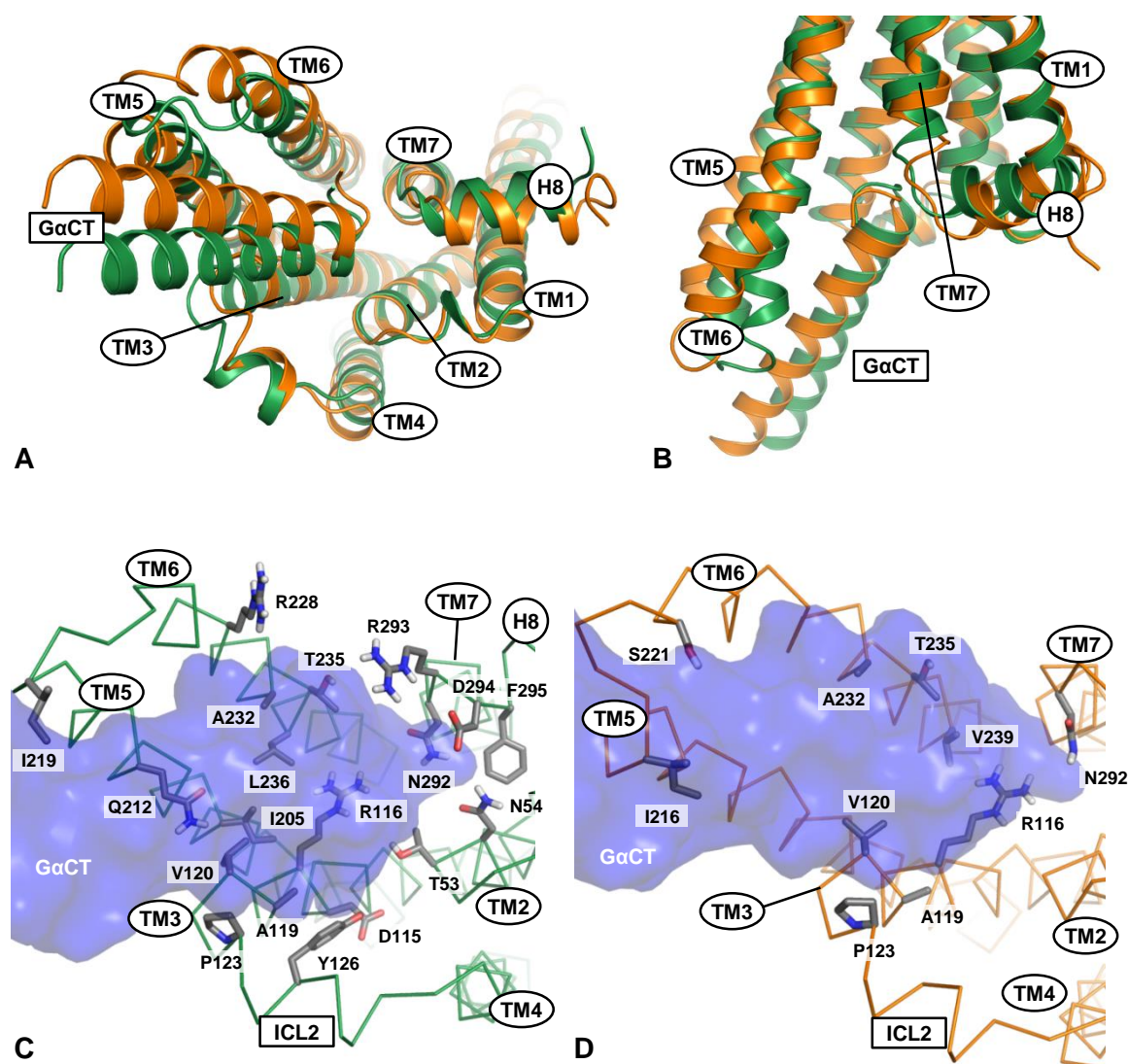


Figure 3.18: Interactions of the cytoplasmic part of the hH₂R with the G_{sa}-protein

A, B, Side view (A) and cytoplasmic view (B) of the interaction of the C-terminus of the G_{sa}-protein (GaCT) with the (aligned) hH₂R_{a1} (■) and hH₂R_{a2} (■) models, shown in ribbons. C, D, Cytoplasmic view on amino acid side chains (shown in sticks) of hH₂R_{a1} (C, C_α trace in green) and hH₂R_{a2} (D, C_α trace in orange) which interact with GaCT (transparent blue surface). Docking of GaCT was performed as described in section 3.2.2

Apart from these hydrophobic interactions some polar contacts occur. In the hH₂R_{a1} model a hydrogen bond between Asn54^{2,40} and Glu392 (GaCT) and a salt bridge between Arg293 in H8 and the charged C-terminus of GaCT, Leu394, are observed. In the hH₂R_{a2} model hydrogen bonds between Ser221^{ICL3} and Asp381 (GaCT), Asn292 (TM7-H8) and Glu392 (GaCT), as well as an ionic interaction between Arg116^{3,50} and Glu392 (GaCT) are formed.

3.4 Summary and conclusion

3.4.1 Template selection and quality of the models

Homology models of the hH₂R in its inactive and active state were generated. At the time of creation of the inactive state model only bovine and squid rhodopsin, the t β ₁AR, the h β ₂AR and the adenosine A_{2A} receptor were available as templates. However, the retrospective analysis of all GPCR structures resolved till the end of 2012 confirmed that using the t β ₁AR (Warne *et al.*, 2008a) was most appropriate. A phylogenetic analysis of class 1 aminergic GPCRs indicated that the H₂R is closest related to the 5-HT₄R and the β ARs, whereas the H₁R, of which a crystal structure is available (Shimamura *et al.*, 2011), is distant from the H₂R and more related to muscarinic acetylcholine receptors (Vassilatis *et al.*, 2003). For the active hH₂R state, two models were generated. At first the crystal structure of the t β ₁AR was used in a multiple template approach, inserting the structural changes crucial for receptor activation by adaption to the structure of opsin (Scheerer *et al.*, 2008). This was attained by the partly adaption of the intracellular segments of the TMs of hH₂R_{a1} to opsin in order to mimic the shift of the helices observed for the transition from rhodopsin to opsin, including opening of the cytoplasmic crevice enabling interaction with the C-terminal part of the G-protein. Therefore, most of the TMs in hH₂R_{a1} could be taken from the t β ₁AR. However, TM7 of opsin shows a different helical turn compared to the inactive conformation of rhodopsin (Okada *et al.*, 2004). In the active state, Tyr306^{7.53} is directed towards the core of the receptor, keeping TM6 away from moving inward. This is the case in the h β ₂AR, too. The corresponding Tyr326^{7.53} moves to the center of the receptor upon activation and keeps TM6 in the outward position (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011a). In order to preserve this structural feature, TM7 and helix 8 were directly taken from opsin. Taking only the latter as whole template for the hH₂R seemed inappropriate since in a previous MD simulation of the hH₂R based on the structure of rhodopsin distance restraints had to be applied to maintain α -helical elements of the receptor (Preuss, 2007). For the second model the structure of the nanobody-stabilized active state of the h β ₂AR was used (Rasmussen *et al.*, 2011a), because of its close relatedness to the hH₂R and the exclusion of remaining active state GPCR crystal structures, such as the adenosine A_{2A} receptor (Lebon *et al.*, 2011; Xu *et al.*, 2011). The stereochemical quality of the hH₂R homology models was confirmed by analyzing the planarity of the peptide bond, aromatic side chains and delocalized π -electron systems in amino acid side chains. The stability of naturally planar moieties during energy minimization is a good measure for the integrity of the receptor with proper side chain conformations. Generally, the deviation from planarity was only marginal. The analysis of backbone dihedral angles (Φ and Ψ) as well as side chain torsions of the models showed a

good agreement with experimental values (Lovell *et al.*, 2000). The majority of residues with dihedral angles located in disallowed regions of the conformational space were obtained from the templates or from protein fragments of the Protein Data Bank received by a loop search. The good stereochemical quality of the models was a prerequisite for the analysis and comparison of hH₂R states.

3.4.2 Comparison of hH₂R states

The overall topology (i.e. the relative positions of TM domains) of both active state models was very similar. Differences between the active hH₂R states occurred at the bottom of the binding site which linked the binding pocket to the rearrangement of TM6. In the hH₂R_{a1} the activation mechanism resembled that of rhodopsin/opsin, whereas in hH₂R_{a2} it was similar to the β_2 AR. The rigid body fitting of t β_1 AR-TM6 to TM6 of opsin in hH₂R_{a1} resulted in close contacts especially between TM5 and TM6 in the middle of the receptor which required the adjustment of side chain torsions. Furthermore, due to the close relatedness of the hH₂R with the β ARs, the hH₂R_{a2} seemed to be a more suitable model for the active hH₂R state. In comparison to the inactive state, the crucial rearrangements especially of the cytoplasmic part of TM6 and TM7 were in accordance with changes observed for the transition of rhodopsin in the ground state to activated opsin (Okada *et al.*, 2004; Scheerer *et al.*, 2008) and of the inactive to the active h β_2 AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011a). These major changes were accompanied by altered distances of nearby TM domains, mainly in the intracellular part of the receptor, and by a different number and kind of interactions between them. Most of the differences of inter-TM contacts were caused by conformational changes around highly conserved motifs. These molecular switches, connected to each other, are suggested to be essential for receptor activation and present in at least six regions of the hH₂R models. A modification of the TM5 position and of side chain orientations, mainly in TM3 and TM6, linked the ligand binding site to the critical outward movement of TM6. A concomitant rearrangement of amino acids contributing to the ionic lock in some inactive state structures contributed to stabilization of the active hH₂R state and to the binding of the G-protein fragment. Further changes during hH₂R activation are: reordering of the water mediated hydrogen bond network around TM7, rearrangement of the highly conserved NPxxY(x)_{5,6}F motif, disruption of a hydrophobic barrier in the center of the TM bundle and an altered TM5-TM6 interface. Most of these changes correspond to those resulting from comparison of inactive and active states of resolved GPCR structures.

3.4.3 Conclusion

The homology models of the hH₂R enabled valuable insights into the differences between inactive and active hH₂R states. However, molecular dynamics simulations are necessary to further validate the inactive and active hH₂R state models. In particular, their relatively stable existences in a natural, dynamic environment and the impact of conformational flexibility on the process of receptor activation have to be analyzed.

3.5 References

- Alewijnse AE, Timmerman H, Jacobs EH, Smit MJ, Roovers E, Cotecchia S and Leurs R. The effect of mutations in the DRY motif on the constitutive activity and structural instability of the histamine H(2) receptor. *Mol Pharmacol* **2000**, 57, 890-898.
- Angel TE, Chance MR and Palczewski K. Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc Natl Acad Sci U S A* **2009**, 106, 8555-8560.
- Angelova K, Fanelli F and Puett D. A model for constitutive lutropin receptor activation based on molecular simulation and engineered mutations in transmembrane helices 6 and 7. *J Biol Chem* **2002**, 277, 32202-32213.
- Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, Gether U and Javitch JA. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* **2001**, 276, 29171-29177.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T and Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **1977**, 112, 535-542.
- CBN. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. *Eur J Biochem* **1970**, 17, 193-201.
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK and Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, 318, 1258-1265.
- Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V and Stevens RC. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **2010**, 330, 1091-1095.
- Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P and Ernst OP. Crystal structure of metarhodopsin II. *Nature* **2011**, 471, 651-655.
- Consortium TU. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **2012**, 40, D71-D75.
- Costanzi S. Homology modeling of class a G protein-coupled receptors. *Methods Mol Biol* **2012**, 857, 259-279.
- Deupi X, Edwards P, Singhal A, Nickle B, Oprian D, Schertler G and Standfuss J. Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc Natl Acad Sci U S A* **2012a**, 109, 119-124.
- Deupi X and Standfuss J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* **2011**, 21, 541-551.
- Deupi X, Standfuss J and Schertler G. Conserved activation pathways in G-protein-coupled receptors. *Biochem Soc Trans* **2012b**, 40, 383-388.
- Dore AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, et al. Structure of the adenosine A(2A)

- receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **2011**, 19, 1283-1293.
- Dror RO, Arlow DH, Borhani DW, Jensen MO, Piana S and Shaw DE. Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc Natl Acad Sci U S A* **2009**, 106, 4689-4694.
- Edison AS. Linus Pauling and the planar peptide bond. *Nat Struct Biol* **2001**, 8, 201-202.
- Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP and Ernst OP. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. *Proc Natl Acad Sci U S A* **2003**, 100, 2290-2295.
- Gantz I, DelValle J, Wang LD, Tashiro T, Munzert G, Guo YJ, Konda Y and Yamada T. Molecular basis for the interaction of histamine with the histamine H2 receptor. *J Biol Chem* **1992**, 267, 20840-20843.
- Ghorai P, Kraus A, Keller M, Götte C, Igel P, Schneider E, Schnell D, Bernhardt Gn, Dove S, Zabel M, Elz S, Seifert R, et al. Acylguanidines as Bioisosteres of Guanidines: NG-Acylated Imidazolylpropylguanidines, a New Class of Histamine H2 Receptor Agonists. *J Med Chem* **2008**, 51, 7193-7204.
- Gonnet G, Cohen M and Benner S. Exhaustive matching of the entire protein sequence database. *Science* **1992**, 256, 1443-1445.
- Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J and Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **2010**, 38, 3.
- Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI and Kobilka BK. Structure of the delta-opioid receptor bound to naltrindole. *Nature* **2012**, 485, 400-404.
- Greasley PJ, Fanelli F, Rossier O, Abuin L and Cotecchia S. Mutagenesis and modelling of the alpha(1b)-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol Pharmacol* **2002**, 61, 1025-1032.
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T and Kobayashi T. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012**, 482, 547-551.
- Hall SE, Roberts K and Vaidehi N. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. *J Mol Graph Model* **2009**, 27, 944-950.
- Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, et al. Crystal structure of a lipid G protein-coupled receptor. *Science* **2012**, 335, 851-855.
- Huang YH and Chen CM. Statistical analyses and computational prediction of helical kinks in membrane proteins. *J Comput Aided Mol Des* **2012**, 26, 1171-1185.
- Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP and Stevens RC. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, 322, 1211-1217.
- Kabsch W and Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577-2637.
- Katritch V, Cherezov V and Stevens RC. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2012**, 8, 8.
- Kelley MT, Burckstummer T, Wenzel-Seifert K, Dove S, Buschauer A and Seifert R. Distinct interaction of human and guinea pig histamine H-2-receptor with guanidine-type agonists. *Mol Pharmacol* **2001**, 60, 1210-1225.
- Kim DC, Choi SY, Kim SH, Yun BS, Yoo ID, Reddy NR, Yoon HS and Kim KT. Isoliquiritigenin selectively inhibits H(2) histamine receptor signaling. *Mol Pharmacol* **2006**, 70, 493-500.
- Kim JM, Altenbach C, Kono M, Oprian DD, Hubbell WL and Khorana HG. Structural origins of constitutive activation in rhodopsin: Role of the K296/E113 salt bridge. *Proc Natl Acad Sci U S A* **2004**, 101, 12508-12513.
- Kobilka BK and Deupi X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol Sci* **2007**, 28, 397-406.

- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, et al. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **2012**, 482, 552-556.
- Langelaan DN, Wieczorek M, Blouin C and Rainey JK. Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. *J Chem Inf Model* **2010**, 50, 2213-2220.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, 23, 2947-2948.
- Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **1993**, 26, 283-291.
- Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AG and Tate CG. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, 474, 521-525.
- Lefkowitz RJ, Cotecchia S, Samama P and Costa T. Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol Sci* **1993**, 14, 303-307.
- Li J, Edwards PC, Burghammer M, Villa C and Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **2004**, 343, 1409-1438.
- Lovell SC, Word JM, Richardson JS and Richardson DC. The penultimate rotamer library. *Proteins* **2000**, 40, 389-408.
- MacArthur MW and Thornton JM. Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* **1996**, 264, 1180-1195.
- Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK and Granier S. Crystal structure of the micro-opioid receptor bound to a morphinan antagonist. *Nature* **2012**, 485, 321-326.
- Mobarec JC, Sanchez R and Filizola M. Modern homology modeling of G-protein coupled receptors: which structural template to use? *J Med Chem* **2009**, 52, 5207-5216.
- Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, 12, 345-364.
- Murakami M and Kouyama T. Crystal structure of squid rhodopsin. *Nature* **2008**, 453, 363-367.
- Nederkoorn PH, van Gelder EM, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. II. Theoretical investigations of histamine binding to receptor models of the seven alpha-helical transmembrane domain. *J Comput Aided Mol Des* **1996a**, 10, 479-489.
- Nederkoorn PH, van Lenthe JH, van der Goot H, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. I. Theoretical investigations of histamine binding to an oligopeptide mimicking a part of the fifth transmembrane alpha-helix. *J Comput Aided Mol Des* **1996b**, 10, 461-478.
- Okada T, Sugihara M, Bondar AN, Elstner M, Entel P and Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* **2004**, 342, 571-583.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M and Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, 289, 739-745.
- Park JH, Scheerer P, Hofmann KP, Choe HW and Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **2008**, 454, 183-187.
- Preuss H. Species-selective Interactions of Histamine H₂ Receptors with Guanidine-type Agonists: Molecular Modelling, Site-directed Mutagenesis and Pharmacological Analysis. Doctoral thesis, Regensburg, Regensburg, **2007**. <http://epub.uni-regensburg.de/10580/>

- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011a**, 469, 175-180.
- Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **2007**, 450, 383-387.
- Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **2011b**, 477, 549-555.
- Rossi KA, Weigelt CA, Nayeem A and Krystek SR, Jr. Loopholes and missing links in protein modeling. *Protein Sci* **2007**, 16, 1999-2012.
- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP and Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, 455, 497-U430.
- Schneider EH, Schnell D, Strasser A, Dove S and Seifert R. Impact of the DRY motif and the missing "ionic lock" on constitutive activity and G-protein coupling of the human histamine H4 receptor. *J Pharmacol Exp Ther* **2010**, 333, 382-392.
- Shapiro DA, Kristiansen K, Weiner DM, Kroeze WK and Roth BL. Evidence for a model of agonist-induced activation of 5-hydroxytryptamine 2A serotonin receptors that involves the disruption of a strong ionic interaction between helices 3 and 6. *J Biol Chem* **2002**, 277, 11441-11449.
- Shi L and Javitch JA. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu Rev Pharmacol Toxicol* **2002**, 42, 437-467.
- Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA and Javitch JA. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J Biol Chem* **2002**, 277, 40989-40996.
- Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, et al. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **2011**, 475, 65-70.
- Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD and Schertler GF. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **2011**, 471, 656-660.
- Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, Huang XP, Trapella C, Guerrini R, Calo G, Roth BL, Cherezov V, et al. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **2012**, 485, 395-399.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **1997**, 25, 4876-4882.
- Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A and Filipek S. Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr Med Chem* **2012**, 19, 1090-1109.
- Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE and Gaitanaris GA. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci U S A* **2003**, 100, 4903-4908.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG and Schertler GF. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **2008a**, 454, 486-491.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG and Schertler GF. Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* **2008b**, 454, 486-U482.

- Wieland K, Laak AM, Smit MJ, Kuhne R, Timmerman H and Leurs R. Mutational analysis of the antagonist-binding site of the histamine H(1) receptor. *J Biol Chem* **1999**, 274, 29994-30000.
- Wieland K, Zuurmond HM, Krasel C, Ijzerman AP and Lohse MJ. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc Natl Acad Sci U S A* **1996**, 93, 9276-9281.
- Worth CL, Kleinau G and Krause G. Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One* **2009**, 4.
- Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, 330, 1066-1071.
- Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI, Mascarella SW, Westkaemper RB, et al. Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* **2012**, 485, 327-332.
- Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V and Stevens RC. Structure of an agonist-bound human A2A adenosine receptor. *Science* **2011**, 332, 322-327.
- Yarnitzky T, Levit A and Niv MY. Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. *Curr Opin Drug Discov Devel* **2010**, 13, 317-325.
- Yohannan S, Faham S, Yang D, Whitelegge JP and Bowie JU. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **2004**, 101, 959-963.
- Zhang J, Qi T and Wei J. Homology modeling and antagonist binding site study of the human histamine H2 receptor. *Med Chem* **2012**, 8, 1084-1092.

Chapter 4

Molecular Dynamics Simulations of Inactive and Active Human Histamine H₂ Receptor States

4.1 Introduction

Advances in crystallization techniques enabled the resolution of several 3D structures of G-protein coupled receptors in the last years (Venkatakrishnan *et al.*, 2013). In TM domains the sequences as well as the structures fit quite well between these receptors (Figure 3.1). In the terminal segments of TMs and especially in the extra- and intracellular loops, sequence identity is reduced. Accordingly, backbone courses of crystallized GPCRs deviate most in these regions, resulting in characteristic structures of each receptor. Only GPCRs with high sequence identity adopt nearly identical structures, e.g. the t β_1 AR and the h β_2 AR with an overall and TM sequence identity of about 54% and 70%, respectively, and a backbone RMSD of 1.4 Å (Cherezov *et al.*, 2007; Warne *et al.*, 2008). Since GPCRs with rather low sequence similarities possess different 3D structures, homology models based on templates with a moderate sequence identity – such as models of the H₂R derived from the β -adrenoceptors (TM sequence identity of 40% to 45%; cf. chapter 3) – are too close to the template from which the TM domains are mostly directly taken. Although the technique of energy minimization is able to relax the protein structure and reduce unfavorable contacts, it is not able to find alternative conformations of amino acid side chains and TM positions. Furthermore, artificial restraints from, e.g., lattice contacts or metal ions present in crystal structures are directly transferred into homology models. Also the protein environment composed of detergents during the crystallization is of impact on the conformation, and structures obtained with x-ray crystallography are of low energy and adopt stable geometries (Deupi and Kobilka, 2010). Thus, it is important to embed the protein into a natural environment, to allow the protein to relax and to enable the adoption of different backbone and side chain conformations. In molecular dynamics (MD) simulations atoms or groups of atoms are able to interact with each other based on forces and potential energies obtained from a molecular mechanics force field. Solving Newton's equations of motion, atoms are capable to move within a distinct time step (Van Der Spoel *et al.*, 2005). MD simulation is an

established computational method to investigate the conformational dynamics of membrane proteins and has been successfully used for studying the structure, activation and dimerization of GPCRs (Taddese *et al.*, 2012; Vanni and Rothlisberger, 2012).

In the present study the putative inactive and active receptor states of the human H₂R (hH₂R_i and hH₂R_{a2}; cf. chapter 3) were embedded in 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) bilayers (Figure 4.1) to simulate a natural environment for the membrane proteins, and subsequently subjected to MD simulations, consisting of 20 ns equilibration phases and 80 ns production runs. The systems were hydrated with water molecules, and counter ions were added to attain a final system charge of zero. The purpose of this study was to characterize the structure and function of the hH₂R in an inactive and active state and to check and upgrade the results obtained from analyzing hH₂R homology models (chapter 3).

4.2 Materials and methods

4.2.1 Materials

MD simulations were performed with the software package GROMACS 4.0.7 (GRONingen MAchine for Chemical Simulations; Hess *et al.*, 2008; Van Der Spoel *et al.*, 2005). The pre-equilibrated lipid bilayer consisting of 128 1,2-dipalmitoyl-sn-glycero-3-phosphocholine molecules (Figure 4.1) including the respective parameters of the GROMOS96 53a6 force field was obtained from Kukol (2009). Visual inspections of the simulation systems were performed with VMD (Visual Molecular Dynamics; Humphrey *et al.*, 1996). For structure validation of the proteins the program described in chapter 5 was used (*gro_validation*). Analysis of the simulation runs was performed with tools included in GROMACS or with the programs *gro_hbonds* and *gro_contacts* described in chapter 5 for the calculation of hydrogen bonds and van der Waals contacts, respectively. Interactions between residues were assigned by using cut-off values for ionic interactions (distance of charged heteroatoms below 5 Å), hydrogen bonds (distance of heteroatoms below 3.5 Å, angle of heteroatom 1, hydrogen and heteroatom 2 at least 120°) and van der Waals interactions (distance of hydrophobic heavy atoms below 5 Å). The area per lipid was calculated with the program *InflateGRO* described by Kandt *et al.* (2007) and provided at the homepage of the Life and Medical Science Center of the University of Bonn, Germany.

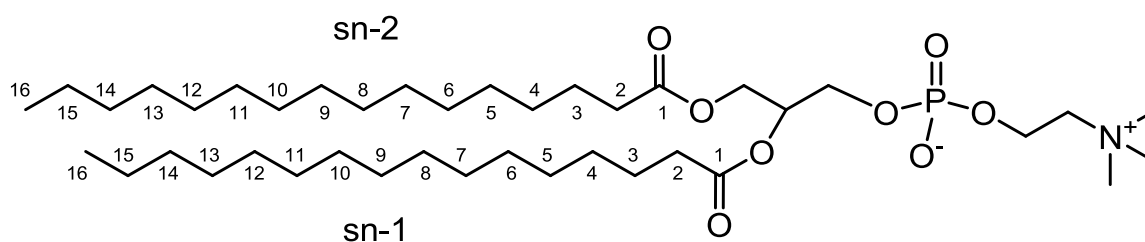


Figure 4.1: Structure of DPPC and numbering of the acyl chains sn-1 and sn-2

4.2.2 Parameters of the MD simulations

All MD simulations were performed applying the GROMOS96 53a6 force field (Oostenbrink *et al.*, 2004). In order to allow a larger time step of 2 fs the LINCS (LINEar Constraint Solver) algorithm (Hess, 2007; Hess *et al.*, 1997) with constraints on all bonds was used. Cut-off values for the generation of the short-range neighbor list for calculating van der Waals interactions, the Lennard-Jones interactions, and the long-range electrostatic interactions were set to 1.4 nm. The latter were calculated with the Particle Mesh Ewald method (Darden *et al.*, 1993; Essmann *et al.*, 1995). The neighbor list was updated every 10th step of the calculation, and the frequency to write coordinates, velocities, forces and energies to output files was set to 500 steps (corresponding to 1 ps). Periodic boundary conditions were applied in all directions of the simulation box. Deviations of the system temperature (323 K) were corrected using the Berendsen algorithm (Berendsen, 1991) with a coupling constant of 0.1 ps. To enable the NPT ensemble, i.e. constant number of particles (N), conserved system pressure (P) and temperature (T), the Berendsen barostat (Berendsen *et al.*, 1984) was used with a semi-isotropic pressure coupling in the x-y plane parallel to the membrane and the z-dimension. A complete list of parameters for MD simulations is provided in the Appendix (section 9.1).

4.2.3 Construction of the hH₂R-DPPC-water systems

The energy optimized homology models hH₂R_i and hH₂R_{a2} (refer to chapter 3 for a description) were transferred to the GROMACS file format (*gro*), including the generation of an appropriate topology file with parameters of the GROMOS96 53a6 force field, defining atom types, bond angles and lengths, as well as dihedral angles and constraints. The

parameters for the ligand histamine, present in the binding pocket of hh_2R_{a2} , were adapted to the parameters of the GROMOS96 53a6 force field (a listing of all parameters used for histamine is given in the Appendix, section 9.2). The receptor models including ten conserved water molecules, respectively, and in case of hH_2R_{a2} histamine and the C-terminal part of the G_{sas} -subunit (GaCT), were energy minimized in two steps with the steepest descent method. In the first minimization backbone atoms were restrained (force constant $1,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) and in the second one all atoms were able to move freely until the maximum force was below $10 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ or the maximum number of minimization steps was reached (50,000). Then the models were inserted in a lipid bilayer consisting of 128 DPPC molecules (Kukol, 2009). This bilayer was built with standard parameters of the GROMOS96 53a6 force field and was equilibrated by Kukol in a 40 ns MD simulation using the software package GROMACS. The embedding of the protein in the membrane was performed according to a method described by Kandt *et al.* (2007). Accordingly, the lipid molecules of the pre-equilibrated bilayer were scaled in their x and y direction parallel to the membrane with a factor of 4. The protein was inserted in the middle of the widened bilayer, and DPPC molecules overlapping with the receptor were deleted manually, taking care that at the extra- and intracellular side of the membrane the same amount of lipid molecules remained. In iterative steps the bilayer was scaled with a factor of 0.95 and energy-minimized with the protein positions restrained (force constant $100,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$), respectively, until the area per lipid converged to the reference value for a DPPC bilayer at 50°C of about 60 to 64 \AA^2 (Kucerka *et al.*, 2008; Kucerka *et al.*, 2011; Nagle, 1993; Nagle and Tristram-Nagle, 2000b). In the next step water molecules (Simple Point Charge water, SPC) were added to the system in an automated procedure, inserting water molecules into every position of the box so that the distance between any atom of the solute molecules (protein, lipid and ligand) and any atom of the solvent molecule was less than the sum of the van der Waals radii of both atoms (command *genbox*). To ensure that water molecules were not included within the bilayer or the hydrophobic interface between the protein and the lipid tails, carbon atoms (common radius of 1.5 \AA) in the DPPC acyl chains were replaced by fluorine atoms (not present in the system) and a radius of 3.5 \AA was assigned. In order to enable a hydration of the polar head groups of DPPC, the carbonyl carbon atom and the two carbon atoms next to the carbonyl group were not exchanged by fluorine atoms (carbon atoms 1 to 3 according to the numbering in Figure 4.1). After addition of water molecules, fluorine atoms were replaced again by the original carbon atoms. Both systems were checked visually, and water molecules were deleted if appropriate. Finally, Cl^- ions were added to the system to achieve a total charge of zero (command *genion*). The two systems were energy-minimized in four steps with the steepest descend method, applying decreasing

position restraints on the system (first step: protein, histamine in case of the hH₂R_{a2}, conserved water molecules, DPPC; second step: protein, histamine in case of the hH₂R_{a2}, conserved water molecules; third step: protein backbone; fourth step: no restraints). Thereafter, the equilibration of the systems was performed in four steps, each with 5 ns duration. In steps one to three position restraints (force constant 1,000 kJ mol⁻¹ nm⁻²) were imposed (1) on the complete protein and the conserved water molecules, (2) on the protein backbone (C_α, C, N), and (3) on the backbone atoms of the TM domains. In the fourth step all atoms were able to move without restrictions. Subsequently, the hH₂R systems were subjected to 80 ns production runs.

4.3 Results

4.3.1 Size and composition of the simulation systems

The dimension of the simulation box with the hH₂R_i was 7.28 x 7.30 x 10.00 nm and consisted of 39,904 atoms or sites (united atoms), respectively, with 292 residues of the hH₂R, 120 DPPC molecules, ten conserved water molecules, 10,305 water molecules as solvent and 11 chloride ions to attain a total charge of zero (Figure 4.2 A). The system of the hH₂R_{a2} was larger in the z-dimension because of the additional GαCT docked at the cytoplasmic side of the hH₂R_{a2} (7.28 x 7.30 x 11.00 nm; Figure 4.2 B). Accordingly, the box contained 44,800 atoms or sites, respectively, with 292 residues of the hH₂R, 28 residues of GαCT, the ligand histamine, 120 DPPC molecules, ten conserved water molecules, 11,828 water molecules as solvent and 13 chloride ions. The number of positively charged moieties in the hH₂R_{a2} was increased because of the protonated amine function of histamine (Eriks *et al.*, 1993) and the uncharged Asp115^{3,49}, which was proposed for active GPCR states (Dror *et al.*, 2011; Ghanouni *et al.*, 2000; Vogel *et al.*, 2008).

4.3.2 Equilibration of the hH₂R-DPPC-water systems

The analysis of system parameters and the inspection of the lipid bilayer during the equilibration phase are crucial since the quality and suitability of an MD simulation depends considerably on the accurateness of the input structures. Furthermore, comparisons with experimental reference values can prove a bilayer structure to be appropriate.

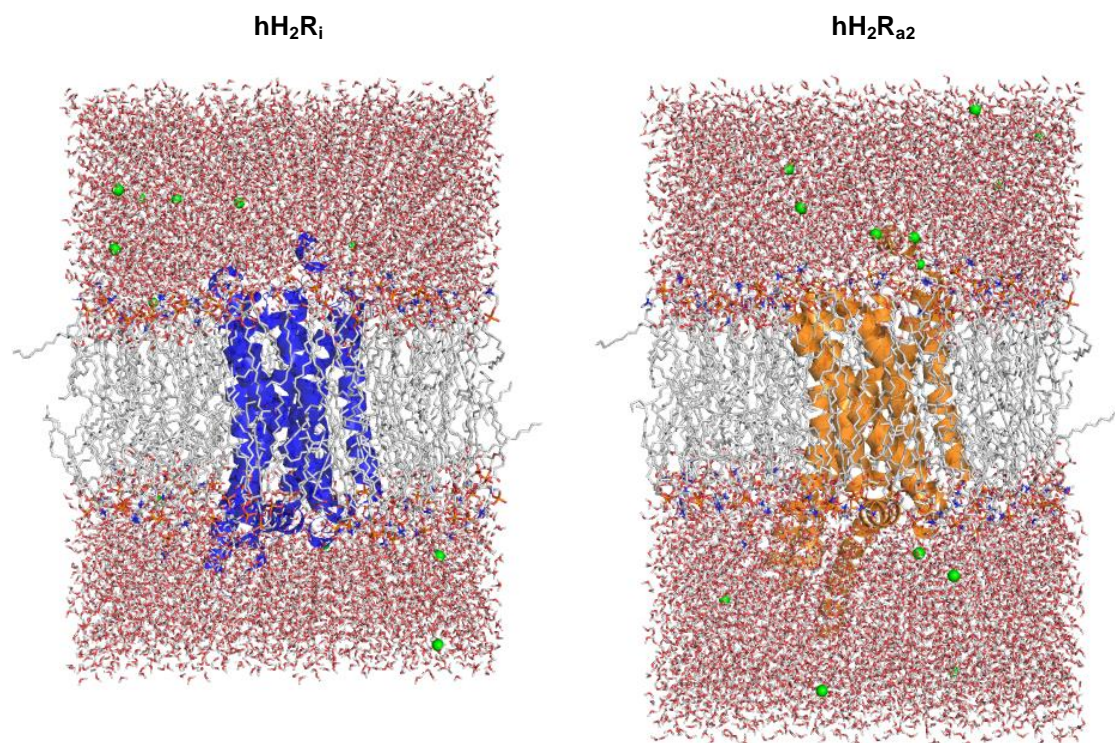


Figure 4.2: Structures of hH₂R models embedded in a solvated DPPC bilayer after 20 ns equilibration

The backbone of the hH₂R_i (blue ribbon and tubes) and hH₂R_{a2} (orange ribbon and tubes) is shown within the DPPC bilayer (carbon atoms in grey sticks). Chloride ions are highlighted in green spheres.

4.3.2.1 System parameters

In order to assure the quality of the simulation system, convergence of parameters such as energy, temperature, pressure, box dimensions and density were checked (Figure 4.3). The total energy of the system is the sum of the potential and the kinetic energy. The latter remained stable during the equilibration (mean \pm SD for hH₂R_i and hH₂R_{a2}: $107 \times 10^3 \pm 381$ kJ mol⁻¹ and $120 \times 10^3 \pm 404$ kJ mol⁻¹) since it is based on the temperature of the system which was restricted by coupling to an external thermostat. The potential energy is the sum of terms related to internal coordinates and non-bonded energy contributions between pairs of atoms and is constant in an equilibrated system. Because of position restraints applied during the equilibration phase it decreased slightly in both systems, mainly at points in time when the constraints were released. However, potential energies nearly converged to a stable value until the end of the respective equilibration period (final values for hH₂R_i and hH₂R_{a2}: -589×10^3 kJ mol⁻¹ and -655×10^3 kJ mol⁻¹). Density remained unchanged in both

systems (mean \pm SD for hH₂R_i and hH₂R_{a2}: $990 \pm 1.8 \text{ kg m}^{-3}$ and $990 \pm 1.7 \text{ kg m}^{-3}$), as well as the pressure which was restricted by the berendsen barostat.

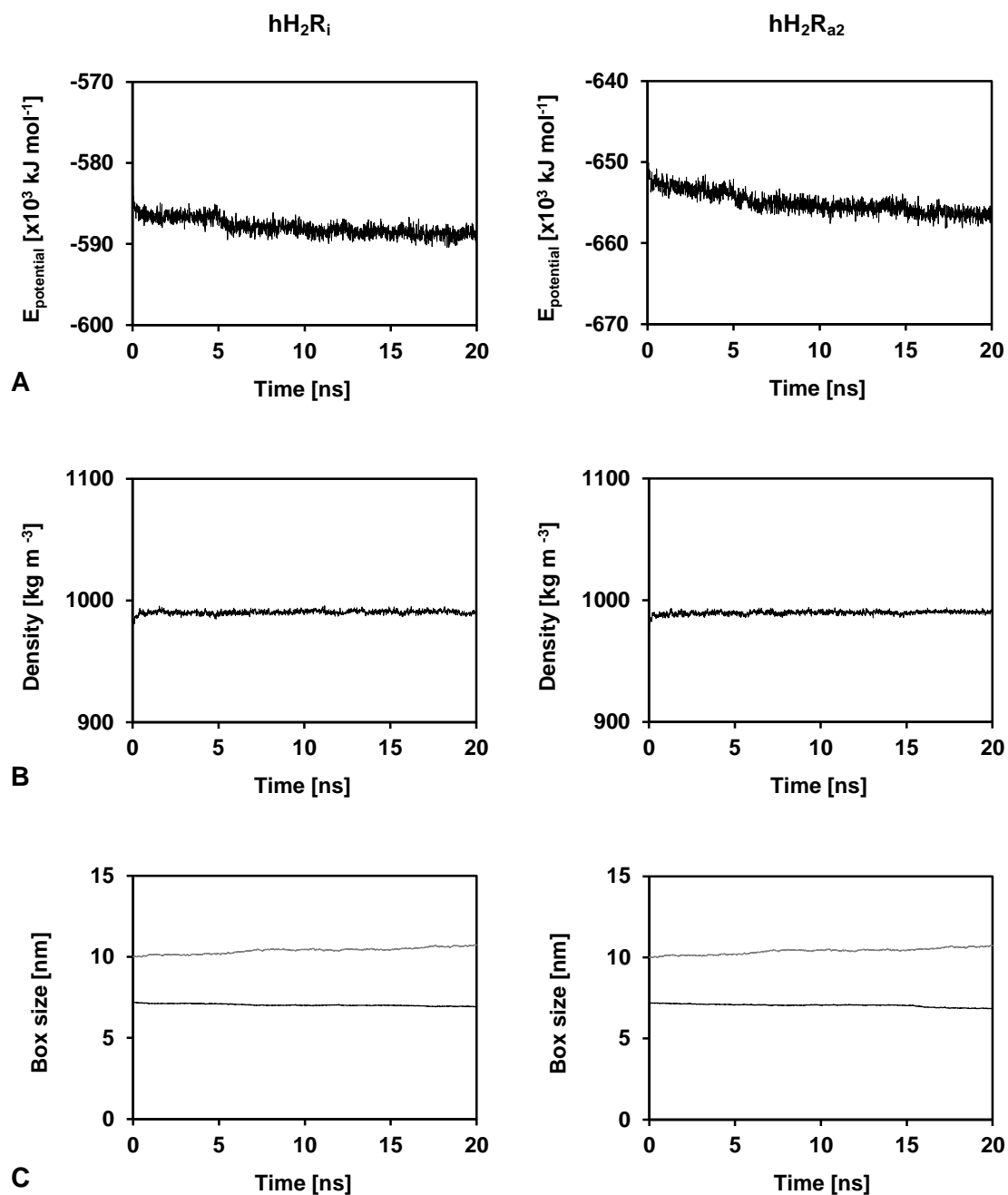


Figure 4.3: System parameters of the 20 ns equilibration phase

Potential energy (A), density (B) and box dimensions (C) in x- (black line) and z-direction (grey line) for the system of the hH₂R_i and hH₂R_{a2}, respectively. The y-vector was similar to the x-vector (semiisotropic pressure coupling) and is not shown. Every 10th frame (10 ps interval) was analyzed.

The box size decreased slightly in the plane parallel to the membrane (x- and y-direction) and increased in the z-dimension, primarily in the hH₂R_{a2} system when all position restraints were released (15 ps). However, the box dimensions stabilized quickly.

To avoid ‘edge effects’ at the border of the simulation box, periodic boundary conditions were applied to the MD simulations, i.e. one box is surrounded by copies of itself in all three dimensions. However, also the periodicity can cause artifacts in the calculation if a molecule in a box and its periodic image in a neighboring box approach each other by less than 1.4 nm. Thus, the distance of the proteins to their nearest periodic images were analyzed. During the 20 ns equilibration phase of the hH₂R_i and the hH₂R_{a2} this distance was above the cut-off value for the calculation of long-range interactions (1.4 nm) at any time, respectively (Figure 4.4).

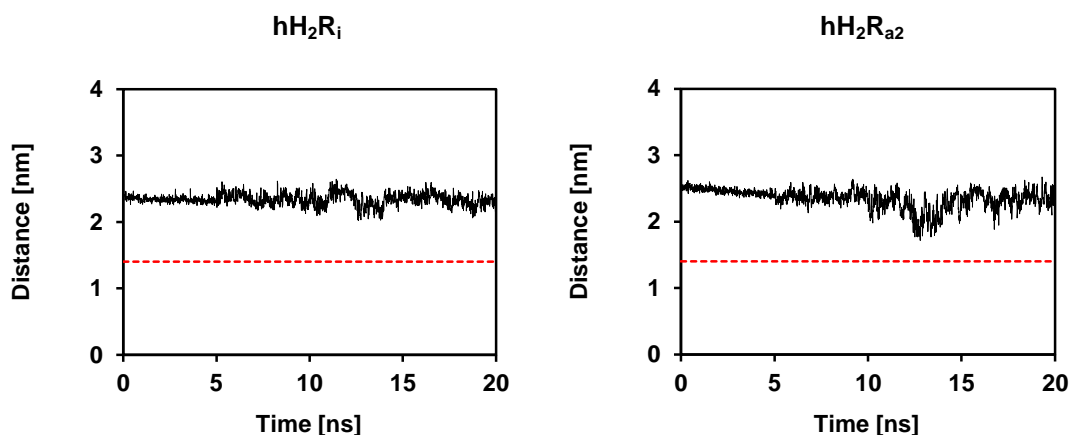


Figure 4.4: Distance of the protein to its nearest periodic image

The red dashed line indicates the cut-off value for non-bonded long range interactions (1.4 nm). Every 10th frame was analyzed (10 ps interval).

4.3.2.2 Lipid bilayer

Hydrophobic, polar and charged amino acids are distributed over GPCRs depending on their exposure to the extra- and intracellular solvent or the lipid membrane (Rees *et al.*, 1989; Wallin *et al.*, 1997). Whereas at the extra- and intracellular surface charged residues accumulate, at the receptor surface which is exposed to the membrane hydrophobic amino acids predominate. The protein was inserted into the lipid bilayer to enable an alignment of the amino acids of this so called hydrophobic belt with the lipophilic tails of the DPPC

molecules (Figure 4.5). Charged amino acids faced the solvent or the lipid head groups (Ballesteros *et al.*, 2001).

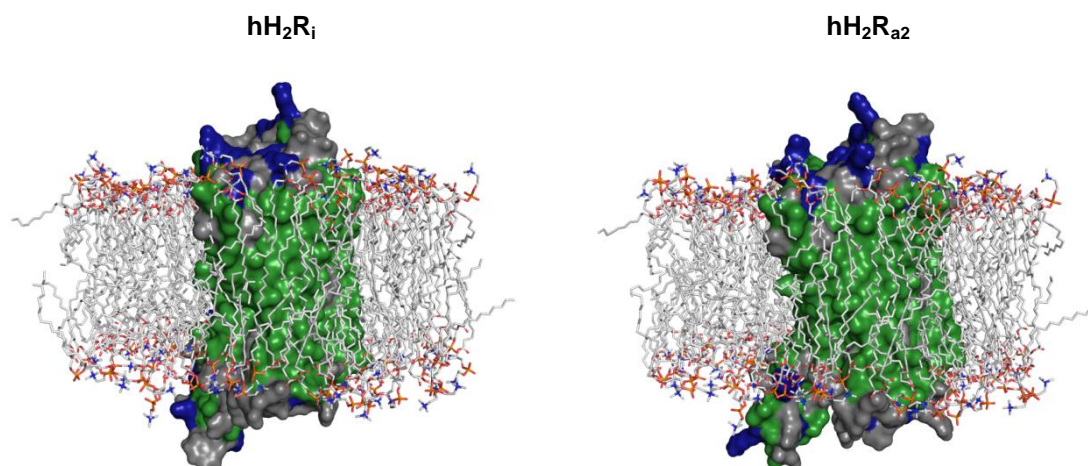


Figure 4.5: Embedding of the hH₂R models in the DPPC bilayer

Side view of the hH₂R_i and hH₂R_{a2} in surface representation within the lipid bilayer (sticks). Hydrophobic amino acids (alanine, glycine, isoleucine, leucine, methionine, phenylalanine, proline and valine) are highlighted in green and charged amino acids (arginine, aspartate, glutamate and lysine) in blue.

An appropriate solvation of the bilayer is confirmed by a close contact of water molecules with polar and charged atoms of the lipid head groups (N, O, P). Thus, the radial distribution functions (RDFs) of water oxygens and the carbonyl oxygen atoms, phosphorus atoms and nitrogen atoms of DPPC, respectively, were calculated (Figure 4.6). The integration of the first peaks gives the average number of water molecules within the respective radius around the atoms (Pandey and Roy, 2011; Poger and Mark, 2010). Based on the calculation of the last 2.5 ns of the equilibration, nitrogen atoms were solvated best (in hH₂R_i and hH₂R_{a2} on average 14.8 and 13.7 water molecules within a radius of 6.2 Å and 6.1 Å, respectively, around the nitrogen atoms; Table 4.1). In a radius of 4.6 Å around phosphorous atoms an average number of about 3.5 water molecules were found in both simulation systems. The oxygen atoms of the carbonyl functions were solvated least because of the shielding effects of the choline moieties of DPPC.

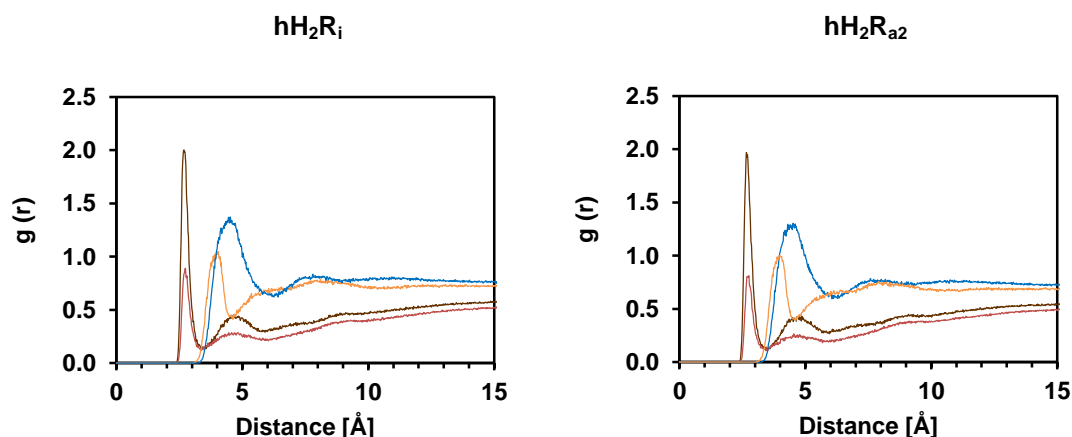


Figure 4.6: Radial distribution functions for clustering of water oxygens with head group nitrogen, phosphorus and carbonyl oxygens

The radial distribution functions are shown for the oxygen atoms of water molecules and the nitrogen (blue), phosphorous (orange), carbonyl oxygen of the sn-1 (brown) and sn-2 (red) chain of DPPC, respectively. The data shown is the average over time (last 2.5 ns of the equilibration phase) and DPPC molecules.

Table 4.1: Average number of water molecules around the lipid head groups of DPPC

Shown is the average number of water molecules received from integrating the radial distribution functions (Figure 4.6). The radiuses around the heteroatoms of DPPC are given in brackets [Å].

	hH ₂ R _i	hH ₂ R _{a2}
Nitrogen	14.8 (6.2)	13.7 (6.1)
Phosphorous	3.5 (4.6)	3.5 (4.6)
Oxygen (sn-1 chain)	1.5 (3.4)	1.5 (3.5)
Oxygen (sn-2 chain)	0.7 (3.4)	0.7 (3.4)

The dynamics of the carbon chains of bilayers strongly depends on the temperature. For MD simulations a temperature about ten degree above the phase transition temperature of the lipid is generally used. In case of the DPPC bilayer, this temperature was determined at 314.4 K (De Young and Dill, 1988; Lewis *et al.*, 1987), and therefore for MD simulations 323 K is commonly used (Patra *et al.*, 2003; Prates Ramalho *et al.*, 2011). A measure for the dynamics of carbon chains is the time-averaged mobility of a particular C-D bond (carbon-deuterium bond) within a lipid molecule, the deuterium order parameters S_{CD} (Vermeer *et al.*, 2007). Experimental values are obtained from deuterium NMR quadrupole splittings. The values calculated for both simulation systems are in good agreement with values reported in literature (Figure 4.7).

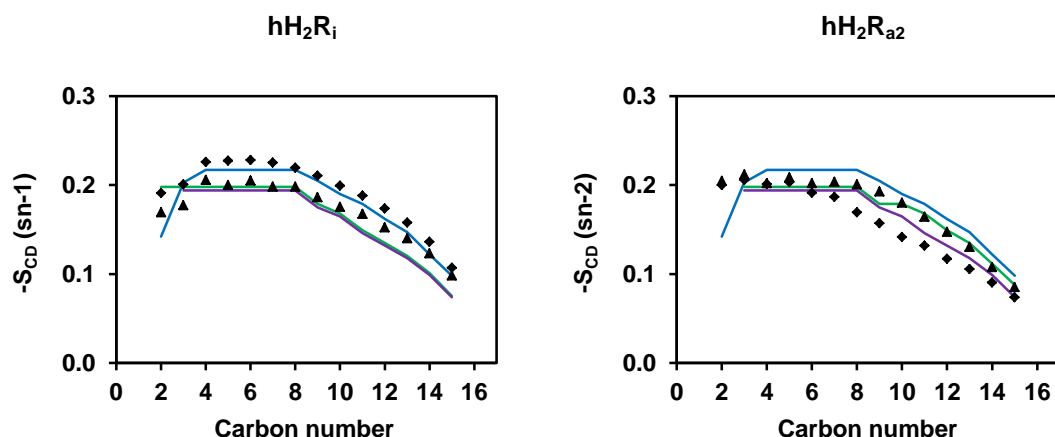


Figure 4.7: Deuterium order parameters for the carbon chains of DPPC

Deuterium order parameters are shown for the sn-1 (▲) and sn-2 (◆) chain for the systems with the hH₂R_i and with the hH₂R_{a2}. Experimental S_{CD} values are from Douliez *et al.* (1995 ; blue line), Leftin and Brown (2011 ; purple line) and Petrache *et al.* (2000 ; green line).

The orientation of the lipid head groups was calculated as probability distribution of the angle between the P→N dipole and the bilayer normal and compared with experimental data, too. In neutron diffraction studies the P→N vector was shown to be nearly parallel to the membrane (Buldt *et al.*, 1978). The peak of the distribution, averaged over the last 2.5 ns of the equilibration phase and all lipid molecules, was at about 79° and 77° for the system with the hH₂R_i and with the hH₂R_{a2}, respectively (Figure 4.8). Published values ranged from 60° to 90° (Jambeck and Lyubartsev, 2012; Poger and Mark, 2010; Ulmschneider and Ulmschneider, 2009) indicating a good agreement with data from the MD simulations.

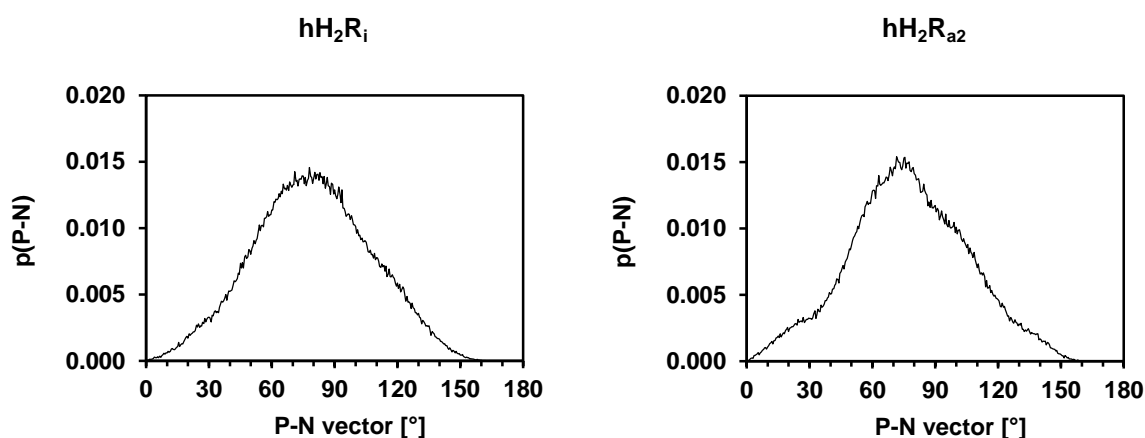


Figure 4.8: Distribution of the angle between the P→N dipole of the DPPC head groups and the normal of the membrane in the last 2.5 ns of the equilibration phase

To further validate the simulated bilayer, the membrane thickness was compared with data from X-ray scattering experiments. The distance between the peaks in the electron density profile, corresponding to the distance between the lipid head group phosphates, is a good estimation of the bilayer thickness (Kucerka *et al.*, 2011). For a DPPC bilayer at 50 °C, 38.0 Å were measured (Kucerka *et al.*, 2008). The thickness calculated during the equilibration phase was below that value for both systems (mean \pm SD for hH₂R_i and hH₂R_{a2}: 33.6 \pm 1.5 Å and 34.7 \pm 1.8 Å; Figure 4.9). However, at the end of the last equilibration step of hH₂R_{a2} the thickness converged to the reference value, probably caused by a slight contraction of the simulation box in the x-y plane and an expansion in z-dimension, i.e. the normal of the bilayer (Figure 4.3 C).

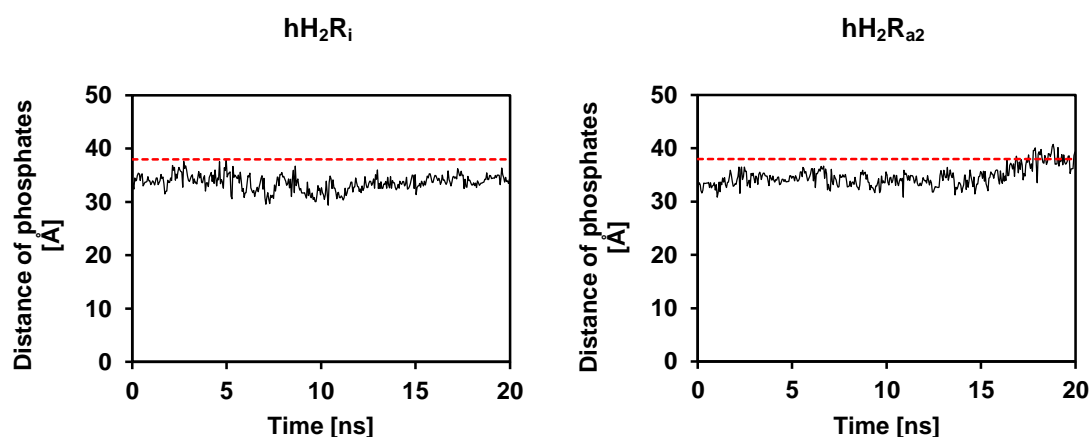


Figure 4.9: Thickness of the DPPC bilayer measured as the average distance between the phosphates of DPPC

The thickness was measured as the average distances between the phosphates of DPPC every 10th frame (10 ps interval). The red line indicates the experimentally determined thickness of a DPPC bilayer at 50 °C.

Finally, the area per lipid was analyzed. In a simulation system consisting only of lipid and water molecules this is just the area of the plane parallel to the membrane divided by the number of lipids in one layer. If a membrane protein is embedded the required area of the protein has to be estimated and subtracted from the membrane plane (Allen *et al.*, 2009). Values for the area per lipid derived from experiments (X-ray methods, NMR and neutron diffraction) and simulations vary considerably. Measured areas per lipid of a DPPC bilayer at about 323 K range from 50 to 72 Å² (Nagle and Tristram-Nagle, 2000a; Poger and Mark, 2010). The values obtained from both simulations with the hH₂R are in the range of data received from literature (Figure 4.10). Comparable to other parameters, values changed after

releasing constraints from simulation components and converged to constant values at the end of the respective equilibration step. The decrease in the area per lipid observed in both simulations, most pronounced at the 4th step of the hH₂R_{a2} system, correlated with the alterations in the box dimensions (Figure 4.3 C).

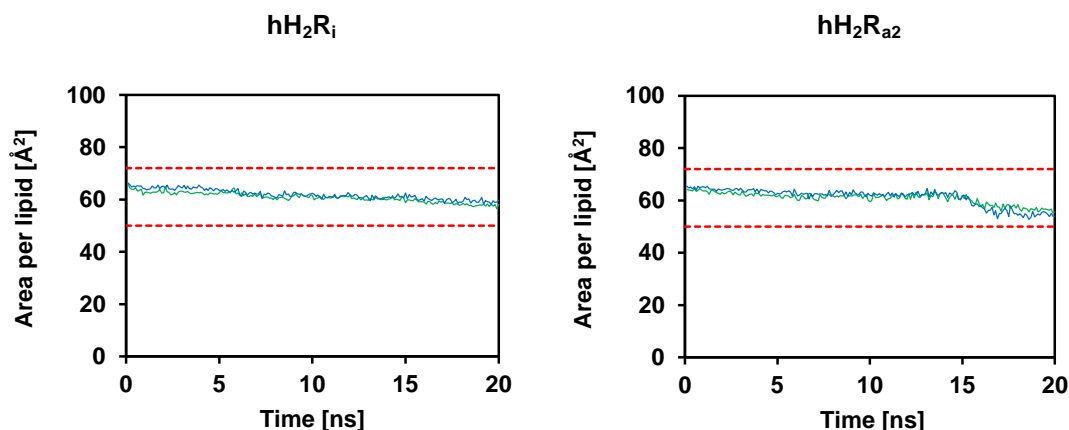


Figure 4.10: Area per lipid of the DPPC bilayer during the equilibration phase

The area per lipid was calculated with *InlateGRO* every 100th ps. Blue line, DPPC monolayer at the extracellular side; Green line, DPPC monolayer at the intracellular side; Red dashed line, Lower and upper limits (50 Å² and 72 Å², respectively) received from literature for the area per lipid of DPPC at 50 °C.

In summary, the convergence of system parameters to stable values and the agreement of the DPPC bilayer structure with experimental data proved both simulation systems to be appropriate for starting productive 80 ns MD simulations which will be analyzed in the following sections.

4.3.3 Analysis of system parameters

The potential energy (mean \pm SD for hH₂R_i and hH₂R_{a2}: $-589 \times 10^3 \pm 608$ kJ mol⁻¹ and $-657 \times 10^3 \pm 619$ kJ mol⁻¹) and the density (mean \pm SD for hH₂R_i and hH₂R_{a2}: 991 ± 1.1 kg m⁻³ and 991 ± 1.0 kg m⁻³) of both simulation systems remained constant during the 80 ns production runs (Figure 4.11 A, B). However, the simulation box of the hH₂R_i was contracted slightly in the plane of the membrane and raised in the z-dimension at the beginning of the simulation and at 35 to 40 ns (about 2%; Figure 4.11 C). The box dimensions of the hH₂R_{a2} system remained stable.

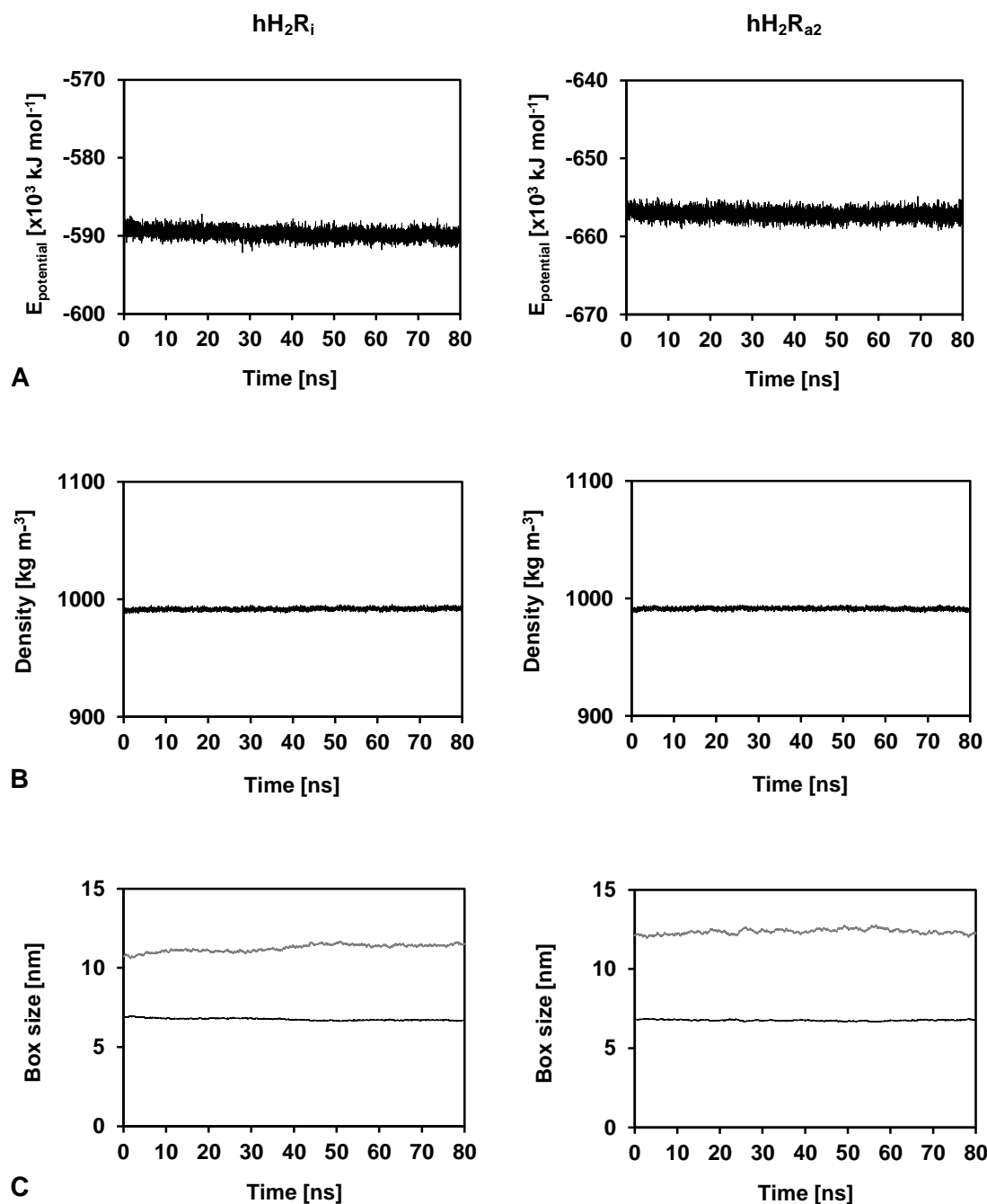


Figure 4.11: System parameters of the 80 ns production run

Shown are the potential energy (A), the density (B) and the box dimensions (C) in the x- (black line) and z-direction (grey line) for the system of the hH_2R_i and hH_2R_{a2} , respectively. The y-vector was similar to the x-vector (semiisotropic pressure coupling) and is not shown. Every 10th frame (10 ps interval) was analyzed.

The distance of the hH_2R_i to its nearest periodic image was below the cut-off value for calculating long-range interactions (1.4 nm) in about 0.4% of all time points analyzed. The side chains of Arg215 and Asn217 in ICL3 approach Cys304 and Cys305 in the C-terminus, respectively (Figure 4.12). However, this long range interaction in the x-y plane was only

temporary, and no artifacts were expected from that. The hH₂R_{a2} was in a proper distance to its periodic images during the whole simulation.

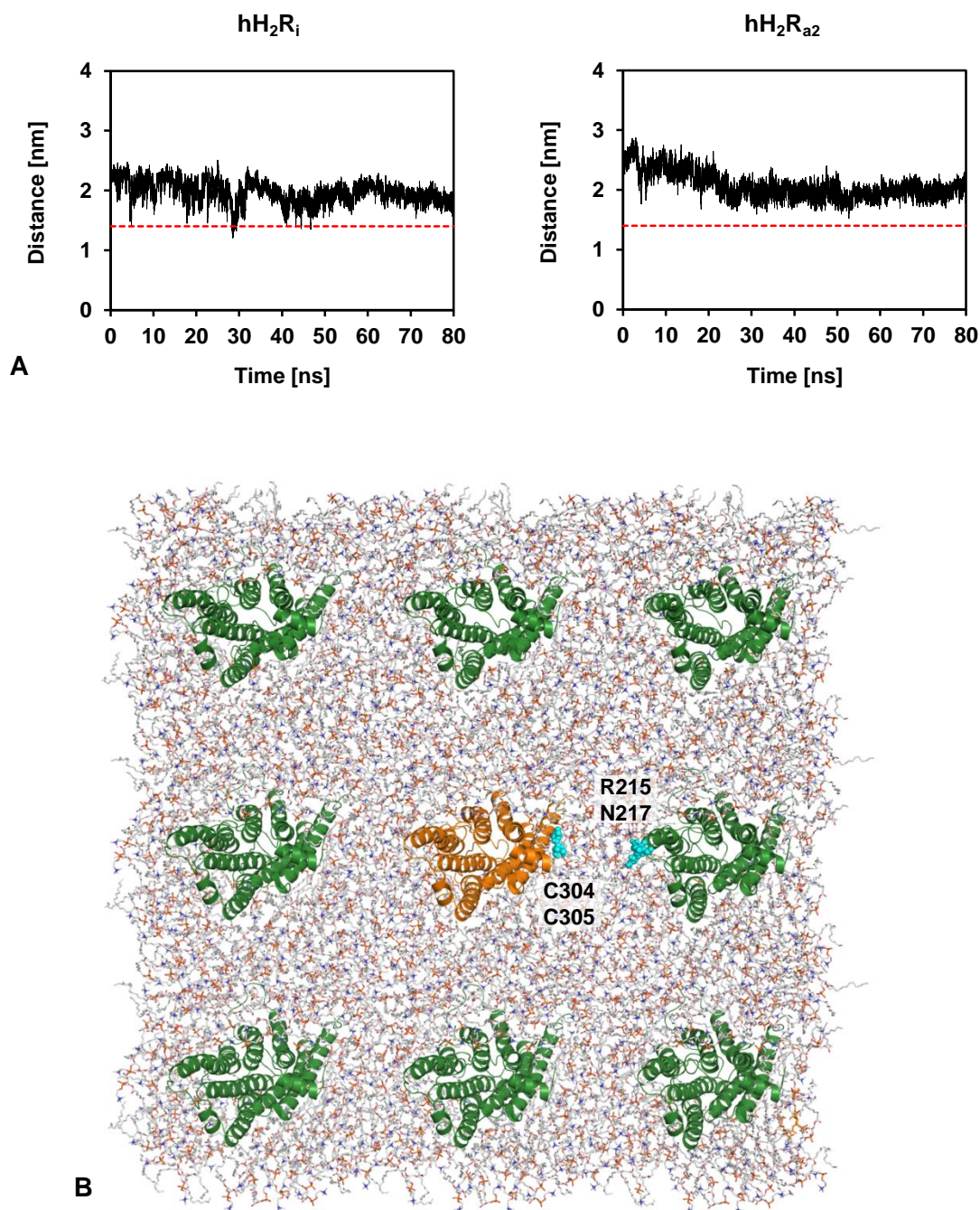


Figure 4.12: Distance and position of the hH₂R to its periodic images

A, Distance of the hH₂R_i and hH₂R_{a2} to its nearest periodic image. The red dashed line indicates the cut-off value for the calculation of non-bonded long range interactions (1.4 nm). Every 10th frame was analyzed (10 ps interval). B, The hH₂R_i in the middle (orange ribbon and tubes) and its eight periodic images (green ribbon and tubes) in the x-y plane, viewed from the intracellular side. Amino acids closer than 1.4 nm in 0.4% of all time points are highlighted with cyan spheres.

4.3.4 Structural analysis of the DPPC bilayer

Analogously to the equilibration phase of the hH_2R_i and hH_2R_{a2} the corresponding parameters of the DPPC bilayer were analyzed during the 80 ns productive MD simulations. The deuterium order parameters (Figure 4.13 A) of the sn-1 and sn-2 chains of both simulation systems were in very good agreement with the values provided by Douliez (1995). The orientations of the lipid head groups, represented by the frequency distributions of the angle formed by the P→N vector and the bilayer normal (peaks at 81° for both hH_2R_i and hH_2R_{a2} ; Figure 4.13 B) were in good agreement to experimental values, too. The bilayer thickness remained nearly constant in the simulation systems (mean \pm SD for hH_2R_i and hH_2R_{a2} : $37.9 \pm 1.6 \text{ \AA}$ and $38.1 \pm 1.4 \text{ \AA}$) and was in excellent agreement to the values obtained from experiments (Figure 4.13 C). As demonstrated (Figure 4.10), the area per lipid for DPPC decreased in both systems during the equilibration phase by about 10 \AA^2 . This was in correlation with the alterations of the box dimensions which were compressed in the plane of the membrane and expanded in the direction parallel to the normal of the membrane in the hH_2R_i and hH_2R_{a2} system, respectively. During the production run (Figure 4.13 D) a slight, continuous decrease of the area per lipid was obvious in the hH_2R_i system. In both systems, the values were still above the lowest limit of the experimental reference values, and the standard deviations were small (mean \pm SD for hH_2R_i and hH_2R_{a2} : $54.3 \pm 1.7 \text{ \AA}^2$ and $53.6 \pm 1.4 \text{ \AA}^2$).

The alignment of the hydrophobic belt of the hH_2R_i and hH_2R_{a2} with the lipophilic DPPC chains was assessed by calculating the fraction of time points in which a hydrophobic moiety of an amino acid was in close distance to the carbon chains of the membrane lipids (distance of hydrophobic heavy atoms $< 5 \text{ \AA}$). Figure 4.14 shows that most amino acids of the GPCR core were in close contact to the lipids during the entire simulation.

The absence of water molecules within the lipid tails of DPPC during the MD simulations of hH_2R_i and hH_2R_{a2} was checked visually and by calculating the density profile of water molecules along the normal of the membrane (Figure 4.15). The density curve of the hH_2R_{a2} system was shifted to the right due to the space filling GαCT below the intracellular part. No water molecules diffused into the hydrophobic core of the bilayer within the 80 ns MD simulations.

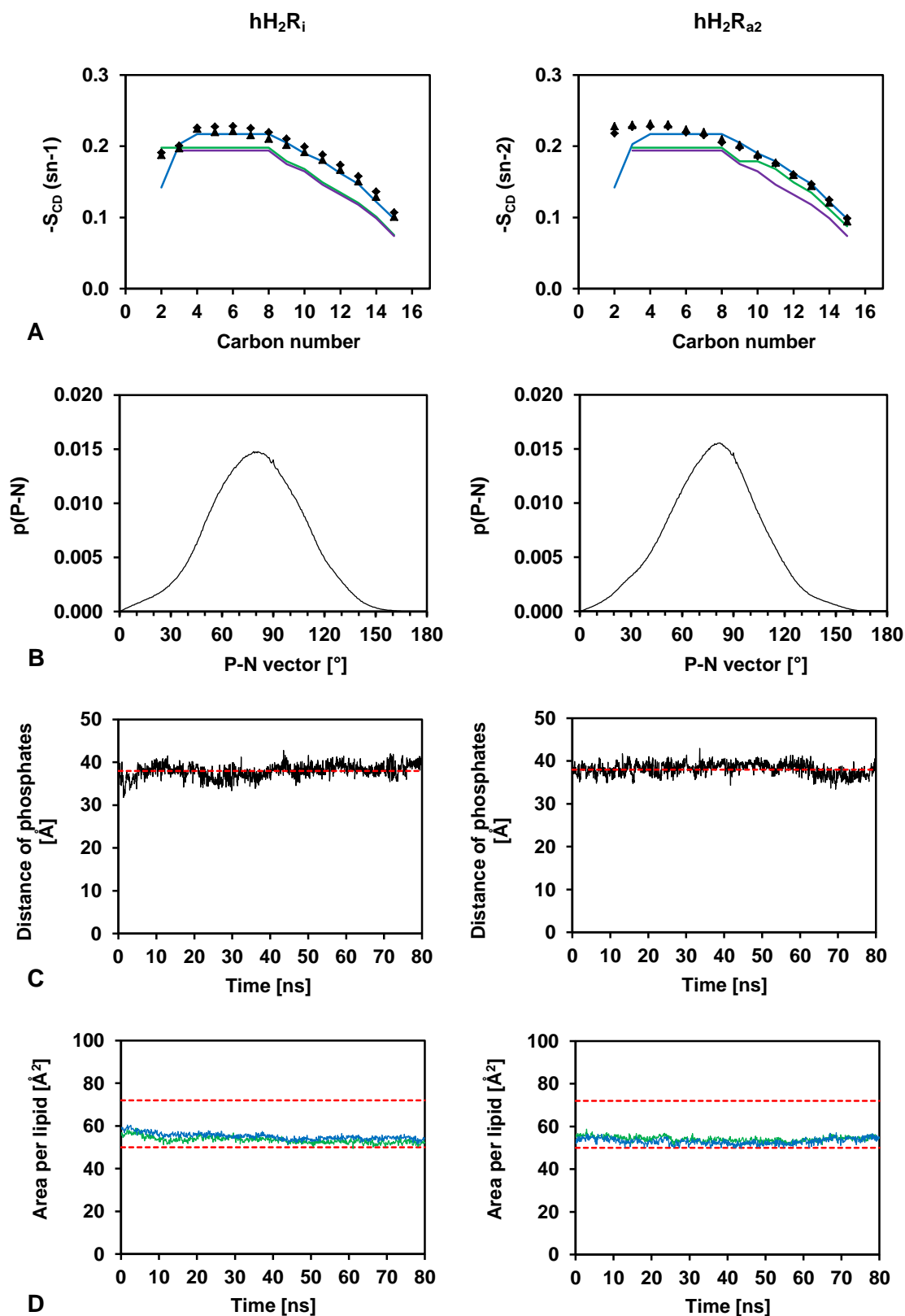


Figure 4.13: Bilayer parameters of the 80 ns MD simulation of the hH₂R_i and hH₂R_{a2}

The deuterium order parameters (A), the orientation of the P→N vector of the lipid head groups in relation to the bilayer normal (B), the bilayer thickness (C) and the area per lipid (D) were determined and compared to reference values as described in Figure 4.7, Figure 4.8, Figure 4.9 and Figure 4.10, respectively.

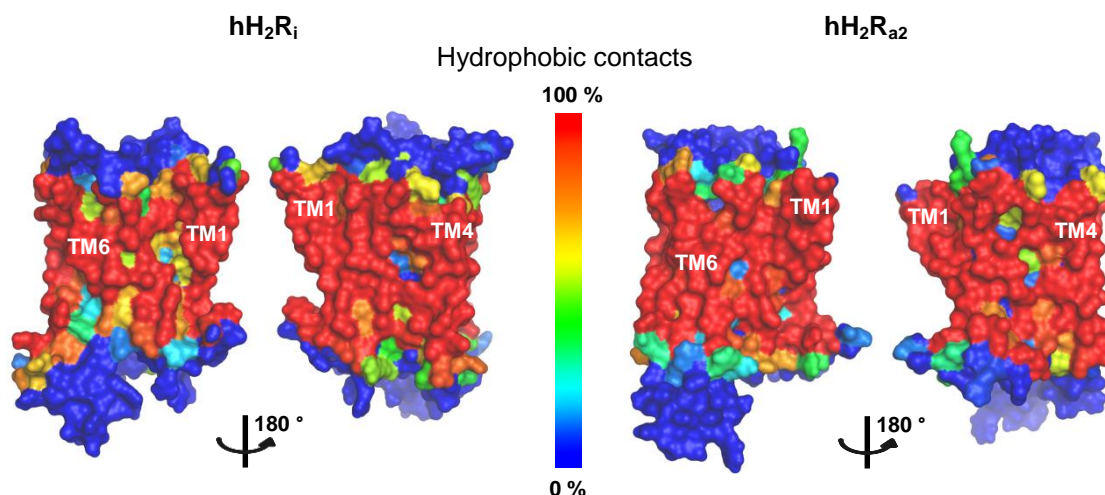


Figure 4.14: Contact of the hH₂R states to the lipophilic tails of DPPC

For each amino acid of the hH₂R_i and hH₂R_{a2}, respectively, the frequency of hydrophobic contacts to the lipophilic tails of DPPC (carbon atom 2 to 16 of the sn-1 and sn-2 chain, respectively, as defined in Figure 4.1) were calculated. The surface of the final structures (after 80 ns MD simulations) of the two receptor states were colored according to these frequencies (rainbow coloring from blue to red corresponding to 0% to 100% frequency).

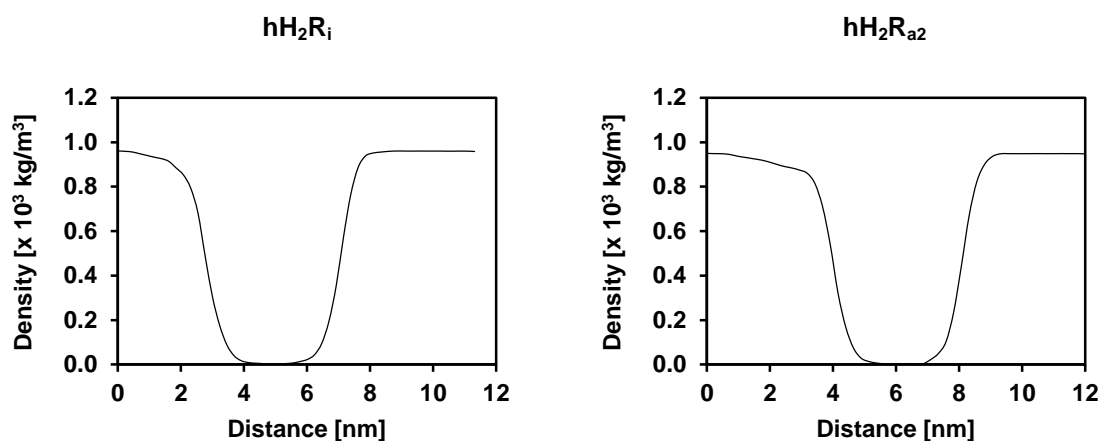


Figure 4.15: Density profile of the water molecules in the simulation boxes of the hH₂R_i and hH₂R_{a2} along the normal of the membrane

4.3.5 Protein structure validation

The quality of the proteins was assessed every 10th picosecond during the entire 80 ns MD simulations of the hH₂R_i and the hH₂R_{a2}, respectively, resulting in 8,000 time points. All data were obtained with the program *gro_validation* and the tool for analyzing the frequency of H-bonds of each heteroatom of a protein included in *gro_hbonds*. Both programs are described in chapter 5.

4.3.5.1 Stereochemistry and planarity

Stereochemical errors in MD simulations may dramatically influence results (Schreiner *et al.*, 2011). The analysis of correct configurations at C_α atoms of amino acids is therefore a fundamental task for proving the quality of an MD approach. In our simulations, the stereochemistry at C_α atoms was correct for all amino acids of the hH₂R_i, hH₂R_{a2} and GαCT. All dihedral angles (more than 2.21 and 2.42 million in hH₂R_i and hH₂R_{a2}-GαCT, respectively) used for testing stereochemistry were within the range of 15° to 47° (Figure 4.16 A), with a mean (and mode) of the distributions of all angles of 32° to 33° (Figure 4.16 B). For proteinogenic L-amino acids with an ideal tetrahedral configuration around the C_α atom this angle is about 35°.

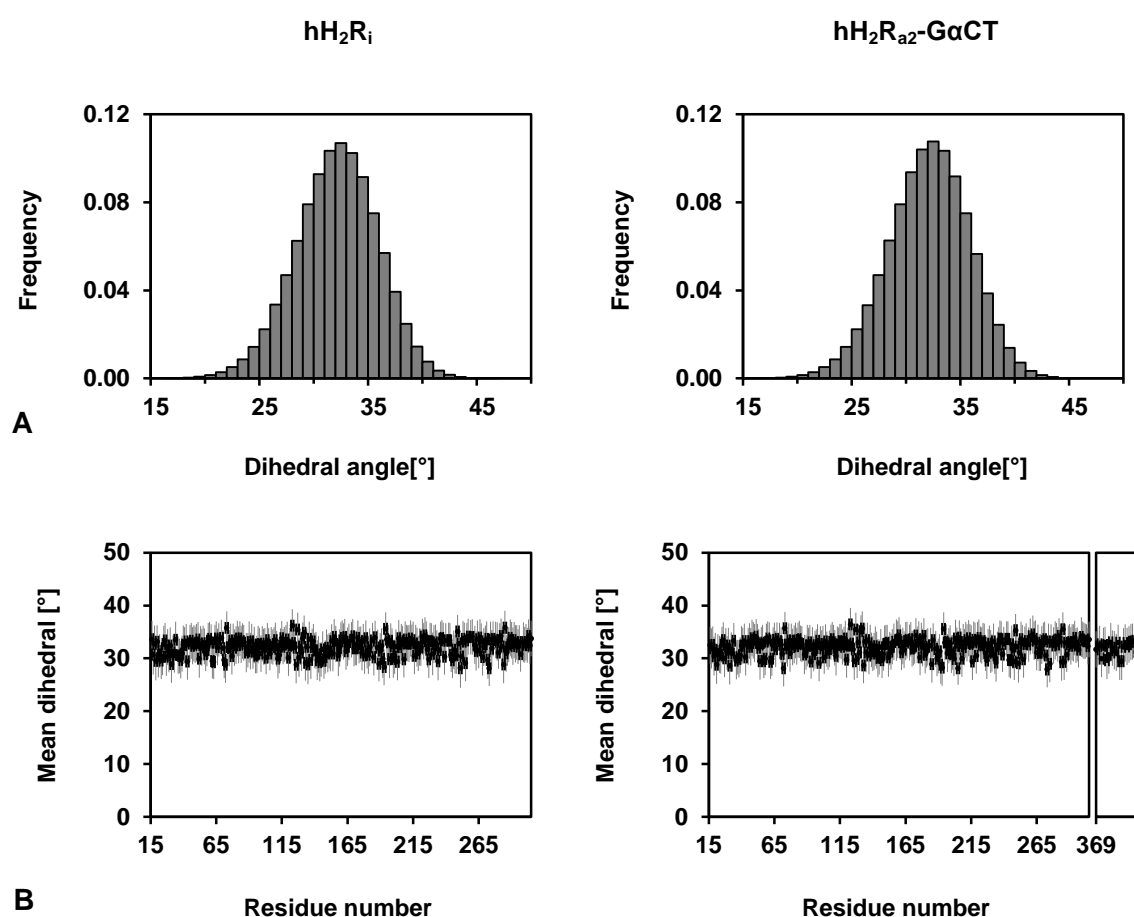


Figure 4.16: Control of the stereochemistry at C_α atoms

A, Frequency distribution of the dihedral angle C_α-N-C-C_β used to check the stereochemistry at the C_α atom of all amino acids of the respective protein over all time points analyzed. B, Mean (black point) ± SD (grey line) of the dihedral angle C_α-N-C-C_β of every residue. The break of the graph in hH₂R_{a2}-GαCT separates hH₂R_{a2} and GαCT.

Aromatic side chains (of His, Phe, Trp and Tyr) as well as polar side chains (of Arg, Asn, Asp, Gln and Glu) containing a delocalized π -electron system are putative planar moieties in proteins. In accurate MD systems their planarity is retained throughout the simulation. To investigate the deviation from planarity, suitable dihedral angles (chapter 5) were measured in the hH₂R_i and hH₂R_{a2}-G α CT.

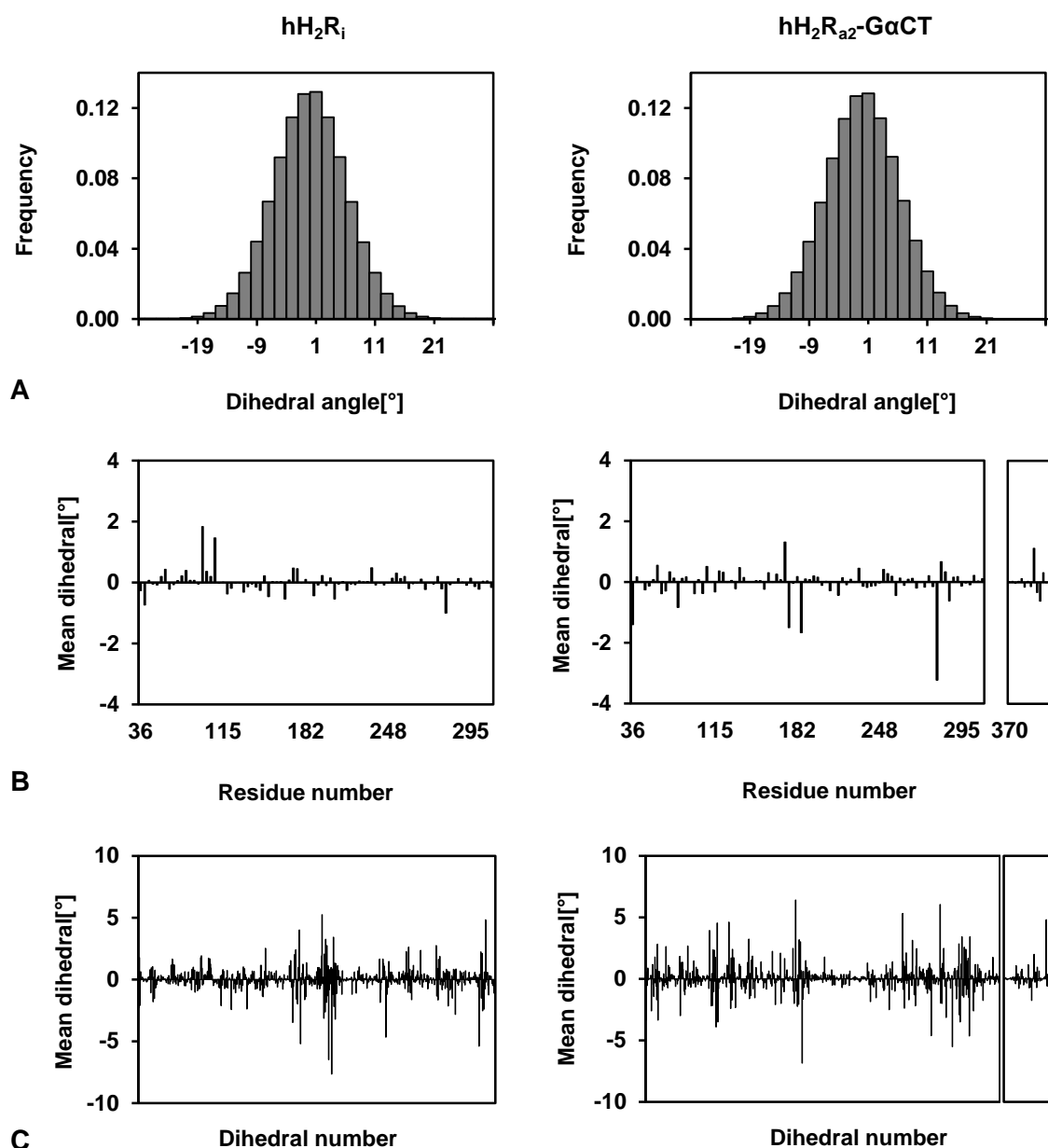


Figure 4.17: Analysis of naturally planar moieties in amino acid side chains

A, Frequency distribution of all dihedral angles used to measure planarity of aromatic side chains (His, Phe, Trp and Tyr) and planar delocalized π -electron systems (Arg, Asn, Asp, Gln and Glu) of amino acid side chains. B, Mean of all dihedral angles used to measure planarity of one specific residue. C, Mean of each analyzed dihedral angle. The break of the graph in hH₂R_{a2}-G α CT separates hH₂R_{a2} and G α CT.

In both simulations only 13 angles (of a total of more than 4.75 and 5.20 million angles analyzed in hH₂R_i and hH₂R_{a2}-GaCT, respectively) deviated more than 30° from planarity (Figure 4.17 A). The mean dihedral angle of each residue with a planar moiety, composed of up to 20 single dihedral angles (e.g. for the side chain of tryptophan), was within $\pm 4^\circ$ in both systems (Figure 4.17 B). The mean of each measured dihedral angle was within $\pm 8^\circ$ (Figure 4.17 C).

In order to assess the geometry of the peptide bond of the proteins in both MD simulations, the omega angle between the planes O_i-C_i-N_{i+1} and C_i-N_{i+1}-H_{i+1} was measured (chapter 5). This angle is 180° in case of a planar trans configuration of the atoms at the amide bond C_i-N_{i+1}.

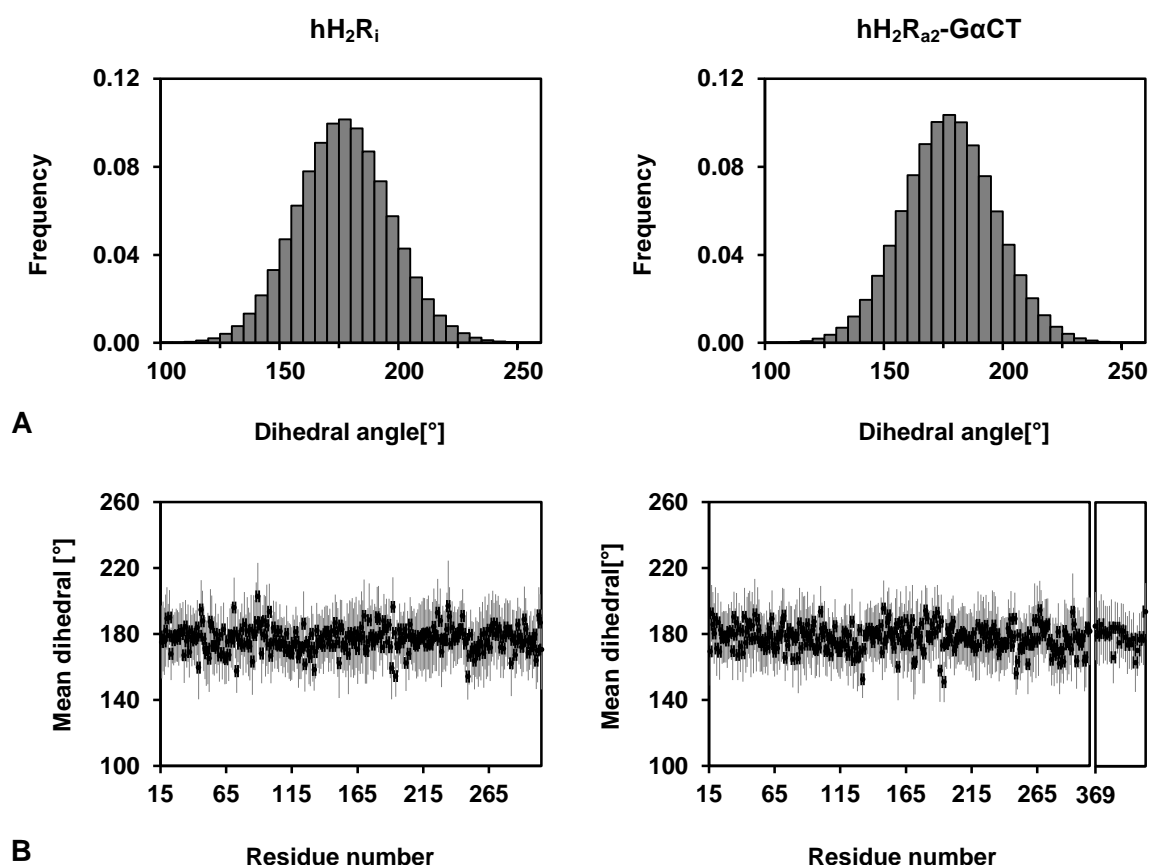


Figure 4.18: Analysis of the peptide dihedral angle

A, Frequency distribution of the omega dihedral angles of all amino acids over all time points analyzed. B, Mean (black point) \pm SD (grey line) of the omega angle of each residue of the hH₂R_i and hH₂R_{a2}-GaCT, respectively. The break of the graph in hH₂R_{a2}-GaCT separates hH₂R_{a2} and GaCT.

Figure 4.18 shows the distribution of all omega angles of the hH₂R_i and the hH₂R_{a2} systems (2.32 and 2.52 million, respectively) as well as the mean omega angles of each residue. The mean of all angles was $\sim 177^\circ$. This is in good agreement with values obtained from literature of about 179° (MacArthur and Thornton, 1996; Morris *et al.*, 1992). However, the width of the distribution was smaller for experimentally derived omega angles, having the majority of values within 170° to 190° (Edison, 2001). Thus, parameters of the GROMOS96 53a6 force field affecting the amide bond C_i-N_{i+1} should be defined more strictly.

4.3.5.2 Ramachandran analysis

The majority of residues of the hH₂R_i and hH₂R_{a2} exhibited backbone Φ/Ψ (phi/psi) combinations located either in favored or allowed regions of the conformational space during the entire simulations (Figure 4.19). Only eleven and eight amino acids in hH₂R_i and hH₂R_{a2}-GaCT, respectively, possessed dihedral angles not frequently occurring in proteins in more than 20% of all time points. Ramachandran plots of residues which have disallowed values in more than 50% of the time points are shown in Figure 4.20. Strikingly, all of these amino acids except Ser96^{3,30} resided in flexible parts of the proteins, namely in loops or at the border of TMs.

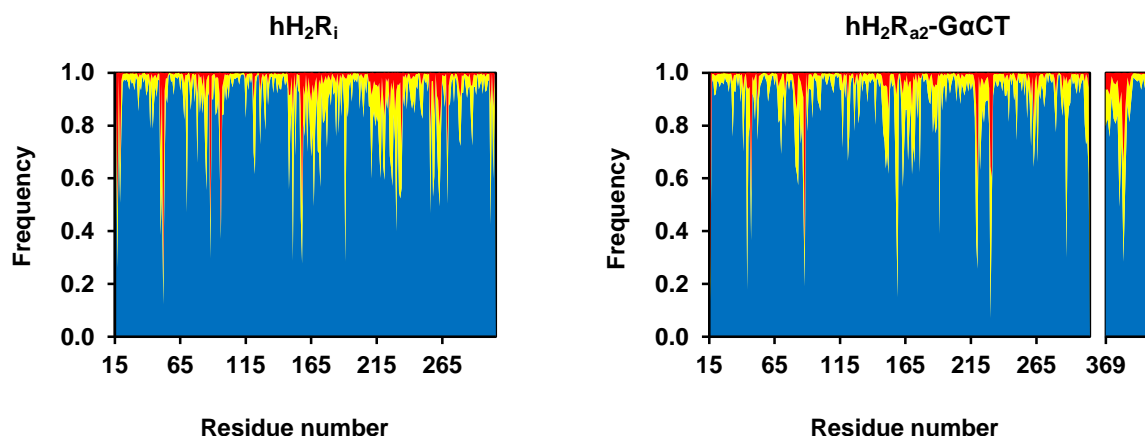


Figure 4.19: Ramachandran analysis of hH₂R_i and hH₂R_{a2}-GaCT in 80 ns MD simulations

For each residue the frequency of Φ/Ψ dihedral combinations located in favored (blue), allowed (yellow) and disallowed (red) regions (as defined in chapter 5) of the Ramachandran plot is shown. The break of the graph in hH₂R_{a2}-GaCT separates hH₂R_{a2} and GaCT.

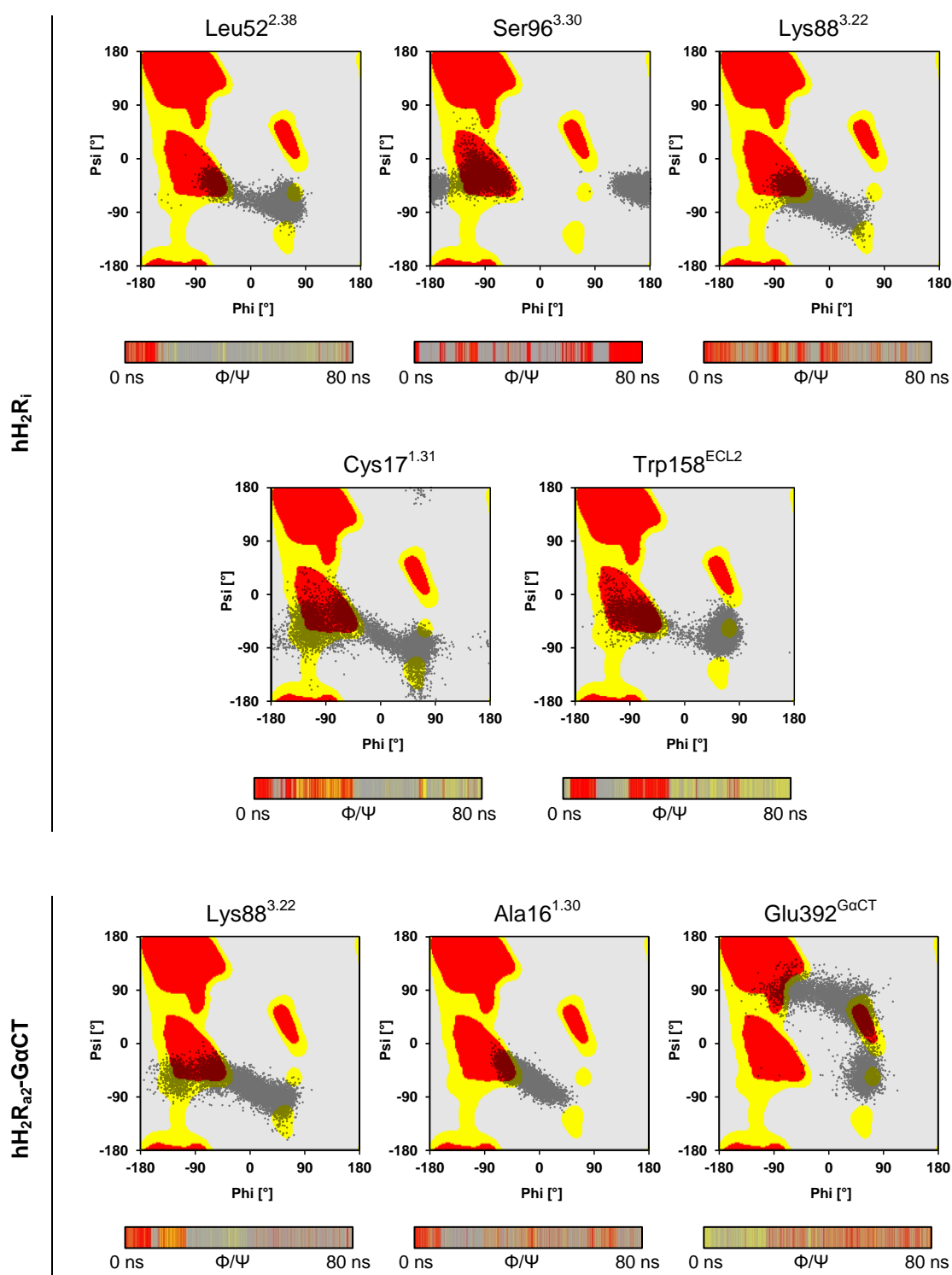


Figure 4.20: Ramachandran plots of residues with a high proportion of time points in which an outlier region was occupied

Ramachandran plots for amino acids of the hH_2R_i and hH_2R_{a2} (including GaCT) with more than 50% of their Φ/Ψ dihedral angles in the outlier region. Below each Ramachandran plot the time dependent distribution of Φ/Ψ pairs is given. The analysis was performed as described in chapter 5. Red, favored region; yellow, allowed region; grey, disallowed region.

4.3.5.3 Side chain rotamers

Frequently occurring side chain conformations (rotamers) of amino acids are accessible from protein crystal structures and were analyzed previously (Lovell *et al.*, 2003; Lovell *et al.*, 2000). The agreement of rotamers obtained in MD simulations with these experimentally derived low-energy conformations accounts for the reliability of the MD results. The majority of residues of the hH₂R_i, hH₂R_{a2} and GαCT had side chain dihedral angles or combinations of dihedral angles (in case of residues with more than one rotatable bond in the side chain), respectively, during the 80 ns MD simulation which are frequently occurring in protein crystal structures (Figure 4.21). In both simulation systems just two amino acids possessed in more than 20% of the simulations improbable conformations. In hH₂R_i the rotamers of His230^{6,31} and Glu180^{5,36} had in 26.2% and 25.4% of the simulation, respectively, disallowed dihedral angles. The side chain torsions of Met100^{3,34} of the hH₂R_{a2} were in 24.8% of the simulation in outlier regions, and of Glu370 (GαCT) in 22.5%.

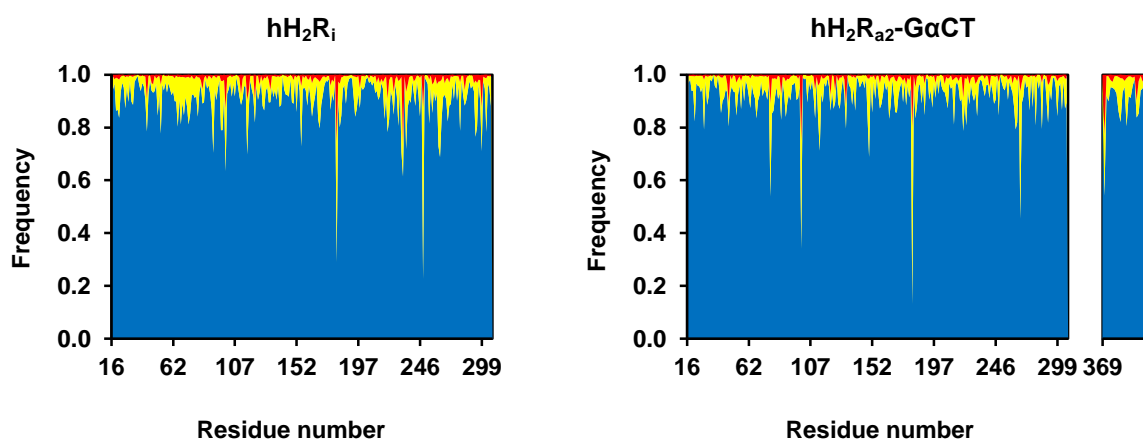


Figure 4.21: Rotamer analysis

For the 80 ns simulation for each residue of the hH₂R_i and hH₂R_{a2} (including GαCT) the number of side chain dihedral angles located in favored (blue), allowed (yellow) and disallowed (red) regions (as defined in chapter 5) is shown. The break of the graph in hH₂R_{a2}-GαCT separates hH₂R_{a2} and GαCT.

4.3.5.4 Saturation of heteroatoms with H-bonds

The control of unsatisfied buried H-bond donors and acceptors in protein structures is a common tool in molecular modeling (Vriend, 1990). The saturation of heteroatoms with H-bonds was checked with the program *gro_hbonds* described in chapter 5. The majority of heteroatoms of hH₂R_i and hH₂R_{a2}-GαCT were H-bonded for most of the simulation time

(Figure 4.22 A, B). Furthermore, most of the side chains of polar amino acids located in TM domains were well H-bonded (Figure 4.22 C, D). The average saturation of polar side chains in TM domains was 96.7% and 94.8% for the simulation of the hH₂R_i and hH₂R_{a2}, respectively.

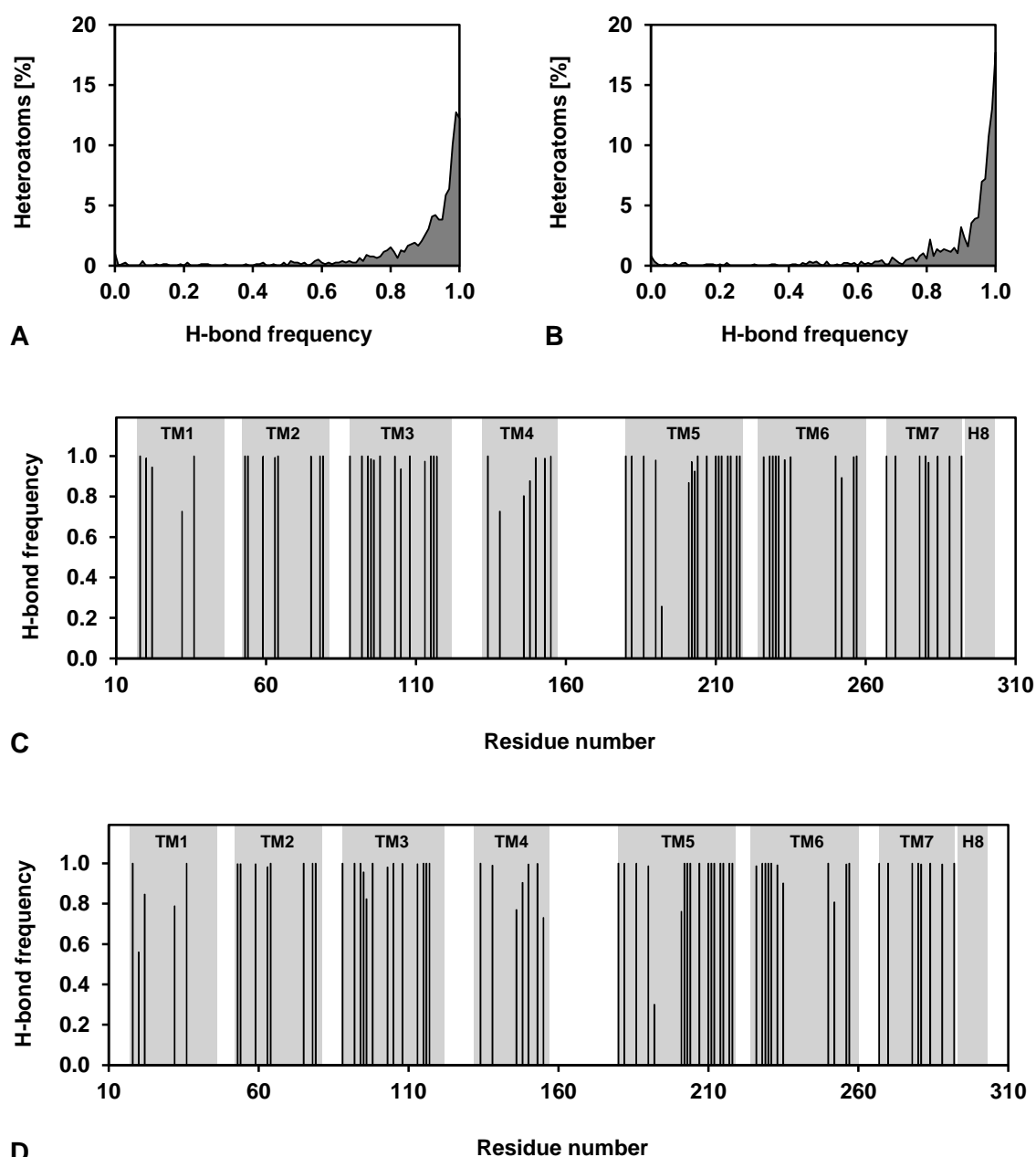


Figure 4.22: Saturation of protein heteroatoms with H-bonds

A, B, Frequency distribution of protein heteroatoms with respect to their fraction of being H-bonded for the hH₂R_i (A) and hH₂R_{a2}-GaCT (B). C, D, Frequency of side chains of polar amino acids within the TM segments of the hH₂R_i (C) and hH₂R_{a2} (D) involved in H-bonds. The grey background indicates the maximal TM domains of hH₂R_i and hH₂R_{a2}.

Tyr192^{5.48} was least saturated with H-bonds in the hH₂R_i (26%) and hH₂R_{a2} (30%), respectively. In both receptor states the side chain of Tyr192^{5.48} was mainly connected to the backbone oxygen of Phe251^{6.52}. This is similar to the structures of the β_1 AR and the β_2 AR, in which the corresponding Tyr217^{5.48} and Tyr209^{5.48} are close to the backbone oxygen of Phe307^{6.52} and Phe290^{6.52}, respectively (Cherezov *et al.*, 2007; Warne *et al.*, 2008). In both structures this oxygen is the only heteroatom in a radius of 4 Å around the Tyr^{5.48} hydroxyl group which is capable of forming H-bonds.

4.3.6 Secondary structure of hH₂R states

The stability of the α -helical structure of TM domains of the hH₂R_i and hH₂R_{a2} was checked with the DSSP (Define Secondary Structure of Proteins) algorithm (Kabsch and Sander, 1983). The prolines in position 2.59 and 5.50 caused an interruption in the α -helix of the two hH₂R_s (Figure 4.23). Additional deviations from an ideal α -helix occurred in TM3, TM4 and TM7 of hH₂R_i and in TM4, TM6 and TM7 of hH₂R_{a2}. However, the secondary structure of these TMs was retained throughout the simulation (Table 4.2).

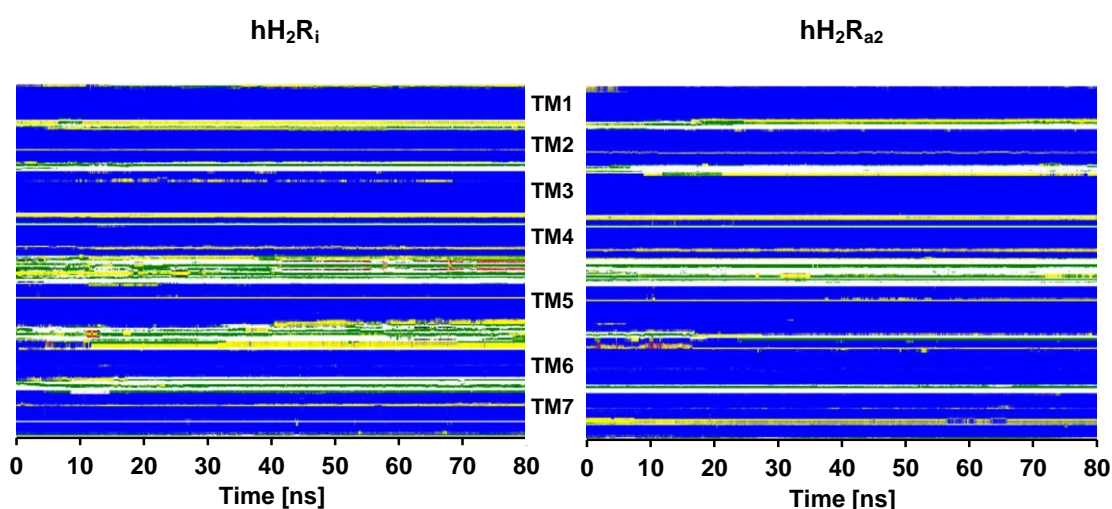


Figure 4.23: DSSP plot of the hH₂R_i and hH₂R_{a2}

DSSP plot illustrating the dynamics of secondary structure elements of the hH₂R models during the 80 ns MD simulation. Blue, α -helix; red, β -strands; yellow, turn; green, bend; white, coil.

The distortion of the α -helix in TM4 is located around position 4.57. The templates t β_1 AR and h β_2 AR of the hH₂R_i and hH₂R_{a2}, respectively, contain a proline kink based on Pro^{4.60} (chapter 3). Distortions of α -helices in TM domains three amino acids before a proline are typical

when analyzing the secondary structure with the DSSP algorithm and were observed for TM2 and TM5, too. The temporary interruption of the α -helix at His230^{6,31} in hH₂R_{a2} is located at the cytoplasmic end of TM6 which belongs to the most flexible parts of TM domains (Figure 4.27).

Table 4.2: Secondary structure analysis of critical residues in TMs of the hH₂R_s

hH ₂ R _i			hH ₂ R _{a2}		
Residue	H ^a	DSSP ^b	Residue	H ^a	DSSP ^b
Thr95 ^{3,29}	62		Leu149 ^{4,56}	51	
Ser96 ^{3,30}	61		Ser150 ^{4,57}	9	
Leu149 ^{4,56}	68		Phe151 ^{4,58}	54	
Ser150 ^{4,57}	26		Glu229 ^{6,30}	78	
Phe151 ^{4,58}	51		His230 ^{6,31}	10	
Tyr278 ^{7,43}	49		Ala279 ^{7,44}	72	
Ala279 ^{7,44}	14				

^a Percentage of time points the respective amino acid is α -helical.

^b Each bar represents the whole 80 ns simulation, black color indicates the presence of an α -helix.

Most secondary structure elements were retained during the simulations of the hH₂R_i and hH₂R_{a2}. However, in both cases the α -helical part of ECL2 was already disrupted during the equilibration phase. By contrast, the α -helix of ICL2 was stable until the end of the simulations. Disorganization of the secondary structure was also observed in the hH₂R_i at the cytoplasmic ends of TM5 and TM6. In the initial model ICL3 was constructed by a loop search (chapter 3), resulting in an elongation of the α -helical segments of TM5 and TM6 into the solvent. Without the stabilizing effect of GaCT in the active hH₂R these segments are obviously instable (Figure 4.24). Only small α -helical fragments separated from TM6 were temporarily present.

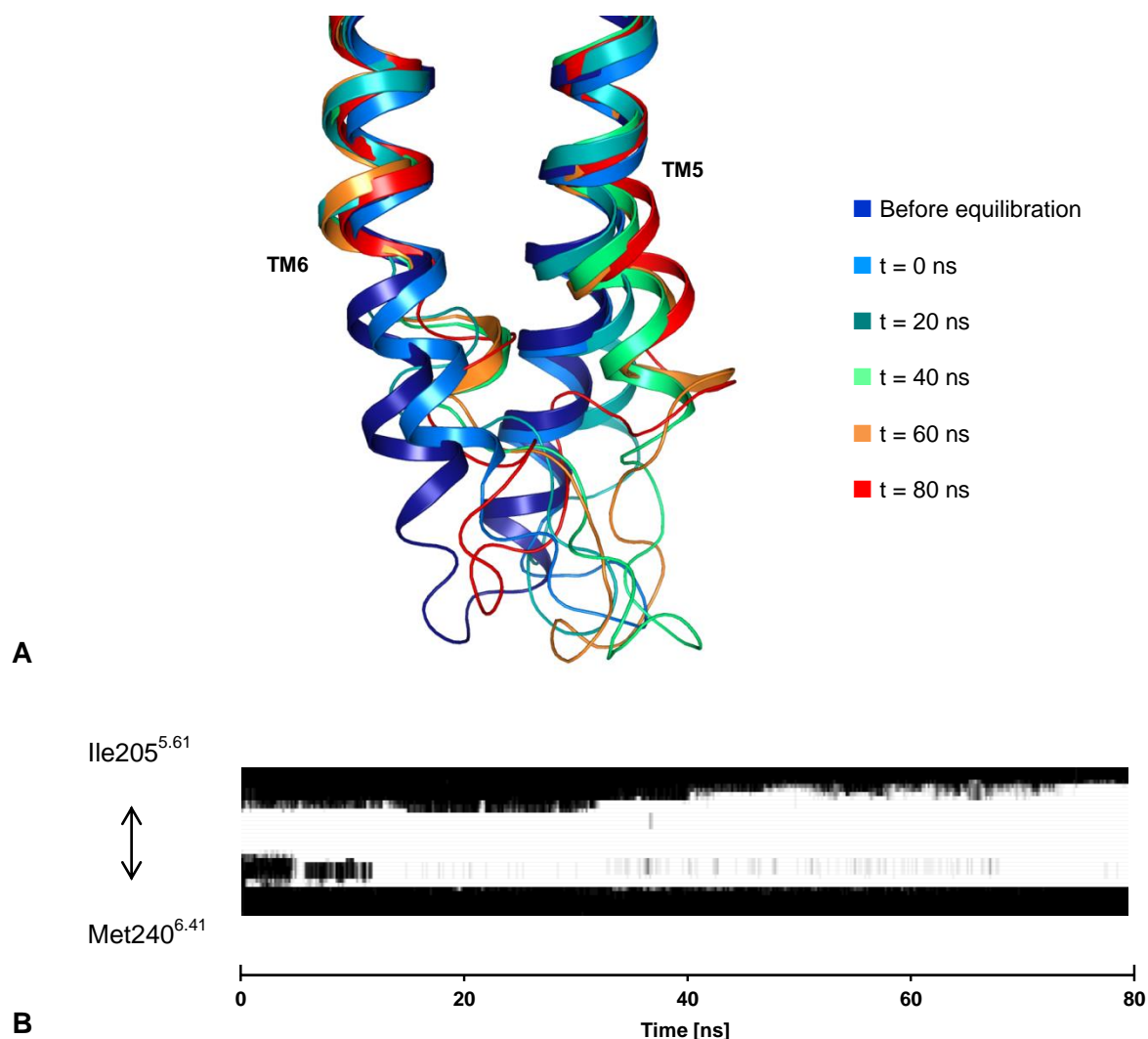


Figure 4.24: Disruption of the cytoplasmic ends of TM5 and TM6 in hH₂R_i

A, Snapshots of the MD simulation of hH₂R_i before the equilibration and at time points 0, 20, 40, 60 and 80 ns. B, DSSP plot from the end of TM5 to the beginning of TM6 (Ile205^{5.61} to Met240^{6.41}); a black line indicates the presence of an α -helix.

Kinks in the α -helical segments of the hH₂R states were present in TM2, TM5, TM6 and TM7 (Figure 4.25). The kink in TM2 of hH₂R_i was constant throughout the simulation (mean \pm SD: $27.9 \pm 2.3^\circ$) and comparable to that in the initial model (29.8° ; cf. Table 3.8). In hH₂R_{a2} the kink rose in the first 10 ns of the simulation from about 30° to 33° . The kink in the initial hH₂R_{a2} model was only 26.8° . The marginal increase of the kink angle could not be ascribed to changes around Pro73^{2.59} or the extra- and intracellular positions of TM2. The kinks in TM5 of hH₂R_i and hH₂R_{a2} fluctuated least and were comparable to the kinks in the initial models (mean \pm SD for hH₂R_i: $14.5 \pm 1.6^\circ$ vs. 12.0° ; mean \pm SD for hH₂R_{a2}: $11.8 \pm 1.4^\circ$ vs. 13.4°).

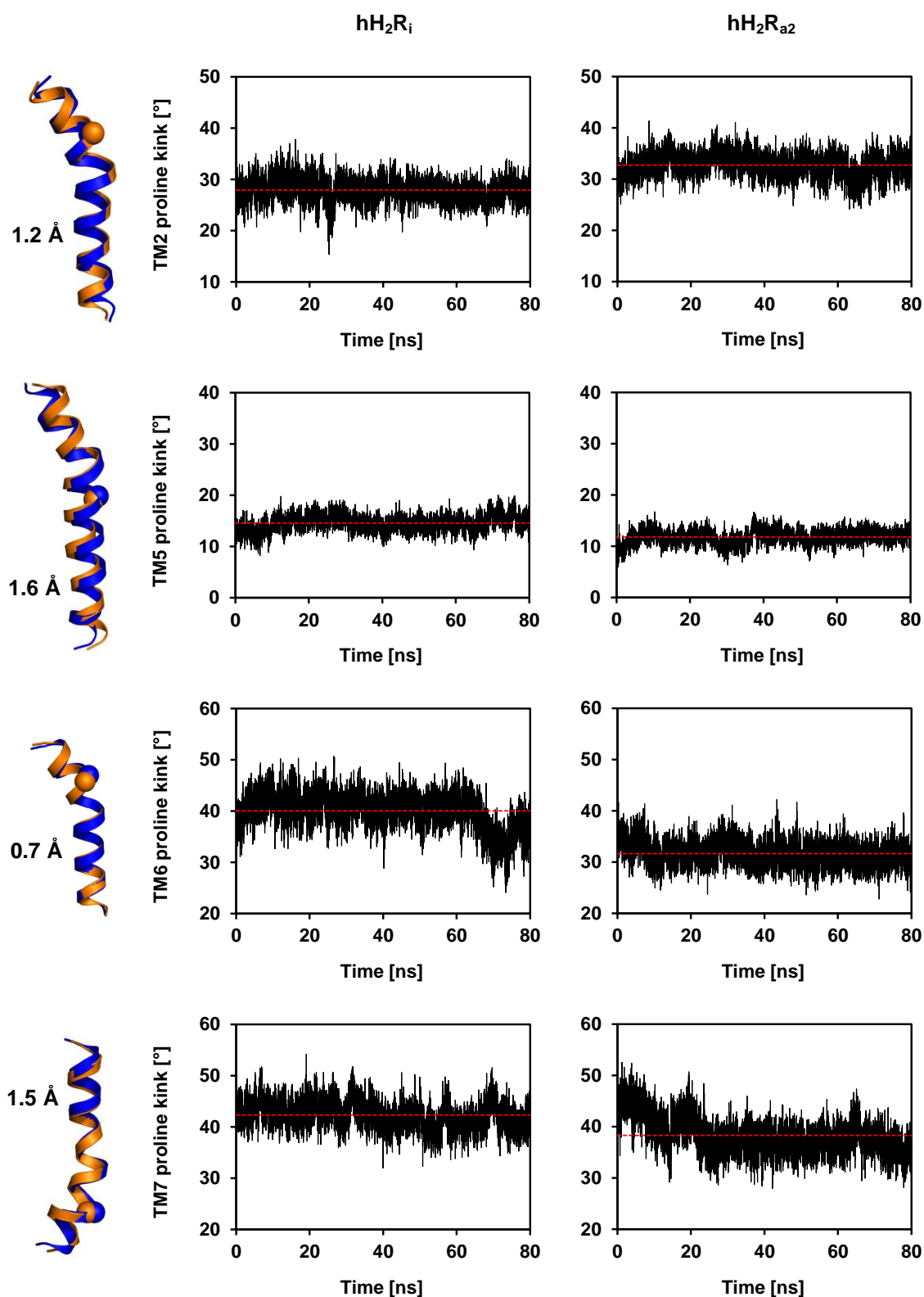


Figure 4.25: Proline kinks of hH₂R states in TM2, TM5, TM6 and TM7

Left, Superimposed TM domains of the hH₂R_i (■) and hH₂R_{a2} (■), using the average structures of the 80 ns MD simulation, respectively (cf. Figure 4.30). C_α atoms of proline are shown as spheres. The numbers on the left are the respective backbone RMSDs. Right, Proline kinks of TM2, TM5, TM6 and TM7. The mean kink is shown with the red dashed line. The kink was calculated every 10th picosecond.

The proline kinks of TM6 in hH₂R_{a2} and of TM7 in hH₂R_i remained stable and were similar to those in the initial models ($31.7 \pm 2.5^\circ$ vs. 29.1° and $42.3 \pm 2.7^\circ$ vs. 40.2° , respectively). The TM6 kink in hH₂R_i rose in the first 10 ns of the simulation from about 38° to 42° , caused by the slight movement of the highly flexible cytoplasmic part of TM6 towards TM7. At 65 ns the kink decreased by about 7° , probably due to the release of a water molecule near Pro249^{6.50} (section 4.3.8.1). The kink in TM7 of hH₂R_{a2} was higher at the beginning of the simulation ($\sim 44^\circ$) compared to the initial model (36.1°) and decreased in the first 25 ns to about 36° . The determination of this kink was not quite accurate for the hH₂R_{a2} since the α -helix of TM7 was permanently present only until Leu287^{7.52}, just two amino acids after Pro285^{7.50} (from Tyr288^{7.53} to Leu291^{7.56}, α -helical structures occurred in $< 10\%$ of time points). Remarkably, the proline kink was not responsible for the inward move of TM7 at the cytoplasmic part during receptor activation. Accordingly, the average TM7 domains of hH₂R_i and hH₂R_{a2} fitted quite well (backbone based RMSD 1.5 Å; Figure 4.25). However, the complete TM7 was shifted, resulting in a more inwards position of its cytoplasmic end. Nevertheless, the change of the kink at the beginning of the hH₂R_{a2} simulation was the result of the distorted α -helix following Pro285^{7.50}. The backbone of Ala289^{7.54} was shifted towards TM1. Whereas H-bonds between the backbone nitrogen of Ala289^{7.54} and the backbone oxygen of Asn284^{7.49} became less frequent, the H-bond frequency between the oxygen of Ala289^{7.54} and the backbone nitrogen of Asn292^{7.57} increased. Furthermore, van der Waals contacts between Ala289^{7.54} and Pro285^{7.50} occurred more frequently (Figure 4.26).

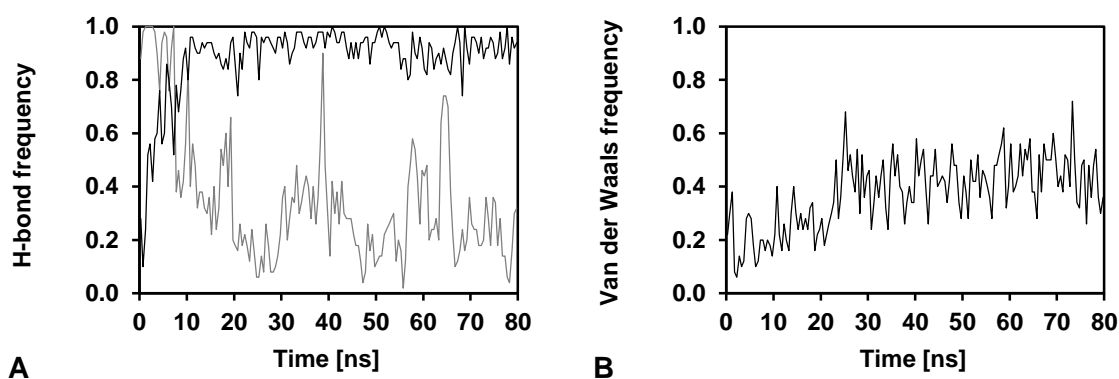


Figure 4.26: Contacts of Ala289^{7.54} in hH₂R_{a2}

A, H-bond frequency between the backbone nitrogen of Ala289^{7.54} and the backbone oxygen of Asn284^{7.49} (grey line) as well as between the backbone oxygen of Ala289^{7.54} and the backbone nitrogen of Asn292^{7.57} (black line). B, Frequency of van der Waals contact between Ala289^{7.54} and Pro285^{7.50}.

In summary, changes of the kinks in TM domains caused by highly conserved prolines seem not to be necessary for the transmission from the inactive to the active hH₂R state. Moreover, the kinks were well comparable when superimposing the respective TM domains of the hH₂R_i and hH₂R_{a2} (Figure 4.25). Instead, receptor activation of the hH₂R is accompanied with movements of complete TMs.

4.3.7 Flexibility of the proteins and comparison of the backbone positions of hH₂R states

The root mean square fluctuation (RMSF) is a measure for the deviation of a special residue or particle averaged over time. It is therefore suitable to determine the mobility of a protein during a simulation (Figure 4.27). Not surprisingly, extra- and intracellular loops, as well as the N- and C-terminus were most flexible. RMSF values based on side chain atoms were generally slightly higher than values based on backbone atoms. A similar curve is presented in Figure 7.1, which shows the crystallographic temperature factors (B-factors; an indicator of the mobility) of an active GPCR conformation of the h β_2 AR (Rasmussen *et al.*, 2011a).

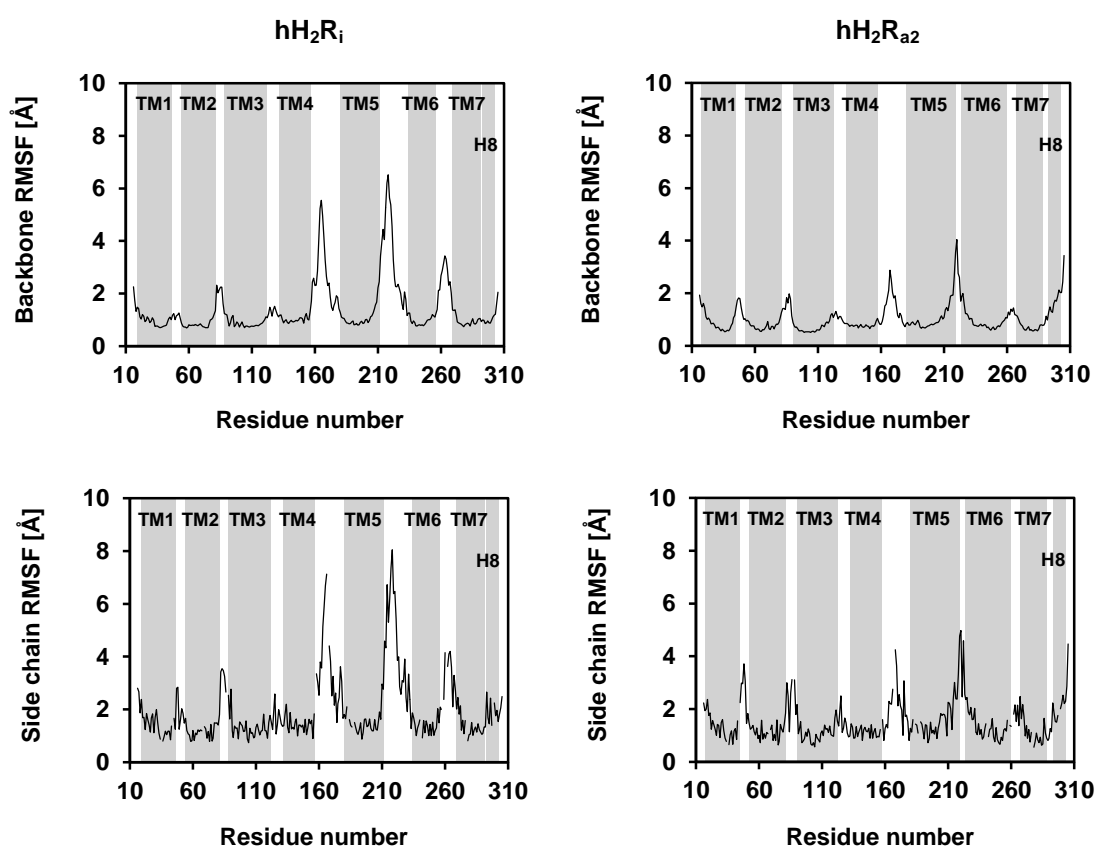


Figure 4.27: Root mean square fluctuation (RMSF) of the hH₂R_i and hH₂R_{a2} during 80 ns of MD simulation

Remarkably, ECL2, ICL3 and ECL3 were more flexible in the hH_2R_i than in the active receptor conformation. Despite similar courses of ECL2 and ECL3, respectively, in the initial models of the hH_2R_i and hH_2R_{a2} (Figure 3.9), during the MD simulation of hH_2R_{a2} constant interactions between ECL2 and ECL3 occurred (section 4.3.9.3), restricting the mobility of ECL3. In the hH_2R_i , ECL3 moved outwards already during the equilibration phase and retained in that position in the production run, broadening the access to the binding site between ECL2 and ECL3 (cf. section 4.3.9.3). In the hH_2R_{a2} , the intracellular ends of TM5 and TM6 as well as ICL3 were stabilized by the parallel arrangement of TM5 and TM6 and contacts to G α CT. In hH_2R_i , the respective parts were able to move freely in the intracellular solvent and therefore exhibited higher mobility.

Analyzing the deviation from the input structure of the MD simulation based on the complete protein backbone, in hH_2R_i stable RMSD values were only obtained temporarily between 20 ns and 40 ns (~ 3.0 Å) and after 45 ns (~ 3.5 Å), respectively (Figure 4.28 A). If only the TM domains are considered, a quite stable RMSD value of about 2.0 Å was reached after 15 ns. However, RMSD values were always lower in the hH_2R_{a2} simulation. The hH_2R_{a2} converged after 15 ns to stable values of about 2.0 Å and 1.2 Å in the case of all backbone atoms and the TM-backbone, respectively (Figure 4.28 B). The higher displacement in the hH_2R_i simulation was possibly caused by different interactions of extracellular loops (section 4.3.9.3), the insertion of the G-protein fragment G α CT in hH_2R_{a2} and by the stabilizing effects of histamine in the binding site of hH_2R_{a2} , connecting TM domains at the extracellular part by generating stable contacts to TM3, TM5, TM6 and TM7 (section 4.3.10).

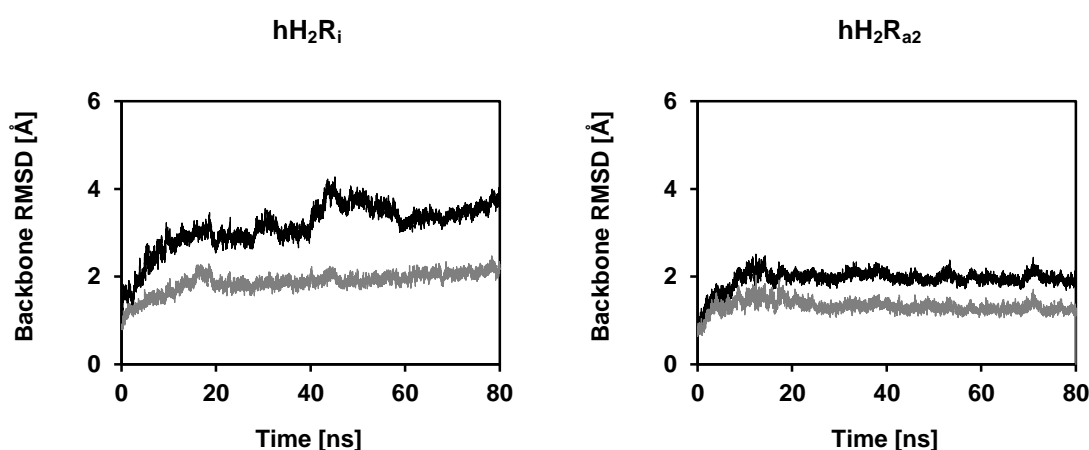


Figure 4.28: Root mean square deviation of the backbone of the hH_2R_i and hH_2R_{a2} from the input structures during 80 ns of simulation time

Root mean square deviations based on backbone atoms (black line) and on TM-backbone atoms (grey line) of the hH_2R_i and hH_2R_{a2} .

In the hH₂R_i, largest displacements of TM domains compared to the input structure of the simulation occurred in TM5 (Figure 4.29). The extracellular part before the kink caused by Pro194^{5.50} (backbone RMSD up to 3.1 Å) moved away from TM4 (~ 1 Å; measured between the C_α atoms of Ile154^{4.61} and Asp186^{5.42}). Thus, the number of interactions at the TM4-TM5 interface was very low (cf. section 4.3.9.2). Instead, hydrophobic amino acids (Leu149^{4.56}, Ile154^{4.61}, Val185^{5.41} and Val189^{5.45}) were in close contact to the lipophilic tails of DPPC. Furthermore, Trp158^{ECL2} interacted with residues at the top of TM4 and TM5 (Ser153^{4.60}, Ile154^{4.61}, Val185^{5.41} and Val189^{5.45}). The center of TM5 (Thr190^{5.46} to Ile205^{5.61}) remained quite stable in its position (RMSD ~ 1.7 Å). The intracellular part of TM5 extending into the cytoplasmic solvent moved away from TM6 (~ 1.5 Å; measured between the C_α atoms of Tyr202^{5.58} and Ala237^{6.38}; RMSD up to 5.3 Å). Together with the disruption of the cytoplasmic end of TM6 this resulted in a reduction of contacts between the cytoplasmic parts of TM5 and TM6. In the hH₂R_{a2} the protruding position of TM6 enabled a parallel positioning of both TMs and thus a high number of interactions (Figure 4.47). TM3 of hH₂R_i moved as a whole slightly into the cytoplasm and at its extracellular part further towards TM6 (by 1.7 Å; measured between the C_α atoms of Val99^{3.33} and Phe251^{6.52}), which also moved towards TM3. Highest RMSD values (~ 2.1 Å) were obtained for amino acids facing the center of the receptor, namely Phe90^{3.24}, Tyr94^{3.28} and Asp98^{3.32}. Such as TM3, also TM4 moved slightly towards the cytoplasm (average RMSD 2.2 Å).

In the hH₂R_{a2} minor changes occurred. At the extracellular part of the receptor TM1 was displaced most (average RMSD 4.6 Å). The α-helix of Thr20^{1.34} to Val23^{1.37} was slightly disrupted during the last equilibration step. However, an ideal 1,4 α-helix was quickly restored in the production run. Furthermore, TM1 and TM2 were separated at their extracellular ends by about 1 Å. The new position of TM1 was closer to the middle of the TM2-TM7 interface and even further away from TM1 in hH₂R_i after superimposing the two hH₂R states (Figure 4.30). ICL3 and the intracellular ends of TM5 and TM6, which both extended into the solvent (Asp211^{5.67} to Val234^{6.35}), moved inwards following the shift of GαCT (section 4.3.10.2) to retain the close interaction between the receptor and the G-protein fragment. Finally, H8 further approached TM1. Comparing the first and the last quarter of the simulation, the frequency of van der Waals contacts rose for Val38^{1.52}-Phe295 (39% → 100%), Val38^{1.52}-Leu302 (7% → 63%), Leu41^{1.55}-Leu302 (41% → 97%), Ala42^{1.56}-Gly298 (19% → 99%), Ala42^{1.56}-Leu302 (65% → 86%) and Val43^{1.57}-Asp294 (7% → 96%).

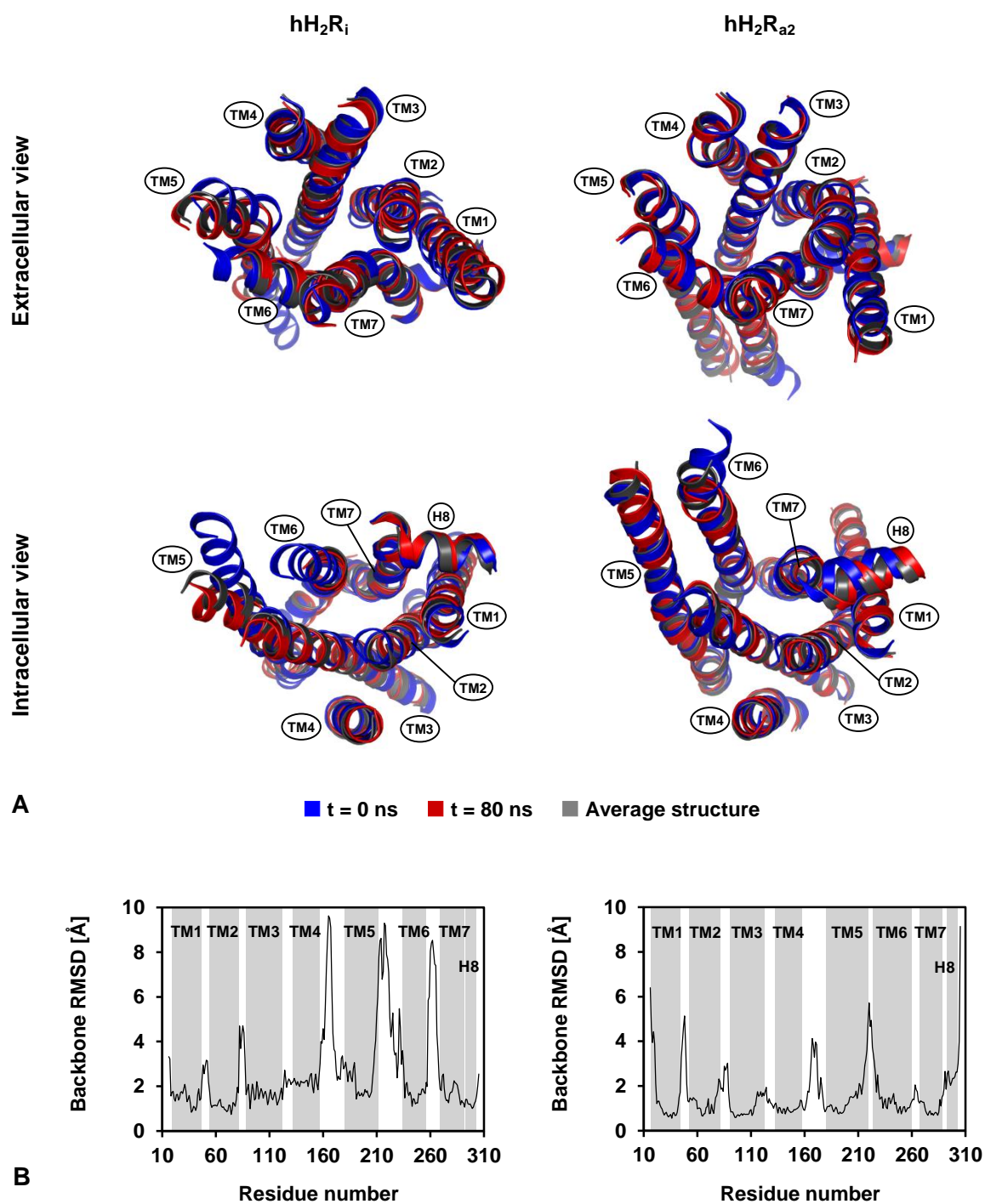


Figure 4.29: Displacement of the backbone of hH₂R_i and hH₂R_{a2} during the 80 ns MD simulation

A, Comparison of the structure at the start of the simulation, the final snapshot after 80 ns and the average structure (average coordinates) of the hH₂R_i and hH₂R_{a2}, respectively. B, Residue based RMSD of all snapshots (8,000 in total) with respect to the structure at the beginning of the simulation.

Based on the superposition of backbone atoms of the common TM domains of the hH₂R_i and hH₂R_{a2}, respectively (Figure 4.30 and Figure 4.31), major differences occurred at the extra- and intracellular end of TM1, which in hH₂R_{a2} was shifted more towards TM7 and located more outwards due to the inwards movement of TM7 and helix 8, respectively. At the intracellular side of the hH₂R_{a2}, TM3 was closer to the core of the receptor and TM4 moved away from TM2, which adopts an outward position in hH₂R_{a2} to allow the inward movement of TM7 and the interaction with GαCT. Furthermore, at the intracellular side the protein fragment from the end of TM5 to the beginning of TM6 was located more outwards in the hH₂R_{a2} as already resulting from the comparison of the initial hH₂R models (chapter 3). At the extracellular side minor changes occurred. TM5 of hH₂R_i was more separated from TM4, and TM6 followed this shift and moved slightly towards TM7. However, the highly flexible loops, especially ECL2, ICL3 and ECL3, deviated most between hH₂R states.

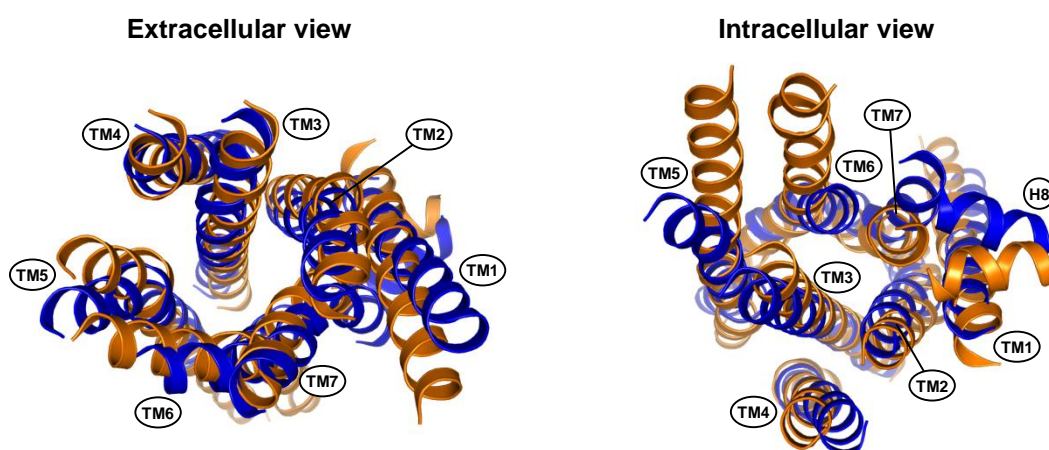


Figure 4.30: Superimposition of the average structures of the hH₂R_i and hH₂R_{a2}

The average backbone coordinates of the hH₂R_i (■) and hH₂R_{a2} (■), respectively, were calculated after superimposing every 10th coordinate file (8,000 in total) of the respective MD simulation to the input structure of the simulation, based on backbone atoms (using *g_rmsf* of GROMACS). The resulting average structures of the two hH₂R states were then superimposed based on their common TM domains. Loops are not shown for reasons of clarity.

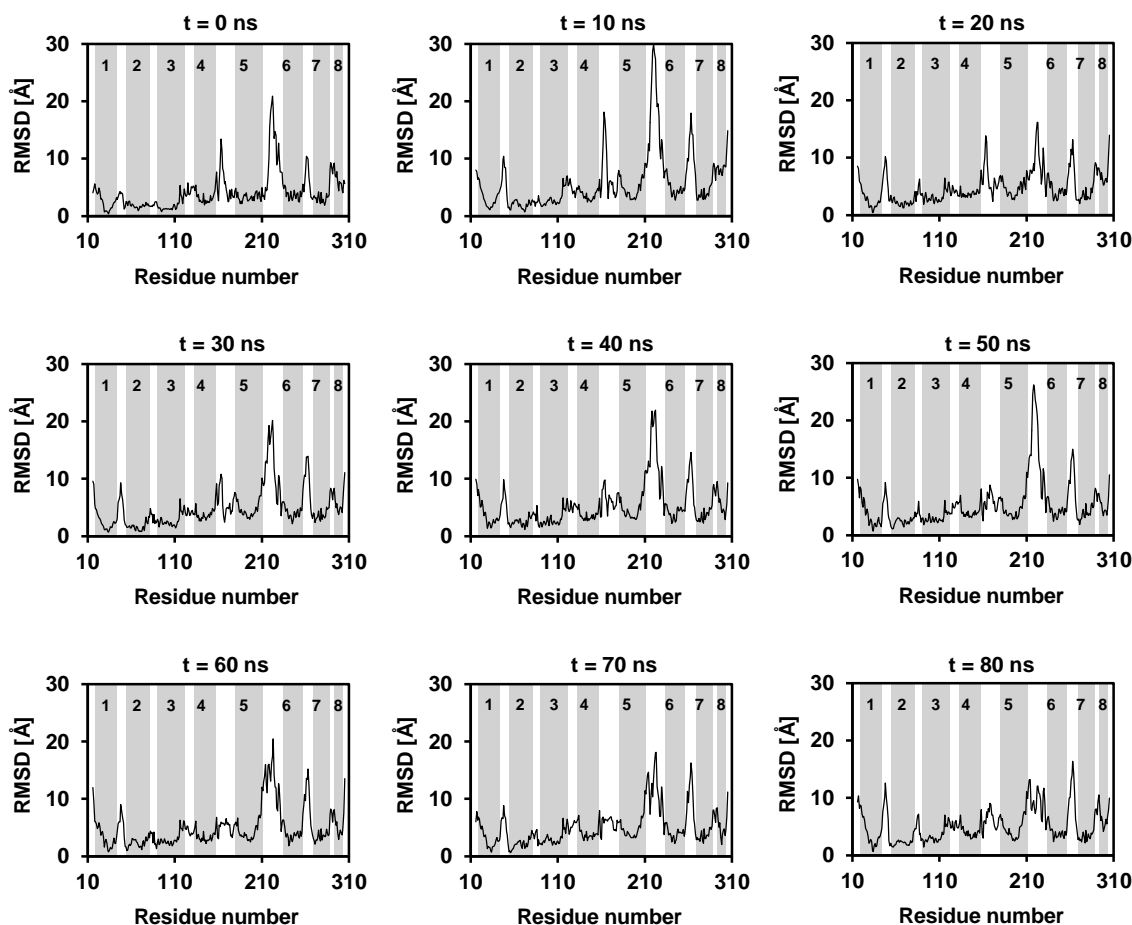


Figure 4.31: Backbone based root mean square deviation of hH₂R states

The structures of the hH₂R_i and hH₂R_{a2} at different simulation times were superimposed based on backbone atoms of their common TM domains. The grey background indicates the common TM domains of the two hH₂R states.

4.3.8 Internal water molecules

4.3.8.1 Stability of putatively conserved water molecules inserted into the initial hH₂Rs

In the initial homology models of the hH₂R_i and hH₂R_{a2}, ten putatively conserved water molecules were inserted (Angel *et al.*, 2009), respectively. Their stability within the TM bundle during the MD simulations was assessed by calculating their distance to the center of mass of the receptor and their hydrogen bonding to the receptor. These two measurements were in good agreement to each other (see example in Figure 4.32 A, B). In the hH₂R_i none of the ten water molecules remained stable in its position during the whole 80 ns simulation period. The water molecule received from the crystal structure of the hβ₂AR (Cherezov *et al.*,

2007) mediates the contacts between Cys246^{6.47}, Tyr250^{6.51} and Val273^{7.38}. It was present in its original position between TM6 and TM7 until 67 ns. Then it moved to the extracellular solvent. Strikingly, this was exactly the time the angle of the proline kink of TM6 decreased from about 40° to 36° (Figure 4.25). The missing water molecule led to the formation of an H-bond between the backbone atoms of Cys246^{6.47} and Tyr250^{6.51}, bringing these residues into closer distance (Figure 4.32 C, D). However, an impact of this conformational change on receptor activation was not obvious.

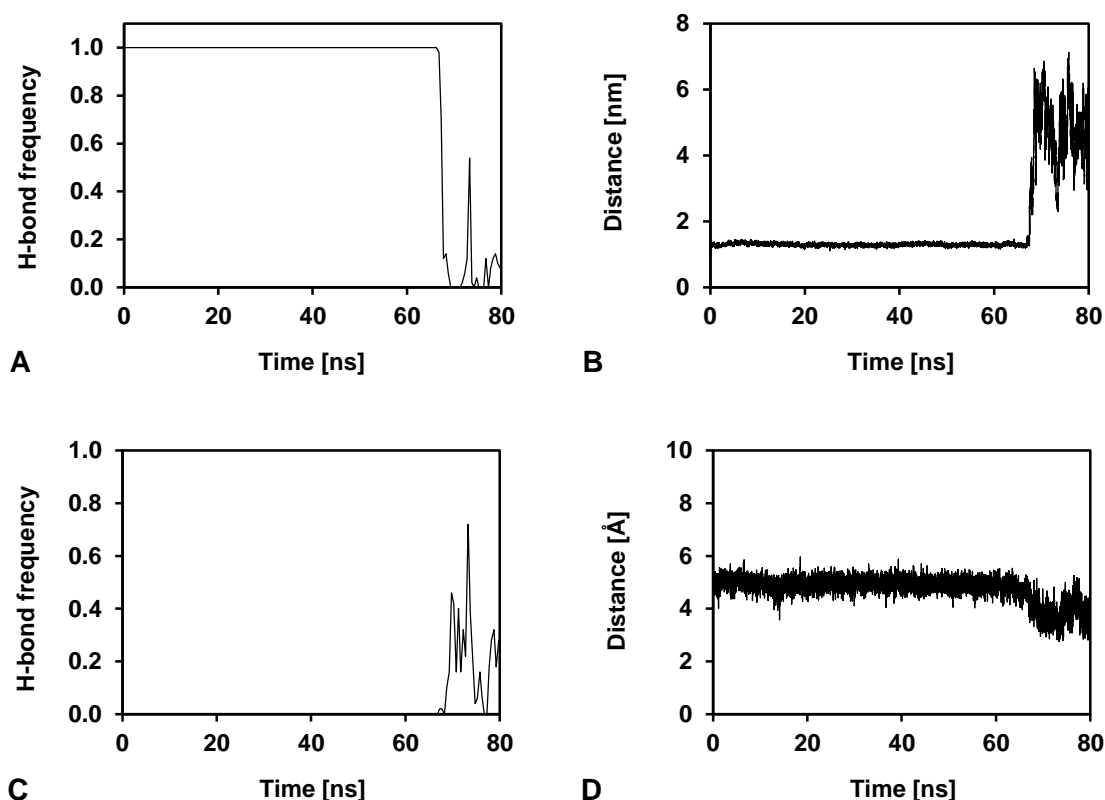


Figure 4.32: Conserved water molecule between TM6 and TM7

A, Frequency of hydrogen bonding of the water molecule inserted near Cys246^{6.47}, Tyr250^{6.51} and Val273^{7.38} to the protein; B, Distance of this water molecule to the center of mass of the receptor. C, Frequency of H-bonds between the backbone oxygen of Cys246^{6.47} and the backbone nitrogen of Tyr250^{6.51}. D, Distance of the backbone oxygen of Cys246^{6.47} and the backbone nitrogen of Tyr250^{6.51}.

Water molecules connecting the side chains of Asn36^{1.50} and Asp64^{2.50} in the initial hH₂R_i model, as well as Val239^{6.40}, Asn280^{7.45} and Asn284^{7.49} were stable until 20 and 10 ns, respectively. They were part of the H-bond network between TM1, TM2, TM3 and TM7 just like five further water molecules leaving the hH₂R_i within the first 3 ns of the simulation. All of

them were replaced by other water molecules from the solvent during the MD simulation. The water molecule inserted between TM4 and TM5 near the orthosteric binding site, obtained from the structure of the $\text{t}\beta_1\text{AR}$ (Warne *et al.*, 2008), was not stable in the receptor and was not replaced. In its initial position it stabilized the kink in TM4 which was introduced by the template $\text{t}\beta_1\text{AR}$ (section 3.3.1). Instead, backbone oxygen atoms of Leu149^{4.56} and Ser150^{5.57} were saturated by forming direct H-bonds to the side chain (frequency of 90%) or backbone (55%) of Ser153^{4.60}. The remaining water molecule in the initial hH_2R_i model (again obtained from the structure of the $\text{h}\beta_2\text{AR}$) connected the side chains of Ser75^{2.61} and Trp275^{7.40} near the extracellular surface. However, diffusion of water molecules from the solvent displaced this molecule and the connection was broken, too. Subsequently the side chains of Ser75^{2.61} (direct H-bonds to Tyr94^{3.28} and Tyr278^{7.43} with a frequency of 83% and 92%, respectively) and Trp275^{7.40} (water mediated H-bond to Gln79^{2.65} in 66% of the simulation time) formed new and stable interactions.

In general conserved water molecules added to the initial model of the hH_2R_{a2} were more stable in their positions. The water molecule between the backbone atoms of Cys246^{6.47}, Tyr250^{6.51} and Val273^{7.38} remained in its position until 59 ns. After that point a direct H-bond was formed, however, without affecting the kink angle at Pro249^{6.50} such as in the inactive state. Remarkably, from 68 ns until the end of the simulation another water molecule from the solvent was present in the respective position, indicating the importance of a water molecule at this site. The connection between TM6 (Cys246^{6.47} and Tyr250^{6.51}) and TM7 (Val273^{7.38}) mediated by these water molecules was weak but stable during the simulation (frequency of ~ 30%). Corresponding to the hH_2R_i , the water molecule added between Ser75^{2.61} and Trp275^{7.40} was quickly replaced by other water molecules from the extracellular solvent, and the side chains did not interact until the end of the simulation. Instead, the two residues formed stable H-bonds with Tyr94^{3.28} (frequency of 90%) and Tyr278^{7.43} (98%) as well as Gln79^{2.65} (95%), respectively. The water molecule near the kink of TM4, saturating mainly the backbone oxygen atoms of Leu149^{4.56} and Ser150^{5.57} was replaced after 25 ns by other water molecules. Stable water mediated H-bonds including the backbone and side chain atoms of Ser153^{4.60} stabilized the kink. The water molecules within the H-bond network of TM1, TM2 and TM7 were stable in their positions at least until 25 ns. All water molecules were subsequently replaced by solvent water, except of two which remained stable in their position until the end of the MD simulation (Figure 4.33). They mediated the contact between Asn36^{1.50} and Asp50^{2.50} and stabilized the intracellular end of TM7 by connecting the backbone nitrogen of Ala289^{7.54} to the H-bond network.

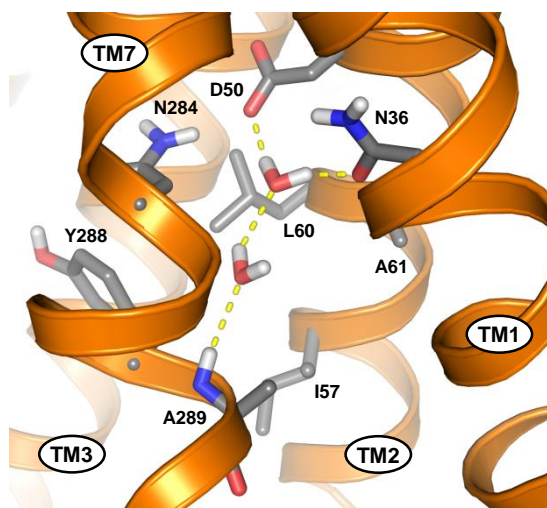


Figure 4.33: Stable conserved water molecules in hH₂R_{a2}

Taken together, the majority of the water molecules inserted in the initial hH₂R models were not stable during the MD simulations. However, most of them were replaced by water molecules from the solvent, suggesting that these putatively conserved water molecules were appropriately placed in the homology models hH₂R_i and hH₂R_{a2}, and that there is a pronounced exchange between water molecules located within and outside the receptor, indicating the permeation of the inter-TM region with water molecules from the solvent.

4.3.8.2 Solvation of polar amino acid side chains within TMs

The permeation of the core of the receptor with water from the solvent was analyzed by calculating the average solvation of amino acids, i.e. the H-bond frequency of each residue with water molecules. Apart from the extra- and intracellular loops, especially residues of TM3, TM5, TM6 and TM7 were solvated despite their location between TM domains (Figure 4.34 A, B), taking into account the space filling orthosteric binding site and the H-bond network around TM7. Of special interest was the solvation of polar amino acid side chains located in TM domains. Their average solvation was similar for the two hH₂R states (70.5% in hH₂R_i and 69.1% in hH₂R_{a2}). However, not all polar side chains were in contact to water molecules during the simulation (Figure 4.34 C, D). Least solvated in the hH₂R_i were side chains of Thr22^{1.36} (0.7%), Ser96^{3.30} (2.3%), Thr103^{3.37} (1.2%), Ser150^{4.57} (4.7%), His155^{4.62} (2.0%) and Tyr192^{5.48} (0.0%). However, all of these buried residues were well saturated with H-bond acceptors and donors of neighboring amino acids (in more than 70% of the simulation H-bonds were detected), except Tyr192^{5.48}, whose hydroxyl moiety was H-bonded in just 26% of all time points analyzed (cf. section 4.3.5.4). In hH₂R_{a2} Thr20^{1.34} (1.2%),

Thr22^{1.36} (1.3%), Thr32^{1.46} (0.4%), Ser96^{3.30} (0.5%), Thr103^{3.37} (3.1%), Ser146^{4.53} (0.0%), Tyr192^{5.48} (1.0%) and Thr252^{6.53} (1.5%) were the most buried residues. Saturation of a polar side chain with H-bonds was again minimal for Tyr192^{5.48} (30%; cf. section 4.3.5.4). All other polar side chains were sufficiently H-bonded as in the case of the hH₂R_i.

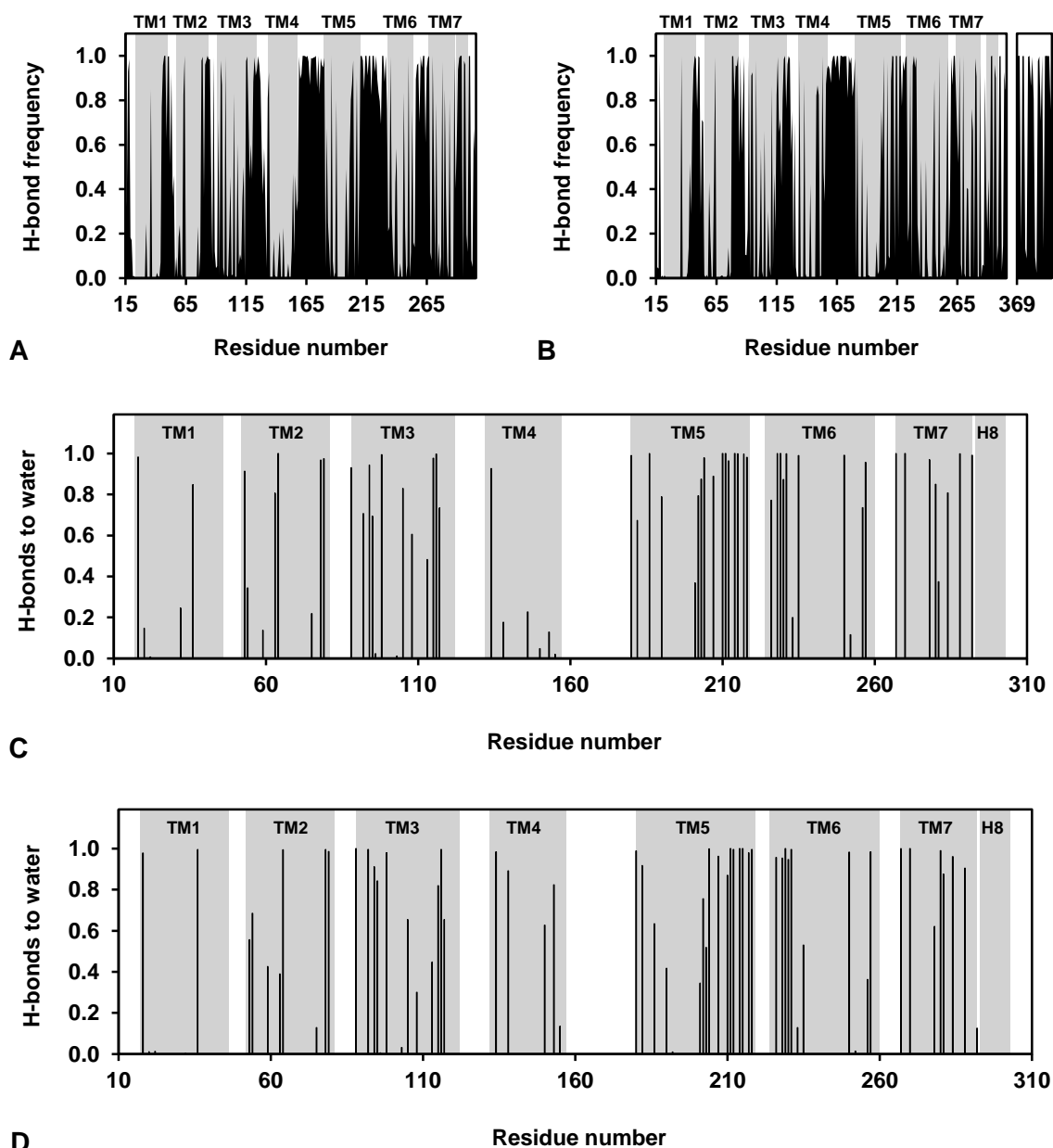


Figure 4.34: Solvation of amino acids in hH₂R_i and hH₂R_{a2}-GaCT

A, B, H-bond frequency of amino acids of the hH₂R_i (A) and hH₂R_{a2}-GaCT (B) to water molecules. The grey background indicates the TM domains. The break of the graph in hH₂R_{a2}-GaCT separates hH₂R_{a2} and GaCT. C, D, Fraction of H-bonds of polar amino acid side chains (Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, Thr, Tyr) to water molecules during the 80 ns MD simulations of the hH₂R_i (C) and hH₂R_{a2} (D). The grey background indicates the maximal TM regions.

Thus, not only parts of the receptor exposed to the extra- and intracellular solvent were surrounded by water molecules, but also residues within TM domains, predominantly polar amino acid side chains. An H-bond network including polar side chains and water molecules was recently reported to be part of the activation mechanism of a constitutively active rhodopsin mutant (Standfuss *et al.*, 2011).

4.3.8.3 Solvation of polar side chains near the H-bond network around TM7

In general, amino acids located between TM1, TM2, TM3 and TM7 and containing a polar side chain such as Thr32^{1.46}, Asn36^{1.50}, Thr63^{2.49}, Asp64^{2.50}, Ser105^{3.39}, Asn280^{7.45}, Ser281^{7.46} and Asn284^{7.49} are potential candidates to contribute to the H-bond network located at the cytoplasmic part of TM7. In the initial models of hH₂R_i and hH₂R_{a2} this network was extended by seven conserved water molecules. However, single receptor conformations do not consider the flexibility related to water mediated H-bonds. Moreover, the interface between TM1, TM2, TM3 and TM7 was able to accept even more water molecules, enabling an extensive hydration of polar amino acid side chains. This solvation, calculated as the frequency of H-bonds between side chains and water molecules, is shown in Figure 4.35.

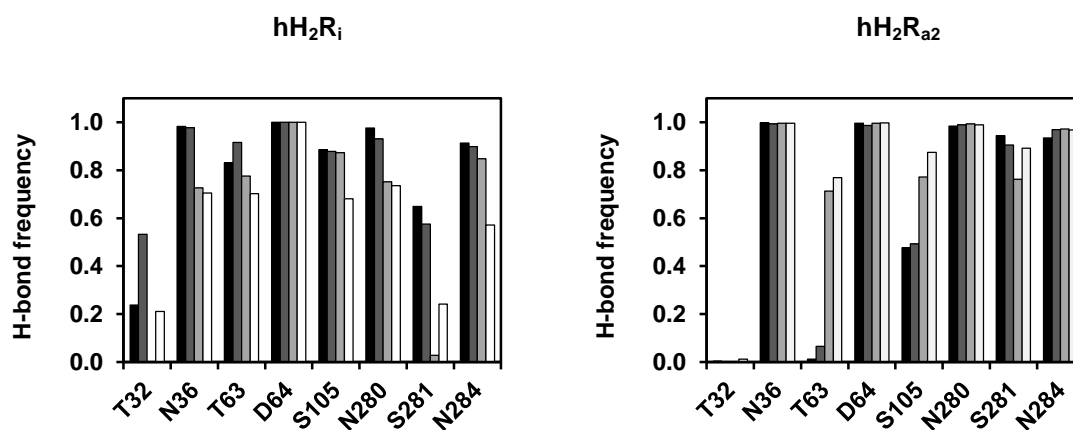


Figure 4.35: Solvation of buried polar amino acid side chains

A, B, Hydrogen bonds of the side chain heteroatoms of amino acids indicated at the abscissa to water molecules during the 80 ns MD simulations of the hH₂R_i and hH₂R_{a2}. Shown is the H-bond frequency for the period 0-20 ns (black bar), 20-40 ns (dark grey bar), 40-60 ns (bright grey bar) and 60-80 ns (white bar).

Least solvated in the hH₂R_i were Thr32^{1.46} and Ser281^{7.46}, in the hH₂R_{a2} Thr32^{1.46}, Thr63^{2.49} and Ser105^{3.39}. During the whole simulation, in the hH₂R_i 194 and in the hH₂R_{a2} 29 water molecules contributed to the H-bond network of these eight amino acid side chains, indicating a higher water diffusion in the inactive hH₂R. The average number of water molecules in the network was slightly higher in the inactive state. In a radius of about ~ 10 Å around the side chain heteroatoms of Asn36^{1.50}, Asp64^{2.50}, Ser105^{3.39} and Asn280^{7.45} in the hH₂R_i and hH₂R_{a2} there were on average 10.0 and 8.6 water molecules, respectively, hence in both hH₂R states more than in the initial models.

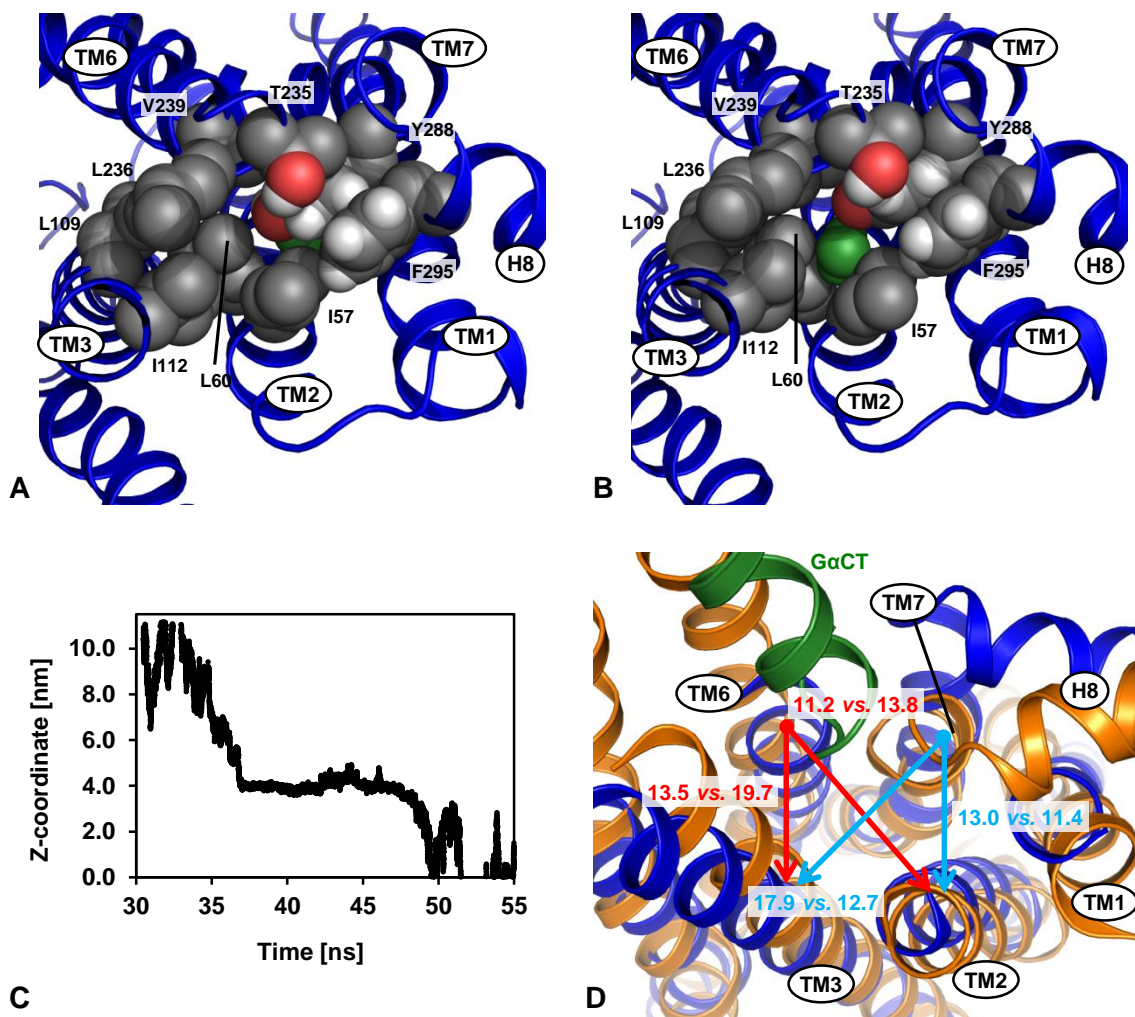


Figure 4.36: Path of water molecule 737 through the hH₂R_i

A, B, View from the intracellular side of the hH₂R_i at time points 47,575 ps (A) and 47,576 ps (B). The water molecule 737 is shown in green. C, All water molecules partly contributing to the H-bond network around TM7 were checked for penetrating the hH₂R_i and the hH₂R_{a2}, respectively, by analyzing their z-coordinates. The z-coordinates of water molecule 737 in the hH₂R_i are shown. D, Superimposition of the average structures (for a description cf. Figure 4.30) of the hH₂R_i (blue ribbons) and hH₂R_{a2} (orange ribbons). The average distances (hH₂R_i vs. hH₂R_{a2}; in Å) during the 80 ns MD simulations are shown for TM2-TM6 and TM3-TM6 (in red) as well as TM2-TM7 and TM3-TM7 (in cyan). The distances were calculated as described in Figure 3.10.

The close contact between TM3 and TM7 at the cytoplasmic part of the hH₂R_{a2} (12.7 Å vs. 17.9 Å in the hH₂R_i) prevented a diffusion of water molecules from the extra- to the intracellular solvent and vice versa. In the hH₂R_i 26 water molecules which contributed to the H-bond network crossed the receptor from one side of the membrane to the other. An example is shown in Figure 4.36. The water molecule migrated from the extra- to the intracellular side of the receptor between Leu60^{2.46}, Ile112^{3.46}, Leu236^{6.37} and Tyr288^{7.53}.

4.3.9 Molecular differences between the inactive and active hH₂R state

4.3.9.1 Detailed analysis of molecular switches and amino acid interactions

Changes occurring during receptor activation of the hH₂R are all directly or indirectly connected to differences in the conformation of the orthosteric binding site between hH₂R_i and hH₂R_{a2}. Whereas the initial homology models described in chapter 3 suggested how the signal is transferred from the binding site to the intracellular part of the receptor, the result of the 80 ns MD simulation of the hH₂R_i and hH₂R_{a2} proved interactions to be stable, only transient or arising during the simulation.

The crucial difference between the ‘inactive’ and ‘active’ binding site occurred at its bottom (Figure 4.37). In the hH₂R_i the side chain of Cys102^{3.36} fills the gap between Phe191^{5.47}, Phe243^{6.44} and Phe251^{6.52}. This was already the case in the initial hH₂R_i model. The 80 ns simulation proved this interaction to be stable (frequency of van der Waals contacts 73%, 92% and 61%, respectively). Cys102^{3.36} prevented the contacts of Ile106^{3.40} with Phe251^{6.52} and Trp247^{6.48}. In the active state Cys102^{3.36} points towards TM7, forming a water mediated H-bond with the backbone of Gly277^{7.42} (continuously during the simulation with a frequency of 35%), an interaction lacking in the initial hH₂R_{a2} model, which did not contain a water molecule in that position. This contact is a first link between the H-bond network and the altered binding site of the active state (see below). A connection between the corresponding residues in the β₂AR (Val117^{3.36} and Gly315^{7.42}; Rasmussen *et al.*, 2011a) and opsin (Gly121^{3.36} and Ala295^{7.42}; Scheerer *et al.*, 2008) was not observed. However, in the crystal structure of the partly activated human adenosine A_{2A} receptor a water molecule is co-crystallized between Thr88^{3.36} and Ser277^{7.42} (PDB ID 2YDO; Lebon *et al.*, 2011). Furthermore, the side chain switch of Cys102^{3.36} enabled the approximation of Ile106^{3.40} to Phe251^{6.52} and Trp247^{6.48} (van der Waals contact frequencies of 100%) below the binding site of hH₂R_{a2}. The latter two residues keep Ile106^{3.40} in a position towards TM6. This allowed the change of the side chain orientations of Ile106^{3.40} and Phe243^{6.44}, recently identified to be an important transition step in some GPCRs (Rasmussen *et al.*, 2011a; Valentin-Hansen *et al.*, 2012b). After receptor activation, Phe243^{6.44} was close to TM5 (van der Waals contact

frequency to Pro194^{5.50} of 64%). Instead, in hH₂R_i Ile106^{3.40} was close to TM5 (van der Waals contacts to Leu195^{5.51} and Met198^{5.54} with a frequency of 86% and 98%, respectively), and Phe243^{6.44} was packed between Trp247^{6.48}, Ser105^{3.39} and Ile106^{3.40} (frequency > 94%, respectively).

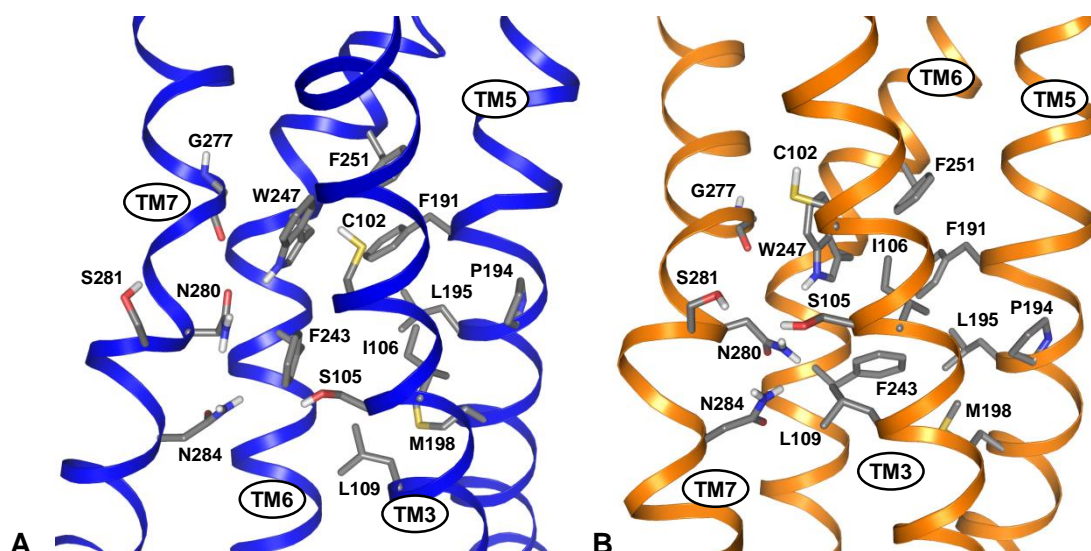


Figure 4.37: Conformational differences below the orthosteric binding site of hH₂R states

Shown are the final snapshots (at 80 ns) of the hH₂R_i (A) and hH₂R_{a2} (B).

Furthermore, in hH₂R_i Phe243^{6.44} was able to contact TM7, forcing the side chain of Asn280^{7.45} by van der Waals contacts (frequency of 85%) towards TM2. By this, the H-bond network at the cytoplasmic part of the receptor is affected, too. In the initial hH₂R_i model Asn280^{7.45} pointed towards TM6. Due to the switch of Phe243^{6.44} in hH₂R_{a2} towards TM5, the van der Waals contacts to Ser105^{3.39} (frequency of 94% in hH₂R_i) and Asn280^{7.45} disappeared. Hence, both residues were able to form new interactions influencing the H-bond network. Instead, in direction TM2/TM7 Leu109^{3.43} flanked Phe243^{6.44}, contributing to restrain it in its 'active' position. Furthermore, in both hH₂R states Phe243^{6.44} is in close contact to Leu195^{5.51}. However, the switch of Phe243^{6.44} forced the break of the Ile106^{3.40}/Leu195^{5.51} contact (frequency of 86% in the inactive vs. 0% in the active state) and a separation of TM3 and TM5 in hH₂R_{a2} to enable the new side chain position of Phe243^{6.44}. The distance between the C_α atoms of Ile106^{3.40} and Leu195^{5.51} increased on average from 6.8 Å in hH₂R_i to 9.8 Å in hH₂R_{a2}. Between the inactive and active β₂AR conformation, the respective increase was about 1.5 Å. To avoid an overlap of the Phe243^{6.44} side chain with

Ile106^{3.40}, Phe191^{5.47}, Pro194^{5.50} and Leu195^{5.51}, the outward position of TM6 is inevitable. Phe243^{6.44} is just the anchor below which the outward movement of TM6 and deviations between hH₂R_i and hH₂R_{a2} became obvious (increasing RMSD values; Figure 4.31).

Additionally, the separation of TM3 from TM5 enabled an ‘upward’ shift of Met198^{5.54}, just below the phenyl moiety of Phe243^{6.44}. Met240^{6.41} moved upwards too and interacted with Met198^{5.54} (van der Waals contacts frequency of 100%). The Met198^{5.54}-Met240^{6.41} contact contributed to keep TM6 in an outward position. The contact between Met240^{6.41} and Leu109^{3.43} (frequency 90% in hH₂R_i) disappeared in hH₂R_{a2} (increased distance between TM3 and TM6). The changes of Met198^{5.54} and Met240^{6.41} were closely linked to the side chain switch of Tyr202^{5.58} (Figure 4.38).

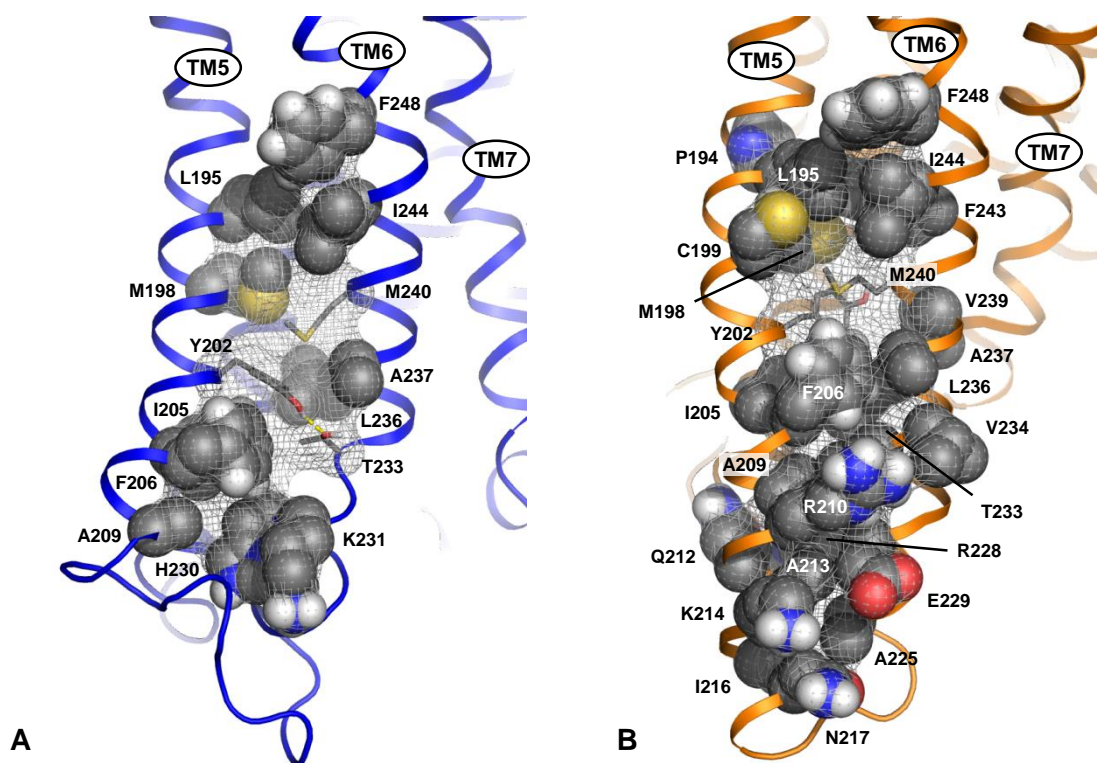


Figure 4.38: The interface between TM5 and TM6 of hH₂R_i and hH₂R_{a2}

Shown are the final snapshots (at 80 ns) of hH₂R_i (A) and hH₂R_{a2} (B). Tyr202^{5.58} and Met240^{6.41} which exchange their position during receptor activation are shown as sticks, as well as Thr233^{6.34} in hH₂R_i which helps to maintain Tyr202^{5.58} in its ‘inactive’ position.

First, the upward shift of Met240^{6.41} enabled that Tyr202^{5.58} may occupy the gap between TM3 and TM6 below Phe243^{6.44}, and second, the upward move of Met198^{5.54} was necessary for the new position of the Tyr202^{5.58} side chain in hH₂R_{a2} where it contacted Ser113^{3.47} (H-bond frequency of 66%), Arg116^{3.50} of the DRY motif (direct and water mediated H-bond frequency 79%) and Tyr288^{7.53} of the NPxxY(x)_{5,6}F motif (water mediated H-bond frequency 53%; Figure 4.39).

These interactions maintained the position of TM6, disrupted the ionic lock and the hydrophobic barrier and stabilized the inward position of TM7. In mutational studies at the β_2 AR Tyr223^{5.58} was shown to play a role in stabilizing the active receptor conformation (Valentin-Hansen *et al.*, 2012a), and in rhodopsin, the mutation of Tyr223^{5.58} to phenylalanine decreases the lifetime of the Meta II intermediate (Goncalves *et al.*, 2010).

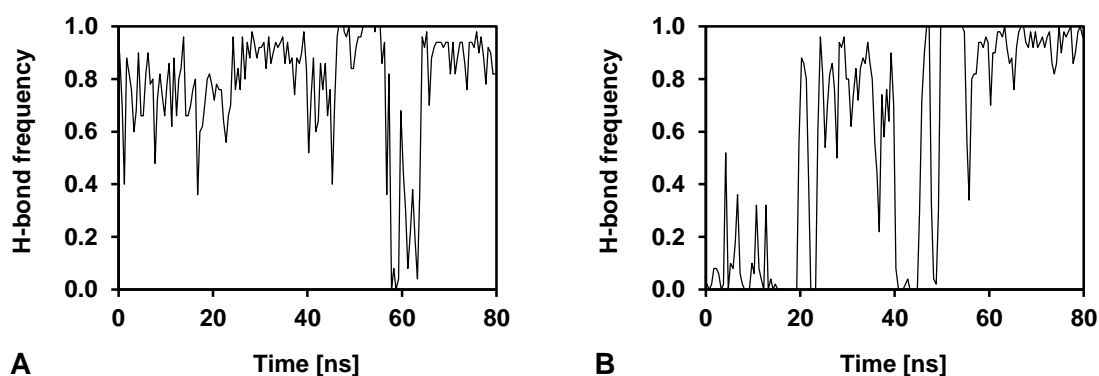


Figure 4.39: Contacts of Tyr202^{5.58} to Arg116^{3.50} and Tyr288^{7.53} in hH₂R_{a2}

A, Summarized H-bond frequency (direct and water mediated) between the side chains of Tyr202^{5.58} and Arg116^{3.50}; B, Water mediated H-bond frequency between the side chains of Tyr202^{5.58} and Tyr288^{7.53}.

An inward shift of TM5 as observed in the initial hH₂R_{a2} model and the structure of the active β_2 AR (Rasmussen *et al.*, 2011a) seemed not to be important for the activation mechanism. Furthermore, a rotamer switch of Trp247^{6.48} which was proposed for some GPCRs (Shi *et al.*, 2002) and which was not observed in the initial hH₂R model did also not occur during the 80 ns MD simulation. Only slight differences were observed (mean \pm SD for the χ_1 and χ_2 dihedral angle: $-70.8^\circ \pm 7.3^\circ$ and $95.4^\circ \pm 11.8^\circ$ in hH₂R_i and $-61.3^\circ \pm 9.2^\circ$ and $106.1^\circ \pm 11.7^\circ$ in hH₂R_{a2}; Figure 4.40). Taken together, a rotamer change of Trp247^{6.48} probably does not contribute to receptor activation of the hH₂R.

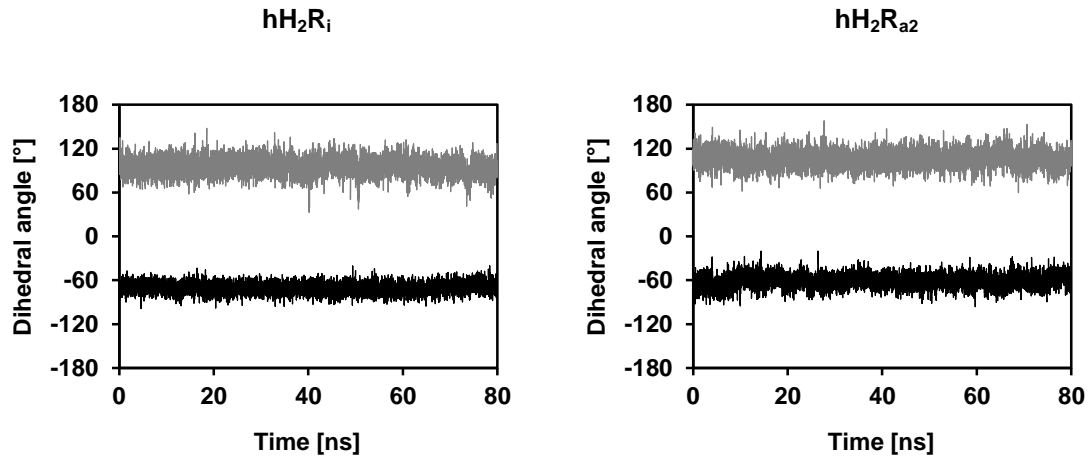


Figure 4.40: Dihedral angles of Trp247^{6.48}

χ_1 (black line) and χ_2 (grey line) dihedral angle of Trp247^{6.48} of the hH₂R_i and hH₂R_{a2}.

The stable position during the 80 ns simulation of a downwards extended TM6 in hH₂R_{a2} enabled a slightly closer distance between TMs 5 and 6 (~ 0.5 Å), a parallel alignment of both helices and a reordering of the TM5-TM6 interface, leading to more contacts between the two TMs than in hH₂R_i (Figure 4.38 and cf. section 4.3.9.2). Due to the flexibility of TM5 and TM6 during the MD simulation, even more residues were part of the TM5-TM6 interface than in the initial homology model of hH₂R_{a2}. Overall 26 amino acids belonged to the interface in hH₂R_{a2}, among which Ala209^{5.65}, Gln212^{5.68}, Ile216^{5.72}, Asn217^{5.73}, Arg228^{6.29}, Val234^{6.35}, Ala237^{6.38} and Ile244^{6.45} were not identified in the initial model. Most of the contacts were hydrophobic interactions. Polar contacts occurred between Arg210^{5.66} and Thr233^{6.34} (H-bond between the side chains in 46% of the simulation), Gln212^{5.68} and Arg228^{6.29} (direct and water mediated H-bond between the side chains in 27% and 24% of the simulation, respectively) and between Glu229^{6.30} and Arg116^{3.50} (see below). In contrast, in hH₂R_i only 14 amino acids of TM5 and TM6 contribute to the interface. All contacts were hydrophobic, except the H-bond between the side chains of Tyr202^{5.61} and Thr233^{6.34} (frequency of 54%) which contributed to keep Tyr202^{5.61} in its ‘inactive’ position.

Strikingly, the ionic lock which is composed of Asp115^{3.49} and Arg116^{3.50} of the highly conserved D/ERY motif and Glu229^{6.30} located at the cytoplasmic end of TM6 (Vogel *et al.*, 2008) showed a relevant change during the simulation of the hH₂R_i (Figure 4.41). Unlike in the initial model indicating an open ionic lock such as observed in the template, the t β ₁AR (Warne *et al.*, 2008), the contact between Arg116^{3.50} and Glu229^{6.30} was temporarily present during the simulation. A hydrogen bond from the side chain of Arg116^{3.50} to the backbone of Glu229^{6.30} occurred in 40.7% of the simulation (Figure 4.42 A). The distance between the

charged side chain of Arg116^{3.50} and Glu229^{6.30} was in 22% of the simulation smaller than 5 Å, indicating a moderate ionic interaction. In the period from 15 to 35 ns, the distance was increased (Figure 4.42 B). Furthermore, the side chain methylene groups of both amino acids were in close contact to each other in 25% of the simulation (Figure 4.42 C). The temporary presence of the ionic lock already supposed in chapter 3 is in accordance to experimental results from a rH₂R mutant (see chapter 3; Alewijnse *et al.*, 2000) and is also discussed for the β₂AR (Dror *et al.*, 2009). In the inactive state further interactions of the Arg116^{3.50} side chain to amino acids of TM6 occurred, namely van der Waals contacts with Thr233^{6.34} and Leu236^{6.37} (frequency of 67% and 84%, respectively; the latter was already present in the initial model) and direct and water mediated H-bonds to the backbone of Ala232^{6.33} (frequency of 20%) and the side chain of Thr235^{6.36} (20%). In summary, the interactions of Arg116^{3.50} did essentially contribute to the strong TM3-TM6 contact in the inactive state (cf. Figure 4.47 in section 4.3.9.2).

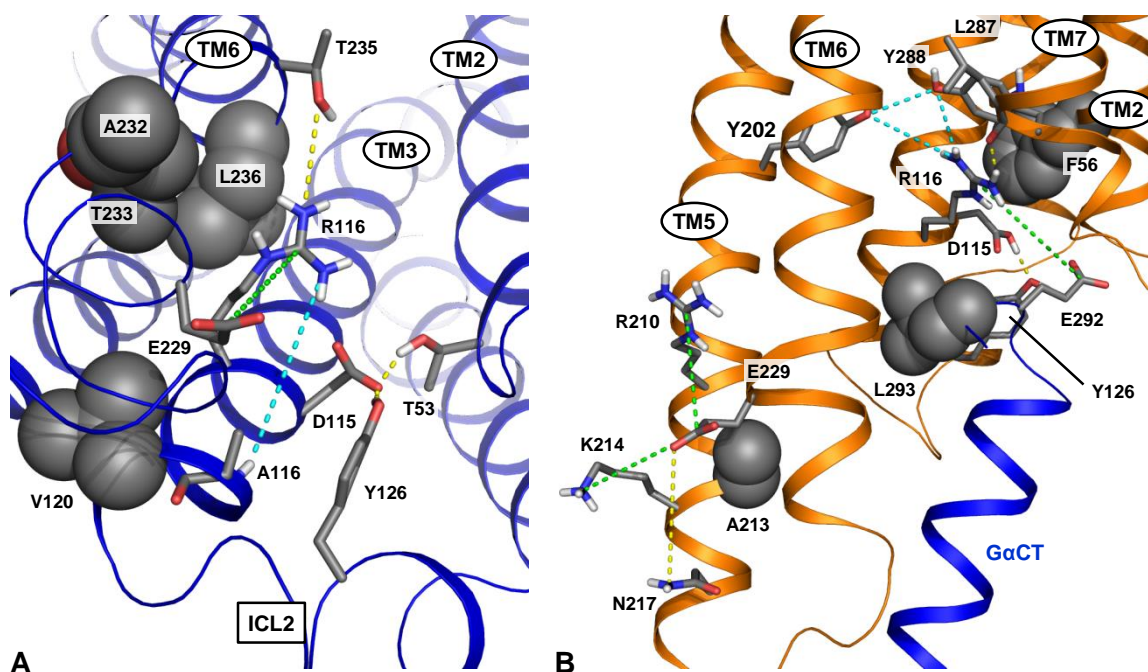


Figure 4.41: The ionic lock in hH₂R_i and its disruption in hH₂R_{a2}

Amino acids forming the ionic lock and their neighboring residues are shown in the hH₂R_i (A; blue ribbons) and hH₂R_{a2} (B; orange ribbons with GαCT in blue ribbons). H-bonds, water mediated H-bonds and ionic interactions are indicated with yellow, cyan and green dashed lines, respectively. Amino acids performing polar contacts are shown in sticks and residues contributing to van der Waals interactions in spheres.

In contrast, in hH₂R_{a2} the distance between Arg116^{3.50} and Glu229^{6.30} was too large during the entire simulation (about 20 Å between the positively charged side chain moieties; Figure 4.42 D). In hH₂R_{a2} Arg116^{3.50} interacted with the C-terminal fragment of the G_{saS} protein (section 4.3.10.2), Tyr202^{2.38} (Figure 4.39), the backbone of Leu287^{7.52} (H-bond frequency of 70%) and Tyr288^{7.53} (see below). Thus, in the hH₂R_{a2} Arg116^{3.50} played an important role in connecting TM3 and TM7 as well as bridging TM5 with TM7, both keeping TM6 in an outward position. The contacts to TM7 were not observed in the initial model.

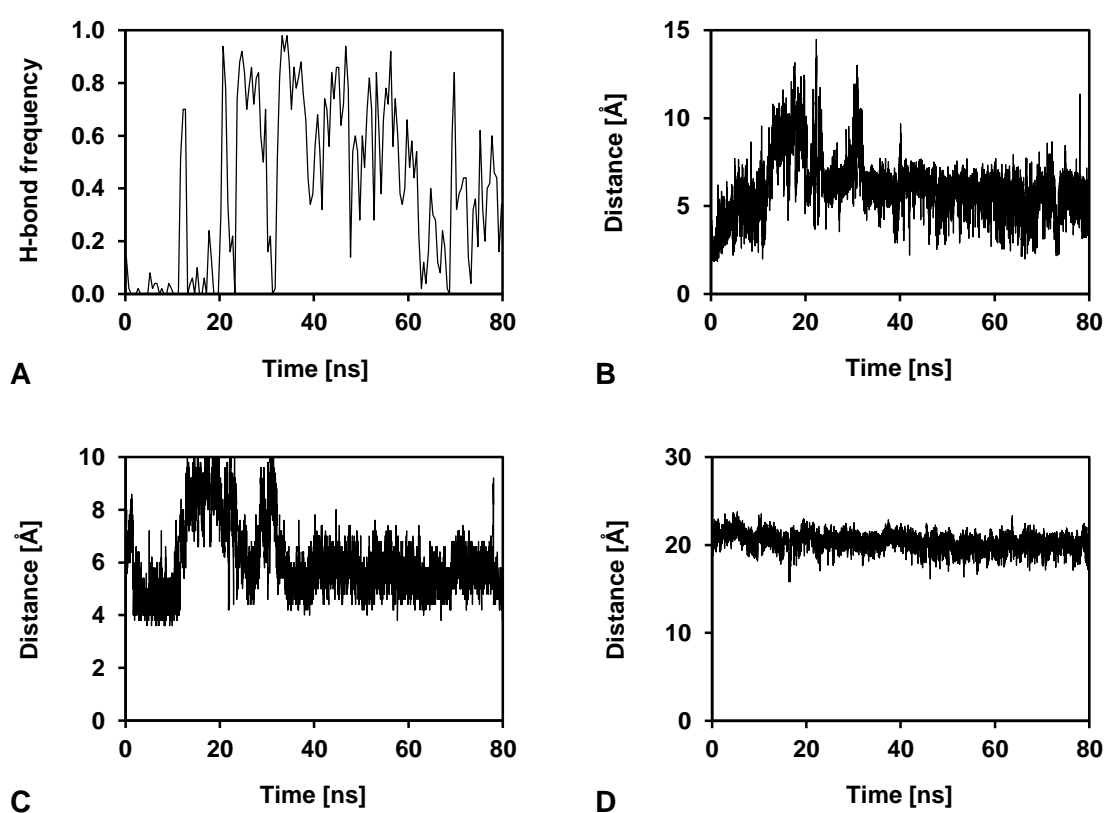


Figure 4.42: Ionic lock – contacts between Arg116^{3.50} and Glu229^{6.30}

A-C, MD simulation of the hH₂R_i; A, H-bond frequency between the side chain of Arg116^{3.50} and the backbone of Glu229^{6.30}. B, Distance between the closest charged side chain heteroatoms of Arg116^{3.50} and Glu229^{6.30}. C, Distance between hydrophobic side chain atoms of Arg116^{3.50} and Glu229^{6.30}; the distance calculation between hydrophobic sites was truncated at 10 Å. D, Distance between closest side chain heteroatoms of Arg116^{3.50} and Glu229^{6.30} in the MD simulation of the hH₂R_{a2}.

Besides Arg116^{3.50} also the neighboring Asp115^{3.49} had a specific function in hH₂R_i. As already obvious in the initial models, it was connected to Arg116^{3.50} via a salt bridge (frequency of 87%) as well as via direct and water mediated H-bonds (in 20% and 54% of the simulation, respectively), keeping Arg116^{3.50} in its 'inactive' position. In both receptor states contacts to TM2 were formed (H-bond to Thr53^{2.39} in hH₂R_i and van der Waals contact to Phe56^{2.42} in hH₂R_{a2} in 92% and 84% of the simulation, respectively). Furthermore, the contact to Tyr126^{ICL2} was retained in both hH₂R states (H-bond frequency > 98%).

In hH₂R_i, Glu229^{6.30} was not only involved in the ionic lock. It formed additional contacts with TM3 (van der Waals interactions with Ala119^{3.53} and Val120^{3.54} in 54% and 84% of the simulation, respectively) and with the neighboring Ala225^{6.26}, Thr226^{6.27}, Ile227^{6.28} and Arg228^{6.29}, maintaining the side chain in an appropriate position to interact with Arg116^{3.50}. In contrast, in hH₂R_{a2} Glu229^{6.30} exclusively interacted with TM5. Polar contacts were formed with Arg210^{5.66} (frequency of 73%, in 50% of the simulation an ionic interaction between the two side chains occurred as already observed in the initial hH₂R_{a2} model), Lys214^{5.70} (frequency of 64%, in 58% ionic interactions), Asn217^{5.73} (frequency of direct and water mediated H-bonds of 50%), and van der Waals contacts with Ala213^{5.69} (93%). Thus, whereas Glu229^{6.30} was a key amino acid in hH₂R_i for connecting TM3 and TM6, it was one of the key residues for the strong TM5-TM6 contact in hH₂R_{a2} (cf. Figure 4.47 in section 4.3.9.2).

The disruption of the packing of hydrophobic amino acids of TM2, TM3 and TM6 in the core of the receptor (the so called hydrophobic barrier; Li *et al.*, 2004; Standfuss *et al.*, 2011) upon receptor activation was already observed by comparing the initial homology models hH₂R_i and hH₂R_{a2}. Like the cleavage of the ionic lock, it was the result of the TM6 movement. The average distance, e.g. between C_α atoms of Leu109^{3.43} and Leu236^{6.37}, increased from 7.8 Å in hH₂R_i to 13.1 Å in hH₂R_{a2}. During the 80 ns MD simulation the hydrophobic barrier was composed of Leu60^{2.46}, Leu109^{3.43}, Ile112^{3.46}, Ser113^{3.47}, Leu236^{6.37}, Val239^{6.40} and Met240^{6.41} (Figure 4.43). In hH₂R_i, van der Waals contacts were formed by Leu60^{2.46} with Val239^{6.40} (frequency of 80%), by Leu109^{3.43} with Leu236^{6.37}, Val239^{6.40} and Met240^{6.41} (frequency of 77%, 94% and 90%, respectively), and by Leu236^{6.37} with Ile112^{3.46} and Ser113^{3.47} (frequency of 99% and 66%, respectively). The contact between Leu236^{6.37}, Thr53^{2.39} and Ile57^{2.43} described in chapter 3 was not stable. In hH₂R_{a2}, the contacts between TM3 and TM7 were closer than in hH₂R_i. Interactions were formed between Leu60^{2.46} and Asn284^{7.49} (van der Waals interaction in 92% of the simulation), Leu109^{3.43} and Asn284^{7.49} as well as Tyr288^{7.53} (39% and 100%, respectively), Ile112^{3.46} and Tyr288^{7.53} (100%), and between Ser113^{3.47} and Tyr288^{7.53} (68%; additionally direct and water mediated H-bonds between the side chains). Thus, all residues of TM2 and TM3 being part of the hydrophobic

barrier in hH₂R_i were connected to Asn284^{7.49} and/or Tyr288^{7.53} in the active state. They contributed to keep TM7 in its 'active', inward position. Furthermore, Leu236^{6.37} interacted with GαCT (section 4.3.10.2) and amplified the interaction with TM5 by a hydrophobic contact to Phe206^{5.62} (frequency of 69%). Also other exclusive interactions of TM6 residues being part of the hydrophobic barrier in hH₂R_i strengthened the TM5-TM6 contact in hH₂R_{a2}, e.g. Val239^{6.40} approached Tyr202^{5.58} (van der Waals contact in 79% of the simulation) and Met240^{6.41} Leu195^{5.51} (72%).

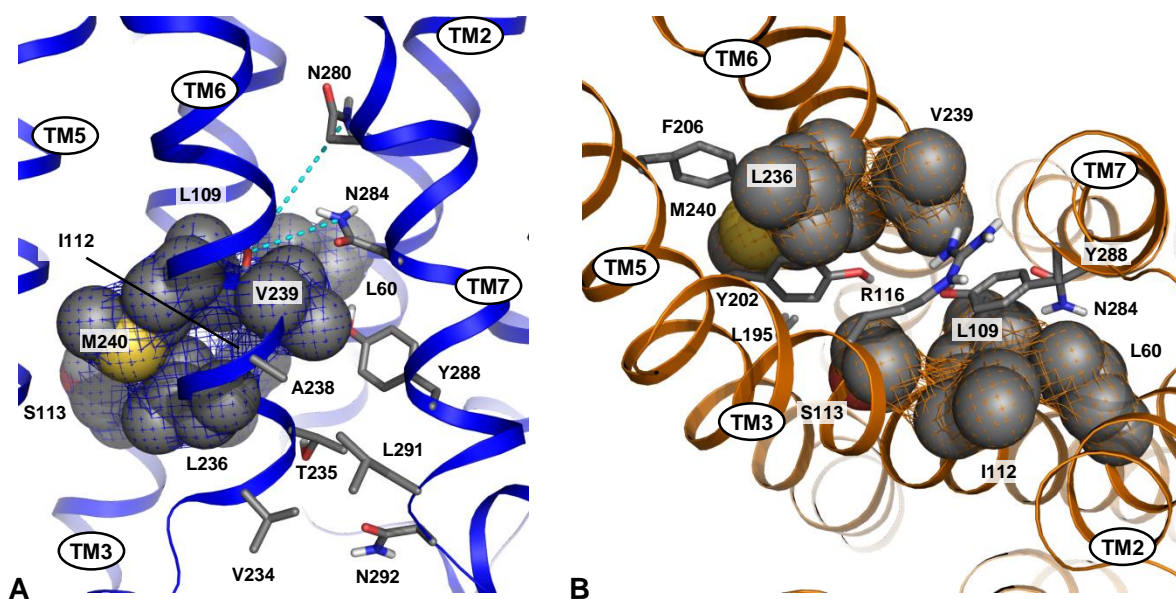


Figure 4.43: Hydrophobic barrier and the interface between TM6 and TM7 at the cytoplasmic part of hH₂R states

Side view of the hH₂R_i (A; blue ribbons) and view from the cytoplasm of the hH₂R_{a2} (B; orange ribbons). Amino acids of the hydrophobic barrier are shown as spheres, and remaining residues connecting TM6 and TM7 (A) or TM5, TM6 and TM7 (B), respectively, are shown as sticks.

The relative positions of TM6 and TM7 in hH₂R_i were stabilized by the van der Waals contacts between Val234^{6.35} and Leu291^{7.56}, Thr235^{6.36} and Tyr288^{7.53}, Thr235^{6.36} and Asn292^{7.57}, Ala238^{6.39} and Leu291^{7.56} as well as Val239^{6.40} and Asn284^{7.49} (in 69%, 99%, 85%, 88% and 60% of the simulation, respectively). The water mediated H-bonds from the backbone of Val239^{6.40} to Asn280^{7.45} and Asn284^{7.49} were only temporarily but present until the end of the 80 ns simulation (26% and 39%, respectively). All these TM6-TM7 contacts occurred only in hH₂R_i. In hH₂R_{a2}, the opening of the TM3-TM6 contact enabled the inward position of TM7. Tyr288^{7.53} is of special importance as reported for several GPCRs (Ahuja and Smith, 2009; Rasmussen *et al.*, 2011a; Scheerer *et al.*, 2008; Vanni *et al.*, 2010; Wess

et al., 2008). In both hH₂R states, the position of Tyr288^{7.53} is stabilized by hydrophobic contacts to Ile57^{2.43}, Leu60^{2.46}, Val239^{6.40}, Asn284^{7.49} and Leu287^{7.52}. However, the different arrangement of TM7 at the cytoplasmic part of the receptors led to unique interactions mainly with TM1 and H8 in the hH₂R_i, as well as TM3 in the hH₂R_{a2}. In the inactive state, the side chain position of Tyr288^{7.53} close to TM1 enabled a stable interaction with Val39^{1.53} (van der Waals contact in 99% of the simulation). Additionally, a van der Waals contact to Ala61^{2.47} and a water mediated H-bond to the side chain of Asp64^{2.50} were present in about 20% of the simulation. A hydrophobic interaction with Thr235^{6.36} was stable (99%). Crucial for the 'inactive' position of Tyr288^{7.53} in hH₂R_i were the contacts to Phe295^{H8} (π-π interaction, frequency of 100%) and Tyr299^{H8} (75%), as well as the interactions with Pro285^{7.50} and Asn292^{7.57} (frequency of 50% and 28%, respectively). In the hH₂R_{a2}, van der Waals contacts to Leu109^{3.43} (frequency of 100%), Ile112^{3.46} (100%), Ser113^{3.47} (68%) and Arg116^{3.50} (63%) were formed. Furthermore, the hydroxyl group of Tyr288^{7.53} formed direct and water mediated H-bonds to the side chain of Ser113^{3.47} (frequency of 28% and 14%, respectively), as well as water mediated H-bonds to the side chains of Arg116^{3.50} (37%) and Tyr202^{5.38} (57%) and to the backbone of Val239^{6.40} (45%). Thus, the rearrangement of Tyr288^{7.53} during receptor activation, already observed in the initial hH₂R_{a2} model, was proved to be stable during the 80 ns MD simulations. In hH₂R_{a2} it was crucial for the TM3-TM7 contact at the intracellular side of the receptor and contributed to maintain the outward position of TM6 by contacts to Tyr202^{5.38} and Arg116^{3.50}. In hH₂R_i, Tyr288^{7.53} was linked to the H-bond network between TM1, TM2, TM3 and TM7. Water mediated H-bonds were formed by the hydroxyl group of Tyr288^{7.53} with the side chains of Asp64^{2.50} and Asn284^{7.49} (frequency of 18% in both cases). Despite the low frequency, these contacts were often formed during the simulation (Figure 4.44). Thus, the H-bond network around TM7 accounted for the outwards position of TM7 in hH₂R_i by directing the side chain of Tyr288^{7.53} towards TM2.

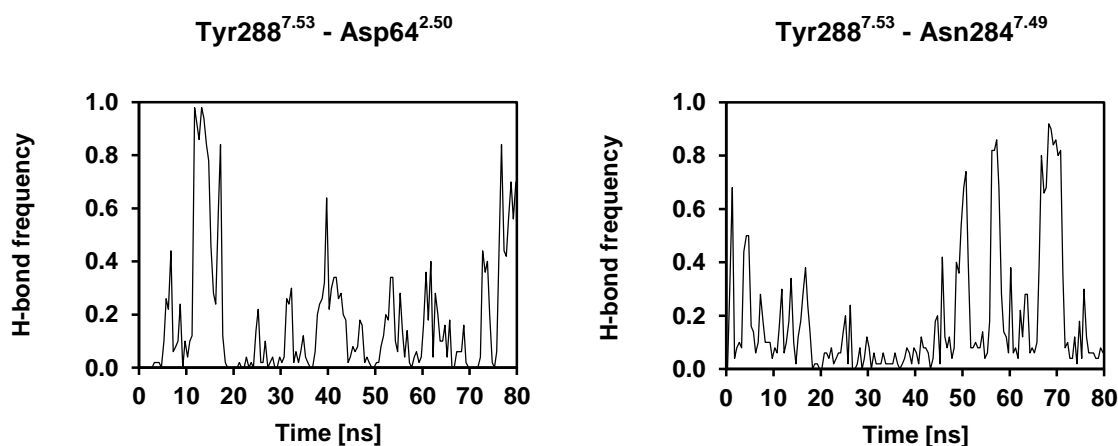


Figure 4.44: Water mediated H-bonds from Tyr288^{7.53} to Asp64^{2.50} and Asn284^{7.49} in hH₂R_i

The H-bond network (Figure 4.45) is furthermore directly influenced by amino acids below the orthosteric binding site of the hH₂R_{a2}. Besides the interaction between Cys102^{3.36} and Gly277^{7.42} (cf. Figure 4.37), Ser105^{3.39} was able to contact Asn280^{7.45}, Ser281^{7.46} and Asn284^{7.49} in overall 58% of the simulation (direct and water mediated H-bonds). Therefore, Cys102^{3.36} and Ser105^{3.39} were able to pull the cytoplasmic part of TM7 towards TM3. In hH₂R_i, the conformation of Asn280^{7.45}, adjusted by the side chain position of Phe243^{6.44} below the binding site, was stabilized in its ‘inactive’ position by direct H-bonds of the side chain to Asp64^{2.50} (frequency of 53%) and to its own backbone (34%). In contrast, in hH₂R_{a2} Asn280^{7.45} is released from this constraint and formed H-bonds with Ser105^{3.39} (18%, mainly in the second part of the simulation), Cys246^{6.47} (17%, decreasing in the second part of the simulation) and Asn284^{7.49} (direct and water mediated H-bond in 46% and 87% of the simulation, respectively). The side chain of Asn284^{7.49} is thereby kept in its close position to TM3 (water mediated H-bond to Ser105^{3.39} in 20% of the simulation). Thus, Asn280^{7.45} seemed to be a further trigger for tearing TM7 in its ‘active’, inward position. Whereas the extracellular part of TM7 fitted quite well between hH₂R_i and hH₂R_{a2}, the shift of TM7 towards TM3 in hH₂R_{a2} ‘started’ at Asn280^{7.45} (cf. Figure 4.37). The crucial role of Asn280^{7.45} and the impact of the bottom of the orthosteric binding site became not clear from the initial homology models of hH₂R_i and hH₂R_{a2}.

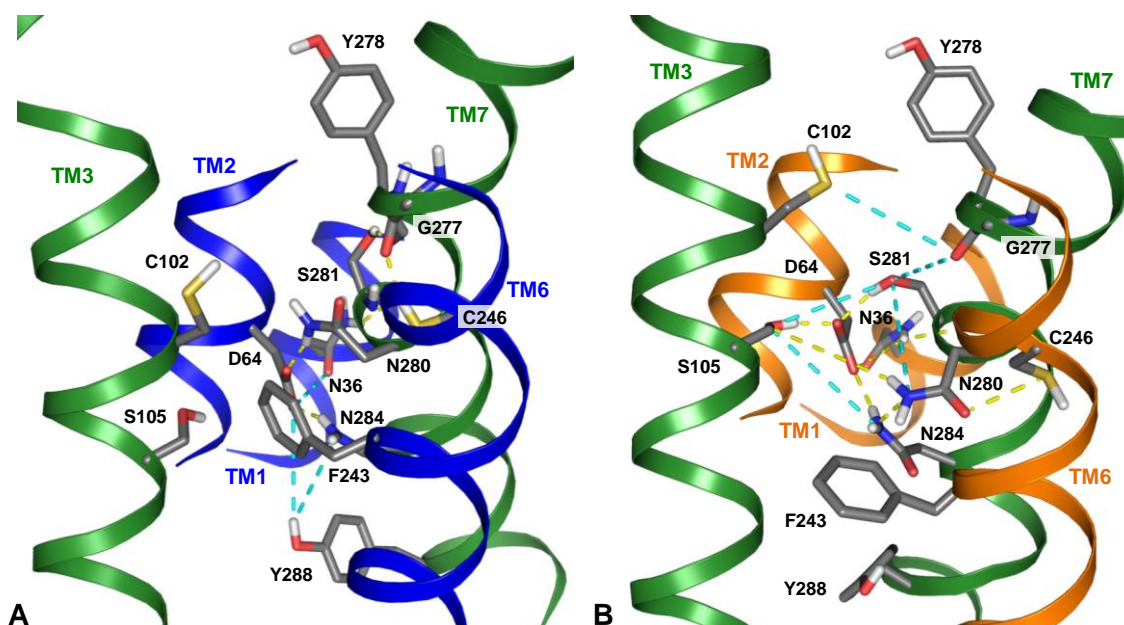


Figure 4.45: H-bond network between TM1, TM2, TM3 and TM7

Most important amino acids of the H-bond network around TM7 are shown in the final snapshot (80 ns) of hH₂R_i (A) and hH₂R_{a2} (B). TM1, TM2 and TM6 (A, blue ribbons; B, orange ribbons) are not shown completely and TM4 and TM5 are omitted for reasons of clarity. H-bonds and water mediated H-bonds are indicated with yellow and cyan dashed lines, respectively.

In the hH₂R_{a2}, the side chains of Asp64^{2.50} and Ser105^{3.39} were linked by a direct H-bond (frequency of 62%; in hH₂R_i only by a water mediated H-bond in 63% of the simulation), leading to a shift of the side chain of Asp64^{2.50} towards TM3. As a consequence, Asn284^{7.49} which was well H-bonded to Asp64^{2.50} in hH₂R_i and hH₂R_{a2} (frequency of 79% and 95%, respectively) followed this shift and thus possibly contributed to the more inward position of the cytoplasmic part of TM7 in the hH₂R_{a2} (Figure 4.45). The movement was further facilitated by the strong interaction between Asp64^{2.50} and Ser281^{7.46} in the hH₂R_{a2} (frequency of 92% vs. 29% in hH₂R_i). In hH₂R_i, the hydroxyl group of Ser281^{7.46} formed an additional water mediated H-bond to the side chain of Asn280^{7.45} (frequency of 24%). However, these contacts of Ser281^{7.46} decreased from the first 20 ns to the last 20 ns of the simulation from 65% to 2% and from 55% to 1%, respectively. Instead, Ser281^{7.46} developed a direct H-bond to the backbone oxygen of Tyr278^{7.43} (frequency increase from 30% to 78%). This stabilized the kink in TM7 and is similar to other GPCRs, e.g. the inactive β_2 AR state (Cherezov *et al.*, 2007), where the hydroxyl group of the corresponding Ser319^{7.46} is in close distance to the backbone oxygen of Tyr316^{7.43} (3.1 Å). By contrast, in the hH₂R_{a2} the hydroxyl group of Ser281^{7.46} is continuously H-bonded to the carboxyl group of Asp64^{2.50} (frequency of 92%), as well as by water mediated H-bonds to the side chains of Ser105^{3.39} (24%) and Asn280^{7.45} (42%) and to the backbone of Gly277^{7.42} (32%). Thus, in the active hH₂R state Ser281^{7.46} fulfills an important role in connecting TM7 to TM2 and TM3 at the cytoplasmic part (Figure 4.47).

Thr63^{2.49} had only a minor role in the H-bond network. Its side chain pointed towards TM3 and is averted from the interface with TM7. Thus, it contributes to the connection TM2-TM3 by H-bonds to Ser105^{3.39} (frequency of about 34% in the two hH₂R states) and Asn108^{3.42} (32%, only in hH₂R_{a2}) and stabilizes the side chain of Asp64^{2.50} by water mediated H-bonds (68% and 29% in hH₂R_i and hH₂R_{a2}, respectively). The backbone oxygen of Ser281^{7.46} is connected to the side chain of Asn36^{1.50} (about 80% of the simulation in both states). Thus, Asn36^{1.50} stabilized the kink in TM7. Furthermore, in both receptor states Asn36^{1.50} stabilized the side chain of Asp64^{2.50} in its respective position by water mediated H-bonds (in 72% and 94% of the hH₂R_i and hH₂R_{a2} simulation, respectively). In the inactive state, Asn36^{1.50} helps to restrain Asn284^{7.49} in a more downward position by direct and water mediated H-bonds (12% and 19%, respectively), decreasing the H-bond frequency with Asp64^{2.50} (see above).

In both hH₂R states the side chain of Thr32^{1.46} did not contribute to the stabilization of other polar side chains in the interface and to bridging TM domains. In both hH₂R states it is mainly H-bonded to the backbone of Leu28^{1.42} (frequency of 40% and 58%, respectively). Nevertheless, according to the multiple sequence alignment analysis described in chapter 7, in 60% of the aminergic GPCRs in position 1.46 a threonine is present. The role of a polar

side chain in this region of the hH₂R remains unclear from the 80 ns MD simulations. Only the backbone oxygen is important for adjusting the side chain position of Asn36^{1.50} via H-bonds in about 90% of the simulations of both hH₂R states.

4.3.9.2 Analysis of TM-TM contacts

The analysis of contacts between TMs is meaningful for detecting differences in hH₂R states (cf. chapter 3). Differences in the number of interactions at the extracellular part of the receptor were obvious around TM7 (in the active state stronger interactions with TM1, TM2 and TM3, and weaker contacts with TM6), between TM4 and TM5 (weak in hH₂R_i) as well as TM2 and TM3 (weaker in hH₂R_{a2}). In general, more changes occurred at the intracellular side. The majority of deviations between hH₂R_i and hH₂R_{a2} could be ascribed to molecular switches occurring during hH₂R activation (Table 4.3). These switches described above are important for rhodopsin and the β_2 AR, too (Trzaskowski *et al.*, 2012). In the active state the contact of TM3 with TM5 and TM6 is weaker and with TM4 and TM7 stronger than in hH₂R_i. TM7 is more attached to TM2 and less connected to TM6 in the active state, where also the TM5-TM6 link is closer. Most TM-TM contacts remained stable during the 80 ns MD simulations of both states.

Extracellular part

One of the most striking differences at the extracellular side of hH₂R_i and hH₂R_{a2} is the position of TM1. Caused by the h β_2 AR as template, TM1 of hH₂R_{a2} is closer to TM7 than in hH₂R_i (Figure 4.30). During the MD simulation TM1 of hH₂R_{a2} further moves towards TM7 (section 4.3.7). The different TM1-TM2 interface and the closer distance to TM7 (about 2 Å at the extracellular ends) resulted in additional contacts. However, the majority of interactions (hydrophobic and van der Waals contacts) occurred between identical residues in both hH₂R states. Some contacts in hH₂R_{a2} were more stable during the 80 ns simulation than in hH₂R_i, e.g. in the TM1-TM7 interface the hydrophobic contacts between Leu28^{1.42} and Trp275^{7.40} (frequency of 98% vs. 22%), Leu28^{1.42} and Ala282^{7.47} (53% vs. 24%) as well as between Ile29^{1.43} and Trp275^{7.40} (97% vs. 18%). Only in the hH₂R_{a2}, Leu28^{1.42} interacted with Ala279^{7.44} (frequency of 65%). The type and the number of interactions are similar at the TM2-TM7 interface in the extracellular part of the hH₂R_i and hH₂R_{a2} (Figure 4.47). Differences concerned mainly the side chain orientation of Gln79^{2.65}, which was more flexible in hH₂R_i and positioned towards the backbone of Ala271^{7.36} enabling an H-bond (frequency of 29%). In contrast, in the active state an H-bond to the side chain of Trp275^{7.40} which was less frequent in hH₂R_i (66% vs. 95%) restrained Gln79^{2.65} in a more downward position,

resulting in an interaction with the hydroxyl group of Tyr278^{7.43} via a water mediated H-bond (frequency of 36%). The average distance between C α atoms of Gln79^{2.65} and Tyr278^{7.43} is smaller for the hH₂R_{a2} (12.7 Å vs. 14.2 Å). The 'downward' shift of Gln79^{2.65} in hH₂R_{a2} is probably a reason for the larger kink in TM2 and may stabilize the rearrangement of the cytoplasmic part of TM7. Additional interactions between TM3 and TM7 in hH₂R_{a2} (2.8 vs. 0.9 in hH₂R_i) were mainly caused by the side chain switch of Cys102^{3.36}, enabling the interaction with Gly277^{7.42} (Figure 4.37) and the H-bond between Asp98^{3.32} and Tyr278^{7.43} (frequency of 100%). Presumably, the latter was enabled by the more inward position of the extracellular part of TM2 (larger proline kink, Figure 4.25) which forced the side chain of Tyr278^{7.43} closer to TM3.

The TM2-TM3 interface at the extracellular side is characterized by a crossing of the TM segments, amplified by the proline kink of TM2 (Figure 4.30) which is more pronounced in the hH₂R_{a2} (Figure 4.25). Moreover, TM3 of hH₂R_i moved towards the core of the receptor during the simulation. Together with a slightly closer distance between TM2 and TM3 this enabled more contacts between TM2 and TM3 in hH₂R_i (Figure 4.47). Apart from a direct H-bond from the side chain of Ser75^{2.61} and a water mediated H-bond from the side chain of Gln79^{2.65}, respectively, to the hydroxyl group of Tyr94^{3.28} (present in hH₂R_i and hH₂R_{a2} with a frequency of more than 83% and 49%, respectively), all contacts occurred between hydrophobic side chains of TM2 and TM3. The shift of TM3 in hH₂R_i strengthened the contacts near the extracellular surface (the average number of interactions increased from 1.2 in the first quarter of the simulation to 5.2 in the last quarter), as obvious for the hydrophobic contacts between Phe74^{2.60} and Ile93^{3.27} (frequency rising from 53% to 100%), Phe74^{2.60} and Leu97^{3.31} (11% to 60%), Ile77^{2.63} and Phe90^{3.24} (18% to 100%), Tyr78^{2.64} and Phe90^{3.24} (25% to 100%) as well as Tyr78^{2.64} and Cys91^{3.25} (0% to 44%), all absent in hH₂R_{a2}. These interactions stabilized the 'inactive' position of TM3 and therefore contributed to the crucial differences between both hH₂R states. The striking difference of the TM4-TM5 interface is related to the movement of TM5 in hH₂R_i. Instead of direct contacts between the two TMs in hH₂R_i they were mainly stabilized by close contacts to DPPC molecules (section 4.3.7). Although the TM5-TM6 interfaces also varied because of the altered TM5 position, no implications on receptor activation were evident.

Intracellular part

The difference of the TM1-TM2 interfaces between hH₂R_i and hH₂R_{a2} becomes obvious by the more outward position of TM1 in the active state (Figure 4.30). However, most interactions occurred between the same pairs of amino acids. In hH₂R_{a2}, an additional van der Waals contact between Cys40^{1.54} and Leu65^{2.51} (frequency of 47%) and a stronger contact between Val39^{1.53} and Ile57^{2.43} (98% vs. 31% in hH₂R_i) was observed. They contribute to stabilize the intracellular part of TM1 in the 'active' position, which allowed the inward move of TM7 upon receptor activation. Instead, in hH₂R_i Cys40^{1.54} is close to Val58^{2.44} (van der Waals contact frequency of 86%), and a direct H-bond from the backbone of Val43^{1.57} to the side chain of Asn54^{2.40} (frequency of 88%) stabilized both TMs in their relative positions. Asn54^{2.40} formed additional polar contacts to ICL1 (Arg50 and Asn51) and helix 8 (Asp294). The outward move of TM1 at the TM1-TM7 interface in hH₂R_{a2} was compensated by the inward move of TM7, thus resulting in similar contacts between the two receptor states. A crucial difference occurred for Val39^{1.53}. In the inactive state it was linked to Tyr288^{7.53}, helping to lock it in its 'inactive' position (van der Waals contact frequency of 99%), whereas in hH₂R_{a2} the inward position of TM7 allowed a contact to Ala289^{7.54} (frequency of 99%).

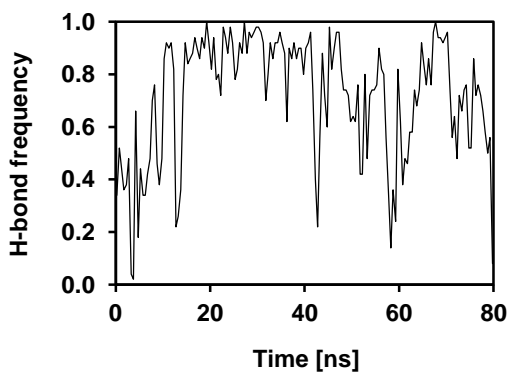
The TM2-TM3 interface was nearly identical in both hH₂R states. Only the interaction between Thr53^{2.36} and Asp115^{3.49} was not possible in hH₂R_{a2} (cf. section 4.3.9.1, ionic lock), and the contact between Phe56^{2.42} and Asn108^{3.42} was more pronounced in hH₂R_i (van der Waals contact frequency 66% vs. 19% in hH₂R_{a2}) due to the shift of TM3 towards the cytoplasm. The closer distance between TM2 and TM4 in hH₂R_i (Figure 4.30) led to the stabilization of the H-bond between Ser59^{2.45} and Trp143^{4.50} (frequency of 97% vs. 33%) and of the van der Waals contact between Leu66^{2.52} and Trp143^{4.50} (frequency of 83% vs. 45%). Both interactions contributed to the more outward position of TM2 in hH₂R_i and allowed the inward move of TM2 during receptor activation, presumably necessary for the interaction with the G-protein.

The interaction between TM2 and TM6 at the intracellular part of the hH₂R_i was caused by the packing of hydrophobic side chains in the center of the receptor (hydrophobic barrier, contact of Leu60^{2.46} with Val239^{6.40}). The stronger contact between TM2 and TM7 in hH₂R_{a2} was also caused by Leu60^{2.46}, which was released from the hydrophobic barrier, and additionally by the van der Waals contact Ile57^{2.43}-Ala289^{7.54} (frequency of 87%) enabled by the inward position of TM7. Additional differences occurred in the H-bond network. Thus, the interfaces TM2-TM6 and TM2-TM7 were mainly affected by alterations in the well-known switches of the hydrophobic barrier and the H-bond network.

Table 4.3: Links between molecular switches and differences in the TM-TM interfaces at the intracellular parts of the hH₂R_i and hH₂R_{a2}

Molecular switch	Affected TM-TM interface
Binding site	TM3-TM6, TM3-TM7
Tyr202 ^{5.58}	TM5-TM7
TM6 outward position	TM3-TM6, TM5-TM6, TM6-TM7
Ionic lock	TM3-TM6, TM5-TM6
Hydrophobic barrier	TM2-TM6, TM2-TM7, TM3-TM6, TM3-TM7, TM5-TM6, TM6-TM7
H-bond network	TM2-TM7, TM3-TM7, TM6-TM7
TM7 rearrangement with Tyr288 ^{7.53}	TM1-TM7, TM3-TM7, TM5-TM7, TM6-TM7

The stronger contact between TM3 and TM4 in hH₂R_{a2} was the result of a different side chain conformation of Arg134^{4.41}. Whereas in hH₂R_i it pointed to the lipid head groups, in hH₂R_{a2} the side chain was only flanked by lipids and additionally pointed to Met111^{3.45}, Leu114^{3.48} and Cys118^{3.52}. In all cases water mediated H-bonds were formed (summarized frequency of 72%; Figure 4.46).

**Figure 4.46: Frequency of water mediated H-bonds of Arg134^{4.41} to TM3**

Ser113^{3.47} played different roles in the two receptor states. In hH₂R_i it contributed to the TM3-TM5 connection by direct and water mediated H-bonds (frequency of 86%) as well as by van der Waals contacts (frequency of 62%) to Thr201^{5.57} and was furthermore close to Ile205^{5.61} (van der Waals frequency of 81%). Instead, in hH₂R_{a2} Ser113^{3.47} performed an H-bond to the hydroxyl group of Tyr202^{5.58} (frequency of 66%), keeping the Tyr202^{5.58} side chain in its position between TM3 and TM6. Furthermore, Ser113^{3.47} was close to Met198^{5.54} (H-bond and van der Waals frequency of 34% and 93%, respectively). The position of Met198^{5.54} was important for keeping TM6 in the outward position. The higher number of TM3-TM5 interactions in hH₂R_i was due to the differences below the binding site (Ile106^{3.40} and Leu109^{3.43}) and a closer contact between TM3 and TM5 at their intracellular ends (~ 3 Å).

The majority of contacts between TM3 and TM6 was already described in section 4.3.9.1 (cf. changes below the binding site, the hydrophobic barrier and the ionic lock). Additionally,

Ala119^{3.53} and Val120^{3.34} contacted residues in hH₂R_i belonging to the intracellular end of TM6 (Ala225^{6.26}, Arg228^{6.29}, Glu229^{6.30} and His230^{6.31}). The contacts between TM3 and TM7 in hH₂R_{a2} and their role – cleavage of the hydrophobic barrier and rearrangement of TM7 including the switch of Tyr288^{7.53} – were also described in section 4.3.9.1. The difference of the TM5-TM6 interface at the cytoplasmic part between hH₂R_i and hH₂R_{a2} during the 80 ns MD simulation was related to the outward position of TM6 in hH₂R_{a2} (Figure 4.38). Due to the high flexibility of TM5 and TM6 at their cytoplasmic ends in hH₂R_i (Figure 4.27), the average number of contacts during the MD simulation between the two domains varied between 13.9 and 17.7. The single contact between TM5 and TM7 in hH₂R_{a2} refers to the water mediated H-bond between Tyr202^{5.58} and Tyr288^{7.53}. The differences of the TM6-TM7 interface between hH₂R_i and hH₂R_{a2} were also described in section 4.3.9.1 (cf. hydrophobic barrier).

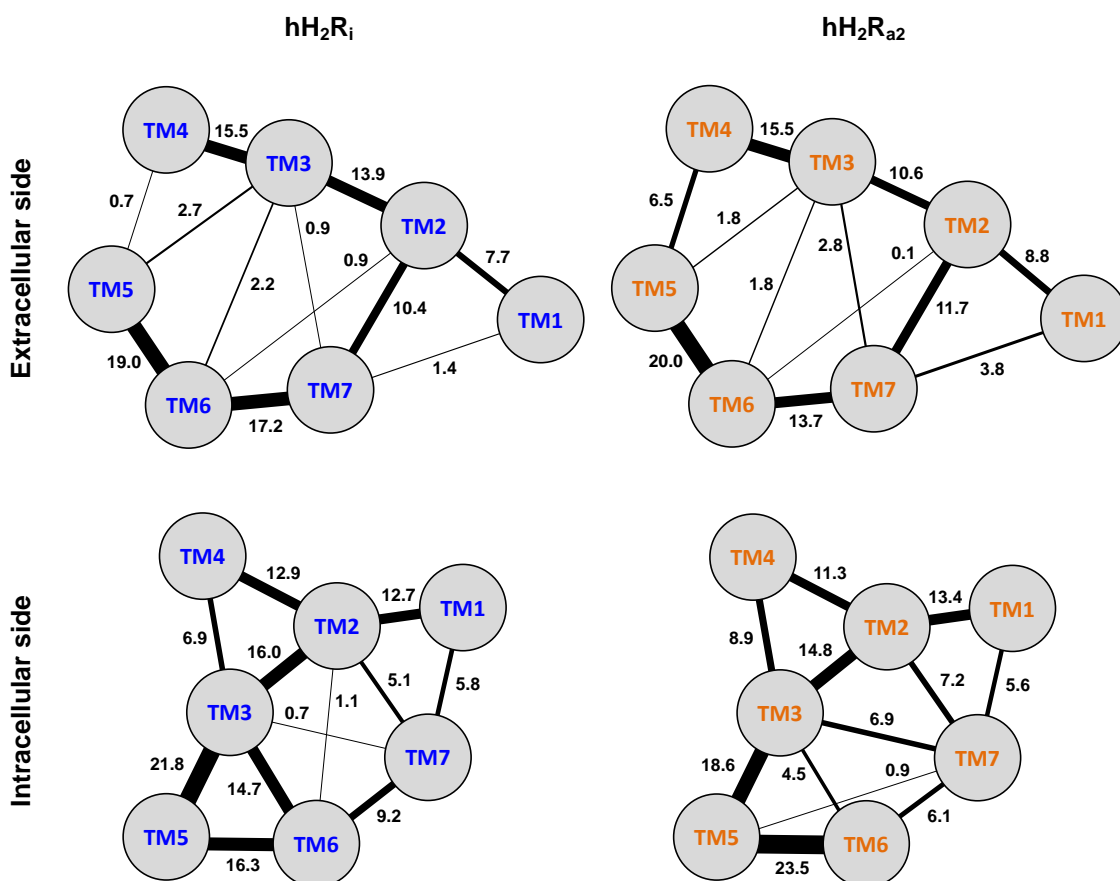


Figure 4.47: Contacts between TM domains at the extra- and intracellular part of the hH₂R states during 80 ns MD simulations

Average number of interactions (salt bridge, hydrogen bond, water mediated hydrogen bond, van der Waals contacts; for definition cf. section 4.2) between TM domains at the extracellular and the intracellular part of hH₂R_i and hH₂R_{a2}, respectively (for definition of extra- and intracellular parts of the hH₂R states, cf. Figure 3.16).

4.3.9.3 Loops and helix 8

The number of interactions of extracellular loops with neighboring domains was generally higher in hH₂R_{a2} (Table 4.4). In contrast, the ECLs of hH₂R_i were more flexible (Figure 4.27) and more displaced compared to the input structure of the simulation (Figure 4.28). The higher number of interaction in hH₂R_{a2} correlated with the restraining of loops in the active hH₂R state. Furthermore, the solvent accessible surface area (SASA, Figure 4.48) was slightly higher in hH₂R_i, indicating that less residues were in contact to the rest of the protein and more residues were exposed to the solvent.

Table 4.4: Average number of contacts to extra- and intracellular loops

Extracellular side				Intracellular side			
ECL	domain	hH ₂ R _i	hH ₂ R _{a2}	ICL	domain	hH ₂ R _i	hH ₂ R _{a2}
ECL1	TM2	6.0	5.9	ICL1	TM1	4.0	5.3
	TM3	4.5	7.4		TM2	2.4	5.0
	ECL2	4.9	6.0		ICL2	0.6	0.3
ECL2	TM2	4.0	4.8		H8	7.9	6.6
	TM3	10.2	8.2	ICL2	TM2	3.4	3.4
	TM4	4.9	7.2		TM3	7.8	9.0
	TM5	13.0	10.3		TM4	6.4	6.6
	TM6	3.8	5.3		TM5	1.5	0.0
	ECL3	0.2	6.9		ICL3	4.0	0.0
	TM7	1.0	3.0		TM6	0.7	0.0
ECL3	TM5	0.4	0.0	ICL3	TM5	4.9	4.7
	TM6	1.9	6.0		TM6	2.6	5.4
	TM7	3.6	4.1				
Helix8	TM1	9.5	9.3		TM6	1.3	0.0
	TM2	1.1	2.7		TM7	6.2	2.9

For the definition of interactions (salt bridge, hydrogen bond, water mediated hydrogen bond and van der Waals contacts) cf. section 4.2.

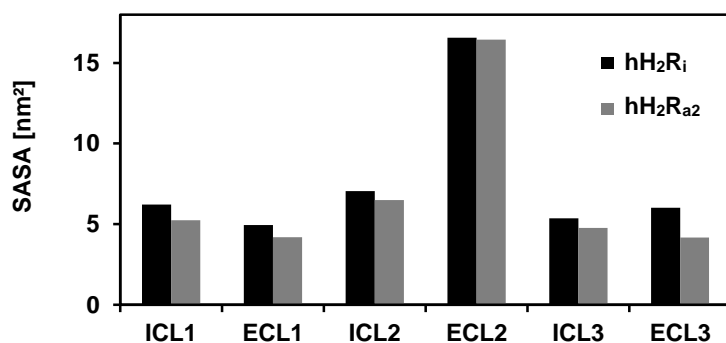


Figure 4.48: Solvent accessible surface area (SASA) for the loops of hH₂R_i and hH₂R_{a2}

The SASA was calculated for the same residues of all loops in hH₂R_i and hH₂R_{a2}.

ECL1 of hH₂R_{a2} was better attached to TM3 and ECL2 than in hH₂R_i (Table 4.4). Van der Waals interactions of Trp84^{ECL1} with Phe90^{3.24}, Cys91^{3.25} and Tyr94^{3.28} were more stable in hH₂R_{a2} (frequency in hH₂R_{a2} of 94%, 99% and 99% vs. 14%, 20% and 20% in hH₂R_i, respectively). The higher number of interactions between ECL1 and ECL2 in hH₂R_{a2} was mainly caused by stronger van der Waals contacts between Cys82^{ECL1} and Lys173^{ECL2} (frequency of 61% in hH₂R_{a2} vs. 0% in hH₂R_i), Trp84^{ECL1} and Thr170^{ECL2} (76% vs. 0%), Trp84^{ECL1} and Cys174^{ECL2} (100% vs. 20%), as well as Gly87^{ECL1} and Thr170^{ECL2} (68% vs. 42%). However, the ECL1 course was only slightly different between the two receptor states (Figure 4.49), as also indicated by the backbone RMSD values of corresponding residues (Figure 4.31). The most remarkable difference in ECL2 occurred in the N-terminal part until Cys174^{ECL2}, connected to Cys91^{3.25} by a disulfide bridge. Whereas in hH₂R_i this part of ECL2 moved more into the extracellular solvent, it was rather located above the C-terminal part of ECL2 in hH₂R_{a2} (beginning with Cys174^{ECL2}) and thus able to contact ECL3 and keep it in an inward position. However, due to the high flexibility of the side chains of ECL2 and ECL3 (Figure 4.27), the average frequency of interactions between pairs of residues was below 30%. Direct and water mediated H-bonds, salt bridges and van der Waals contacts occurred for Arg161, Thr164, Lys166, Lys173, Val176, Gln177 and Val178 (ECL2) as well as Arg260, Gly261, Asp263, Asp262 and Ile265 (ECL3). Instead, in hH₂R_i ECL3 moved outwards, opened the crevice between ECL2 and ECL3 and thus possibly enabled the accession of ligands to the orthosteric binding site. Not surprisingly, the differences of ECL2 and ECL3 between hH₂R_i and hH₂R_{a2} belonged to the most prominent ones during the simulation (see RMSD values in Figure 4.31). Comparing the crystal structures of inactive and active β_2 AR states, the differences at the extracellular side were lower (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011b). However, the courses of loops in GPCR crystals should generally interpreted with caution because of possible interfering lattice contacts with adjacent molecules (Warne *et al.*, 2008).

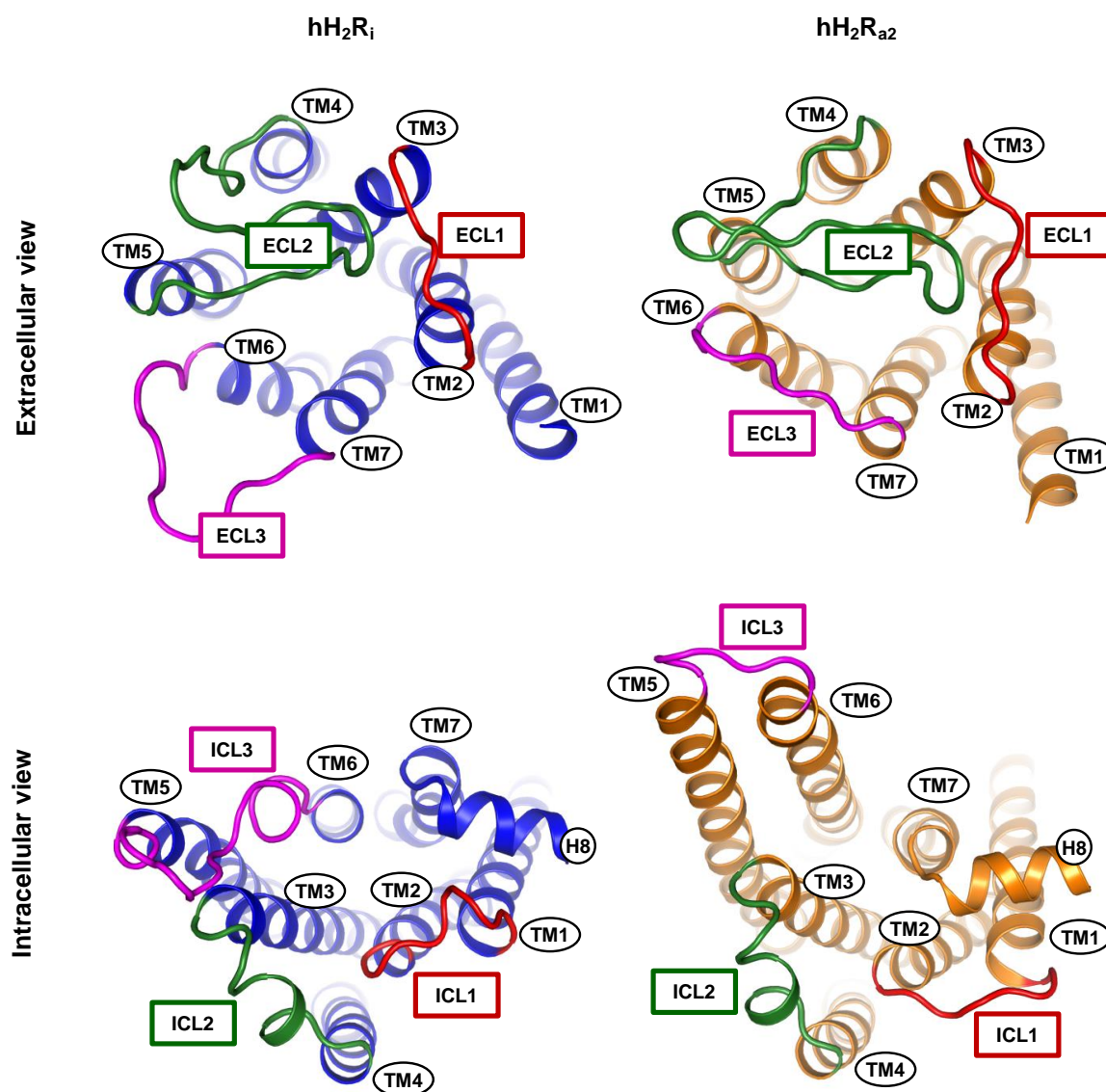


Figure 4.49: Extra- and intracellular loops of hH₂R_i and hH₂R_{a2}

The average simulation structures of the hH₂R_i and hH₂R_{a2} are shown (for the determination refer to Figure 4.30).

The generally higher SASA of the extracellular loops of hH₂R_i may facilitate the adhesion of approaching ligands. This remarkable difference was not obvious in the initial hH₂R_i and hH₂R_{a2} models. The course of both ECL2 and ECL3 in the two receptor states was quite similar (Figure 3.8). Moreover, an ionic interaction occurred between Glu163^{ECL2} and Arg260^{ECL3} in hH₂R_i (distance of the closest side chain heteroatoms 3.2 Å vs. 15.6 Å in hH₂R_{a2}).

The different contacts of intracellular loops in hH₂R_i and hH₂R_{a2} were due to the rearrangements of TM6 and TM7 and to the docking of GαCT into hH₂R_{a2}. Due to the interaction with GαCT the SASA of all intracellular loops was lower in hH₂R_{a2} (Figure 4.48). The position of ICL1 was already distant in the initial hH₂R models (Figure 3.8). This distance was even increased in the simulations, since helix 8 of hH₂R_{a2} further approached TM1 and ICL1. Thus, the RMSD between corresponding residues of hH₂R_i and hH₂R_{a2} rose until about 10 ns. This was one of the most remarkable differences between the two receptor states (Figure 4.31). Different interactions with TM1, TM2 and ICL2 could not be attributed to receptor activation. However, the van der Waals contact between Leu49^{ICL1} and Phe295^{H8} only present in hH₂R_i (frequency of 97%) is crucial for the 'inactive' position of Phe295^{H8} and thus for the stabilizing contact to Tyr288^{7.53} (π-π-stacking). The higher distances from ICL2 to ICL3, TM5 and TM6 did not allow an interaction of ICL2 with these domains in hH₂R_{a2}. Furthermore, the presence of GαCT prevented a direct contact. Instead, ICL2 of hH₂R_{a2} was able to interact with the G-protein fragment (section 4.3.10.2). In hH₂R_i, the H-bond between Asp122^{ICL2} and Trp222^{ICL3} (frequency of 58%) and the van der Waals contacts between Pro123^{ICL2} and Trp222^{ICL3}, Pro123^{ICL2} and Lys223^{ICL3}, Leu124^{ICL2} and Trp222^{ICL3} as well as Leu124^{ICL2} and Lys223^{ICL3} (frequency of 64%, 54%, 65% and 41%, respectively) contributed to the inward position of TM5, ICL3, and TM6 and to the prevention of G-protein binding. In hH₂R_{a2}, the strong interaction of ICL3 with GαCT (average number of interactions ~ 12) probably caused the reduced flexibility (Figure 4.27) and the less replacement from its initial position (Figure 4.28). Contacts of ICL2 to TM5 and TM6 in hH₂R_i were not consistent during the simulation (Asp122^{ICL2}, Pro123^{ICL2}, Lys214^{5.70}, Ala225^{6.25} and Als225^{6.26}).

Aromatic residues at the end of TM7 and helix 8 were connected to each other and performed conformational changes during receptor activation. In hH₂R_i, Tyr288^{7.53} and Phe295^{H8} (both part of the NPxxY(x)_{5,6}F motif) were arranged parallel and established aromatic π-π interactions during the entire simulation. This strong interaction restrained the side chain of Tyr288^{7.53} in its 'inactive' conformation (Figure 4.50 A). Phe295^{H8} was packed between the hydrophobic side chains of Val39^{1.53}, Ala42^{1.56}, Val43^{1.57}, Leu49^{ICL1} and Ile57^{2.43} (contact frequency of 92%, 59%, 100%, 97% and 99%, respectively). It was furthermore in close contact to Thr235^{6.36} (frequency of 71%), thus contributing to keep TM6 in its inward position. Phe295^{H8} forced Asn292^{7.57} (van der Waals contact frequency 100%) to show towards TM6 and to interact with Val234^{6.35} (van der Waals contact frequency 26%) and Thr235^{6.36} (water mediated H-bond and van der Waals contacts in 13% and 85% of the simulation, respectively), again linking TM6 to TM7. The side chain of Arg228^{6.29} was connected to the side chain of Asp294^{H8} by a salt bridge and a water mediated H-bond in 22% and 17% of the simulation, respectively. Tyr288^{7.53} and Phe295^{H8}, both pointing towards

TM2 (resulting in higher distances of TM2 to TM7 and H8; cf. Figure 4.49), were key amino acids for keeping TM7 in its outward position in hH₂R_i.

In hH₂R_{a2}, the Tyr288^{7.53}-Phe295^{H8} contact was broken, which has been suggested to be necessary for GPCR activation (Fritze *et al.*, 2003; Prioleau *et al.*, 2002; Scheerer *et al.*, 2008). Tyr288^{7.53} switched towards the TM3-TM6 interface (Figure 4.43), and Phe295^{H8} underwent a conformational change and pointed towards the second half of H8. Therefore, the side chain positions of Tyr288^{7.53} and Phe295^{H8} in hH₂R_{a2} allowed the approximation of TM7 to TM2 and TM3. Similar to Phe295^{H8}, Tyr299^{H8} rotated away from the TM1-TM7 interface, resulting in a parallel arrangement of Phe295^{H8}, Tyr299^{H8} and Phe303 (van der Waals contacts > 99%; Figure 4.50 B). Residues stabilizing Phe295^{H8} were Val38^{1.52}, Val39^{1.53}, Ala42^{1.56}, Val43^{1.57}, Ala289^{7.54}, Ala290^{7.55} and Asn292^{7.57} (van der Waals contact frequency of 84%, 100%, 100%, 100%, 99%, 88% and 91%, respectively). In hH₂R_i, the switch of Phe295^{H8} was prevented by Tyr299^{H8}, which was packed between Val38^{1.52}, Val39^{1.53}, Ala42^{1.56}, Pro285^{7.50}, Ala289^{7.54} and Gly298^{H8} (van der Waals contact frequency of 99%, 81%, 83%, 47%, 100% and 91%, respectively). Furthermore, Tyr299^{H8} was close to Tyr288^{7.53} and performed H-bonds with its hydroxyl group to the backbone of Pro285^{7.50} and Ala289^{7.54} (frequency of 70% and 53%, respectively). The higher number of interactions between TM7 and H8 in hH₂R_i was mainly caused by the side chain positions of Phe295^{H8} and Tyr299^{H8}. In hH₂R_{a2}, amino acids of helix 8 were part of the interaction site of GaCT (section 4.3.10.2).

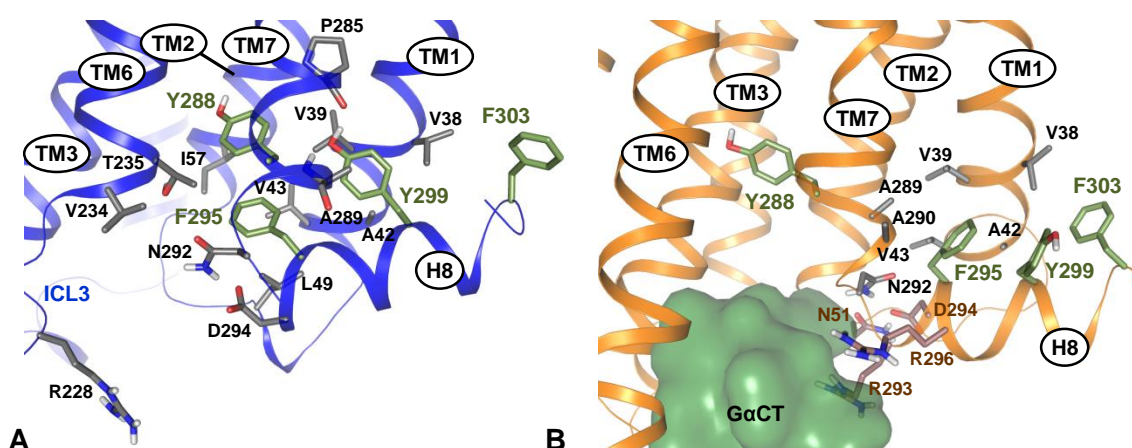


Figure 4.50: The role of helix 8 in hH₂R activation

Residues of the hH₂R_i (A; blue ribbons) and hH₂R_{a2} (B; orange ribbons with GaCT in green surface representation) which contributed to the interactions of the aromatic amino acids Tyr288^{7.53}, Phe295, Tyr299 and Phe303 (green sticks). Residues interacting with GaCT in hH₂R_{a2} are shown in brown sticks.

4.3.10 Ligand and GαCT interactions with the active hH₂R state

4.3.10.1 Interactions of histamine with the hH₂R binding pocket

Histamine remained stable in its position in the orthosteric binding site (average RMSD 1.12 Å), indicating proper docking of the ligand in the initial hH₂R_{a2} model (chapter 3). The well-known contacts to Asp98^{3.32} and Asp186^{5.42} (Gantz *et al.*, 1992), which were presumed for docking histamine were retained (Figure 4.51). The charge assisted H-bond from Asp98^{3.32} to the positively charged side chain amine of histamine and the H-bond from Asp186^{5.42} to N^π of the imidazole moiety were present in 95% and 99% of the simulation, respectively. Van der Waals contacts to Val99^{3.33}, Tyr250^{6.51}, Phe251^{6.52} and Phe254^{6.55} already observed in the initial hH₂R_{a2} model were preserved during the complete simulation (frequency of 100% for all residues).

Furthermore, histamine was linked to Tyr182^{5.38} in 77% of the simulation (hydrophobic contact of the imidazole moiety and the phenyl ring). Additionally, a water mediated H-bond between the hydroxyl moiety of Tyr182^{5.38} and N^π of the imidazole moiety of histamine (frequency of 30%) was observed. However, this interaction is probably not essential. The data obtained at an hH₂R mutant in which Tyr182 was replaced by Phe (chapter 6) indicated a minor impact of the hydroxyl group of Tyr182^{5.38} on histamine activity, contrary to previous suggestions (Nederkoorn *et al.*, 1996a; Nederkoorn *et al.*, 1996b). Instead, the hydroxyl group of Tyr182^{5.38} was mainly H-bonded to Asp186^{5.42} (frequency of 85%), adjusting a proper side chain conformation of the latter. The N^π nitrogen of the imidazole moiety of histamine was primarily connected to Tyr250^{6.51} by a direct H-bond (frequency of 91%). Nevertheless, Tyr182^{5.38} seemed to play a more important role in histamine binding than Thr190^{5.46}, which was also suggested to contribute to binding (Gantz *et al.*, 1992) but was not connected to histamine during the whole MD simulation.

Polar contacts occurred also between the amine nitrogen of histamine and Cys102^{3.36} (direct H-bond in 78% of the simulation) as well as the backbone oxygen of Leu274^{7.39} (frequency of 37%). In several GPCRs, ECL2 is part of the ligand binding site (Avlani *et al.*, 2007; Jaakola *et al.*, 2008; Klco *et al.*, 2005; Scarselli *et al.*, 2007; Shi and Javitch, 2002). However, for histamine binding ECL2 seems to be less important. The N^π nitrogen of the imidazole moiety of histamine and Lys175^{ECL2} were connected via a water mediated H-bond in only 16% of the simulation of hH₂R_{a2}. Finally, van der Waals contacts of histamine not present in the initial hH₂R_{a2} model were formed with Thr103^{3.37}, Ile106^{3.40}, Trp247^{6.48} and Tyr278^{7.43} (frequency of 98%, 49%, 99% and 26%, respectively).

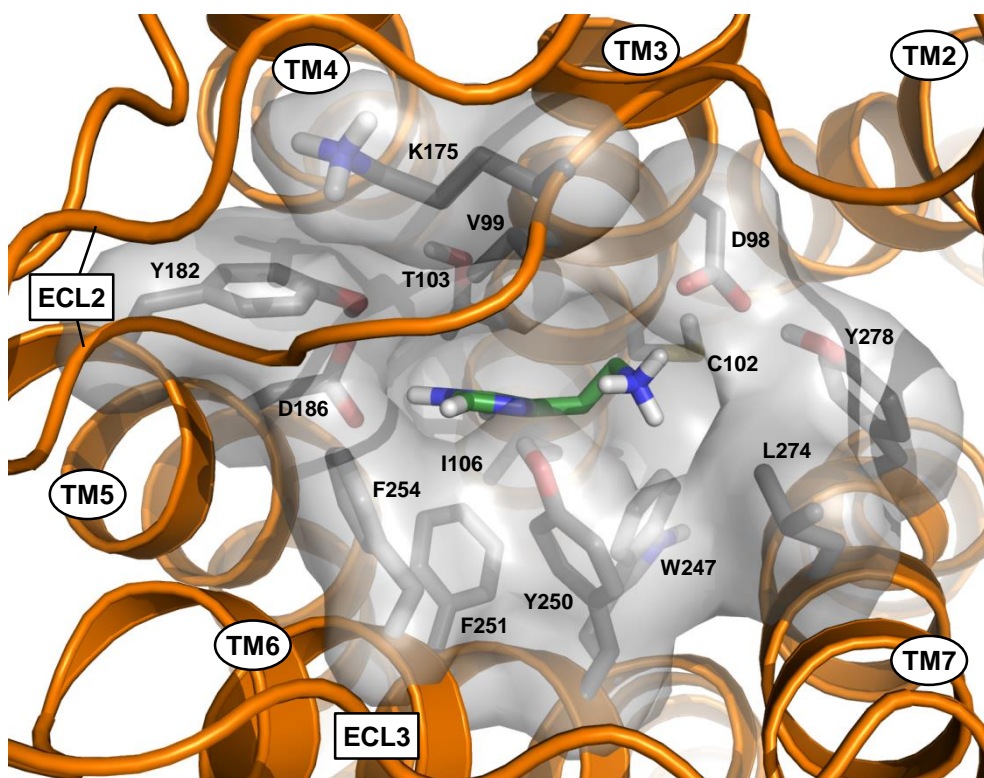


Figure 4.51: Histamine in the orthosteric binding site of the hH₂R_{a2}

Histamine (green carbon atoms) in the binding site of the hH₂R_{a2} (TMs in orange ribbons) after 80 ns MD simulation. All residues which interacted with histamine during the MD simulation are shown (sticks with grey transparent surface).

4.3.10.2 Receptor – GαCT interaction

The C-terminal fragment of the G_{sαS}-protein, GαCT, interacted with the most flexible parts of hH₂R_{a2}, the cytoplasmic surface which forms an aligned pocket. Amino acids of all intracellular loops, the intracellular ends of TM2, TM3, TM5, TM6 and TM7 as well as helix 8 were part of the receptor-GαCT interface (30 amino acids in total). Thus, all residues of the cytoplasmic crevice which was formed by the outward move of TM6 were part of the interaction site (Figure 4.52 A). However, the high flexibility of the hH₂R_{a2} in this region and of parts of GαCT led to varying pairs of interacting residues. All in all 67 contacts with an average frequency of 66% were detected. Compared to the hH₂R_{a1}-GαCT complex (chapter 3), the number of interactions between the initial hH₂R_{a2} and GαCT was lower, but rose significantly during the 80 ns simulation (Figure 4.52 B). GαCT moved towards helix 8 and ICL1, implying a concomitant shift of the cytoplasmic ends of TM5 and TM6 (Figure 4.52 C).

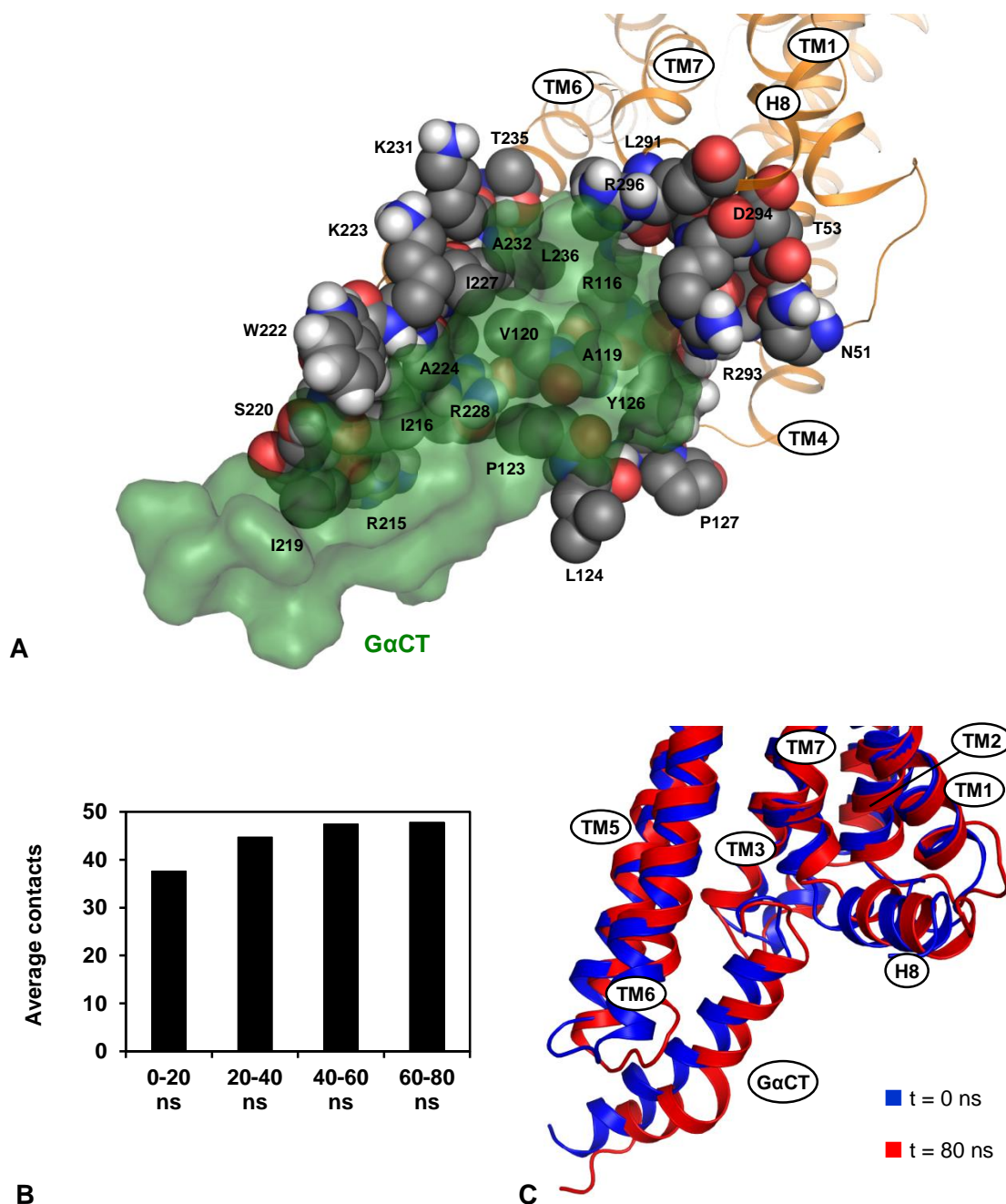


Figure 4.52: Interactions between hH₂R_{a2} and GαCT

A, Interaction between hH₂R_{a2} and GαCT (green transparent surface representation). All residues (in spheres) of the hH₂R_{a2} (TMs in orange ribbons) which contributed to the binding of the G-protein fragment are shown, except Asp115^{3,49}, Gln212^{5,68}, Ser221^{ICL3} and Asn292^{7,57} (not visible). B, Average number of interactions between hH₂R_{a2} and GαCT. C, Superposition of the hH₂R_{a2} at the start and the end of the 80 ns MD simulation.

Among the interactions, several polar contacts were obvious, for the most frequent see Table 4.5. The salt bridge between the side chains of Arg116^{3.50} and Glu392 described in chapter 3 was present in 41% throughout the simulation. Additionally, several hydrophobic contacts linked the GαCT protein to the hH₂R_{a2}. The most frequent contacts are listed in Table 4.6. Also repulsive ionic interactions between amino acids of the hH₂R_{a2} and GαCT occurred (Arg293^{H8}-Arg389, frequency of 18%; Asp294^{H8}-Glu392, 88%; Arg296^{H8}-Arg389, 19%; Lys223^{ICL3}-Arg385, 35%).

Table 4.5: Polar contacts between hH₂R_{a2} and GαCT

hH ₂ R _{a2}	Type ¹	GαCT	Type ¹	Contact ²	Frequency [%]	
Thr53 ^{2.39}	sc	Glu392	sc	H-bond	63	
Arg116 ^{3.50}	sc	Glu392	bb	SOL	49	4, 5
Ser220 ^{ICL3}	sc	Asn377	sc	H-bond	55	5
Ser221 ^{ICL3}	sc	Asp381	sc	H-bond ³	71	5
Trp222 ^{ICL3}	bb	Asp381	sc	H-bond	64	
Lys223 ^{ICL3}	bb	Asp381	sc	H-bond	73	
Ala224 ^{6.25}	bb	Asp381	sc	H-bond	79	
Arg228 ^{6.29}	sc	Gln384	sc	H-bond	69	
Asn292 ^{7.57}	sc	Glu392	sc	H-bond ³	94	
Arg293 ^{H8}	bb	Glu392	sc	H-bond	80	
Arg293 ^{H8}	sc	Glu392	sc	Ionic	78	
Asp294 ^{H8}	sc	Glu392	sc	SOL	57	4
Arg296 ^{H8}	sc	Leu394	bb	Ionic	93	

¹ sc, side chain; bb, backbone.

² SOL, water mediated H-bond.

³ This H-bond was already present in the initial hH₂R_{a2}-GαCT complex (cf. chapter 3).

⁴ The corresponding residue of opsin interacts with the GαCT peptide in the crystal structure of opsin (Scheerer *et al.*, 2008).

⁵ The corresponding residue of the hβ₂AR interacts with the G-protein in the crystal structure of the β₂ adrenergic receptor–G_s protein complex (Rasmussen *et al.*, 2011b).

Table 4.6: Frequency of van der Waals contacts between hH₂R_{a2} and GαCT

hH ₂ R _{a2}	GαCT	F [%]		hH ₂ R _{a2}	GαCT	F [%]	
Arg116 ^{3.50}	Leu393	50	1, 2	Lys223 ^{ICL3}	Arg385	81	
Ala119 ^{3.53}	Tyr391	97	1, 2	Ala224 ^{6.25}	Asp381	76	1
Val120 ^{3.54}	Leu388	54	1, 2	Ala224 ^{6.25}	Gln384	77	1
Val120 ^{3.54}	Leu393	96	1, 2	Ala224 ^{6.25}	Arg385	51	1
Pro123 ^{ICL2}	His387	97	2	Ile227 ^{6.28}	Arg385	81	
Pro123 ^{ICL2}	Tyr391	100	2	Ile227 ^{6.28}	Leu388	74	
Tyr126 ^{ICL2}	Tyr391	95	2	Ile227 ^{6.28}	Leu394	95	
Ile216 ^{5.72}	Phe376	87		Arg228 ^{6.29}	Leu388	87	1
Ile216 ^{5.72}	Arg380	63		Lys231 ^{6.32}	Leu394	68	
Ile216 ^{5.72}	Gln384	43		Ala232 ^{6.33}	Leu388	80	1, 2
Ile219 ^{ICL3}	Ile372	54		Ala232 ^{6.33}	Leu393	75	1, 2
Ile219 ^{ICL3}	Phe376	76		Ala232 ^{6.33}	Leu394	52	1, 2
Ser220 ^{ICL3}	Phe376	66		Leu236 ^{6.37}	Leu393	91	2
Ser220 ^{ICL3}	Asn377	38		Leu291 ^{7.56}	Leu393	59	
Trp222 ^{ICL3}	Asn377	42		Arg293 ^{H8}	Glu392	70	
Trp222 ^{ICL3}	Asp381	51					

¹ The corresponding residue of opsin interacts with the GαCT peptide in the crystal structure of opsin (Scheerer *et al.*, 2008).

² The corresponding residue of the hβ₂AR interacts with the G-protein in the crystal structure of the β₂ adrenergic receptor–G_s protein complex (Rasmussen *et al.*, 2011b).

4.4 Discussion

4.4.1 Quality of the MD simulations

The analysis of system parameters during the 20 ns equilibration phase and the 80 ns production run of the hH₂R_i and hH₂R_{a2} showed that the parameters attained stable values as expected for accurate MD simulation systems. Highest variability was measured for the box dimensions. In both simulation systems the plane parallel to the membrane was contracted and the z-dimension expanded. The reason is possibly related to the bilayer (see below). However, in the 80 ns simulation period the box vectors converged. The distance of the hH₂R_i to its nearest periodic image was in a marginal fraction of the simulation below the cut-off value for long-range interactions (1.4 nm). Therefore, during these phases the protein

could manipulate itself resulting in artifacts of the simulation. However, no bad effects were observed. The structure of the DPPC bilayer was generally in very good agreement to experimentally characterized bilayers. Only the area per lipid was slightly above the lowest reference value. However, the quality of a lipid bilayer does not only depend on this parameter (Poger and Mark, 2010). The reduction of the area per lipid correlated with the alterations of the box vectors. The insertion of the protein with the tool *InflateGRO* (Kandt *et al.*, 2007) in the DPPC bilayer was performed as recommended (refer to the website of the Life and Medical Science Center of the University of Bonn, <http://www.csb.bit.uni-bonn.de/inflategro.html>). Furthermore, parameters potentially influencing the box dimensions such as the pressure coupling were adapted to previous simulations with the same lipid parameters (Kukol, 2009). Thus, possibly the estimation of the protein size performed by *InflateGRO* was critical since this is a well-known challenge in MD system preparations (Allen *et al.*, 2009). Nevertheless, the area per lipid was still in accordance to reference values in both simulation systems.

The quality of the proteins hH₂R_i, hH₂R_{a2} and GαCT were extensively examined with the program *gro_validation* described in chapter 5. Stereochemistry at C_α atoms and planarity of aromatic side chains and delocalized π-electron systems in polar side chains were as expected for accurate amino acids, substantiating the quality of the initial hH₂R models and of the force field used. Errors of the stereochemistry could originate from inappropriate protein structures or from wrong force fields parameters. However, the distribution of the peptide dihedral angle ω was broadened compared to experimentally derived data (Edison, 2001). The correspondence of the dihedral angles Φ and Ψ of the protein backbone to values obtained from protein crystal structures (Lovell *et al.*, 2000) was very good. The same applied to the side chain dihedral angles. They were in very good agreement to frequently occurring rotamer states. The correctness of stereochemical parameters and the reproducibility of experimental values of the proteins were the prerequisite of using the data obtained from these MD simulations to draw reliable conclusions on the conformational differences between inactive and active hH₂R states and thus activation of the hH₂R.

4.4.2 Activation of the hH₂R

In general two different mechanisms for the activation of GPCRs are discussed (Deupi and Kobilka, 2010; Trzaskowski *et al.*, 2012). According to the 'induced fit' theory, ligand binding is accompanied with an energy transfer inducing a conformational change of the protein. Thus, the ligands had to bind to the receptor with high affinity in order to provide enough energy to enable the 'jump' of the GPCR over the first, high energy barrier. In rhodopsin, the isomerization of the light sensitive and (via a protonated Schiff base with Lys296^{7,43})

covalently bound chromophore 11-cis-retinal to all-trans retinal after light absorption (Li *et al.*, 2004) provides such high energy. In contrast, relatively low binding affinities and fast dissociation rates, e.g. of β_2 AR agonists (Gether *et al.*, 1995), are not in agreement with an induced fit mechanism. Furthermore, several crystal structures of the putative inactive state of the β_1 AR and β_2 AR (Hanson *et al.*, 2008; Rasmussen *et al.*, 2007; Rosenbaum *et al.*, 2007; Warne *et al.*, 2008) suggest that the internal flexibility allows the receptor to explore different conformations (Ghanouni *et al.*, 2001; Peleg *et al.*, 2001). Thus, in case of the β ARs and probably also of the related hH₂R, the mechanism of ‘conformational selection’ is preferred, in which a ligand binds to a suitable receptor conformation (e.g. an agonist to an active receptor conformation) which is then stabilized (Deupi and Kobilka, 2010). Accordingly, in the hH₂R the activation process is not necessarily started at the orthosteric binding site and then transferred to the intracellular surface of the receptor via subsequent or parallel steps, enabling the interaction with the G-protein. Most likely, different conformations exist in equilibrium and are stabilized by different ligands. Therefore, the activation mechanism of the hH₂R may be analyzed by the comparison of inactive and active receptor states based on the systematic calculation of ionic interactions, H-bonds, water mediated H-bonds and van der Waals contacts. The examination of helical contacts was of special interest since the relative positions of TMs deviated between hH₂R_i and hH₂R_{a2}, similar to the differences between the inactive and active state of rhodopsin (Okada *et al.*, 2004; Scheerer *et al.*, 2008) and the β_2 AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011b). Residues with different interaction partners in hH₂R_i and hH₂R_{a2} or changed positions during the MD simulations were especially important. Most of the deviations between the hH₂R states occurred around molecular switches already described in chapter 3. However, the 80 ns MD simulations of hH₂R_i and hH₂R_{a2} allowed a more detailed analysis of dynamic interactions and the detection of linkages between switches. A lot of contacts were not obvious in the initial homology models but emerged during the simulations, or they were unstable at the beginning and more stable at the end of the MD simulations.

The pivotal difference between hH₂R states was the outward position of the cytoplasmic part of TM6, adopted from the template h β_2 AR (Rasmussen *et al.*, 2011a) and previously described in literature to be a key alteration during GPCR activation (Altenbach *et al.*, 2008; Farrens *et al.*, 1996). This change was connected to all other molecular switches, namely the parallel arrangement of TM5 and TM6 (stable during the simulation of hH₂R_{a2}) including the conformational switch of Tyr202^{5.58}, the cleavage of the ionic lock and the hydrophobic barrier, the inward move of TM7 with the change of Tyr288^{7.53}, the rearrangement of the H-bond network between TM1, TM2, TM3 and TM7, and the alterations at the bottom of the binding site which was linked by the key amino acid Phe243^{6.44} to the movement of TM6.

Phe243^{6.44} was close to some residues of the orthosteric binding site (Cys102^{3.36}, Ile106^{3.40}, Trp247^{6.48} and Phe251^{6.52}). By transition from the inactive to the active receptor conformation, Phe243^{6.44} was released from amino acids of TM3 and TM7 (Cys102^{3.36}, Ser105^{3.39} and Asn280^{7.45}) and turned towards TM5, residing in a pocket of several hydrophobic amino acids (Ile106^{3.40}, Leu109^{3.43}, Phe191^{5.47}, Pro194^{5.50}, Leu195^{5.51} and Met198^{5.54}). Thereby it changed its position with Ile106^{3.40}. This switch was first observed for corresponding amino acids in the crystal structure of the nanobody-stabilized active state of the β_2 AR (Rasmussen *et al.*, 2011a). In a recent study, Phe^{6.44}Ala mutants of the ghrelin receptor, GPR119, the β_2 AR and the NK1 tachykinin receptor, respectively, showed strongly impaired or no spontaneous and/or agonist-mediated signaling (Valentin-Hansen *et al.*, 2012b). It was suggested that an aromatic ‘microswitch’ of Phe^{6.44} stabilizes the outward conformation of TM6 by sliding of the side chain into a hydrophobic pocket between TM3 and TM5, whereas the residue in position 3.40 is the gate for this transition. Furthermore, in the hH₂R Phe243^{6.44} affected the H-bond network around TM7. After Phe243^{6.44} approached TM5 in the active hH₂R conformation, Ser105^{3.39} was able to interact with the NPxxY(x)_{5,6}F motif, and Asn280^{7.45} was released from its position towards TM2. Experimental investigations, e.g. on mutants, would be worthwhile for confirming the pivotal role of Phe243^{6.44} in the hH₂R and would furthermore extend the knowledge about this ‘microswitch’ in GPCRs. Besides Phe243^{6.44}, amino acids near the intracellular surface of the hH₂R performed crucial conformational changes, most prominent in the case of Tyr202^{5.58}, Arg116^{3.50} and Tyr288^{7.73}. The switch of Tyr202^{5.58}, first observed in the crystal structures of opsin and the active β_2 AR (Rasmussen *et al.*, 2011a; Scheerer *et al.*, 2008), led to sticking between TM3 and TM6. In the active hH₂R state, Tyr202^{5.58} interacted with Arg116^{3.50} and Tyr288^{7.73}. The interaction with Arg^{3.50} was not possible in the β_2 AR (distance between side chain heteroatoms > 9 Å). Tyr219^{5.58} of the β_2 AR probably plays a role in stabilizing the active receptor conformation since a Tyr^{5.58}Ala mutant showed a loss of G_s-signaling, arrestin mobilization and receptor internalization (Valentin-Hansen *et al.*, 2012a). Also in rhodopsin, the analogous Tyr223^{5.58} was shown to be necessary for G_i-activation (Elgeti *et al.*, 2011). Again mutants of Tyr202^{5.58} at the hH₂R are suitable to further explore the importance of this residue. Apart from these key amino acids in hH₂R activation, several other residues were part of maintaining the receptor in its inactive and active state, respectively. A schematic overview is given in Figure 4.53.

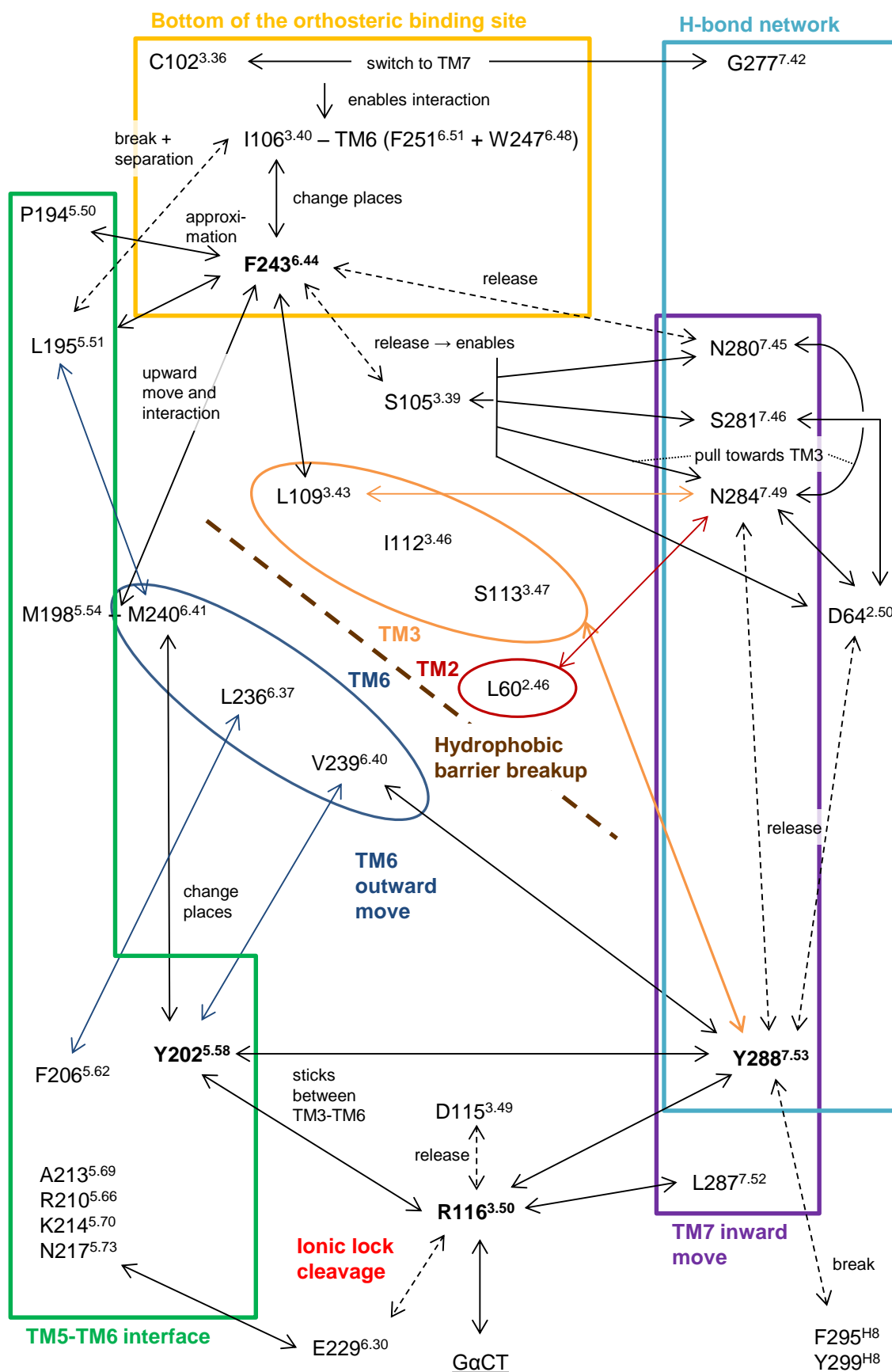


Figure 4.53: Schematic representation of the hH₂R activation based on 80 ns MD simulations of inactive and active hH₂R states

4.4.3 Conclusion

Homology models of the hH₂R in its inactive and active state were investigated by MD simulations, explicitly considering the natural environment consisting of a lipid bilayer (DPPC) and water molecules. Comparison of the simulated membrane structure with parameters derived from experiments, as well as the structure validation of the proteins (hH₂R_i, hH₂R_{a2} and GαCT) during the entire simulations with the program described in chapter 5 (*gro_validation*) ensured the quality and reliability of the MD simulations and of the results which proved the initial homology models of the hH₂R_i and hH₂R_{a2} to be stable and reasonable. In addition to the confirmation of rotamer states and interactions already observed in the initial models (chapter 3), the MD simulations enabled to identify new contacts, the linkage of several molecular switches important for receptor activation of the hH₂R and the role of water molecules in connecting residues. This analysis was possible by the systematic exploration of polar and hydrophobic contacts in the hH₂R states, calculated with the programs *gro_hbonds* and *gro_contacts* provided in chapter 5. The MD simulations confirmed the importance of ‘microswitches’ during hH₂R activation, such as the rearrangement below the orthosteric binding site with Phe243^{6,44} as a key amino acid, the cleavage of the ionic lock and the hydrophobic barrier, conformational changes of Tyr202^{5,58} and Tyr288^{7,53}, rearrangement of the H-bond network between TM1, TM2, TM3 and TM7 as well as the inward move of the cytoplasmic part of TM7. All changes were related to the pivotal outward move of the cytoplasmic part of TM6 upon receptor activation. The results of these theoretical investigations have to be confirmed by experimental methods such as in-vitro site-directed mutagenesis studies.

4.5 References

- Ahuja S and Smith SO. Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol Sci* **2009**, 30, 494-502.
- Alewijnse AE, Timmerman H, Jacobs EH, Smit MJ, Roovers E, Cotecchia S and Leurs R. The effect of mutations in the DRY motif on the constitutive activity and structural instability of the histamine H(2) receptor. *Mol Pharmacol* **2000**, 57, 890-898.
- Allen WJ, Lemkul JA and Bevan DR. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *J Comput Chem* **2009**, 30, 1952-1958.
- Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP and Hubbell WL. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc Natl Acad Sci U S A* **2008**.
- Angel TE, Chance MR and Palczewski K. Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc Natl Acad Sci U S A* **2009**, 106, 8555-8560.
- Avlani VA, Gregory KJ, Morton CJ, Parker MW, Sexton PM and Christopoulos A. Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. *J Biol Chem* **2007**, 282, 25677-25686.

- Ballesteros JA, Shi L and Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. *Mol Pharmacol* **2001**, 60, 1-19.
- Berendsen HJC. Transport Properties Computed by Linear Response through Weak Coupling to a Bath. In *Computer Simulation in Materials Science*, Meyer M and Pontikis V, Eds., Springer Netherlands: **1991**; Vol. 205, pp 139-155.
- Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A and Haak JR. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, 81, 3684-3690.
- Buldt G, Gally HU, Seelig A, Seelig J and Zaccai G. Neutron diffraction studies on selectively deuterated phospholipid bilayers. *Nature* **1978**, 271, 182-184.
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK and Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, 318, 1258-1265.
- Darden T, York D and Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, 98, 10089-10092.
- De Young LR and Dill KA. Solute partitioning into lipid bilayer membranes. *Biochemistry* **1988**, 27, 5281-5289.
- Deupi X and Kobilka BK. Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology* **2010**, 25, 293-303.
- Douliez JP, Leonard A and Dufourc EJ. Restatement of order parameters in biomembranes: calculation of C-C bond order parameters from C-D quadrupolar splittings. *Biophys J* **1995**, 68, 1727-1739.
- Dror RO, Arlow DH, Borhani DW, Jensen MO, Piana S and Shaw DE. Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc Natl Acad Sci U S A* **2009**, 106, 4689-4694.
- Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, Borhani DW and Shaw DE. Activation mechanism of the β_2 -adrenergic receptor. *Proc Natl Acad Sci U S A* **2011**, 108, 18684-18689.
- Edison AS. Linus Pauling and the planar peptide bond. *Nat Struct Biol* **2001**, 8, 201-202.
- Elgeti M, Kazmin R, Heck M, Morizumi T, Ritter E, Scheerer P, Ernst OP, Siebert F, Hofmann KP and Bartl FJ. Conserved Tyr223(5.58) plays different roles in the activation and G-protein interaction of rhodopsin. *J Am Chem Soc* **2011**, 133, 7159-7165.
- Eriks JC, van der Goot H and Timmerman H. New activation model for the histamine H₂ receptor, explaining the activity of the different classes of histamine H₂ receptor agonists. *Mol Pharmacol* **1993**, 44, 886-894.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG. A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, 103, 8577-8593.
- Farrens DL, Altenbach C, Yang K, Hubbell WL and Khorana HG. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **1996**, 274, 768-770.
- Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP and Ernst OP. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. *Proc Natl Acad Sci U S A* **2003**, 100, 2290-2295.
- Gantz I, DelValle J, Wang LD, Tashiro T, Munzert G, Guo YJ, Konda Y and Yamada T. Molecular basis for the interaction of histamine with the histamine H₂ receptor. *J Biol Chem* **1992**, 267, 20840-20843.
- Gether U, Lin S and Kobilka BK. Fluorescent labeling of purified beta 2 adrenergic receptor. Evidence for ligand-specific conformational changes. *J Biol Chem* **1995**, 270, 28268-28275.
- Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR and Kobilka BK. Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. *J Biol Chem* **2001**, 276, 24433-24436.

- Ghanouni P, Schambye H, Seifert R, Lee TW, Rasmussen SG, Gether U and Kobilka BK. The effect of pH on beta(2) adrenoceptor function. Evidence for protonation-dependent activation. *J Biol Chem* **2000**, 275, 3121-3127.
- Goncalves JA, South K, Ahuja S, Zaitseva E, Opefi CA, Eilers M, Vogel R, Reeves PJ and Smith SO. Highly conserved tyrosine stabilizes the active state of rhodopsin. *Proc Natl Acad Sci U S A* **2010**, 107, 19861-19866.
- Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, Velasquez J, Kuhn P and Stevens RC. A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure* **2008**, 16, 897-905.
- Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **2007**, 4, 116-122.
- Hess B, Bekker H, Berendsen HJC and Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **1997**, 18, 1463-1472.
- Hess B, Kutzner C, van der Spoel D and Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, 4, 435-447.
- Humphrey W, Dalke A and Schulten K. VMD: visual molecular dynamics. *J Mol Graph* **1996**, 14, 33-38.
- Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP and Stevens RC. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, 322, 1211-1217.
- Jambeck JP and Lyubartsev AP. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J Phys Chem B* **2012**, 116, 3164-3179.
- Kabsch W and Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577-2637.
- Kandt C, Ash WL and Tieleman DP. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, 41, 475-488.
- Klco JM, Wiegand CB, Narzinski K and Baranski TJ. Essential role for the second extracellular loop in C5a receptor activation. *Nat Struct Mol Biol* **2005**, 12, 320-326.
- Kucerka N, Nagle JF, Sachs JN, Feller SE, Pencer J, Jackson A and Katsaras J. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys J* **2008**, 95, 2356-2367.
- Kucerka N, Nieh MP and Katsaras J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim Biophys Acta* **2011**, 1808, 2761-2771.
- Kukol A. Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *J Chem Theory Comput* **2009**, 5, 615-626.
- Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AG and Tate CG. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, 474, 521-525.
- Leftin A and Brown MF. An NMR database for simulations of membrane dynamics. *Biochim Biophys Acta* **2011**, 3, 818-839.
- Lewis RN, Mak N and McElhaney RN. A differential scanning calorimetric study of the thermotropic phase behavior of model membranes composed of phosphatidylcholines containing linear saturated fatty acyl chains. *Biochemistry* **1987**, 26, 6118-6126.
- Li J, Edwards PC, Burghammer M, Villa C and Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **2004**, 343, 1409-1438.
- Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS and Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003**, 50, 437-450.
- Lovell SC, Word JM, Richardson JS and Richardson DC. The penultimate rotamer library. *Proteins* **2000**, 40, 389-408.
- MacArthur MW and Thornton JM. Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* **1996**, 264, 1180-1195.

- Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, 12, 345-364.
- Nagle JF. Area/lipid of bilayers from NMR. *Biophys J* **1993**, 64, 1476-1481.
- Nagle JF and Tristram-Nagle S. Lipid bilayer structure. *Curr Opin Struct Biol* **2000a**, 10, 474-480.
- Nagle JF and Tristram-Nagle S. Structure of lipid bilayers. *Biochim Biophys Acta* **2000b**, 10, 159-195.
- Nederkoorn PH, van Gelder EM, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. II. Theoretical investigations of histamine binding to receptor models of the seven alpha-helical transmembrane domain. *J Comput Aided Mol Des* **1996a**, 10, 479-489.
- Nederkoorn PH, van Lenthe JH, van der Goot H, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. I. Theoretical investigations of histamine binding to an oligopeptide mimicking a part of the fifth transmembrane alpha-helix. *J Comput Aided Mol Des* **1996b**, 10, 461-478.
- Okada T, Sugihara M, Bondar AN, Elstner M, Entel P and Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* **2004**, 342, 571-583.
- Oostenbrink C, Villa A, Mark AE and Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **2004**, 25, 1656-1676.
- Pandey PR and Roy S. Headgroup mediated water insertion into the DPPC bilayer: a molecular dynamics study. *J Phys Chem B* **2011**, 115, 3155-3163.
- Patra M, Karttunen M, Hyvonen MT, Falck E, Lindqvist P and Vattulainen I. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys J* **2003**, 84, 3636-3645.
- Peleg G, Ghanouni P, Kobilka BK and Zare RN. Single-molecule spectroscopy of the beta(2) adrenergic receptor: observation of conformational substates in a membrane protein. *Proc Natl Acad Sci U S A* **2001**, 98, 8469-8474.
- Petrache HI, Dodd SW and Brown MF. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by (2)H NMR spectroscopy. *Biophys J* **2000**, 79, 3172-3192.
- Poger D and Mark AE. On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J Chem Theory Comput* **2010**, 6, 325-336.
- Prates Ramalho JP, Gkeka P and Sarkisov L. Structure and Phase Transformations of DPPC Lipid Bilayers in the Presence of Nanoparticles: Insights from Coarse-Grained Molecular Dynamics Simulations. *Langmuir* **2011**, 27, 3723-3730.
- Prioleau C, Visiers I, Ebersole BJ, Weinstein H and Sealfon SC. Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT_{2C} receptor. Identification of a novel "locked-on" phenotype and double revertant mutations. *J Biol Chem* **2002**, 277, 36577-36584.
- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011a**, 469, 175-180.
- Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **2007**, 450, 383-387.
- Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **2011b**, 477, 549-555.

- Rees DC, DeAntonio L and Eisenberg D. Hydrophobic organization of membrane proteins. *Science* **1989**, 245, 510-513.
- Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC and Kobilka BK. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **2007**, 318, 1266-1273.
- Scarselli M, Li B, Kim SK and Wess J. Multiple residues in the second extracellular loop are critical for M3 muscarinic acetylcholine receptor activation. *J Biol Chem* **2007**, 282, 7385-7396.
- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP and Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, 455, 497-U430.
- Schreiner E, Trabuco LG, Freddolino PL and Schulten K. Stereochemical errors and their implications for molecular dynamics simulations. *BMC Bioinformatics* **2011**, 12, 1471-2105.
- Shi L and Javitch JA. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu Rev Pharmacol Toxicol* **2002**, 42, 437-467.
- Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA and Javitch JA. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J Biol Chem* **2002**, 277, 40989-40996.
- Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD and Schertler GF. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **2011**, 471, 656-660.
- Taddese B, Simpson LM, Wall ID, Blaney FE, Kidley NJ, Clark HS, Smith RE, Upton GJ, Gouldson PR, Psaroudakis G, Bywater RP and Reynolds CA. G-protein-coupled receptor dynamics: dimerization and activation models compared with experiment. *Biochem Soc Trans* **2012**, 40, 394-399.
- Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A and Filipek S. Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr Med Chem* **2012**, 19, 1090-1109.
- Ulmschneider JP and Ulmschneider MB. United Atom Lipid Parameters for Combination with the Optimized Potentials for Liquid Simulations All-Atom Force Field. *J Chem Theory Comput* **2009**, 5, 1803-1813.
- Valentin-Hansen L, Groenen M, Nygaard R, Frimurer TM, Holliday ND and Schwartz TW. The Arginine of the DRY Motif in Transmembrane Segment III Functions as a Balancing Micro-switch in the Activation of the beta 2-Adrenergic Receptor. *J Biol Chem* **2012a**, 287, 31973-31982.
- Valentin-Hansen L, Holst B, Frimurer TM and Schwartz TW. PheVI:09 (Phe6.44) as a sliding microswitch in seven-transmembrane (7TM) G protein-coupled receptor activation. *J Biol Chem* **2012b**, 287, 43516-43526.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, 26, 1701-1718.
- Vanni S, Neri M, Tavernelli I and Rothlisberger U. A conserved protonation-induced switch can trigger "ionic-lock" formation in adrenergic receptors. *J Mol Biol* **2010**, 397, 1339-1349.
- Vanni S and Rothlisberger U. A closer look into G protein coupled receptor activation: X-ray crystallography and long-scale molecular dynamics simulations. *Curr Med Chem* **2012**, 19, 1135-1145.
- Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF and Babu MM. Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, 494, 185-194.
- Vermeer LS, de Groot BL, Reat V, Milon A and Czaplicki J. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations

- and comparison with H-2 NMR experiments. *Eur Biophys J Biophys* **2007**, 36, 919-931.
- Vogel R, Mahalingam M, Ludeke S, Huber T, Siebert F and Sakmar TP. Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. *J Mol Biol* **2008**, 380, 648-655.
- Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **1990**, 8, 52-56.
- Wallin E, Tsukihara T, Yoshikawa S, von Heijne G and Elofsson A. Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. *Protein Sci* **1997**, 6, 808-815.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG and Schertler GFX. Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* **2008**, 454, 486-U482.
- Wess J, Han SJ, Kim SK, Jacobson KA and Li JH. Conformational changes involved in G-protein-coupled-receptor activation. *Trends Pharmacol Sci* **2008**, 29, 616-625.

Chapter 5

Computational Tools for Analyzing Molecular Dynamics Simulations

5.1 Introduction

Depending on the duration of a molecular dynamics (MD) simulation and the time step, the amount of data can rise tremendously since coordinates, velocities, forces, etc. are saved to hard disc. The simulation itself is time-consuming, and the analysis of the generated data is challenging for the user. With a time step of 2 fs and a frequency of 500 steps for writing data to the output files, a 80 ns simulation of a protein in a natural environment (consisting of a lipid bilayer and water molecules) as described in chapter 4, produces 80,000 coordinate sets with about 3.2 billion coordinates. This corresponds to files with more than 100 GB of data. It is of course not possible to examine every single time point visually in order to detect changes in the system. Furthermore, focusing on special contacts in a protein observed in the starting structure of the simulation or on expected interactions could result in ignoring interactions or changes arising during the simulation. Therefore it is essential to analyze MD simulations systematically, i.e. to measure all possible interactions of a residue or even atom over the whole duration of the simulation. Programs are necessary which are able to handle a huge amount of data and to produce output in a user friendly format which is easily understandable.

In section 5.3 the program *gro_hbonds* is introduced which systematically calculates hydrogen bonds. H-bonds in proteins belong to the most important types of interactions, stabilizing the secondary and tertiary structure and enabling contacts to small molecules interacting with the proteins (Panigrahi and Desiraju, 2007). Besides direct H-bonds, it is also important to analyze water mediated H-bonds. For example, in GPCRs conserved water mediated H-bond networks could play an important role in receptor activation (Nygaard *et al.*, 2010). According to the IUPAC Recommendation of 2011 (Arunan *et al.*, 2011) six criteria are capable of assigning H-bonds. For molecular modeling, criteria are necessary which are

directly or indirectly available from the output of coordinate files. The hydrogen bond angle $X-H\cdots Y$ (X and Y , heteroatoms; $-$ covalent bond; \cdots hydrogen bond) and the distance between participating atoms fit best to this requirement. In molecular modeling software there are different limits (cut-off values) of angles and distances for assigning an H-bond (Table 5.1).

Table 5.1: Default cut-off values for assigning H-bonds in molecular modeling software

Software	Distance cutoff [Å] ¹	Angle cutoff [°] ²
Sybyl ³	2.8 ⁴	> 120
VMD ⁵	3.0	> 160
GROMACS ⁶	3.5	< 30 ⁷
Amber ⁸	3.0	> 135
Charmm ⁹	4.5	> 90
Chimera ¹⁰	4.0	> 160

¹ Distance between participating heteroatoms.

² Angle between heteroatom 1 – hydrogen atom – heteroatom 2.

³ Sybyl-X 1.3 (Tripos, St. Louis, MO, USA).

⁴ Sybyl measures the length of the H-bond itself (distance between the heteroatom and the hydrogen bonded H-atom).

⁵ VMD User's Guide (Humphrey *et al.*, 1996).

⁶ GROMACS User Manual (Hess *et al.*, 2008).

⁷ GROMACS defines the angle in a different way (cf. Figure 5.1).

⁸ AmberTools12 Manual (Salomon-Ferrer *et al.*, 2013).

⁹ Charmm documentation (Brooks *et al.*, 2009).

¹⁰ Chimera User Guide (Pettersen *et al.*, 2004).

The molecular dynamics software GROMACS (Hess *et al.*, 2008; Van Der Spoel *et al.*, 2005) provides an opportunity to calculate H-bonds (command *g_hbond*). However, compared to other molecular modeling programs the angle calculated for assigning H-bonds is not $X-H-Y$, but $Y-X-H$ (Figure 5.1). Furthermore, *g_hbond* is not able to perform a systematic calculation, and water mediated H-bonds are not considered in a convenient way. The speed of the calculation and the format of the output are further drawbacks of this tool.

Hydrophobic contacts between amino acid side chains are a major factor for folding proteins and the packing of α -helices (Kellis *et al.*, 1988; Matthews, 2001). Hydrophobic interactions could be analyzed with the program *g_mdmat* of GROMACS. Unfortunately, the output is inadequate for an intensive and effective analysis (cf. Figure 5.7 in section 5.4.1). Using this tool of GROMACS as a basis, the output is transformed and analyzed with the program *gro_contacts* described in section 5.4.

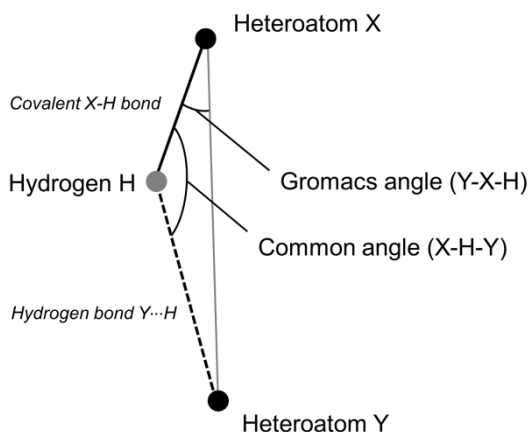


Figure 5.1: Definition of the angles used for the calculation of H-bonds

For single protein structures, the verification of the stereochemical quality is performed with tools like PROCHECK (Laskowski *et al.*, 1993). It proves the availability of the model and checks its comparability with experimentally determined protein structures. Conformations generated during MD simulations originate from artificial systems. Simple interaction models, e.g. Hooke's law for the calculation of bond lengths, are used. With respect to accurateness, force field methods (molecular mechanics) are not comparable to time-consuming *ab initio* quantum mechanics calculations. Nevertheless, reliable structures must be produced by the simulation for drawing authentic conclusions from the results. The examination of the protein structures during a complete MD simulation is performed with the program *gro_validation* described in section 5.5. It analyzes important parameters such as stereochemistry of amino acids, geometry of the peptide bond and planarity of aromatic side chains and of delocalized π -electron systems in side chains. Additionally, backbone (Φ and Ψ angles) and side chain dihedral angles (rotamers) are measured and compared to experimentally derived values.

5.2 Materials and methods

The analysis was performed with data obtained from MD simulations using the software package GROMACS 4.0.7 (Van Der Spoel *et al.*, 2005) and the GROMOS96 53a6 force field (Oostenbrink *et al.*, 2004). The programs were composed of shell scripts and subroutines written in the computer language C. All calculations were performed on a DELL Precision T3500 (Intel® Xeon® Processor W3565, 3.2 GHz and 6 GB RAM) with the operating system CentOS 5.6. Figures were created with Microsoft® Excel® 2010.

5.3 Systematic calculation of direct and water mediated H-bonds: *gro_hbonds*

5.3.1 Skills of the program

5.3.1.1 Systematic calculation of hydrogen bonds

The program is able to calculate H-bonds from each heteroatom of a MD simulation system to all remaining heteroatoms at any time. The criteria for assigning an H-bond are the distance between two heteroatoms and the angle formed by the first heteroatom, a hydrogen atom (bonded to one of the two heteroatoms) and the second heteroatom. Default values (maximum distance of 3.5 Å and an angle of at least 120°) may be adjusted in the parameter file (parameters *var-cutoffdist* and *var-cutoffang*; Table 5.2). To save computation time and avoid repeats in the output files, only interactions of a specific heteroatom to subsequent heteroatoms (to heteroatoms with a higher atom number) in the molecular structure file of GROMACS are calculated. If there are more than one hydrogen atom part of an H-bond at different time points, these interactions are not treated separately but summarized to one result. For example, H-bonds of the side chain of lysine are not classified for each hydrogen atom at the amine nitrogen. The calculation is performed for each heteroatom starting at residue number one until the residue number corresponds to the value of the variable *residues* defined in the parameter file (Table 5.2). The program also allows the determination of H-bonds which are mediated by one water molecule. Every water molecule is separately checked for being H-bonded to two heteroatoms at the same time. Interactions between the same couple of heteroatoms mediated by different water molecules at different time points are combined to one result. So far, every proteinogenic amino acid listed in the residue topology file of the GROMOS96 53a6 force field (*ffG53a6.rtp*) is included in the program. Additionally considered are the residues for blocking the N- and C-terminus (ACE and NH2, respectively), the phospholipid 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), histamine, SPC (Simple Point Charge) water and chloride ions. To include more residues, see Figure 5.2 for an example.

5.3.1.2 Analysis of helical structures

As an alternative to the DSSP (Define Secondary Structure of Proteins) algorithm (Kabsch and Sander, 1983), the helical structure analysis tool may be used to test if assumed α -helical parts of proteins such as TM segments of GPCRs are really helical according to the IUPAC definition using H-bonds (rule 6.3; CBN, 1970). Each backbone oxygen atom is

tested if it is hydrogen bonded to the backbone nitrogen localized four residues farther in the sequence ($i+4 \rightarrow i$ hydrogen bonding). Additionally, the secondary structure can be checked for a 3_{10} helix (the backbone nitrogen of an amino acid forms an H-bond with the backbone oxygen of the amino acid three residues before, corresponding to a $i+3 \rightarrow i$ hydrogen bonding) and a π -helix ($i+5 \rightarrow i$ hydrogen bonding).

5.3.1.3 Structure validation of H-bonds

To validate a protein structure, it is useful to check for unoccupied buried H-bond donors and acceptors like performed by the molecular modeling package WHATIF (Vriend, 1990). The frequency of any heteroatom (nitrogen, oxygen or sulfur) being H-bonded during a MD simulation as well as the time course of the H-bonds are calculated.

5.3.1.4 Detailed analysis of specific hydrogen bonds

For specific interactions, a detailed output can be generated with time dependent distances between heteroatoms and angles of the respective H-bond during the entire simulation, as well as a time-resolved graph of the hydrogen bond presence (*jpg* format).

5.3.2 Structure of the program

For the calculation of H-bonds three files are necessary. The shell script *gro_hbonds.sh* organizes the analysis of H-bonds. It extracts variables from the parameter file (*gro_hbonds-para.txt*) which is necessary to provide information from the simulation and enables to adjust the program by the user. Table 5.2 lists all parameters with a short description of the meaning. An exemplary parameter file is shown in the Appendix (section 9.3.1). Furthermore, the script transfers computationally intensive calculations to the C program *gro_hbonds-calc*. The shell script and the C program are divided in different sections and functions, respectively (Table 5.3). The source code of the programs is also shown in the Appendix (sections 9.3.2 and 9.3.3). The program is started with the command *./gro_hbonds.sh*.

Table 5.2: Content of the parameter file for calculating H-bonds

Parameter name	Description
<i>var-sim</i>	Name of the simulation (all output files start with this term).
<i>var-filename</i>	Name of the <i>gro</i> , <i>top</i> and <i>xtc</i> file of the MD simulation. All three files must have the same name. Otherwise the respective lines in the shell script have to be adjusted.
<i>var-duration</i>	The duration of the simulation in picoseconds.
<i>var-timestep</i>	The time step of the analysis in picoseconds. E.g. a value of 10 means that the analysis is performed every 10 th ps.
<i>var-residues</i>	Last residue of which H-bonds to the rest of the system should be calculated.
<i>var-cutoffdist</i>	Cut-off value for the distance heteroatom 1 – heteroatom 2 for H-bonds in picometer (integer).
<i>var-cutoffang</i>	Cut-off value for the angle (in degree) heteroatom 1 – hydrogen – heteroatom 2 for H-bonds (integer).
<i>var-histo</i>	The number of data points (time points) to be summarized in one average value for the frequency distribution in the output file (integer). Assuming a value of 50 for <i>var-histo</i> and a time step of 10 ps, one data point in the output is the average of 500 ps.

5.3.3 Adjusting the program

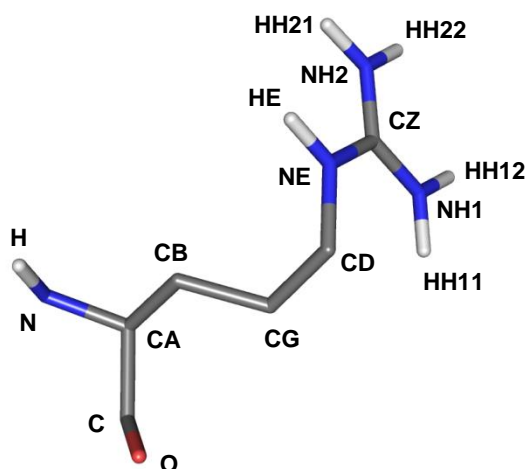
The C program *gro_hbonds-calc.c* can be adjusted to allow the H-bond calculation for more than the residues mentioned in section 5.3.1.1. To add a molecule or residue, the array *AAname* (line 43 ff.) containing all residue names has to be supplemented. The variable *no_def_res* (line 11) indicating the number of molecules/residues which can be analyzed by the program has to be raised by one. Furthermore, a description of the newly added residue/molecule has to be assigned in *AAcode* (line 87 ff.). An example for the usage is shown for the amino acid arginine in Figure 5.2.

For the H-bond calculation of charged N- and C-terminal amino acids, the program has to be adjusted. In contrast to the residue topology file of GROMACS (*ffG53a6.rtp*) such residues contain an additional atom. For example, in the MD simulation of the hH₂R_{a2} (chapter 4) the G-protein fragment GαCT contains at its C-terminus a charged leucine (residue number 319). In this case an additional entry in *AAcode* has to be made and the number of residues the program can handle has to be raised by one (variable *no_def_res*). Furthermore, three

sections (lines 266, 282 and 979) of the source code have to be changed. In case of more than one charged terminal residue corresponding lines have to be added to the respective section of the program. All parts of the program which have to be adjusted are identified easily by the term *charge_termini*.

Table 5.3: Sections of the shell script *gro_hbonds.sh*

Section	Description
1	The variables are received from the parameter file and a list with the residue sequence is generated.
2	Calculation of H-bonds. A file containing the coordinates of 1,000 time points of the whole MD system is created from the binary <i>xtc</i> file (compressed version of the trajectory of the MD simulation with time, coordinates and box vectors) using the command <i>trjconv</i> . This file is used to calculate H-bonds with <i>gro_hbonds-calc</i> (function 1). Subsequently, all these subfiles containing the H-bond information of 1,000 frames are combined (function 2 of <i>gro_hbonds-calc</i>) to one results file (<i>hbond-Het_int.txt</i>). This file contains all information necessary for subsequent analysis (sections 3, 4, 5 and 7). Five further files are created (cf. section 5.3.4) using function 4 of the C program which summarizes type (backbone and side chain) and residue based interactions from the heteroatom based input file (for a description see Table 5.5).
3	Calculation of water mediated H-bonds. Extract all data from <i>hbond-Het_int.txt</i> where a water molecule is part of the H-bond. Uses the function 3 of <i>gro_hbonds-calc</i> . Six files are written.
4	Helix analysis. All records containing interactions between backbone atoms are extracted from <i>hbond-Het_int.txt</i> . With <i>gro_hbonds-calc</i> (function 5) it is checked if an H-bond occurs between the carbonyl oxygen atom of an amino acid and the nitrogen atom three, four or five amino acids further in the sequence, respectively. Six files are created.
5	Calculation of the fraction of time points of the MD simulation a specific heteroatom (N, O and S) is hydrogen bonded (function 6 of <i>gro_hbonds-calc</i>). Using <i>hbond-Het_int.txt</i> as input, two output files are generated.
6	Detailed output (distance and angle over time) for one specific H-bond (function 8 of <i>gro_hbonds-calc</i>). This section could be used separately from the rest of the script. The atom numbers (of the <i>gro</i> file) of the heteroatoms and the hydrogens are needed. One file is written.
7	Graph of the presence of one H-bond during the entire simulation. The H-bonding data are received from one record in <i>hbond-Het_int.txt</i> (corresponds to one H-bond between two heteroatoms). Using the function 9 of the C program, an <i>xpm</i> file is produced which is then converted to a <i>jpg</i> picture.



```
[ ARG ]
[ atoms ]
```

N	N	-0.31000	0
H	H	0.31000	0
CA	CH1	0.00000	1
CB	CH2	0.00000	1
CG	CH2	0.00000	1
CD	CH2	0.09000	2
NE	NE	-0.11000	2
HE	H	0.24000	2
CZ	C	0.34000	2
NH1	NZ	-0.26000	2
HH11	H	0.24000	2
HH12	H	0.24000	2
NH2	NZ	-0.26000	2
HH21	H	0.24000	2
HH22	H	0.24000	2
C	C	0.450	3
O	O	-0.450	3

Code for arginine in *AAcode* of *gro_hbonds-calc.c*: {17,5,0,1,6,1,9,2,12,2,16,0}

Parameter	Atom name	Description
17	-	Number of atoms/sites of the residue.
5	N, NE, NH1, NH2, O	Number of heteroatoms.
0	N	Position of the first heteroatom in the residue topology file (<i>rtf</i>) of GROMACS, starting with zero.
1	H	Number of hydrogen atoms bonded to the preceding heteroatom.
6	NE	Position of the second heteroatom.
1	HE	Number of hydrogen atoms bonded to the preceding heteroatom.
9	NH1	Position of the third heteroatom.
2	HH11, HH12	Number of hydrogen atoms bonded to the preceding heteroatom.
12	NH2	Position of the fourth heteroatom.
2	HH21, HH22	Number of hydrogen atoms bonded to the preceding heteroatom.
16	O	Position of the 5 th heteroatom.
0	-	Number of hydrogen atoms bonded to the preceding heteroatom.

Figure 5.2: Code for H-bond calculation of arginine

Arginine with labeled atom names and the atom definition section such as listed in the residue topology file of the GROMOS96 53a6 force field (*ffG53a6.rtf*) is shown. The parameters of the variable *AAcode* used for the calculation of H-bonds with arginine are explained in the table.

5.3.4 Output files

5.3.4.1 Types of output

Two folders are created to store the output data (text files). In *output_sum* (Table 5.4), all files contain line by line a specific H-bond pair, the percentage of the simulation this H-bond was present as well as the time-resolved H-bond information (frequency distribution). The results for direct and water mediated H-bonds are given in three different ways, depending on the precision of the output (Table 5.5).

Table 5.4: Files created for analyzing H-bonds

File	Description
hbond-Het_-SOL.txt ¹	Direct H-bonds ⁴ .
hbond-type_-SOL.txt ²	
hbond-res_-SOL.txt ³	
hbond_SOL-Het.txt ¹	Water mediated H-bonds.
hbond_SOL-type.txt ²	
hbond_SOL-res.txt ³	
hbond-helix-1-3-res.txt ⁵	Helix analysis. Each line contains the data of one residue pair. Columns one to three: residue numbers of the two amino acids and fraction of time points an H-bond occurs. Column 4 ff.: data for the frequency distribution.
hbond-helix-1-4-res.txt ⁵	
hbond-helix-1-5-res.txt ⁵	
hbonds-amount.txt	Fraction of time points a specific heteroatom is hydrogen bonded. Column one to four: amino acid number, heteroatom number in that amino acid, backbone (0) or side chain (1) heteroatom and fraction of time points the heteroatom is H-bonded. Column 5 ff.: data for the frequency distribution.
hbond-dist-ang-het1-het2-H.txt	Distance and angle information of a specific H-bond for each time point (in lines).
hbond-existance_example.jpg	A black line indicates the presence of an H-bond, using the cut-off values for H-bond distance and angle in the parameter file.

¹ Heteroatom based information, cf. Table 5.5.

² Type based information, cf. Table 5.5.

³ Residue based information, cf. Table 5.5.

⁴ The three files provided contains all hydrogen bonding data except those where water molecules are part of the H-bond (term *-SOL*) in order to reduce the size of the output files.

⁵ H-bonds from O_i to N_{i+3}, O_i to N_{i+4} and O_i to N_{i+5}, respectively.

Table 5.5: Different types of output files for direct and water mediated H-bonds

Term:	Het	type	res
Description:	Heteroatom based H-bond information: H-bond between two heteroatoms	Type based H-bond information: H-bonds between one side chain or backbone and another side chain or backbone are summarized	Residue based H-bond information: H-bonds between residues are summarized
Example:	Investigate which heteroatom of a ligand is H-bonded to the protein	To analyze if an H-bond occurs between an amino acid side chain and the backbone of another amino acid	Analyze if two residues are connected via H-bonds
Column in the output file	1	Number of the residue which contains the first heteroatom	First residue
	2	Number of the residue which contains the second heteroatom	Second residue
	3	Code for the first heteroatom which indicates its location [‡]	Fraction of time points a H-bond occurs between the respective residues
	4	Code for the second heteroatom which indicates its location [‡]	Frequency distribution of the H-bond(s) over time
	5	Number of the first heteroatom in the respective residue	
	6	Number of the second heteroatom in the respective residue	
	7	Atom number in the <i>gro</i> file of the first heteroatom	
	8	Atom number in the <i>gro</i> file of the second heteroatom	
	9	Fraction of time points a H-bond occurs between the heteroatoms	
	10 - end	Frequency distribution of the H-bond over time	

[‡] 0, backbone; 1, side chain; 2, terminus ACE or NH₂; 3, lipid; 4, water; 5, ion; 6, ligand

Each line record starts with two to eight numbers which identify the heteroatoms or groups of heteroatoms which contribute to this respective contact and could be used for extracting information about specific interaction partners. All numbers are separated by tabulators. Furthermore, in the folder *output_detail* the complete data as obtained from the H-bond calculation are provided (with the term *_int*), listing every time point (in picoseconds) an H-bond occurs without summing up like in the frequency distribution. These files could be used for a detailed and further analysis.

5.3.4.2 Direct and water mediated H-bonds

An example of the output for direct H-bonds is shown in Figure 5.3. In 78.4% of the simulation of the inactive hH₂R state (chapter 4) an H-bond between Thr226^{6.27} and Glu229^{6.30} occurred (Figure 5.3 A). This data is obtained from the file *hbond-res_-SOL.txt*. The contact could be split into the interaction of the backbone atoms of both residues, of the backbone of Thr226^{6.27} with the side chain of Glu229^{6.30} and of the side chains of both amino acids (14.5%, 77.4% and 48.4%, respectively; Figure 5.3 B-D). This information was extracted from *hbond-type_-SOL.txt*. The latter two interactions could again be split depending on the carboxyl oxygen of the Glu229^{6.30} side chain which is part of the H-bond (file *hbond-Het_-SOL.txt*, Figure 5.3 E, F). Water mediated H-bonds could be analyzed equivalent, using the data from the files *hbond_SOL-Het.txt*, *hbond_SOL-type.txt* and *hbond_SOL-res.txt*, respectively.

5.3.4.3 Helix analysis

Figure 5.4 shows an example for the analysis of α -helices from the simulation of hH₂R_i (chapter 4). The data provided in the output files could be used to investigate the stability of TM domains. Furthermore, the existence of a special H-bond could be observed over time.

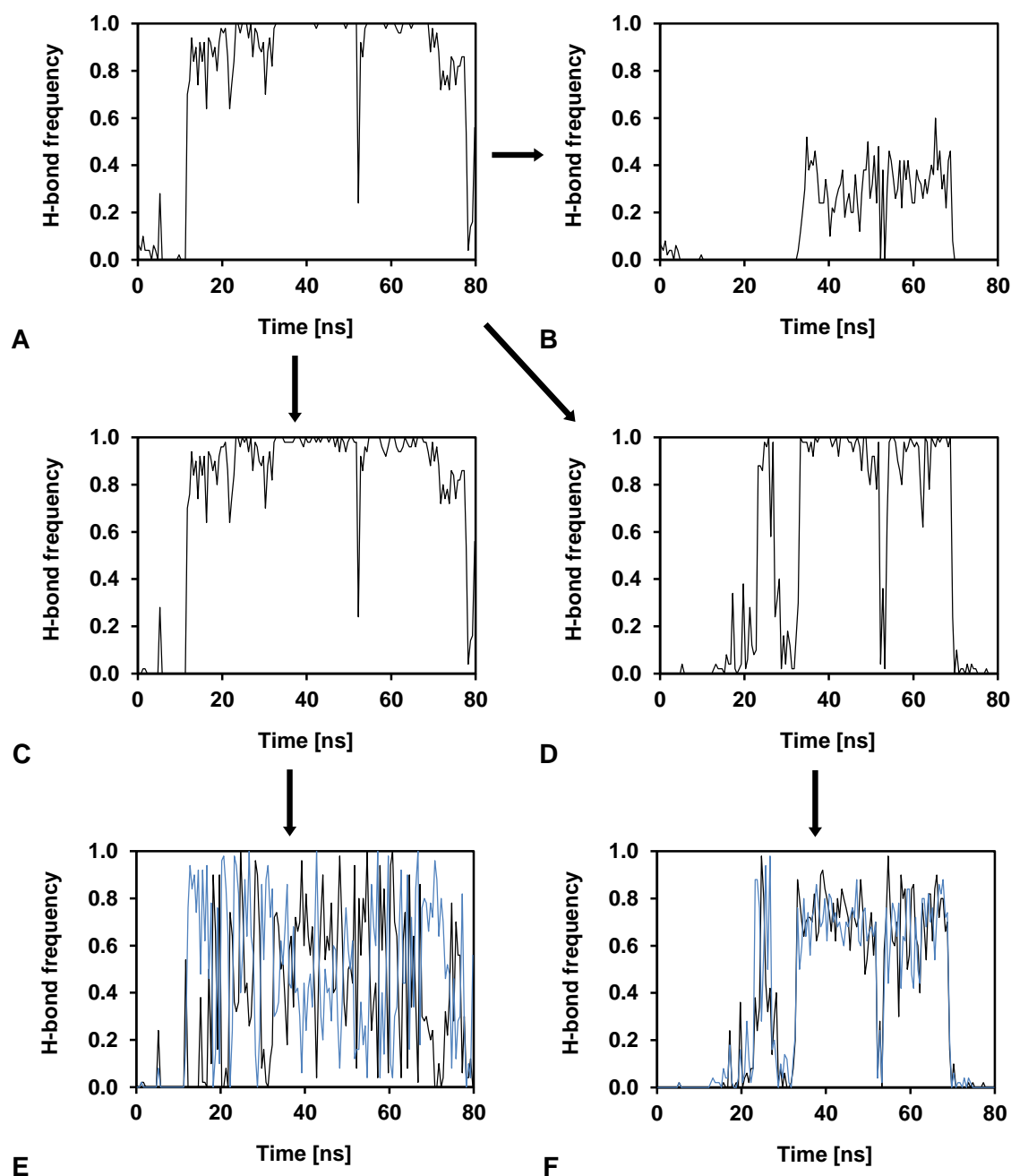


Figure 5.3: Example of the output for analyzing direct H-bonds

Interaction between Thr226^{6.27} and Glu229^{6.30} during the MD simulation of hH₂R_i (chapter 4). A, H-bond frequency between both residues; B, H-bond from the backbone of Thr226^{6.27} to the backbone of Glu229^{6.30}; C, H-bond from the backbone of Thr226^{6.27} to the side chain of Glu229^{6.30}; D, H-bond between the side chains of both residues; E, H-bond from the backbone oxygen of Thr226^{6.27} to the carboxyl oxygen atoms of Glu229^{6.30} (OE1, blue line; OE2, black line); F, H-bond from the side chain oxygen of Thr226^{6.27} to the carboxyl oxygen atoms of Glu229^{6.30} (OE1, blue line; OE2, black line).

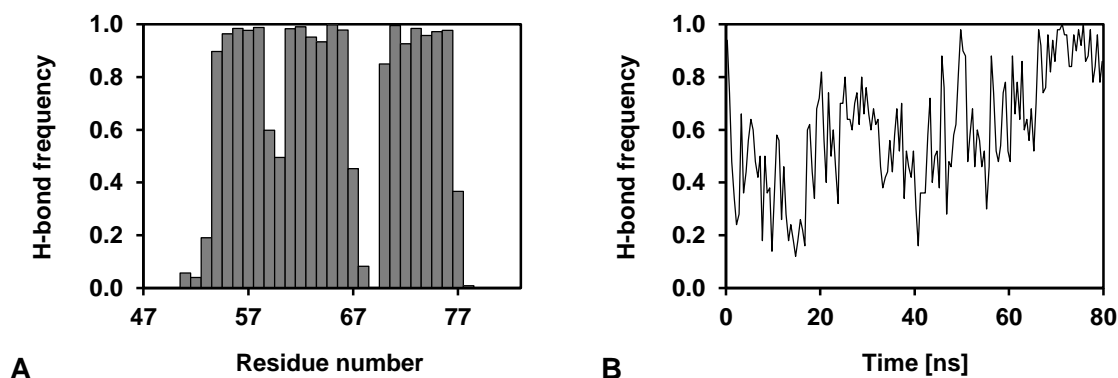


Figure 5.4: Helix analysis

A, H-bond frequency between O_i and N_{i+4} in TM2 of hH_2R_i during the 80 ns MD simulation (chapter 4); B, The reduced helicity of residue 59 is analyzed in detail.

5.3.4.4 Structure validation of H-bonds

The output file *hbonds-amount.txt* contains line by line the data for the heteroatoms of the simulation system, i.e., the fraction of time points the respective heteroatom is H-bonded and the time course (Figure 5.5).

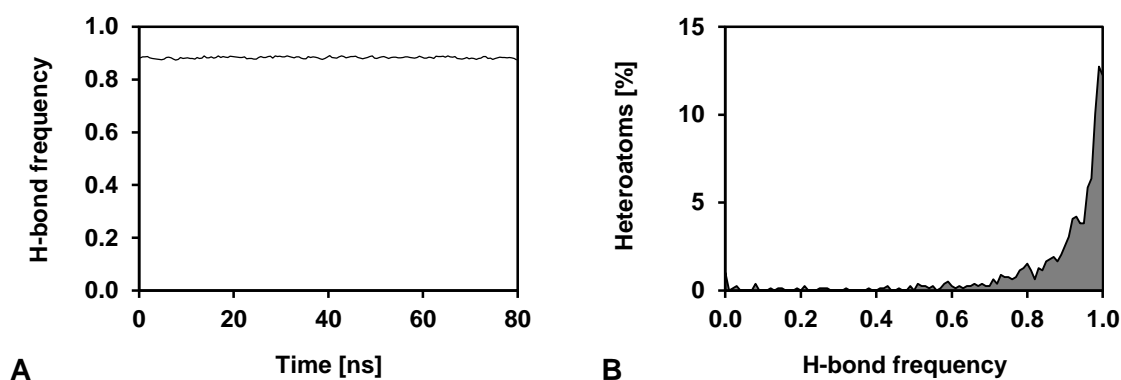


Figure 5.5: Saturation of H-bond donors and acceptors

A, Average frequency heteroatoms of hH_2R_i are H-bonded during the 80 ns MD simulation described in chapter 4. It is also possible to analyze the frequency distribution over time for one specific heteroatom. B, Percentage of heteroatoms which are H-bonded in the fraction of time points given on the abscissa during the whole MD simulation.

5.3.4.5 Detailed output for a specific interaction

An example for a detailed H-bond analysis is shown in Figure 5.6.

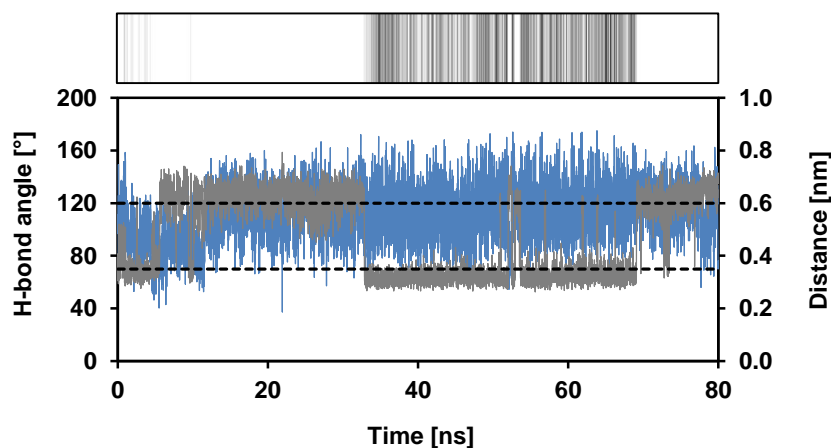


Figure 5.6: Detailed analysis of hydrogen bonds

H-bond angle (blue) and heteroatom distance (grey) between the backbone oxygen of Thr226^{6.27} and the nitrogen of Glu229^{6.30} during the MD simulation of hH₂R_i (chapter 4). The black dashed lines at 120° and 0.35 nm indicate the cut-off values for the H-bond calculation. Above: Associated H-bond existence map (*jpg* file). Different grey tones indicate, from dark to light, higher and lower frequency of H-bonds, respectively.

5.3.5 Performance

The 80 ns MD simulation of the hH₂R_i embedded in a DPPC bilayer and solubilized with water molecules (chapter 4) was used for testing the H-bonds analysis program. The simulation system contains about 1,750 heteroatoms of the protein and the lipid, and additionally about 10,300 oxygen atoms of the water molecules. Thus, calculating H-bonds for just one time point includes the analysis of more than 12,000 heteroatoms and coordinates, respectively. Consequently, for more than 19.4 million possibly interacting heteroatoms distances have to be measured. In case of the 80 ns MD simulation where every 10th picosecond was analyzed (corresponding to 8,000 frames in total) this results in more than 155 billion comparisons which were performed by the function 1 of *gro_hbonds-calc*. The analysis of the hH₂R_i simulation ran for about 3:30 hours. Most time-consuming are the transformation of the binary *xtc* file into the *gro* file using the tool *trjconv* of GROMACS and the calculations performed with the function 1 of *gro_hbonds-calc*.

5.4 Analysis of hydrophobic interactions: *gro_contacts*

5.4.1 Skills of the program

The tool *g_mdmat* of GROMACS analyzes the contacts between residues of a MD simulation. However, the output is rather useful for visual inspections than for a detailed analysis of interactions. The investigation of hydrophobic contacts in the 80 ns MD simulation of the hH₂R_i (chapter 4) at every 10th picosecond in the analysis (i.e. 8,000 coordinate sets in total) results in an *xpm* image file with a size of more than 1.39 GB. This file is divided into 8,000 parts, each representing a contact map for a specific time point (Figure 5.7). Every pixel of such map denotes an interaction between two residues (distance between the closest atoms of two residues), encoded by an alphabetic letter in the file. Depending on the parameters submitted with the command *g_mdmat* e.g. these letters range from 'A' (distance smaller than 0.2 Å; white color in the picture) to 'y' (distance more than 10 Å; black color in the picture). This kind of data is not suitable for a systematic and fast graphical analysis.

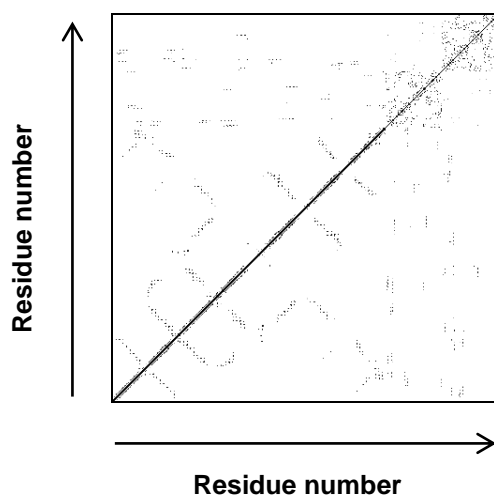


Figure 5.7: Hydrophobic contact map for one specific time point created by *g_mdmat*

For reason of clarity deviating from the original output of *g_mdmat* the colors (decreasing grey scale) are inverted (black, close distance; white, distance more than 10 Å).

Thus, after a short modification, the *xpm* file containing all contact data is read with the C program *gro_contacts-calc* and analyzed for pairs of residues which are at least at one time point closer than the cut-off value defined in the parameter file (*var-cutoff*, Table 5.6). For these interaction pairs the time dependent distances as well as the fraction of time points the contact is closer than the cut-off value are printed to the output file. The generation of an appropriate index file for the command *g_mdmat* allows the selection of atoms (e.g. hydrophobic atoms) which should be considered for the contact analysis of each residue type. An example which was used for analysis of the MD simulations in chapter 4 is given in the Appendix (section 9.4.2).

5.4.2 Structure of the program

The program *gro_contacts* is composed of a shell script (*gro_contacts.sh*; cf. section 9.4.3) receiving information from the parameter file *gro_contacts-para.txt* (Table 5.6). It creates an index file with the atom numbers of a specific MD simulation, considering a reference file *contacts-atoms.txt* which contains in each line a residue and atom name. In the analysis of residue-residue contacts only the atoms given in the index file are considered. If the definition of atoms to be included in the calculation has to be modified, line 96 of the C program must be adjusted (cf. section 9.4.4). The variable *AAcode* contains a number indicating if the respective residue is included in the analysis. So far all proteinogenic amino acids of the GROMOS96 53a6 force field are listed in the file. Furthermore, it contains atom definitions of the phospholipid 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) and of histamine.

Table 5.6: Content of the parameter file for the calculation of hydrophobic contacts

Parameter name	Description
<i>var-sim</i>	Name of the simulation (all output files start with this term).
<i>var-filename</i>	Name of the <i>gro</i> , <i>tpx</i> and <i>xtc</i> files which must have the same name. Otherwise the respective lines in the shell script have to be changed.
<i>var-duration</i>	The duration of the simulation in picoseconds.
<i>var-timestep</i>	The time step of the analysis in picoseconds. E.g. a value of 10 means that the analysis is performed every 10 th ps.
<i>var-cutoff</i>	Cut-off value for contacts of two residues in picometer (the number has to be divided by 10 without a rest).
<i>var-last</i>	Last residue to be analyzed (residue number in the <i>gro</i> file).

An example parameter file as well as the source code of *gro_contacts.sh* and *gro_contacts-calc.c* is shown in the Appendix (section 9.4.1, 9.4.3 and 9.4.4, respectively). The program is started with the command *./gro_contacts.sh*.

5.4.3 Output files

Two files are generated. In a short summary file (*contacts_sum.txt*) each line includes the tabular information of a specific interaction between two residues (Figure 5.8). The first two columns contain the residue numbers and the third one the fraction of time points in which the contact between the corresponding residues is below the cut-off value given in the parameter file (*var-cutoff*). The detailed file (*contacts.txt*) additionally contains the distance of two residues at every time point analyzed. For distances greater than 1.0 nm, a value of 2.0 is assigned. This can be modified in line 168 of *gro_contacts-calc.c*.

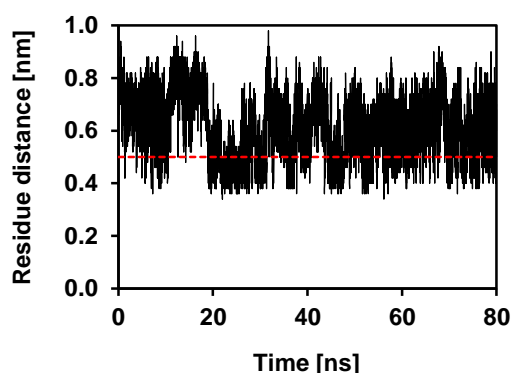


Figure 5.8: Hydrophobic contact analysis

Minimum distance between the side chains of Leu28^{1.42} and Leu72^{2.58} in the MD simulation of hH₂R_i (chapter 4). The red line indicates the cut-off (0.5 nm) for assigning a contact between residues. In 24.1% of the 8,000 time points analyzed the distance was below that limit. In no frame the distance was larger than 1.0 nm.

5.4.4 Performance

The 80 ns MD simulation of the hH₂R_i embedded in a DPPC bilayer and solubilized with water molecules (chapter 4) was used for testing the analysis of hydrophobic contacts with *gro_contacts*. The contact map produced with the command *g_mdmat* of GROMACS which was used as input for the C program *gro_contacts-calc* contained about 672 million data points (when analyzing 8,000 time points), each representing the shortest distance between two atoms of two residues in alphabetic letters. The complete program runs 31 minutes, whereas the analysis of the contact map with *gro_contacts-calc* needed only 6 minutes.

5.5 Structure validation of molecular dynamics simulations: *gro_validation*

5.5.1 Skills of the program

This structure validation tool for MD simulations performed with the simulation package GROMACS (Van Der Spoel *et al.*, 2005) searches for proper stereochemical configurations at C_α atoms of amino acids, checks the planarity of aromatic side chains (His, Phe, Trp, and Tyr) and delocalized π -electron systems in side chains of Arg, Asn, Asp, Gln and Glu, and controls the planarity of the peptide bond. It further calculates the backbone Φ/Ψ distribution (Ramachandran analysis) and compares the side chain rotamers with a dataset obtained from crystalized proteins (Lovell *et al.*, 2003; Lovell *et al.*, 2000). All dihedral angles were calculated with subprograms of GROMACS (*g_angle*, *g_chi*, *g_rama*). Summarization of angles and comparison to experimental values were performed with *gro_validation.sh* and *gro_validation-calc*.

5.5.1.1 Chirality check

The dihedral angle defined by the atoms C_α , N, C and C_β represents the rotation about the virtual bond N-C. It checks the stereochemistry of an amino acid at the C_α atom and distinguishes L- from D-stereoisomers (Morris *et al.*, 1992). In an ideal tetrahedral arrangement of the four atoms around the C_α atom of L-amino acids, the angle between the planes C_α -N-C and N-C- C_β is about 35° (Figure 5.9). A correct stereochemistry is assumed if the dihedral angle C_α -N-C- C_β is between 0° and 70° . These cut-off values can easily be modified by adjusting line 229 of *gro_validation-calc.c* (section 9.5.3).

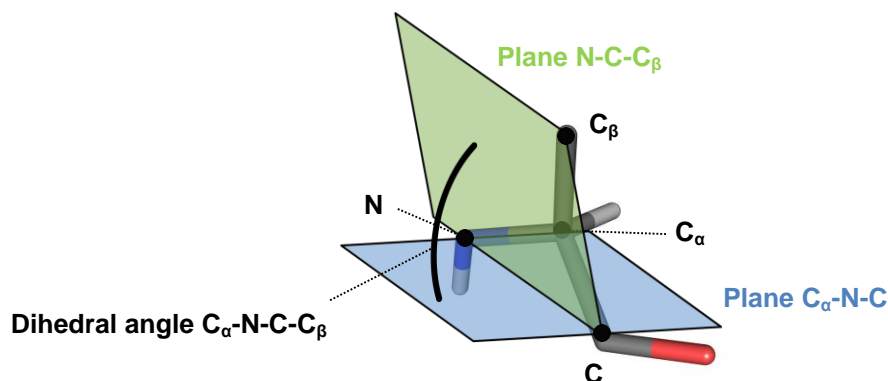


Figure 5.9: Dihedral angle definition for checking the stereochemistry of amino acids

The torsion angle C_α -N-C- C_β is calculated for each amino acid and time point requested by the tool *g_chi* of GROMACS. The data are transferred into and analyzed by the C program *gro_validation-calc*.

5.5.1.2 Planarity check

There are different ways to test normally planar moieties of amino acids such as aromatic side chains (His, Phe, Trp and Tyr) or polar groups containing a delocalized π -electron system (Arg, Asn, Asp, Gln and Glu) for their deviation from planarity. In the structure validation tool PROCHECK (Laskowski *et al.*, 1993), checking single protein structures for their stereochemical quality, the root mean square distance of atoms to a plane is calculated. Another possibility also used in *gro_validation* is the measurement of dihedral angles defined by atoms of the putatively planar substructures (Figure 5.10). Depending on the number of atoms in the respective fragments, different numbers of dihedral angles are considered in the analysis (Table 5.7). These improper torsion angles are also used by GROMACS to control planarity and are defined in the residue topology file (*ffG53a6.rtp*). Dihedral angles are calculated with *g_angle* and passed to *gro_validation-calc* to summarize the data.

Table 5.7: Number of dihedrals per amino acid used for calculating the deviation from planarity

Amino acid ¹	Number of dihedrals ²	Amino acid ¹	Number of dihedrals ²
ARG	4	HISA ⁶	9
ARGN ³	3	HISB ⁶	9
ASN	2	HISH ⁶	10
ASN1 ⁴	2	HIS1 ⁶	9
ASP	1	HIS2 ⁶	9
ASPH ⁵	1	PHE	12
GLN	2	TRP	20
GLU	1	TYR	12
GLUH ⁵	1		

¹ All amino acids with planar moieties listed in the residue topology file of the GROMOS96 53a6 force field (*ffG53a6.rtp*).

² The number of dihedrals were taken from the residue topology file *ffG53a6.rtp* of GROMACS.

³ Neutral arginine.

⁴ Contains another type of N in the side chain compared to ASN.

⁵ Protonated side chain.

⁶ Different protonation states of nitrogen atoms in the side chain of histidine.

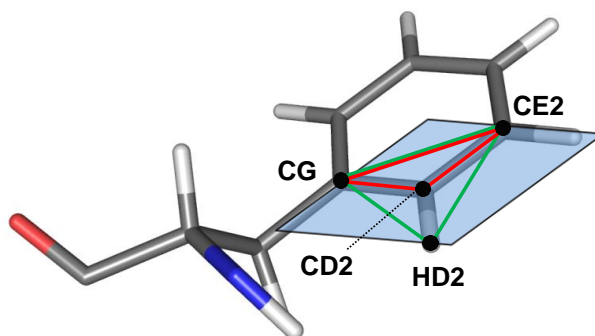


Figure 5.10: Checking for planarity with dihedral angles

One of the 12 dihedral angles used to control the planarity of the phenylalanine side chain. The angle between the planes CD2-CG-CE2 (red lines) and CG-CE2-HD2 (green lines) is measured. For an ideal planar aromatic moiety the angle is 0° and both planes coincide (blue). The atom names were adapted from GROMACS.

5.5.1.3 Peptide bond analysis

According to the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) the peptide omega (ω) angle of residue i is defined by the atom sequence $C^\alpha_i - C_i - N_{i+1} - C^\alpha_{i+1}$, i.e. the angle between the planes $C^\alpha_i - C_i - N_{i+1}$ and $C_i - N_{i+1} - C^\alpha_{i+1}$ (CBN, 1970). Ideally this angle amounts to 180° (trans configuration of the peptide bond $C_i - N_{i+1}$) or 0° (cis configuration), so that the atoms C^α_i , C_i , O_i , N_{i+1} , H_{i+1} and C^α_{i+1} lie in the same plane (Figure 5.11). By contrast, the tool *g_rama* of GROMACS uses the atoms $O_i - C_i - N_{i+1} - H_{i+1}$ for the calculation. For an ideal planar geometry and a trans configuration of the atoms, also an angle of 180° results. The data received from *g_rama* are summarized with *gro_validation-calc*.

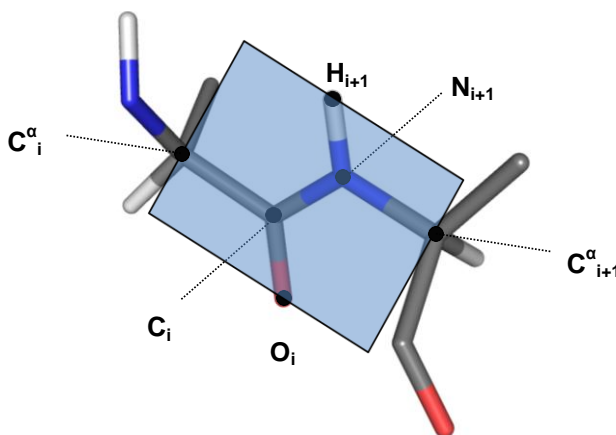


Figure 5.11: The peptide bond and its atoms located in one plane

5.5.1.4 Ramachandran analysis

The protein backbone dihedral angles Φ (phi) and Ψ (psi) of residue i are defined by the sequence of atoms $C_{i-1}-N_i-C_i^\alpha-C_i$ and $N_i-C_i^\alpha-C_i-N_{i+1}$, respectively (CBN, 1970). Depending on the type of an amino acid, different combinations of Φ/Ψ angles are preferred. A suitable classification is the formation of the four groups glycine, proline, pre-proline (amino acids preceding proline in the sequence) and 'general' amino acids (all remaining amino acids).

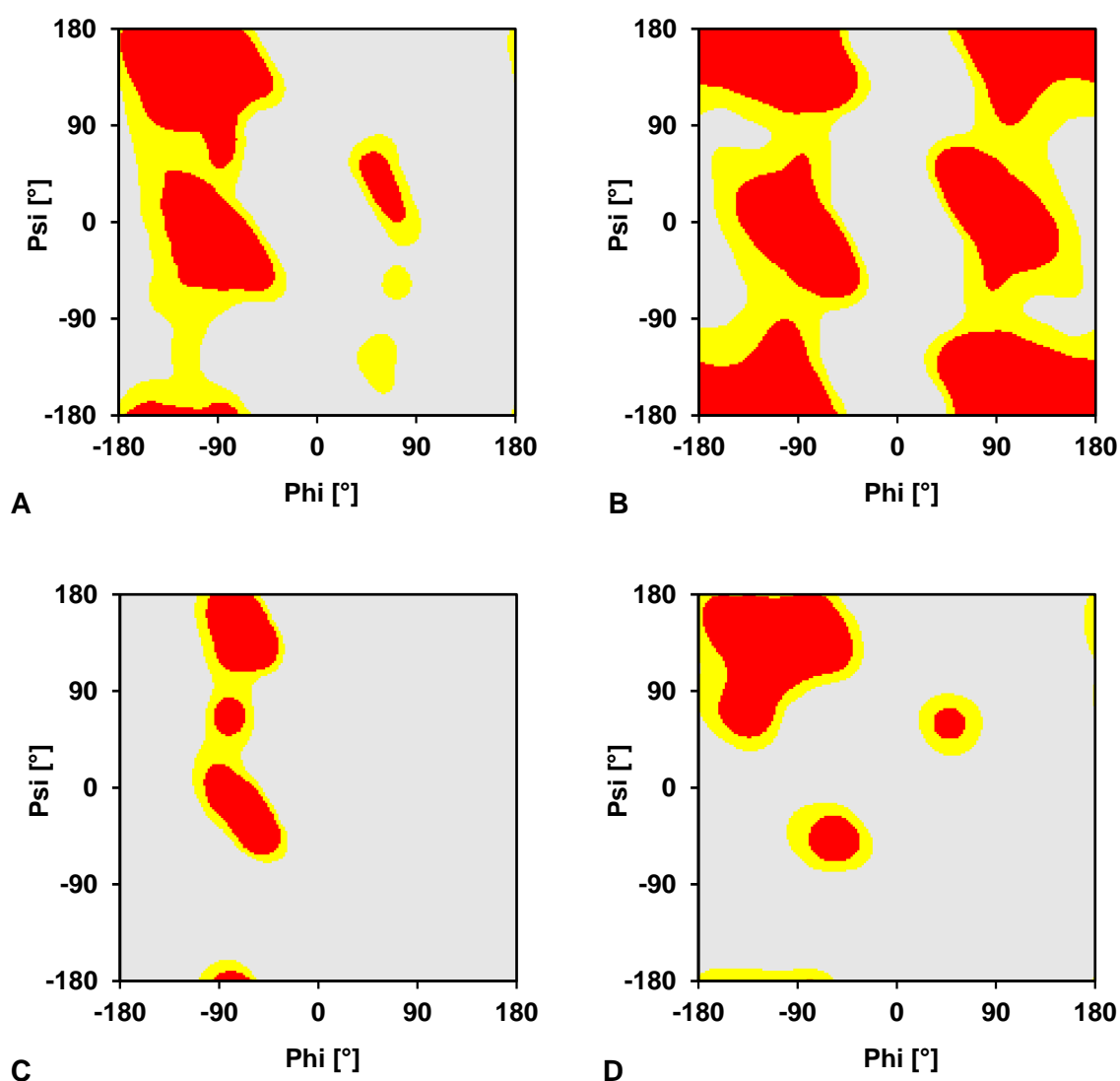


Figure 5.12: Ramachandran plots

A, 'General' amino acids (all except the following); B, Glycine; C, Proline; D, Pre-proline (amino acids before proline). All plots were created with data obtained from Lovell *et al.* (2003). Level limits between favored (red) and allowed (yellow) and between allowed and disallowed (grey) are 0.02 and 0.002, respectively.

Φ and Ψ angles are calculated with the tool *g_rama* of GROMACS. Pairs of values are then compared to reference data derived from protein crystal structures. Accordingly, each Φ/Ψ combination, classified in intervals with an increment of 2° , has a particular 'probability value' calculated from the probability distribution of the Φ/Ψ torsions in proteins (Lovell *et al.*, 2003; Lovell *et al.*, 2000). The dihedral space is divided into three regions. Φ/Ψ angles with a value below 0.002 belong to the disallowed region (outlier), values above 0.02 indicate probable, favored backbone conformations, and intermediate values are part of the allowed regions (Figure 5.12). In the program *gro_validation-calc* all combinations of Φ and Ψ dihedrals received from a MD simulation are assigned to the proper reference Φ/Ψ interval and marked as favored, allowed or disallowed due to the 'probability value'.

5.5.1.5 Side chain rotamers

Similar to Φ and Ψ angles, also side chain dihedral angles and their combination in case of more than one dihedral angle in the side chain occur with varying frequency, leading to different distributions of rotamer states. Databases containing such information could be used to compare values obtained from MD simulations with experimental values. The side chain dihedral angles are calculated with *g_chi* of GROMACS and are then transferred to *gro_validation-calc*, which compares the data of simulations with the reference data and classifies it in the groups favored, allowed or disallowed. The same cut-off values (0.002 and 0.02) as for the Ramachandran analysis are used. Figure 5.13 shows the conformational space of amino acids with one or two dihedral angles.

5.5.2 Structure of the program

The program *gro_validation* is composed of three files (parameter file, shell script, C program) and one folder containing the reference data provided by Lovell *et al.* (2000), as well as files with the definition of dihedral angles adapted from the GROMOS96 53a6 force field. The interplay of programs and text files is described in Figure 5.14.

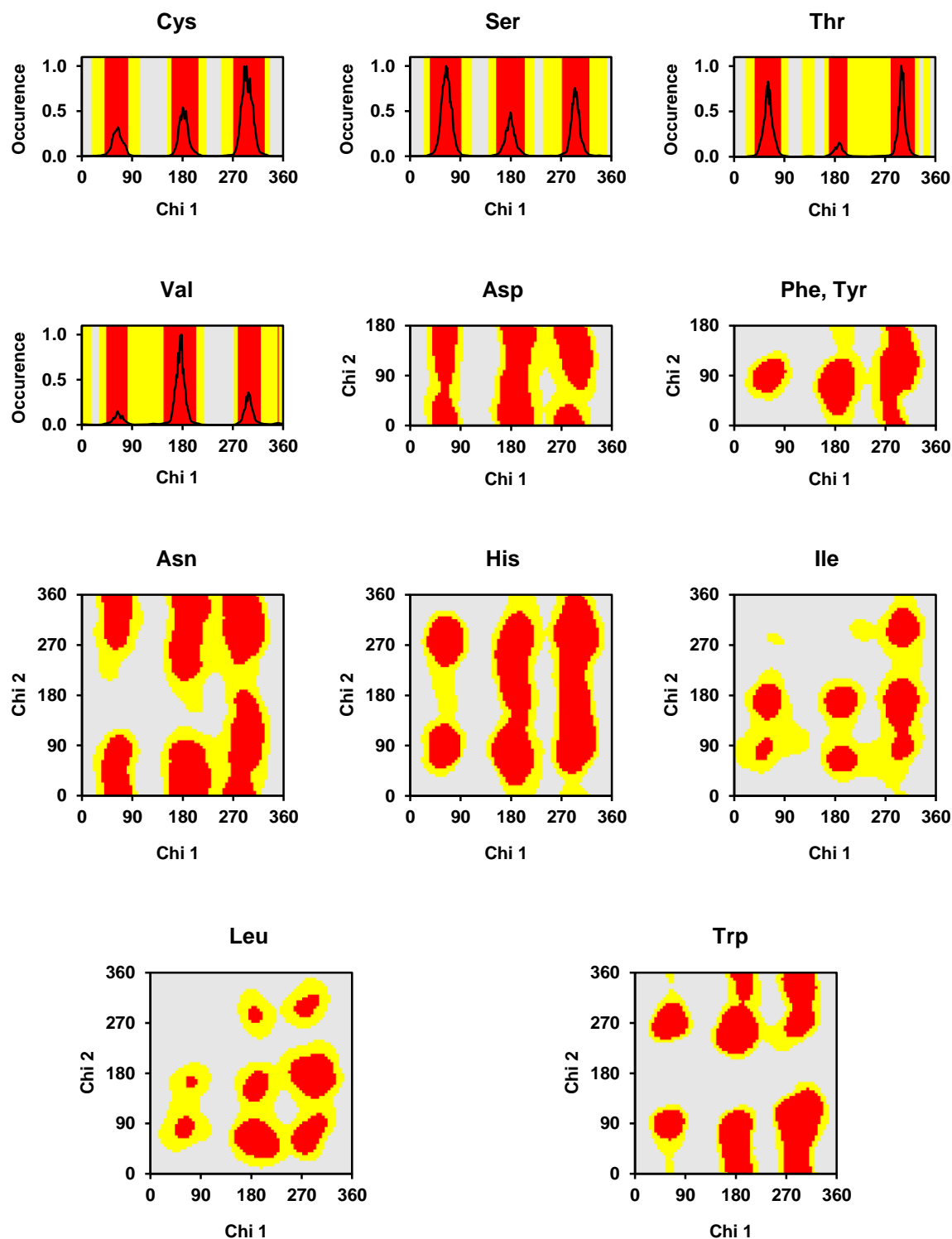
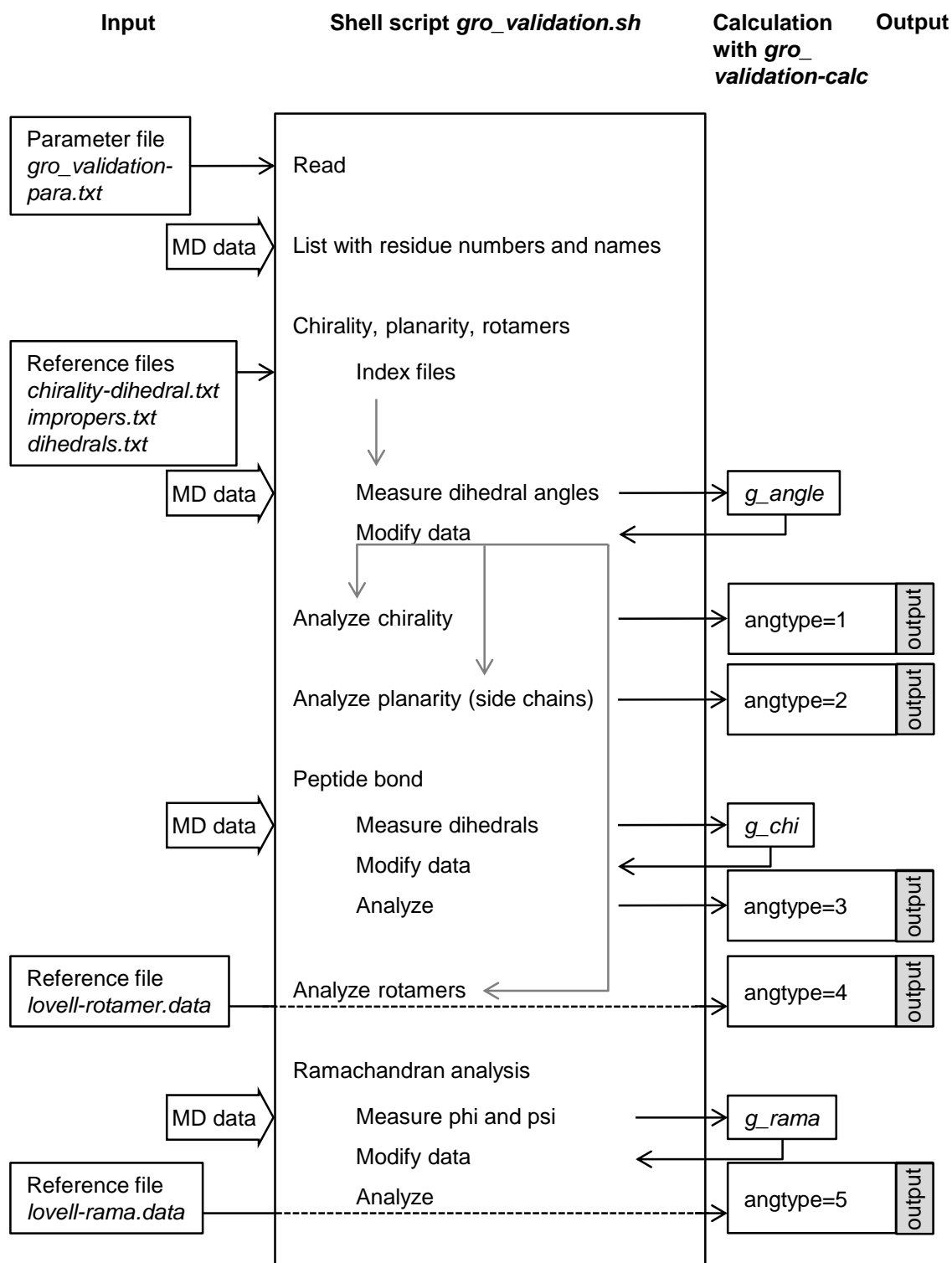


Figure 5.13: Side chain rotamer probability plots

The 'probability values' of side chain dihedrals obtained from Lovell *et al.* (2000) are shown for Cys, Ser, Thr and Val. The red, yellow and grey background indicates favored, allowed and disallowed regions of the conformational space, respectively, as defined in section 5.5.1.5. For amino acids with two dihedral angles in the side chain (Asn, Asp, His, Ile, Leu, Phe, Trp and Tyr) the color coded 'probability values' are shown.

Figure 5.14: Flowchart for *gro_validation*

5.5.2.1 Parameter file *gro_validation-para.txt*

The parameter file *gro_validation-para.txt* contains information necessary for the shell script to access the correct files for a specific MD simulation and allows the user to modify the amount of data analyzed. Table 5.8 lists all parameters needed with a description. An example parameter file is given in the Appendix (section 9.5.1).

Table 5.8: Content of the parameter file for the protein structure validation

Parameter name	Description
var-sim	Name of the simulation (all output files start with this term).
var-filename	Name of the <i>gro</i> , <i>tpr</i> and <i>xtc</i> files which must have the same name. Otherwise the respective lines in the shell script have to be changed.
var-duration	The duration of the simulation in picoseconds.
var-timestep	The time step of the analysis in picoseconds. E.g., a value of 10 means that the analysis is performed every 10 th ps.
var-first	Residue number of the first amino acid to analyze (residue number in the <i>gro</i> file).
var-last	Last amino acid to analyze.

5.5.2.2 Shell script *gro_validation.sh*

The shell script *gro_validation.sh* is the core of the program, receiving information from the parameter file and transferring computational intensive applications to subprograms of GROMACS or to the C program *gro_validation-calc* described below. The script is divided in several sections. Table 5.9 gives a summary of them and a short description of their functions. The complete script is listed in the Appendix (section 9.5.2). The program is started with the command *./gro_validation.sh*.

Table 5.9: Sections of the shell script *gro_validation.sh*

Section	Description
1	Variables are extracted from the parameter file, and new variables are calculated. A file is created containing amino acid names including sequence number, atom types and atom numbers. This file is needed for writing an appropriate index file which works with atom numbers. The atom numbers correspond to amino acid atom types (provided in the folder <i>reference</i>) defining the dihedral angles for the chirality analysis, the control of planarity and side chain rotamers.
2	The index file for the chirality check is constructed. The file containing the definition of the respective dihedral angles in terms of amino acids and atom types is located in the folder <i>reference</i> (<i>chirality-dihedral.txt</i>).
3	Generation of the index file for the planarity check. The file containing the definition of the respective dihedral angles is also located in the folder <i>reference</i> (<i>impropers.txt</i>).
4	Creation of the index file for calculating side chain rotamers. The corresponding reference file is stored in the folder <i>reference</i> (<i>dihedrals.txt</i>).
5	Dihedral angles are calculated using the GROMACS tool <i>g_angle</i> and the index files written in section 2 to 4.
6	Chirality check. With the command line argument 1, the corresponding subfunction of <i>gro_validation-calc</i> is started which controls the dihedral angle defining the chirality of C _α atoms. The respective values are extracted from the file created in section 5. The output file <i>chirality.txt</i> is written.
7	Checks aromatic side chains and polar groups with π-electron systems for planarity (subfunction 2 of <i>gro_validation-calc</i>). A part of the file produced in section 5 contains the dihedrals needed. One file (<i>planarity.txt</i>) is generated.
8	Checks backbone ω angles. The dihedrals are calculated with <i>g_chi</i> of GROMACS. The subprogram 4 of <i>gro_validation-calc</i> controls for planarity of ω angles. The output file <i>omega.txt</i> is created.
9	Compares the side chain rotamers with experimental values (subfunction 3 of <i>gro_validation-calc</i>). The file which is produced with <i>g_angle</i> in section 5 contains the side chain dihedrals. Three files (<i>rotamer-regions.txt</i> , <i>rotamer-time.txt</i> , <i>rotamer-graph.txt</i>) are written.
10	Backbone Φ/Ψ dihedrals (Ramachandran plot) are compared to experimentally determined values using the subfunction 5 of <i>gro_validation-calc</i> . The tool <i>g_rama</i> of GROMACS is used to calculate the Φ and Ψ angles. Four output files (<i>rama-regions.txt</i> , <i>rama-data.txt</i> , <i>rama-time.txt</i> , <i>rama-graph.txt</i>) are printed.

5.5.2.3 C-Program *gro_validation-calc*

Computational challenging tasks which could not be solved with subprograms of GROMACS are transferred to the C program *gro_validation-calc*. It is divided in several sections addressed with the variable *angtype* (Table 5.10). For a detailed description, refer to the

comments included in the source code presented in the Appendix (section 9.5.3). In lines 12 to 26 variables determine the range and increment of the frequency distributions of dihedral angles. In the variable *AAname* (line 46 ff) all amino acid names are stored, and *no_def_res* contains the number of residues (line 11). In order to add additional amino acids not present in the analysis, these lines have to be modified. So far all standard amino acids of GROMACS, present in the residue topology file *ffG53a6.rtp* are integrated, including ACE (acetyl moiety to block the N-terminus) and NH2 (for C-terminal ends of proteins).

Table 5.10: Sections of the C-program *gro_validation-calc*

Section (<i>angtype</i>)	Description
1	Controls the stereochemistry. In <i>AAcode</i> (line 113) it is stored whether a residue contains a chiral C _α atom. The sequence is equal to <i>AAname</i> (line 46 ff).
2	Check planarity in side chain elements. <i>AAcode</i> (line 294) contains the number of dihedrals per residue in the same sequence as <i>AAname</i> (line 46 ff).
3	Summarizes data of the peptide dihedral angle.
4	Compares side chain torsions with an experimentally determined dataset. In lines 550 and 551, the variables <i>cutoff_mf</i> (default value 0.02) and <i>cutoff_dis</i> (default value 0.002) separate the regions favored, allowed and disallowed (see section 5.5.2.4) of the conformational space. The variable <i>AAcode</i> (line 577) contains the number of dihedrals per side chain, the first line of the respective residue in the reference file <i>lovell-rotamer.data</i> , and the ranges and increments of the respective torsion angles (Table 5.11).
5	Compares Φ/Ψ pairs with experimental values. In line 813 and 814 the variables <i>cutoff_mf</i> (default value 0.02) and <i>cutoff_dis</i> (default value 0.002) separate the regions favored, allowed and disallowed.

5.5.2.4 Reference data

Dihedral angle definition

The definition of dihedral angles needed for checking the stereochemistry at C_α atoms, planarity of side chain elements and the calculation of side chain rotamers are deposited in the files *chirality-dihedral.txt*, *impropers.txt* and *dihedrals.txt*, respectively, all stored in the folder *reference*. Each line contains a residue and atom name. Starting at the beginning of each file, four consecutive lines define one dihedral angle. An example for the side chain torsions of arginine is given in Figure 5.15.

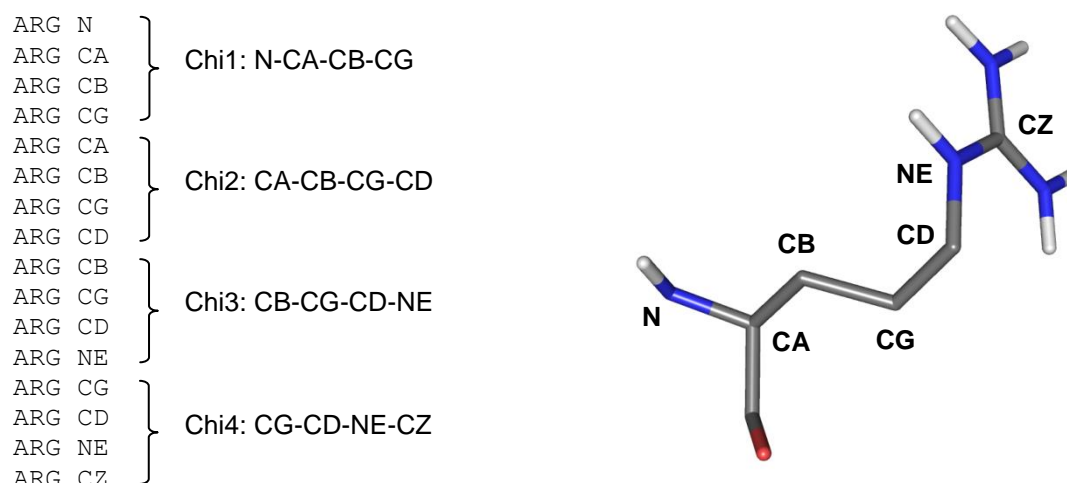


Figure 5.15: Definition of the four side chain dihedral angles of arginine in *dihedrals.txt*

The definition of the side chain dihedral angles of arginine in *dihedrals.txt* is shown on the left. Four consecutive lines, each containing the residue name and an atom name, define one dihedral angle. The dihedral angle definition is equivalent in *chirality-dihedral.txt* and *impropers.txt*.

Preparation of the library with Φ and Ψ dihedral angles

The experimental reference data specifying the frequency of a special Φ/Ψ combination were taken from the website of the Richardson Laboratory of the Duke University (Durham, UK). A zip folder of the ‘(son of) Penultimate Rotamer Library’ is provided there, containing the Ramachandran density traces in 2° increments ranging from $-179^\circ/-179^\circ$ to $179^\circ/179^\circ$ (Lovell *et al.*, 2003; Lovell *et al.*, 2000). Φ and Ψ angles as well as the corresponding frequency value are listed in columns one to three. The files *rama500-general.data*, *rama500-gly-sym.data*, *rama500-pro.data* and *rama500-prepro.data* of the folder *pct/rama* were combined to one file in this sequence (*reference/lovell-rama.data*), resulting in 129,600 data points.

Preparation of the rotamer library

Accordingly, data with most probable side chain conformations were received from the folder *pct/rota*. Respective files were combined to one file (*reference/lovell-rotamer.data*) in the sequence given in Table 5.11. The step size and the range of the dihedral angles depend on the individual residue. If present, dihedral angles χ_1 to χ_4 are listed in column one to four. The frequency value is in the last column. All in all, 3,620,601 data points are summarized in that file.

Table 5.11: Side chain rotames

Residue	No ^a	Angle range from 0° to ^b				Step size [°] ^b			
		Chi1	Chi2	Chi3	Chi4	Chi1	Chi2	Chi3	Chi4
Arg	4	360	360	360	360	10	10	10	10
Asn	2	360	360			5	5		
Asp	2	360	180			5	5		
Cys	1	360				1			
Gln	3	360	360	360		8	8	8	
Glu	3	360	360	180		8	8	7.8 ^c	
His	2	360	360			5	5		
Ile	2	360	360			5	5		
Leu	2	360	360			5	5		
Lys	4	360	360	360	360	10	10	10	10
Met	3	360	360	360		8	8	8	
Phe, Tyr	2	360	180			5	5		
Ser	1	360				1			
Thr	1	360				1			
Trp	2	360	360			5	5		
Val	1	360				1			

^a Number of side chain dihedrals of the respective amino acid.

^b The angle range and step size was determined by the reference data obtained from the Richardson Laboratory (Duke University, Durham, UK).

^c 23 values ranging from 0° to 180° were provided, corresponding to an increment of ~ 7.8°.

5.5.3 Output files

5.5.3.1 Chirality check

One file is saved in the result folder (*output*). In each line the information for one amino acid is stored, namely the residue number, residue name, the mean angle, the standard deviation and a number (zero or one) expressing if all dihedral angles measured were within the limits ($\pm 35^\circ$) around the reference angle of 35° or not, respectively. Additionally the values for a frequency distribution are given, covering values of 10° to 60° of the dihedral angle at the C_α atom. An example is given in Figure 5.16.

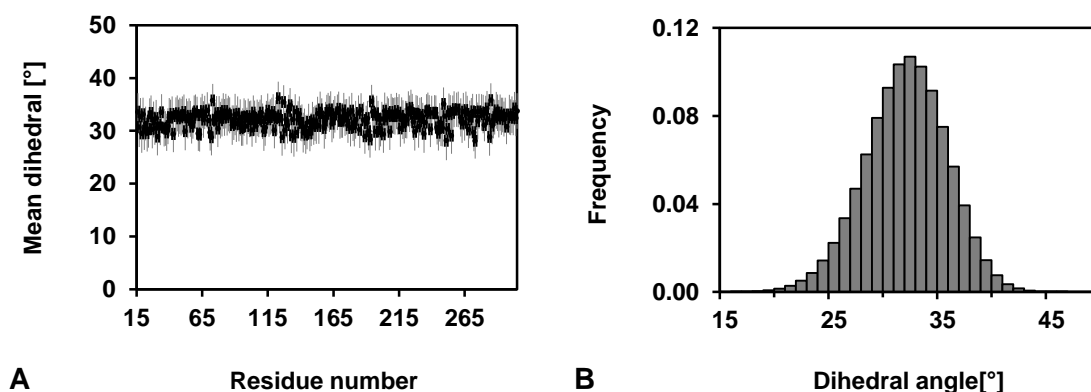


Figure 5.16: Distribution of the mean dihedral angle at C_{α} atoms and probability distribution of one dihedral angle

A, Mean dihedral angle (black points) of each residue of the MD simulation of the hH_2R_i (chapter 4). The standard deviation is shown in grey bars. B, Probability distribution of the dihedral angle defining the chirality at the C_{α} atom of Asn36^{1.50} during the 80 ns MD simulation of hH_2R_i (chapter 4). Every 10th frame was analyzed (8,000 in total).

5.5.3.2 Planarity check

The file *planarity.txt* contains all measured dihedral angles with the amino acid number and name, the mean, standard deviation and the fraction of values within the interval of two standard deviations around the mean. The frequency distribution of the angles ranging from -30° to 30° (a dihedral angle of 0° represents ideal planarity) with an increment of 2° can be used to create a probability distribution. Furthermore, means and histogram data of dihedral angles defining the planarity of one special residue are summarized (Figure 5.17).

5.5.3.3 Peptide bond analysis

In the output file (*omega.txt*) for each amino acid, the residue number, the mean ω angle, the standard deviation and the percentage of values within the interval of two standard deviations around the mean are given. Furthermore a frequency distribution ranging from 100° to 260° is provided (Figure 5.18). Ideal planarity is achieved at a dihedral angle of 180° .

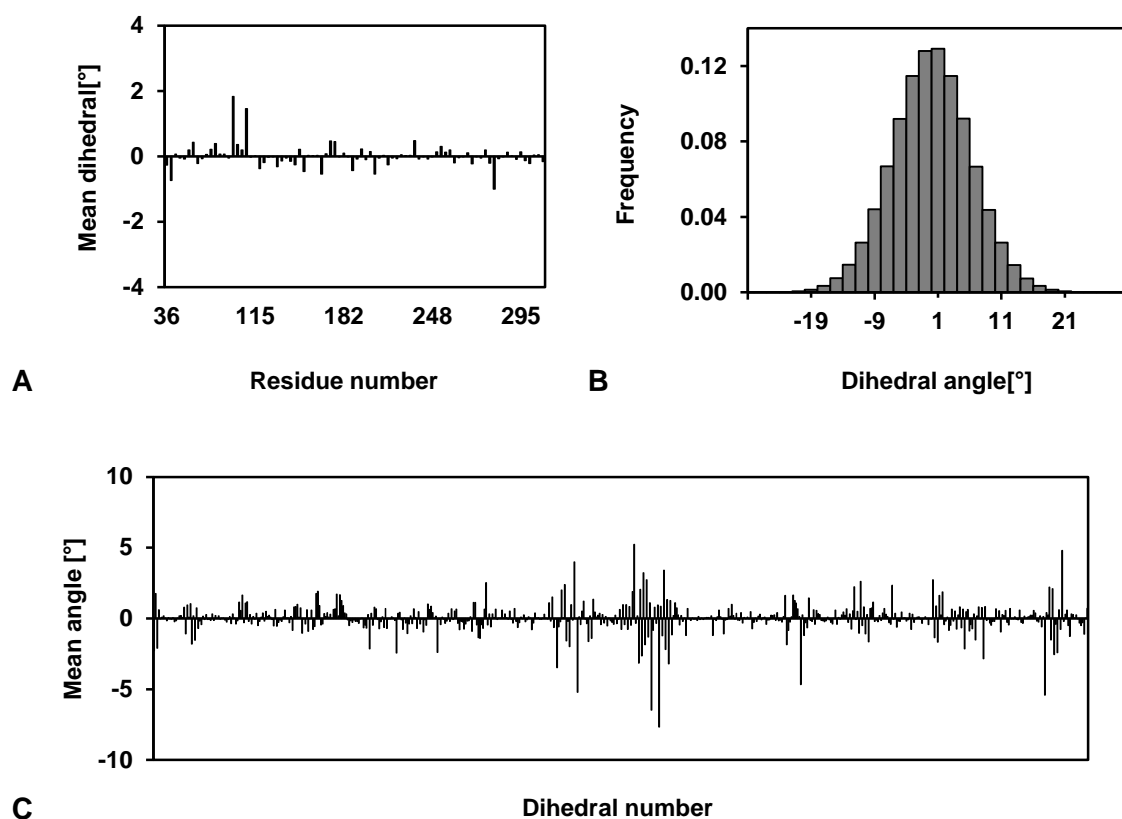


Figure 5.17: Planarity control

A, Mean dihedral angle of each residue containing an aromatic or normally planar polar side chain moiety, including all dihedral angles used to define planarity for the respective residue. B, Distribution of the two dihedral angles defining the planarity of Asn36^{1.50}. C, Mean of each single dihedral angle measured. The data are taken from the simulation of hH₂R_i described in chapter 4, analyzing every 10th frame (8,000 in total).

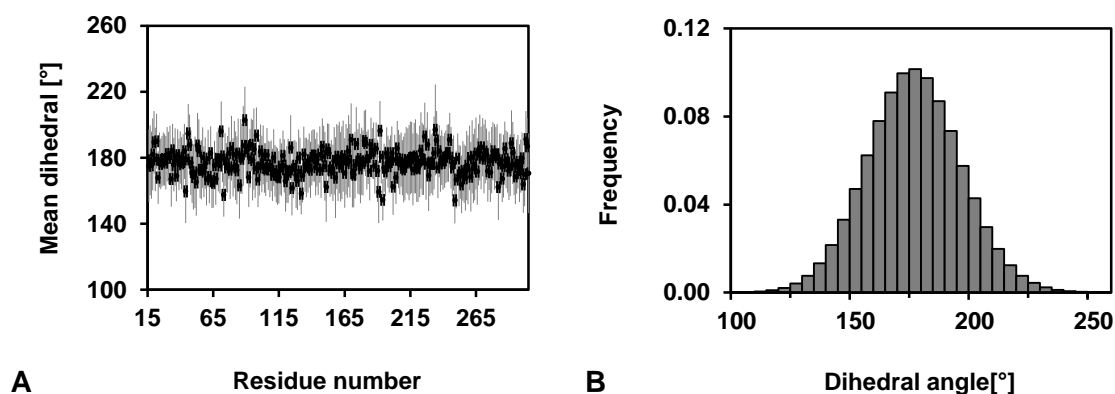


Figure 5.18: Analysis of the backbone omega angles

A, Mean omega angle (black point) of each residue and its standard deviation (grey bar). B, The frequency distribution is shown for Ile31^{1.45}. The data are taken from the simulation of the hH₂R_i described in chapter 4, analyzing every 10th frame (8,000 in total).

5.5.3.4 Ramachandran analysis

Four files are created. For a description see Table 5.12.

Table 5.12: Output files of the Ramachandran analysis

File name	Description
<i>rama-regions.txt</i>	For each amino acid the fraction of favored, allowed and disallowed Φ/Ψ combinations are given. Figure 5.19 was generated from these data.
<i>rama-data.txt</i>	The file contains all single Φ/Ψ values calculated for each amino acid. Figure 5.20 shows four examples of its usage.
<i>rama-time.txt</i>	Each line starts with the amino acid number and contains a character for every time point analyzed (f, favored; a, allowed; d, disallowed). With the script <i>rama-time.sh</i> presented in the Appendix (section 9.6) a time-resolved analysis can be performed (Figure 5.21).
<i>rama-graph.txt</i>	For each residue a 180 x 180 matrix is deposited in this file, containing Φ dihedrals in columns and Ψ values in lines. The Φ/Ψ pairs are categorized in 2° increments, ranging from -180° to 180°. The values represent the percentage of dihedrals being in a special interval (in total 100%). The data are used to create a surface plot (Figure 5.22).

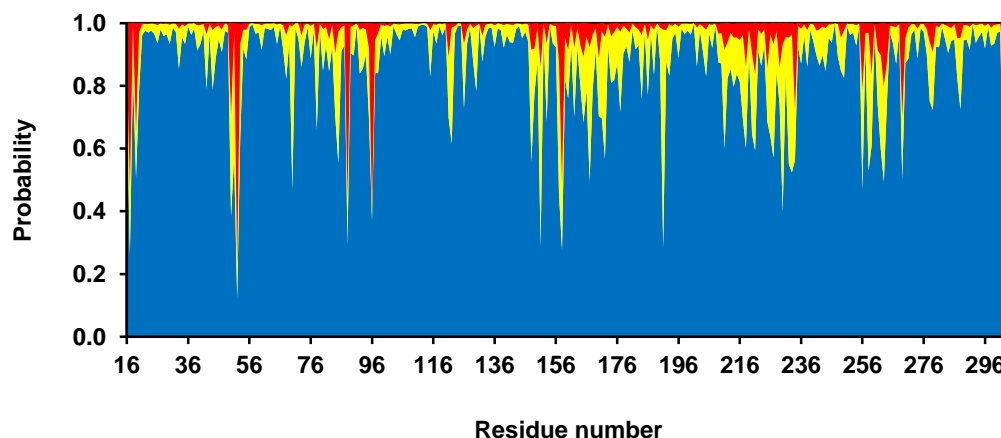


Figure 5.19: Ramachandran analysis

For each residue the fraction of Φ/Ψ dihedral combinations located in favored (blue), allowed (yellow) and disallowed (red) regions of the ramachandran plot is indicated. The data were taken from the MD simulation of the hH_2R_i described in chapter 4. Every 10th frame was analyzed.

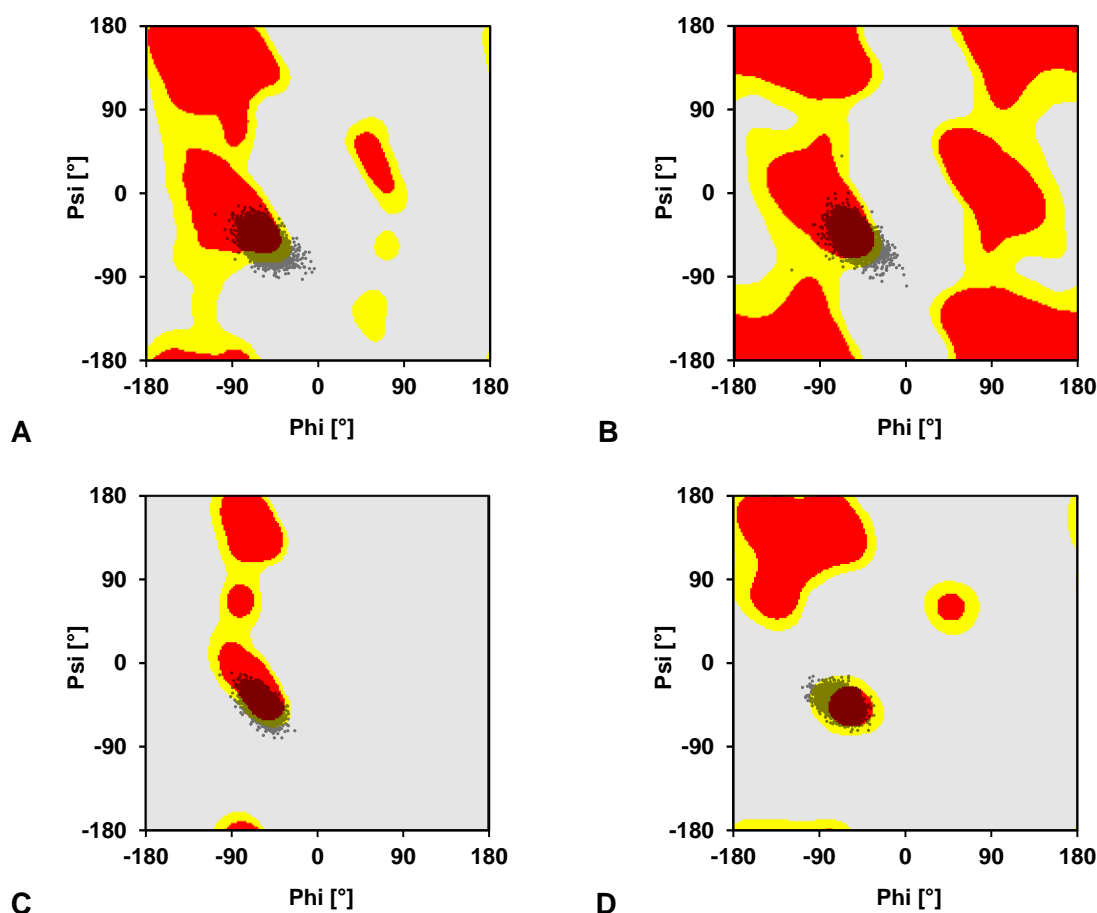


Figure 5.20: Ramachandran plots

Ramachandran plots (grey points) for Asp98^{3.32} (A), Gly183^{5.39} (B), Pro194^{5.50} (C) and Leu193^{5.49} (D) from the MD simulation of the hH₂R_i (chapter 4), analyzing 8,000 time points (time step 10 ps). In the background the Ramachandran plots for 'general' amino acids (A; all amino acids except glycine, proline and amino acids preceding proline), glycine (B), proline (C) and pre-proline (D; amino acids preceding proline) created from the reference data of Lovell *et al.* (2000) and defined in section 5.5.1.4 are shown.

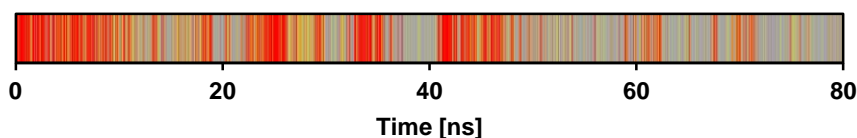


Figure 5.21: Time resolution of a Φ/Ψ dihedral combination

The graph shows the occurrence of favored, allowed and disallowed Φ/Ψ dihedrals of Lys88^{3.22} from the MD simulation of the hH₂R_i described in chapter 4, indicated by a red, yellow and grey line, respectively. In 56% of the simulation the Φ/Ψ combination adopted a disallowed value.

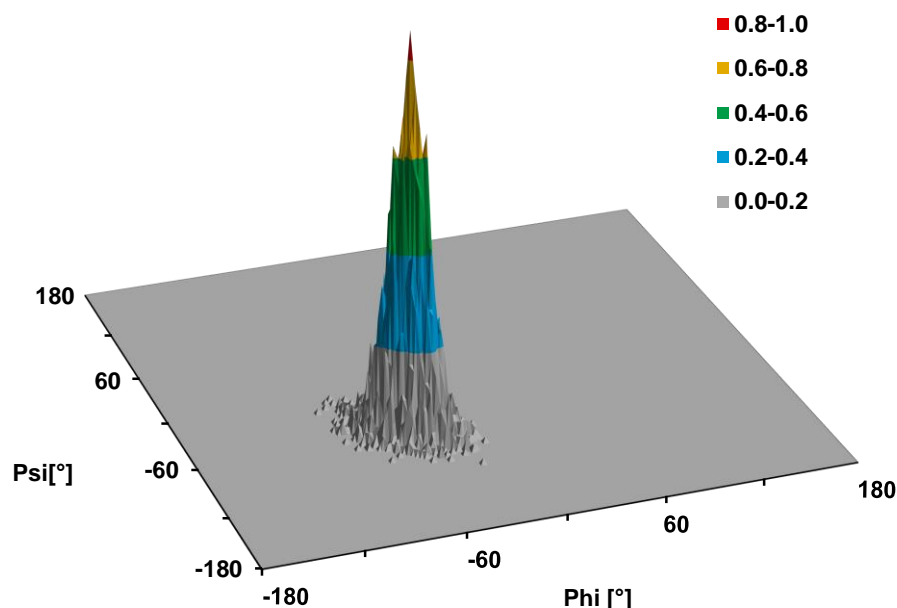


Figure 5.22: 3D probability Ramachandran graph

The distribution is shown for Asp98^{3,32} from the MD simulation of the hH₂R_i described in chapter 4. 8,000 frames were analyzed.

5.5.3.5 Side chain rotamers

For the analysis of the side chain rotamers three files are produced. Table 5.13 gives an overview with a short description of the contents. Figure 5.23 to Figure 5.25 show examples of the usage.

Table 5.13: Output files for rotamer analysis

File name	Description
<i>rotamer-regions.txt</i>	For each amino acid the fraction of favored, allowed and disallowed side chain dihedral combinations are given. Figure 5.23 was generated from these data.
<i>rama-time.txt</i>	Each line starts with the amino acid number and contains a character for every time point analyzed (f, favored; a, allowed; d, disallowed). With the script <i>rama-time.sh</i> presented in the Appendix (section 9.6) a time-resolved analysis can be performed (Figure 5.24).
<i>rama-graph.txt</i>	For residues with one side chain dihedral angle (Cys, Ser, Thr, Val) the frequency distribution is given in one line. For amino acids with two or more dihedral angles a 2-dimensional matrix is deposited, containing the χ_1 and χ_2 dihedrals in lines and columns, respectively. The increment depends on the reference data used (see section 5.5.2.4). The values represent the percentage of dihedral angles within a special interval (in total 100%). The data can be used to create a surface plot (Figure 5.25).

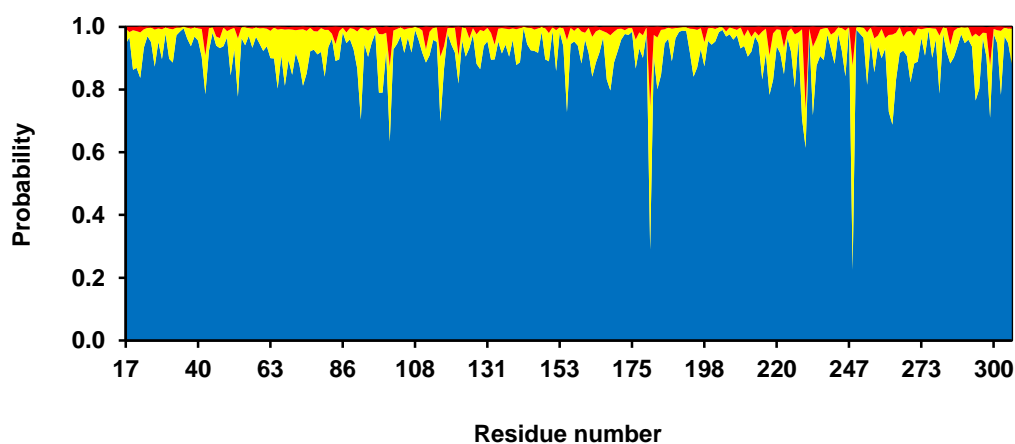


Figure 5.23: Side chain rotamer analysis

For each proteinogenic amino acid except Ala, Gly and Pro the fraction of side chain dihedral angles located in favored (blue), allowed (yellow) and disallowed (red) regions as defined in section 5.5.2.4 is indicated. The data were taken from the MD simulation of the hH_2R_i described in chapter 4. Every 10th frame was analyzed (8,000 in total).

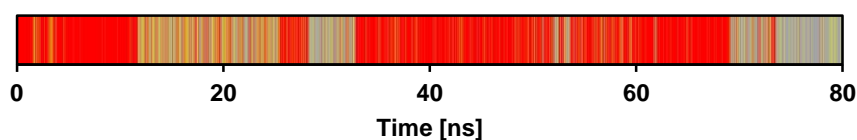


Figure 5.24: Time resolution of a side chain dihedral angle

The graph shows the occurrence of favored, allowed and disallowed side chain conformations of His230^{6,31} from the simulation of the hH_2R_i described in chapter 4, indicated by a red, yellow and grey line, respectively. In 26.2% of the 8,000 time points analyzed, the conformation adopted a disallowed combination of χ_1 and χ_2 dihedrals.

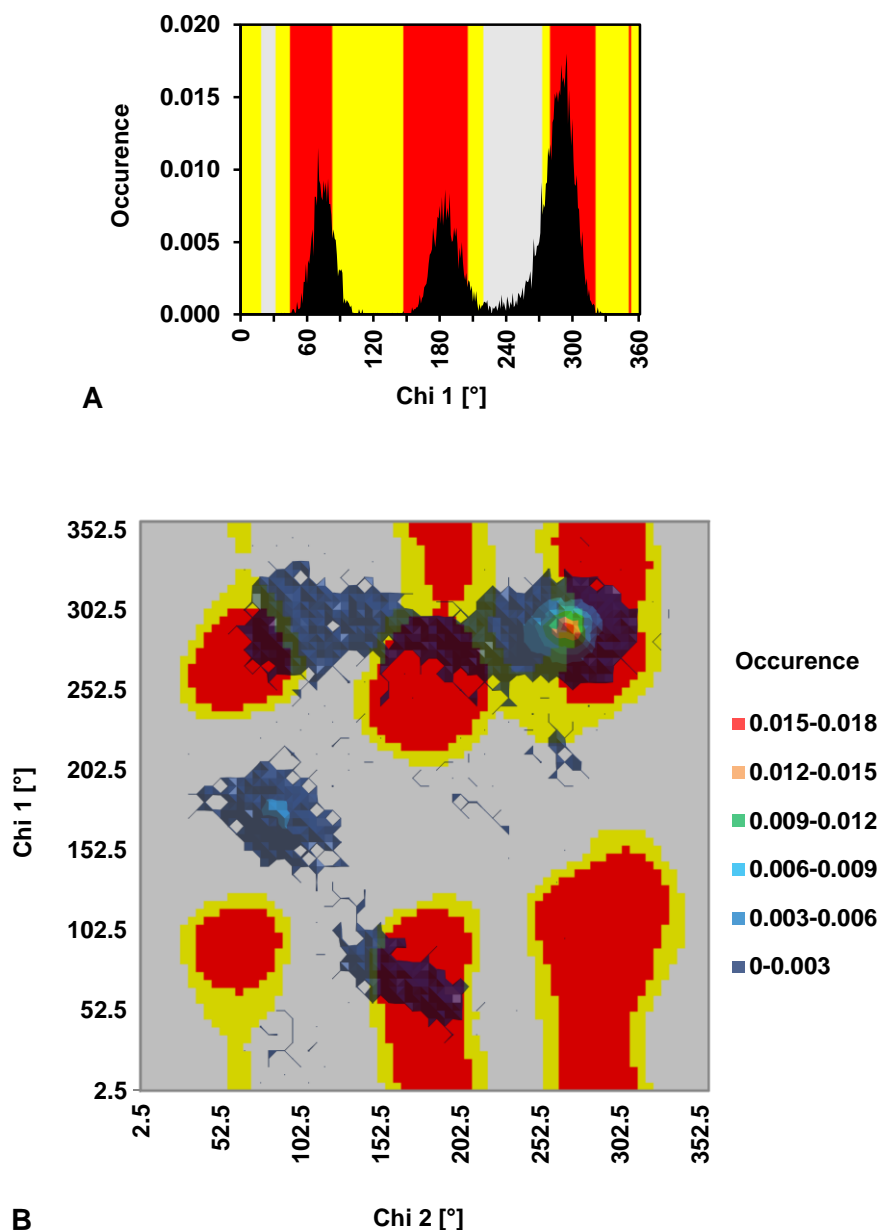


Figure 5.25: Frequency distribution of side chain rotamers and comparison to reference values

A, Distribution of the dihedral angle of Val43^{1.57} from the MD simulation of the hH₂R_i (chapter 4). In the background the favored, allowed and disallowed regions are shown in red, yellow and grey, respectively. 78.5% of the dihedrals were in favored, 12.4% in allowed and 9.1% in disallowed regions. B, Occurrence of the dihedral angles χ_1 and χ_2 of Trp84^{ECL1} from the MD simulation of the hH₂R_i (chapter 4). In the background the favored, allowed and disallowed regions are shown in red, yellow and grey, respectively. 73.7% of the dihedrals were in favored, 9.9% in allowed and 16.5% in disallowed regions. A, B, The analysis was performed every 10th time point (8,000 frames in total).

5.5.4 Performance

The MD simulation of the hH₂R_i embedded in a DPPC bilayer and solubilized with water molecules (chapter 4) was used for testing this structure validation tool. The simulation system contained about 40,000 atoms and ran for 80 ns. Every 10th coordinate file saved to hard disc was analyzed, corresponding to every 10th picosecond or 8,000 frames in total, respectively. All in all, more than 15.4 million dihedral angles were measured and compared to more than 3.7 million reference values. The calculation time was 32 minutes. When analyzing all frames (80,000), more than 154.4 million dihedral angles were measured and compared to the reference data, lasting 1:41 hours.

5.6 Summary and conclusion

Three programs were written in order to systematically analyze direct and water mediated H-bonds (*gro_hbonds*) and van der Waals contacts (*gro_contacts*) occurring in MD simulation systems and in order to verify the quality of a MD simulation by calculating stereochemical parameters of proteins and compare them to experimentally derived reference values (*gro_validation*).

GROMACS provides opportunities to calculate H-bonds and contacts between residues, but several disadvantages are associated with these tools (*g_hbond* and *g_mdmat*, respectively). The program *g_hbond* is rather suitable to analyze single H-bonds. A systematic analysis is not possible without additional tools. Furthermore, *g_hbond* is much more time-consuming compared to the C program *gro_hbonds-calc*. The crucial accelerating element in the program is the introduction of a code for each amino acid (Figure 5.2). The analysis of water mediated H-bonds is not possible with *g_hbond* without additional scripts. A further advantage of *gro_hbonds* is the user-friendly handling. In general, only the parameter file *gro_hbonds-para.txt* (containing eight variables) has to be edited and the complete analysis of all possible H-bonds - direct and water mediated - is performed in less than four hours (in case of a 80 ns MD simulation described in chapter 4). The format of the output files allows a comfortable extraction of individual interactions as well as a transformation to spreadsheet applications (tabulator-separated values). Finally, the calculation method of H-bonds in *gro_hbonds* was adjusted to different molecular modeling packages (cf. Figure 5.1).

The analysis of residue contacts with *g_mdmat* works quite well in GROMACS. However, depending on the number of contacts to be investigated, the creation of an appropriate index file which defines the residues and atoms to be considered in the calculation may be time-consuming for the user. Furthermore, the output of *g_mdmat* is inappropriate for easy

analysis (Figure 5.7). Thus, *gro_contacts* provides the opportunity to analyze all possible contacts in a MD simulation system (in about 30 min. for the 80 ns simulation of hH₂R_i described in chapter 4). In general, only the parameter file *gro_contacts-para.txt* (six variables) has to be adjusted. The output can be scanned for individual interactions and transferred to spreadsheet applications. Since both programs (*gro_hbonds* and *gro_contacts*) calculate all possible interactions, a MD simulation can be completely analyzed without losing any contacts.

The analysis of homology models with programs like PROCHECK (Laskowski *et al.*, 1993) is a standard procedure to probe their stereochemical quality. To verify the quality of MD simulations, a program is necessary which may handle thousands of coordinate sets and compare the measured stereochemical parameters with experimental values. The calculation of dihedral angles within *gro_validation* is performed by tools of GROMACS (*g_angle*, *g_chi* and *g_rama*). However, appropriate index files created in *gro_validation.sh* from reference files (*chirality-dihedral.txt*, *impropers.txt* and *dihedrals.txt*) are necessary to identify the proper dihedral angles. The C program *gro_validation-calc* performs the calculation of these angles and their comparison with reference values. The rapid assignment of the measured backbone dihedrals Φ and Ψ and of the side chain rotamers to the corresponding reference values in the files *lovell-rama.data* and *lovell-rotamer.data*, respectively, is enabled by direct conversion of the value of a dihedral angle into the proper line in the reference files. The complete structure validation of a 80 ns MD simulation as described in chapter 4 is performed in about 30 minutes. Again, in general only a parameter file (*gro_validation-para.txt* with six variables) has to be modified by the user. The output of the program is provided in a user-friendly format and can be transferred into spreadsheet applications to analyze and visualize the data.

5.7 References

- Arunan E, Desiraju GR, Klein RA, Sadlej J, Scheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, Kjaergaard HG, Legon AC, *et al.* Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl Chem* **2011**, 83, 1637-1641.
- Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, 30, 1545-1614.
- CBN. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. *Eur J Biochem* **1970**, 17, 193-201.
- Hess B, Kutzner C, van der Spoel D and Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, 4, 435-447.
- Humphrey W, Dalke A and Schulten K. VMD: visual molecular dynamics. *J Mol Graph* **1996**, 14, 33-38.

- Kabsch W and Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577-2637.
- Kellis JT, Nyberg K, Sail D and Fersht AR. Contribution of hydrophobic interactions to protein stability. *Nature* **1988**, 333, 784-786.
- Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **1993**, 26, 283-291.
- Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS and Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003**, 50, 437-450.
- Lovell SC, Word JM, Richardson JS and Richardson DC. The penultimate rotamer library. *Proteins* **2000**, 40, 389-408.
- Matthews BW. Hydrophobic Interactions in Proteins. In *eLS*, John Wiley & Sons, Ltd: **2001**.
- Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, 12, 345-364.
- Nygaard R, Valentin-Hansen L, Mokrosinski J, Frimurer TM and Schwartz TW. Conserved Water-mediated Hydrogen Bond Network between TM-I, -II, -VI, and -VII in 7TM Receptor Activation. *J Biol Chem* **2010**, 285, 19625-19636.
- Oostenbrink C, Villa A, Mark AE and Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **2004**, 25, 1656-1676.
- Panigrahi SK and Desiraju GR. Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins-Structure Function and Bioinformatics* **2007**, 67, 128-141.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, 25, 1605-1612.
- Salomon-Ferrer R, Case DA and Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, 3, 198-210.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, 26, 1701-1718.
- Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **1990**, 8, 52-56.

Chapter 6

Point Mutation in the Orthosteric Binding Site of the Human Histamine H₂ Receptor: the Role of Tyr182 in TM5

6.1 Introduction

In the early 1990s the first site-directed mutagenesis studies were performed at the histamine H₂ receptor to identify amino acids which contribute to ligand binding (Gantz *et al.*, 1992). Accordingly, Asp98^{3.32} in TM3 is necessary for H₂R binding and agonism of histamine. This highly conserved amino acid was also found to be important in several other aminergic GPCRs, e.g. the β_2 -adrenergic receptor, where mutation of the respective amino acid resulted in lower affinities of agonists and antagonists (Strader *et al.*, 1988), as well as in the dopamine D₂ receptor (Mansour *et al.*, 1992) and the muscarinic acetylcholine receptor M₂ (Schwarz *et al.*, 1995). Replacing Asp186^{5.42} in TM5 of the H₂R caused the loss of binding of radiolabeled tiotidine, but retained the cAMP generation upon stimulation with histamine. However, the cAMP content was reduced compared to the wild-type receptor, and the histamine-stimulated cAMP production was only partially inhibited by cimetidine. The authors concluded that Asp186^{5.42} in TM5 must be present as a hydrogen bond acceptor for the interaction with the imidazole moiety of histamine in order to produce full response to ligand binding. In other GPCRs this residue was shown to play a role in ligand binding and receptor activation as well, e.g. in α -adrenergic and dopaminergic receptors (Liapakis *et al.*, 2000). The exchange of Thr190^{5.46} in the H₂R by alanine or cysteine resulted in a receptor mutant still able to bind radiolabeled tiotidine, but with an altered cAMP response to histamine stimulation (Gantz *et al.*, 1992). Based on these results, a three site interaction model for histamine at the H₂ receptor was proposed, consisting of Asp98^{3.32} in TM3 and the couple Asp186^{5.42}/Thr190^{5.46} in TM5, interacting with the protonated amine function and the heterocycle, respectively (Figure 6.1 A). By contrast, Nederkoorn *et al.* (1996a; 1996b)

suggested the couple Tyr182^{5.38}/Asp186^{5.42} in TM5 to interact with the heterocycle of H₂R ligands (Figure 6.1 B). This conclusion was drawn from molecular mechanics and ab initio data at an oligopeptide, mimicking a part of the fifth transmembrane alpha helix, and the fact that the mutation of Thr190^{5.46} does not completely abolish the cAMP response to agonist stimulation. In the meantime, either Tyr182^{5.38} or Thr190^{5.46} was preferred for the docking of H₂R ligands (Kelley *et al.*, 2001; Sun *et al.*, 2011).

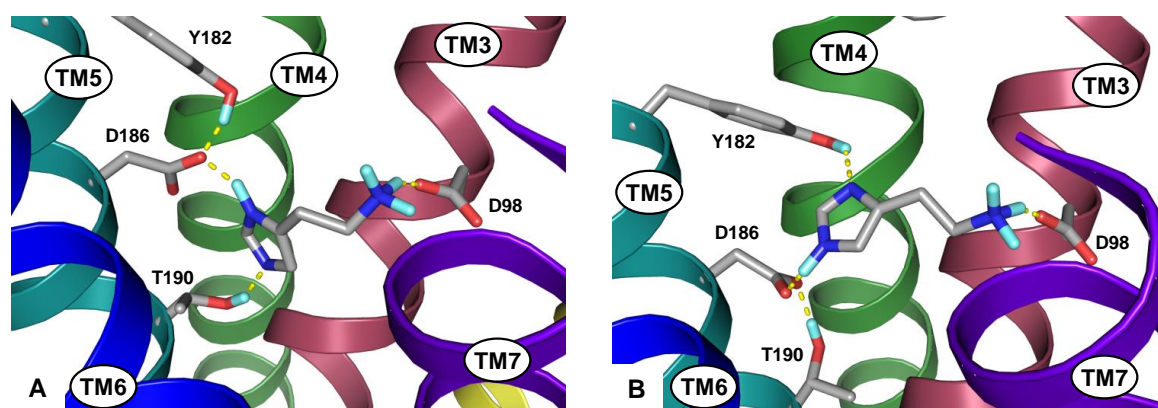


Figure 6.1: Docking modes of histamine in the orthosteric binding pocket of an hH₂R homology model

A, Histamine in the N^TH tautomeric form interacting with the couple Asp186^{5.42}/Thr190^{5.46}. B, Histamine in the N^H tautomeric form interacting with Asp186^{5.42}/Tyr182^{5.38}. Hydrogen bonds are indicated with yellow dashed lines. ECL3 is omitted for reasons of clarity. The homology model is based on the crystal structure of the nanobody-stabilized active state of the β_2 -adrenoceptor (PDB ID 3P0G; Rasmussen *et al.*, 2011) described in chapter 3.

Since the relevance of Tyr182^{5.38} was solely investigated with theoretical methods, the objective of this study was to analyze the role of this residue with pharmacological methods by constructing a Tyr182→Phe182 hH₂R mutant. To explore the impact of Tyr182^{5.38} on ligands varying in their heterocyclic moiety and the distance between the side chain amine/guanidine moiety and the heterocycle, structurally diverse ligands were investigated in the GTPase activity assay and the [³⁵S]GTPγS binding assay (Figure 6.2).

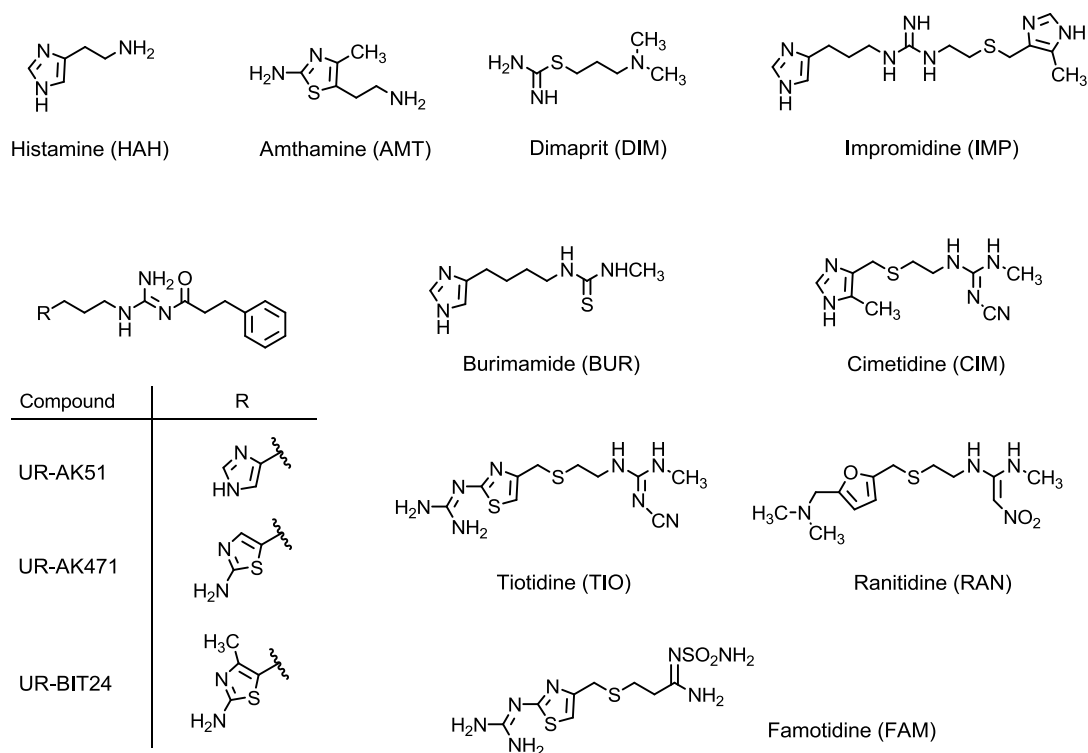


Figure 6.2: Structures of H₂R agonists and antagonists

6.2 Materials and methods

6.2.1 Materials

The generation of pGEM-3Z-SF-hH₂R-His₆-G_{saS} was described previously (Kelley *et al.*, 2001). The DNA primers for polymerase chain reaction (PCR) were synthesized by Eurofins MWG Operon (Ebersberg, Germany). Pfu Ultra II Fusion HS DNA Polymerase was obtained from Agilent Technologies (Böblingen, Germany). Restriction enzymes and T4 DNA ligase were from New England Biolabs (Ipswich, MA, USA). The monoclonal anti-FLAG M1 antibody (F3040) and the anti-mouse IgG-peroxidase antibody (A0168) were purchased from Sigma-Aldrich (Taufkirchen, Germany). [γ -³³P]GTP was synthesized using guanosine 5'-diphosphate and [γ -³³P]P_i (3,000 Ci/mmol orthophosphoric acid) by analogy to the synthesis of [γ -³²P]GTP described before (Walseth and Johnson, 1979). [γ -³³P]P_i and [³⁵S]GTP γ S (\geq 1000 Ci/mmol, radiochemical purity > 95%) were from Hartmann Analytic (Braunschweig, Germany). GF/C filters were from Perkin Elmer (Rodgau, Germany). Histamine, amthamine, dimaprit and tiotidine were purchased from Tocris (Avonmouth, Bristol, UK). Cimetidine, famotidine and ranitidine were from Sigma-Aldrich. Impromidine (Durant *et al.*, 1978) was

synthesized as described (Durant *et al.*, 1985). Burimamide was from W. Schunack (Free University of Berlin, Germany). UR-AK51, UR-AK471 and UR-BIT24 were synthesized as described (Kraus *et al.*, 2009). Stock solutions were prepared with Millipore water. All other reagents were from standard suppliers and of the highest purity available.

6.2.2 Construction of the cDNA encoding the hH₂R-Y182F-G_{sdS} fusion protein

To generate the cDNA for hH₂R-Y182F-G_{sdS}, pGEM-3Z-SF-hH₂R-His₆-G_{sdS} was used as template. As a point mutation of only one amino acid was intended, a method deviated from the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene Cloning Systems, La Jolla, CA, USA) was used. Two complementary primers, 5'-CCCATCCACCAGCCC-GAACACTTCATTGACCTG and 5'-CAGGTCAATGAAGTGTTCGGGCTGGTGGATGGG (mismatching base pairs underlined), both containing a single mismatch for the Tyr182→Phe182 exchange, annealed with the template DNA in a PCR. The product was treated with *Dpn I* endonuclease (target sequence 5'-Gm6ATC), which is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template. This originates from *E. coli* and is therefore susceptible to *Dpn I* digestion. Transformation into competent *E. coli Top10* cells, single clone selection and plasmid preparation led to the desired hH₂R-Y182F-G_{sdS} sequence in the pGEM vector. After double-digestion with *Sac I* and *Xba I* and cloning into the pVL1392-SF-gpH₂R-His₆-G_{sdS} plasmid digested with the same enzymes, the generated DNA sequence of pVL1392-SF-hH₂R-Y182F-His₆-G_{sdS} was confirmed by agarose gel electrophoresis and sequencing (Entelechon, Regensburg, Germany).

6.2.3 Cell culture, generation of recombinant baculoviruses and membrane preparation

As described recently (Schnell *et al.*, 2010), Sf9 insect cells were cultured in 250 ml disposable Erlenmeyer flasks at 28 °C under rotation at 150 rpm in Insect-Xpress medium (Lonza, Walkersville, MD, USA), supplemented with 5% (v/v) fetal calf serum (Biochrom, Berlin, Germany) and 0.1 mg/ml gentamicin (Cambrex Bio Science, Walkersville, MD, USA). Cells were maintained at a density of 0.5 to 4.0 x 10⁶ cells/ml. Baculoviruses encoding recombinant proteins were generated in Sf9 cells using the BaculoGOLD transfection kit (BD Biosciences, San Diego, CA, USA) according to the manufacturer's instructions. After initial transfection with the pVL1392 plasmid containing the DNA sequence of the hH₂R-Y182F-G_{sdS} mutant and the linearized baculovirus DNA, high-titer virus stocks were generated by two sequential virus amplifications. In the first step, cells were seeded at 2.0 x 10⁶ cells/ml

and infected with a 1:50 dilution of the supernatant from the initial transfection. Cells were cultured for 7 days, resulting in the death of virtually the entire cell population. In the second amplification, cells were seeded at 3.0×10^6 cells/ml and infected with a 1:20 dilution of the supernatant fluid from the first amplification. After 48 h the majority of cells showed signs of infections (e.g. altered morphology, viral inclusion bodies), but most of the cells were still intact. The supernatant fluid was harvested and stored under light protection at 4 °C. It was used as routine virus stock for membrane preparations. Before transfection, Sf9 cells were centrifuged, resuspended in fresh medium and seeded at 3.0×10^6 cells/ml. After infection with a 1:100 dilution of the high-titer baculovirus stocks, cells were cultured for 48 h and Sf9 membranes were then prepared as described previously (Seifert *et al.*, 1998), using 1 mM EDTA, 0.2 mM phenylmethylsulfonyl fluoride, 10 µg/ml benzamidine and 10 µg/ml leupeptin as protease inhibitors. Membranes were suspended in binding buffer (12.5 mM MgCl₂, 1 mM EDTA and 75 mM Tris-HCl, pH 7.4) and stored at -80 °C until use.

6.2.4 Immunoblot analysis

Membrane proteins were diluted in Laemmli buffer and separated on SDS polyacrylamide gels containing 12% (w/v) acrylamide. Proteins were then transferred onto 0.2 µm Nitrocellulose Blotting-Membrane (Pierce, Erlangen, Germany). Membranes were reacted with the M1 antibody (10 µg/ml) and immunoreactive bands were visualized with the Pierce ECL Western Blotting Substrate (Thermo Scientific, Rockford, IL, USA) using goat anti-mouse IgG coupled to peroxidase. Immunoblots were scanned with a GS-710-calibrated imaging densitometer (Bio-Rad Laboratories).

6.2.5 Steady-state GTPase activity assay

Steady-state GTPase activity assays, using [γ -³³P]GTP as radioligand, were essentially performed as already described (Preuss *et al.*, 2007c). Assay tubes contained membranes expressing hH₂R-Y182F-G_{ssS} fusion protein (10 µg of protein/tube), 1.0 mM MgCl₂, 100 µM EDTA, 100 µM ATP, 100 nM GTP, 100 µM adenylyl imidodiphosphate, 1.2 mM creatine phosphate, 1 µg creatine kinase, 0.2% (w/v) bovine serum albumin in 50 mM Tris/HCl, pH 7.4, and ligands in various concentrations. For the determination of pK_B values histamine was added to the reaction mixtures (final concentration of 1 µM). The mixtures (80 µl) were incubated for 2 min at 25 °C before the addition of 20 µl of [γ -³³P]GTP (0.05 µCi per tube) and subsequently allowed to react for 20 min at 25 °C. The reaction was terminated by the addition of 900 µl of a slurry consisting of 5% (w/v) activated charcoal (absorbs nucleotides but not P_i) and 50 mM NaH₂PO₄, pH 2.0. Reaction mixtures were centrifuged at room

temperature (7 min, 15,000 g). 600 μ L of the supernatant were removed and $^{33}\text{P}_i$ was determined by liquid scintillation counting using OptiPhase Supermix Cocktail (PerkinElmer, Groningen, The Netherlands). Enzyme activities were corrected for spontaneous hydrolysis of $[\gamma\text{-}^{33}\text{P}]\text{GTP}$ determined in tubes containing all components described above plus a high concentration of unlabeled GTP (1 mM) to prevent enzymatic cleavage of the labeled nucleotides in the presence of Sf9 membranes. Spontaneous $[\gamma\text{-}^{33}\text{P}]\text{GTP}$ hydrolysis was < 1% of the total amount of radioactivity added. The experimental conditions chosen ensured that not more than 20% of the total amount of added $[\gamma\text{-}^{33}\text{P}]\text{GTP}$ was converted to $^{33}\text{P}_i$.

6.2.6 $^{35}\text{S}[\text{GTP}\gamma\text{S}]$ binding assay

Membranes were thawed, sedimented by centrifugation at 4 °C and 13,000 g for 10 min, and carefully resuspended in binding buffer (12.5 mM MgCl_2 , 1 mM EDTA and 75 mM Tris/HCl, pH 7.4) to remove residual endogenous guanine nucleotides. Experiments were performed in 96-well plates in a total volume of 100 μ L per well, containing 10 μ g membrane protein, 1 μ M GDP, 0.05% (w/v) bovine serum albumin (BSA), 20 nCi of $^{35}\text{S}[\text{GTP}\gamma\text{S}]$ (≥ 0.2 nM) and the investigated ligands at various concentrations in binding buffer, added as 10-fold concentrated stock solutions. For the determination of K_B values of neutral antagonists, histamine was added to the reaction mixture at a concentration corresponding to the 10-fold EC_{50} value at the respective receptor (final concentration 1 μ M). Nonspecific binding was determined in the presence of 10 μ M unlabeled $\text{GTP}\gamma\text{S}$. After incubation for 90 min at 25 °C and shaking at 200 rpm, bound $^{35}\text{S}[\text{GTP}\gamma\text{S}]$ was separated from free $^{35}\text{S}[\text{GTP}\gamma\text{S}]$ by filtration through GF/C filters using a 96-well Brandel harvester (Brandel Inc., Unterföhring, Germany), followed by three washes with 2 mL of binding buffer (4 °C). Filters were dried overnight, MeltiLex solid scintillator was melted onto the filtermats and luminescence was measured with the MicroBeta² 1450 Plate Counter (Perkin Elmer, Rodgau, Germany).

6.2.7 Miscellaneous

Protein concentrations were determined using the DC protein assay kit (Bio-Rad, Hercules, CA, USA). Analyses of experimental data were performed with the Prism 5.0 software (GraphPad Software, San Diego, CA) by nonlinear regression and best fit to sigmoidal concentration-response curves. K_B values were calculated using the Cheng-Prusoff equation (Cheng and Prusoff, 1973). The intrinsic activity (E_{max}) is referred to the maximal response to histamine (set to 1.0) at the respective receptor. Data are presented as mean of at least 3 independent experiments \pm standard error of the mean (SEM), performed in duplicate or

triplicate. Significance was calculated using the unpaired two-tailed t-test and a confidence interval of 95% ($p < 0.05$) or 99% ($p < 0.01$).

6.3 Results

6.3.1 Immunological detection of hH₂R-Y182F-G_{saS} in Sf9 cell membranes

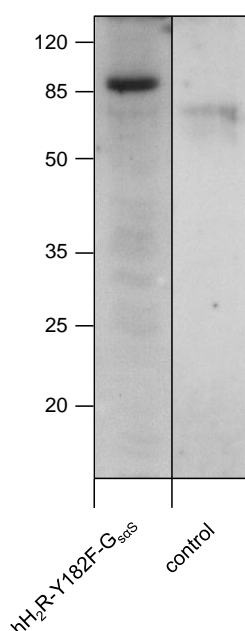


Figure 6.3: Immunological detection of the expression of hH₂R-Y182F-G_{saS} in Sf9 cells

Membranes of Sf9 cells expressing hH₂R-Y182F-G_{saS} and membranes of uninfected Sf9 cells (control) were prepared, separated by SDS-PAGE on gels containing 12% (w/v) acrylamide, transferred onto nitrocellulose membranes and probed with the anti-FLAG (M1) antibody. In each lane, 7.5 µg of membrane protein was loaded onto the gel. Labels on the left designate masses of marker proteins in kDa.

In Sf9 cells hH₂R-Y182F-G_{saS} was well expressed (Figure 6.3). The calculated mass of the non-glycosylated hH₂R fusion protein is 87.8 kDa. Bands of fusion proteins separated by SDS-PAGE on gels containing 12% (w/v) acrylamide were detected at about 80 kDa (Houston *et al.*, 2002; Kelley *et al.*, 2001; Preuss *et al.*, 2007b; c). SDS-PAGE analysis of membranes expressing hH₂R-Y182F-G_{saS} yielded intense bands at 85-90 kDa recognized by the anti-FLAG antibody. This is in good agreement with the result obtained previously. A slight band observed in uninfected Sf9 membranes at 80-85 kDa is also present in membranes expressing hH₂R-Y182F-G_{saS}. In the control there is no protein detectable at 85-90 kDa which could pretend the presence of the fusion protein hH₂R-Y182F-G_{saS}.

6.3.2 Agonistic activities at hH₂R-G_{sαS} and hH₂R-Y182F-G_{sαS} in the GTPase assay

In comparison to the wild-type (wt) hH₂R potencies of HIS, BUR, UR-AK471 and UR-BIT24 at hH₂R-Y182F-G_{sαS} remained nearly unchanged (EC_{50} ratio of wt and Y182F hH₂R 0.7 to 1.0) (Table 6.1). Although significantly, also pEC_{50} values of AMT, IMP and UR-AK51 ($p < 0.05$, respectively) were only slightly changed (EC_{50} ratios of 0.5 to 1.9). DIM exhibited the largest change in potency (3.1 fold increase; $p < 0.01$). For most agonists investigated in the GTPase assay a slight decrease of the pEC_{50} value was observed at the mutant hH₂R. Only DIM and UR-AK51 were more potent at hH₂R-Y182F-G_{sαS}. Intrinsic activities were similar at both receptors. A significant increase was measured for AMT ($p < 0.01$; +0.07) and DIM ($p < 0.05$; +0.04), a decrease for BUR ($p < 0.01$; -0.06).

Table 6.1: Agonistic activities at hH₂R-G_{sαS} and hH₂R-Y182F-G_{sαS} fusion proteins in the GTPase assay

Cpd.	hH ₂ R-G _{sαS} ^a		hH ₂ R-Y182F-G _{sαS}	
	$pEC_{50} \pm SEM$	$E_{max} \pm SEM$ ^b	$pEC_{50} \pm SEM$ ^c	$E_{max} \pm SEM$ ^{b, c}
HIS	6.00 ± 0.02	1.00	5.92 ± 0.08	1.00
AMT	6.74 ± 0.07	0.91 ± 0.01	6.47 ± 0.03 *	0.98 ± 0.01 **
DIM	6.08 ± 0.09	0.85 ± 0.01	6.57 ± 0.05 **	0.89 ± 0.02 *
IMP	6.82 ± 0.06	0.82 ± 0.01	6.57 ± 0.02 *	0.85 ± 0.01
BUR	5.00 ± 0.25	0.16 ± 0.00	4.87 ± 0.03	0.10 ± 0.00 **
UR-AK51	7.00 ± 0.06	0.84 ± 0.03	7.28 ± 0.09 *	0.80 ± 0.01
UR-AK471	7.63 ± 0.03	0.82 ± 0.02	7.64 ± 0.10	0.85 ± 0.03
UR-BIT24	7.68 ± 0.12	0.79 ± 0.02	7.55 ± 0.03	0.81 ± 0.01

GTPase activity on Sf9 membranes was determined as described in section 6.2. Data shown are the means ± SEM of three to five experiments performed in duplicate. Intrinsic activities and potencies, respectively, of ligands at hH₂R-Y182F-G_{sαS} were compared with the corresponding parameters at hH₂R-G_{sαS} using the t-test.

^a Data at the wild-type receptor were taken from Kraus *et al.* (2009), Preuss *et al.* (2007a) and Xi *et al.* (2006).

^b Intrinsic activity relative to the maximal response of histamine ($E_{max} = 1.00$).

^c Comparison with pEC_{50} / E_{max} at hH₂R-G_{sαS}; * $p < 0.05$; ** $p < 0.01$.

6.3.3 Potencies and intrinsic activities at hH₂R-G_{saS} and hH₂R-Y182F-G_{saS} in the [³⁵S]GTPγS binding assay

In contrast to the GTPase assay, pEC₅₀ values from the [³⁵S]GTPγS assay surprisingly increased for all full and strong partial agonists at hH₂R-Y182Y-G_{saS} (Table 6.2). The ligands were 1.5- to 2.9-fold more potent at hH₂R-Y182F-G_{saS}. The increase in potency was significant (p<0.05) for DIM, IMP, UR-AK51 and UR-AK471. The weak partial agonist BUR was slightly less potent at the mutant compared to the wild-type.

Table 6.2: Histamine H₂R agonism and antagonism at hH₂R-G_{saS} and hH₂R-Y182F-G_{saS} fusion proteins in the [³⁵S]GTPγS assay

Cpd.	hH ₂ R-G _{saS}		hH ₂ R-Y182F-G _{saS}	
	pEC ₅₀ or (pK _B) ± SEM	E _{max} ± SEM ^a	pEC ₅₀ or (pK _B) ± SEM ^b	E _{max} ± SEM ^{a, b}
HIS	6.35 ± 0.04	1.00	6.52 ± 0.12	1.00
AMT	6.89 ± 0.03	1.00 ± 0.07	7.11 ± 0.09	0.96 ± 0.03
DIM	6.27 ± 0.09	0.95 ± 0.07	6.73 ± 0.05 *	0.91 ± 0.01
IMP	6.73 ± 0.03	0.73 ± 0.03	6.94 ± 0.04 *	0.91 ± 0.01 **
BUR	5.51 ± 0.12	0.14 ± 0.02	5.18 ± 0.38	0.18 ± 0.06
UR-AK51	7.06 ± 0.05	0.75 ± 0.07	7.31 ± 0.08 *	0.91 ± 0.03
UR-AK471	7.34 ± 0.02	0.71 ± 0.01	7.70 ± 0.12 *	0.89 ± 0.02 **
UR-BIT24	7.86 ± 0.12	0.74 ± 0.06	8.17 ± 0.04	0.84 ± 0.04
CIM	(5.67 ± 0.02)	-0.09 ± 0.00	(5.63 ± 0.06)	-0.10 ± 0.00
FAM	(7.16 ± 0.01)	-0.08 ± 0.01	(7.03 ± 0.03)*	-0.10 ± 0.01
RAN	(5.98 ± 0.05)	-0.09 ± 0.01	(5.21 ± 0.05)**	-0.10 ± 0.01
TIO	(6.96 ± 0.05)	-0.09 ± 0.01	(7.01 ± 0.07)	-0.10 ± 0.01

[³⁵S]GTPγS binding was determined as described in section 6.2. Data shown are the means ± SEM of three to six experiments performed in triplicate. For the determination of K_B values, reaction mixtures contained membranes of Sf9 cells expressing fusion proteins, 1 μM HA as agonist and antagonists at concentrations from 1 nM to 1 μM as appropriate to generate saturated competition curves. To determine the intrinsic activities of inverse agonists, the effects of antagonists at 100 μM on basal GTPγS binding were assessed and referred to the effect of 100 μM HA (set to E_{max} = 1.00). Intrinsic activities and potencies/pK_B values, respectively, of ligands at hH₂R-Y182F-G_{saS} were compared with the corresponding parameters at hH₂R-G_{saS} using the t-test.

^a Intrinsic activity relative to the maximal response of histamine (E_{max} = 1.00).

^b Comparison with pEC₅₀/pK_B / E_{max} at hH₂R-G_{saS}; * p<0.05; ** p<0.01.

Except for AMT and DIM, E_{\max} values of all ligands were increased at the mutant hH_2R ; this was significant ($p < 0.01$) for IMP and UR-AK471 (Table 6.2). Intrinsic activities of partial inverse agonists remained unaffected by the mutation Tyr182→Phe182. Respective pK_B values obtained in competition experiments with 1 μM histamine remained unchanged for CIM and TIO, and decreased slightly for FAM ($p < 0.05$). Remarkably, the K_B value of RAN decreased to a larger extent (6-fold) at hH_2R -Y182F- G_{saS} ($p < 0.01$).

6.3.4 Comparison of data obtained in the GTPase and [^{35}S]GTP γ S assay

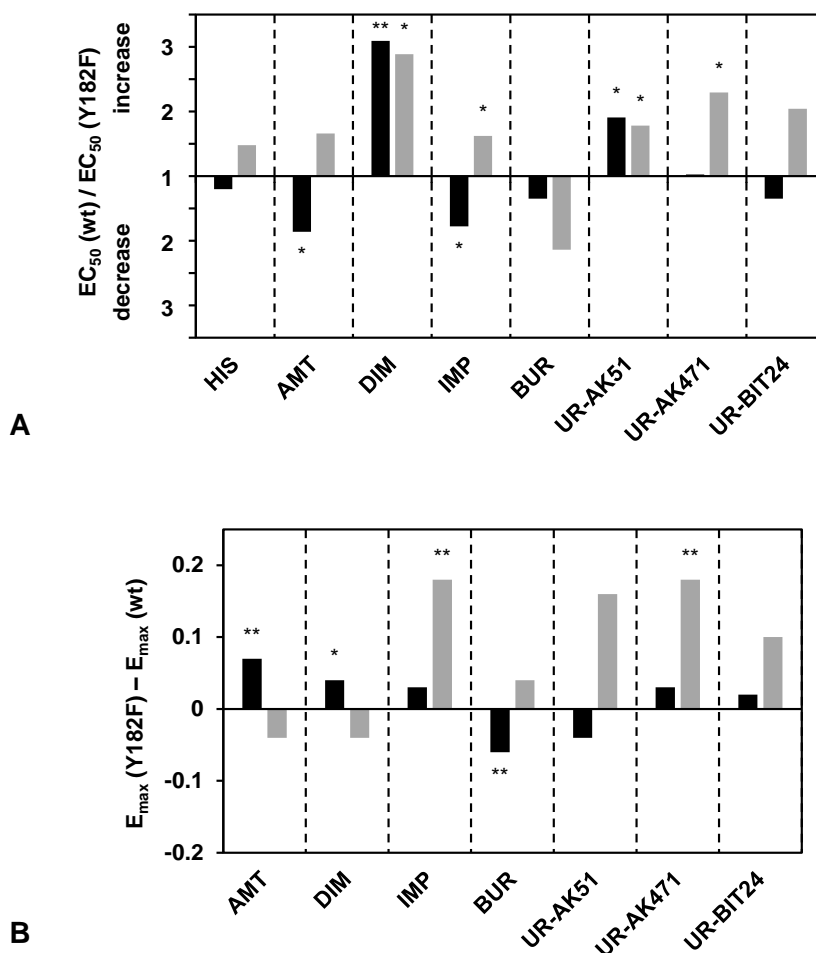


Figure 6.4: Potency ratio and E_{\max} shift of wild-type and mutated hH_2R

Ratio $EC_{50}(hH_2R) / EC_{50}(Y182F)$ (A) and difference of E_{\max} values (B) of mutated and wild-type hH_2R , obtained in the GTPase assay (black) and the [^{35}S]GTP γ S assay (grey). E_{\max} changes and potency ratios, respectively, of ligands at hH_2R -Y182F- G_{saS} were compared with the corresponding parameters at hH_2R - G_{saS} using the t-test (* $p < 0.05$; ** $p < 0.01$).

Potencies increased upon Tyr182→Phe182 mutation in both assays for DIM and UR-AK51 (Figure 6.4 A). UR-AK471 exhibited an unchanged (GTPase) and an increased (GTPγS assay) pEC₅₀ value. BUR was less potent in both assays at the mutated receptor. For the remaining ligands (HIS, AMT, IMP and UR-BIT24) contrary tendencies were observed: a decrease in potency in the GTPase and an increase in the GTPγS assay at the mutated hH₂R. In most cases only minor changes of intrinsic activities (± 0.1) upon mutation of Tyr182^{5,38} were observed. However, in the GTPγS assay, greater increases of E_{max} values resulted for IMP, UR-AK51 and UR-AK471 (Figure 6.4 B).

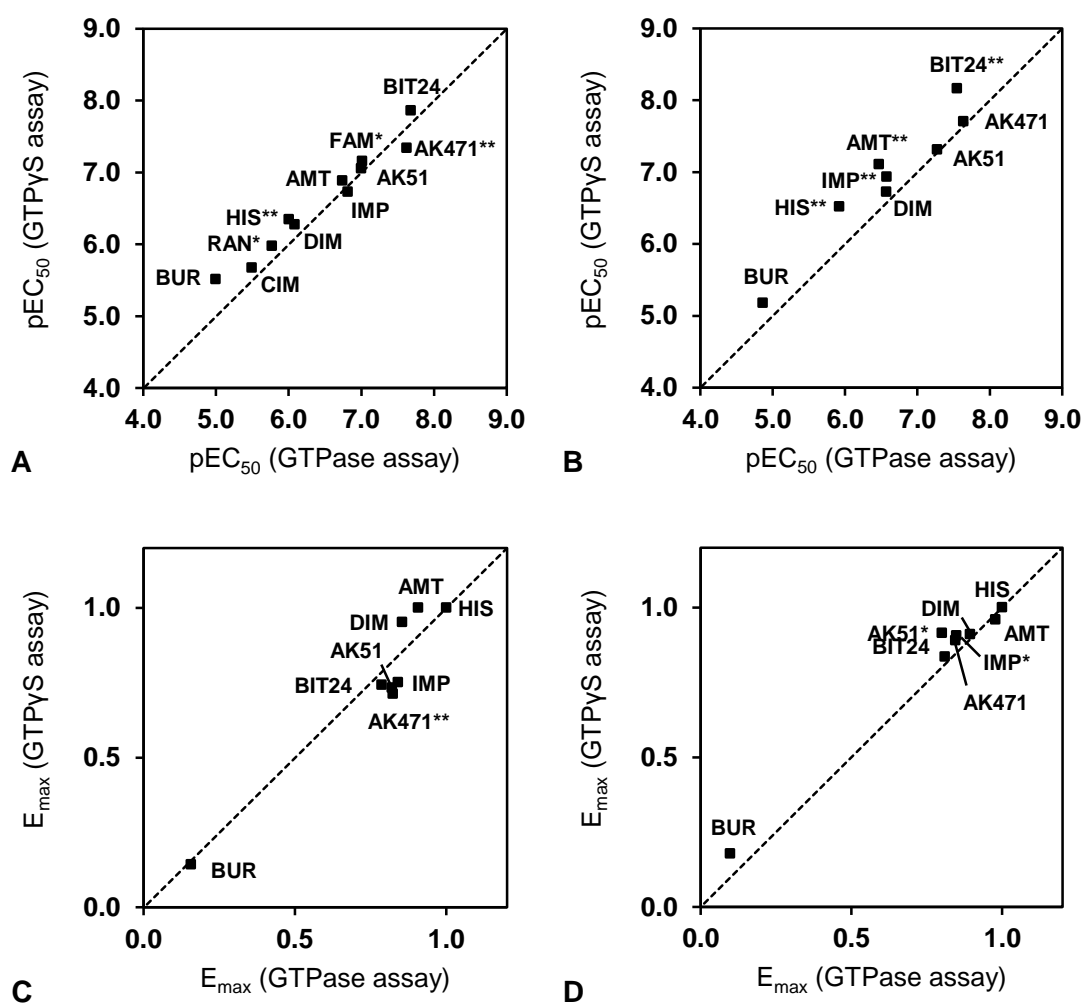


Figure 6.5: Intrinsic activities and potencies obtained in the GTPase assay in comparison to the [³⁵S]GTPγS assay

Potencies at hH₂R-G_{saS} (A) and hH₂R-Y182F-G_{saS} (B) as well as intrinsic activities at hH₂R-G_{saS} (C) and hH₂R-Y182F-G_{saS} (D) as determined in the GTPase assay and the [³⁵S]GTPγS assay. The dotted lines represent the line of identity. E_{max} and pEC₅₀ (pK_B in case of partial inverse agonists CIM, FAM and RAN) values, respectively, of ligands determined in the GTPγS assay were compared with the corresponding parameters determined in the GTPase assay using the t-test (* p<0.05; ** p<0.01).

In principle, for various GPCR systems the GTPase and [^{35}S]GTP γ S binding assays can be used interchangeably to assess the potencies and intrinsic activities of agonists and inverse agonists (Wieland and Seifert, 2006). In the case of the hH₂R-G_{sαS} and hH₂R-Y182F-G_{sαS}, all ligands exhibited higher pEC₅₀ or pK_B values in the [^{35}S]GTP γ S assay than in the GTPase assay except for UR-AK471 (p<0.01) and IMP at the wild-type receptor (Figure 6.5 A). Whereas in most cases a minor difference appeared, for HIS (wt and Y182F), AMT (Y182F), IMP (Y182F), UR-BIT24 (Y182F) (p<0.01) and BUR (wt and Y182F) a 2- to 4-fold increase in the pEC₅₀ value was obtained, respectively. Intrinsic activities were comparable for both assays and no general trend was detected (Figure 6.5 B).

6.4 Discussion

Nederkoorn *et al.* (1996a; 1996b) suggested that Tyr182^{5.38} in TM5 of the H₂R contributes to ligand binding via a hydrogen bond of the phenolic hydroxyl group. To test this hypothesis, a Tyr182→Phe182 mutation at the hH₂R was performed in the present work. Structurally different ligands were investigated in the GTPase and [^{35}S]GTP γ S assay. However, for the majority of these compounds a contribution of Tyr182^{5.38} was not obvious. Only the weak partial agonist burimamide exhibited slightly, but not significantly (significance level 0.05) reduced pEC₅₀ values in both assays. Moreover, contrary to expectations, dimaprit and UR-AK51 showed increased potencies. The decrease in pK_B value was most pronounced in case of ranitidine. In contrast to the other partial inverse agonists CIM, FAM and TIO, ranitidine differs in its physicochemical properties. Whereas the amidine and guanidine moieties of all four inverse agonists have a pK_a value smaller than 2.7 due to an electron withdrawing substitution (Carey *et al.*, 1981; Durant *et al.*, 1977), the tertiary amine of RAN possesses a pK_a value of about 8.2 (Carey *et al.*, 1981). For the corresponding substructures in CIM, FAM and TIO (imidazole, 2-guanidinothiazole) pK_a values of 6.8 to 7.1 were reported (Button *et al.*, 1985; Islam and Narurkar, 1993; Shankley *et al.*, 1988; Vochten *et al.*, 1980). Accordingly, the fraction of charged molecules at a physiological pH of 7.4 is about 86% for RAN and only 20% to 33% for the remaining ones. Possibly, the amine moiety of RAN interacts with the hydroxyl group of Tyr182^{5.38} via a charge assisted hydrogen bond in the wild-type hH₂R. Elimination of this interaction by Tyr182→Phe182 mutation results in decreased affinity. Furthermore, Tyr182^{5.38} is able to form a hydrogen bond to the side chain of Asp186^{5.42}, potentially stabilizing the latter in a favored position. Asp186^{5.42} was shown to be important for ligand binding in the H₂R (Gantz *et al.*, 1992). Besides the removal of a direct interaction between Tyr182^{5.38} and some ligands, the loss of the attachment of Asp186^{5.42} to Tyr182^{5.38} could be a reason for a decrease (BUR, RAN) or increase in potency

(DIM, UR-AK51). Nevertheless, Tyr182^{5,38} has only minor effects on binding of a few ligands at the human H₂R.

6.5 References

- Button RG, Cairns JP and Taylor PJ. Tautomeric ratio in 4-methylthiazol-2-ylguanidine, a model guanidinoheterocycle. *J Chem Soc, Perkin Trans 2* **1985**, 0, 1555-1558.
- Carey PF, Martin LE and Owen PE. Determination of ranitidine and its metabolites in human urine by reversed-phase ion-pair high-performance liquid chromatography. *J Chromatogr* **1981**, 225, 161-168.
- Cheng Y and Prusoff WH. Relationship between Inhibition Constant (K₁) and Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I₅₀) of an Enzymatic-Reaction. *Biochem Pharmacol* **1973**, 22, 3099-3108.
- Durant GJ, Duncan WA, Ganellin CR, Parsons ME, Blakemore RC and Rasmussen AC. Impromidine (SK&F 92676) is a very potent and specific agonist for histamine H₂ receptors. *Nature* **1978**, 276, 403-405.
- Durant GJ, Emmett JC, Ganellin CR, Miles PD, Parsons ME, Prain HD and White GR. Cyanoguanidine-thiourea equivalence in the development of the histamine H₂-receptor antagonist, cimetidine. *J Med Chem* **1977**, 20, 901-906.
- Durant GJ, Ganellin CR, Hills DW, Miles PD, Parsons ME, Pepper ES and White GR. The histamine H₂-receptor agonist impromidine: synthesis and structure activity considerations. *J Med Chem* **1985**, 28, 1414-1422.
- Gantz I, DelValle J, Wang LD, Tashiro T, Munzert G, Guo YJ, Konda Y and Yamada T. Molecular basis for the interaction of histamine with the histamine H₂ receptor. *J Biol Chem* **1992**, 267, 20840-20843.
- Houston C, Wenzel-Seifert K, Burckstummer T and Seifert R. The human histamine H₂-receptor couples more efficiently to Sf9 insect cell Gs-proteins than to insect cell Gq-proteins: limitations of Sf9 cells for the analysis of receptor/Gq-protein coupling. *J Neurochem* **2002**, 80, 678-696.
- Islam MS and Narurkar MM. Solubility, stability and ionization behaviour of famotidine. *J Pharm Pharmacol* **1993**, 45, 682-686.
- Kelley MT, Burckstummer T, Wenzel-Seifert K, Dove S, Buschauer A and Seifert R. Distinct interaction of human and guinea pig histamine H₂-receptor with guanidine-type agonists. *Mol Pharmacol* **2001**, 60, 1210-1225.
- Kraus A, Ghorai P, Birnkammer T, Schnell D, Elz S, Seifert R, Dove S, Bernhardt G and Buschauer A. N-G-Acylated Amino-thiazolylpropylguanidines as Potent and Selective Histamine H₂ Receptor Agonists. *ChemMedChem* **2009**, 4, 232-240.
- Liapakis G, Ballesteros JA, Papachristou S, Chan WC, Chen X and Javitch JA. The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the beta 2-adrenergic receptor. *J Biol Chem* **2000**, 275, 37779-37788.
- Mansour A, Meng F, Meador-Woodruff JH, Taylor LP, Civelli O and Akil H. Site-directed mutagenesis of the human dopamine D₂ receptor. *Eur J Pharmacol* **1992**, 227, 205-214.
- Nederkoorn PH, van Gelder EM, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. II. Theoretical investigations of histamine binding to receptor models of the seven alpha-helical transmembrane domain. *J Comput Aided Mol Des* **1996a**, 10, 479-489.
- Nederkoorn PH, van Lenthe JH, van der Goot H, Donne-Op den Kelder GM and Timmerman H. The agonistic binding site at the histamine H₂ receptor. I. Theoretical investigations of histamine binding to an oligopeptide mimicking a part of the fifth transmembrane alpha-helix. *J Comput Aided Mol Des* **1996b**, 10, 461-478.

- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Constitutive activity and ligand selectivity of human, guinea pig, rat, and canine histamine H-2 receptors. *J Pharmacol Exp Ther* **2007a**, 321, 983-995.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Mutations of Cys-17 and Ala-271 in the human histamine H-2 receptor determine the species selectivity of guanidine-type agonists and increase constitutive activity. *J Pharmacol Exp Ther* **2007b**, 321, 975-982.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Point mutations in the second extracellular loop of the histamine H-2 receptor do the species-selective activity of not affect guanidine-type agonists. *Naunyn Schmiedebergs Arch Pharmacol* **2007c**, 376, 253-264.
- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011**, 469, 175-180.
- Schnell D, Burleigh K, Trick J and Seifert R. No Evidence for Functional Selectivity of Proxyfan at the Human Histamine H-3 Receptor Coupled to Defined G(i)/G(o) Protein Heterotrimers. *J Pharmacol Exp Ther* **2010**, 332, 996-1005.
- Schwarz RD, Spencer CJ, Jaen JC, Mirzadegan T, Moreland D, Tecle H and Thomas AJ. Mutations of aspartate 103 in the Hm2 receptor and alterations in receptor binding properties of muscarinic agonists. *Life Sci* **1995**, 56, 923-929.
- Seifert R, Lee TW, Lam VT and Kobilka BK. Reconstitution of beta(2)-adrenoceptor-GTP-binding-protein interaction in Sf9 cells - High coupling efficiency in a beta(2)-adrenoceptor-G(s alpha) fusion protein. *Eur J Biochem* **1998**, 255, 369-382.
- Shankley NP, Black JW, Ganellin CR and Mitchell RC. Correlation between log POCT/H2O and pKB estimates for a series of muscarinic and histamine H2-receptor antagonists. *Br J Pharmacol* **1988**, 94, 264-274.
- Strader CD, Sigal IS, Candelore MR, Rands E, Hill WS and Dixon RA. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *J Biol Chem* **1988**, 263, 10267-10271.
- Sun X, Li Y, Li W, Xu Z and Tang Y. Computational investigation of interactions between human H2 receptor and its agonists. *J Mol Graph Model* **2011**, 29, 693-701.
- Vochten R, Remaut G and Huybrechts W. Physicochemical properties of the H2-receptor antagonist cimetidine. *J Pharm Pharmacol* **1980**, 32, 863-866.
- Walseth TF and Johnson RA. The enzymatic preparation of [alpha-(32)P]nucleoside triphosphates, cyclic [32P] AMP, and cyclic [32P] GMP. *Biochim Biophys Acta* **1979**, 562, 11-31.
- Wieland T and Seifert R. Methodological Approaches. In *G Protein-Coupled Receptors as Drug Targets*, Wiley-VCH Verlag GmbH & Co. KGaA: **2006**; pp 81-120.
- Xie SX, Kraus A, Ghorai P, Ye QZ, Elz S, Buschauer A and Seifert R. N1-(3-cyclohexylbutanoyl)-N2-[3-(1H-imidazol-4-yl)propyl]guanidine (UR-AK57), a potent partial agonist for the human histamine H1- and H2-receptors. *J Pharmacol Exp Ther* **2006**, 317, 1262-1268.

Chapter 7

The Role of Acidic Amino Acids in the Third Extracellular Loop of the Guinea Pig Histamine H₂ Receptor

7.1 Introduction

The three extracellular loops in G-protein coupled receptors, ECL1, ECL2 and ECL3, are very different regarding sequence and length (Wheatley *et al.*, 2012), and belong to the most flexible elements in GPCRs (Peeters *et al.*, 2011). For example, in the crystal structures of the human adenosine A_{2A} receptor (Jaakola *et al.*, 2008), despite the presence of three restricting disulfide linkages between ECL2 and ECL1, the second extracellular loop is not completely resolved due to weak electron density. Figure 7.1 shows the crystallographic temperature factors (B-factors) of an active GPCR conformation of the hβ₂AR (PDB ID 3P0G; Rasmussen *et al.*, 2011a). Besides the N- and C-terminus, the extra- and intracellular loops possess the highest flexibility. With the recent success in GPCR crystallization, NMR spectroscopy (Tikhonova and Costanzi, 2009), and indirect methods like site-directed mutagenesis (Hawtin *et al.*, 2006) or the substituted cysteine accessibility method (SCAM) (Shi and Javitch, 2004), it has become clear that not only the ligand binding transmembrane domains and the intracellular loops, interacting with downstream effectors, are important for receptor activation. Several amino acids in the extracellular loops were identified to be important for ligand binding and the activation mechanism of class A GPCRs (Wheatley *et al.*, 2012). This is also supported by the fact that some mutations in the extracellular part of GPCRs are associated with diseases (Schöneberg *et al.*, 2004; Thompson *et al.*, 2008). Whereas ECL2 is best investigated and seems to play a major role in receptor activation, ECL3 is probably the most neglected loop (Peeters *et al.*, 2011). Nevertheless, ECL3 is involved in GPCR function, e.g. signal transduction and G-protein activation (Claus *et al.*, 2005), cell surface expression (Hawtin *et al.*, 2006) and ligand binding (Harerich *et al.*,

2008). However, most reports on the loop region are concerned with peptidergic GPCRs, whereas corresponding information on aminergic GPCRs is scarce (Lawson and Wheatley, 2004; Peeters *et al.*, 2012; Wheatley *et al.*, 2007; Wheatley *et al.*, 2012). ECL3 is part of an allosteric site in muscarinic acetylcholine receptors (Gregory *et al.*, 2010; Huang *et al.*, 1999; Jager *et al.*, 2007; Nawaratne *et al.*, 2010) and contributes to ligand binding in the adenosine A_{2A} receptor (Jaakola *et al.*, 2008; Peeters *et al.*, 2012). A hybrid β_2 AR, integrating the ECL3 of the α_{1A} AR, exhibited higher binding affinity for agonists and a higher agonist-independent basal (constitutive) adenylyl cyclase activity than the wild-type β_2 AR (Zhao *et al.*, 1998).

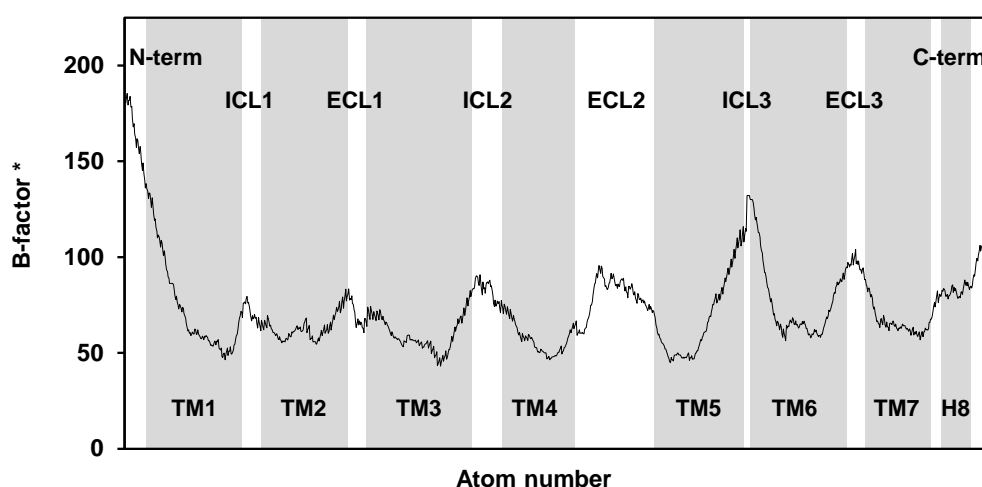


Figure 7.1: B-factors of backbone heavy atoms of the crystal structure of an active h β_2 AR conformation

* The more flexible an atom the larger is the displacement from the mean position, expressed as temperature factors or B-factors. B-factors were taken from PDB entry 3P0G (Rasmussen *et al.*, 2011a).

To study the third extracellular loop, a multiple sequence alignment analysis of 340 aminergic GPCRs was performed (section 7.3.2). The results revealed that the ECL3 in the gpH₂R is exceptional due to a high percentage of negatively charged amino acids. Furthermore, amino acids which potentially interact with the basic acylguanidine moiety of the second pharmacophore of bivalent acylguanidine-type agonists were identified in ECL3 of the H₂R (section 7.3.1). Physicochemical properties and the location of this loop relative to the orthosteric binding site might explain the high potency of bivalent compounds compared to monovalent ones, as well as their dependence on spacer length (chapter 1).

Thus, to investigate the importance of acidic amino acids in ECL3 for ligand binding, ligand attraction and receptor activation, as well as their capability to compose a second binding site for bivalent H₂R agonists, in a first gpH₂R mutant Asp262, Asp263, Glu267 and Glu270 were replaced by serine. Serine was selected to preserve the polar properties of ECL3, able to form hydrogen bonds with the solvent or ligands, but without the possibility of forming strong ionic interactions. Standard H₂R ligands (histamine, amthamine and dimaprit), monovalent (acyl-)guanidines (impromidine and UR-BIT24) and bivalent acylguanidines (UR-AK381, UR-BIT106) (Figure 7.2) were tested at the wild-type and the mutant gpH₂R in the [³⁵S]GTPγS binding assay and in the GTPase activity assay (HIS, AMT, DIM). Based on these results, two additional receptor mutants were generated. In the first one, Asp262, Asp263 and Glu267 were replaced by serine, and in the second one, the single point mutation Glu270→Ser270 was created. Ligands mentioned above were again investigated in the [³⁵S]GTPγS binding assay.

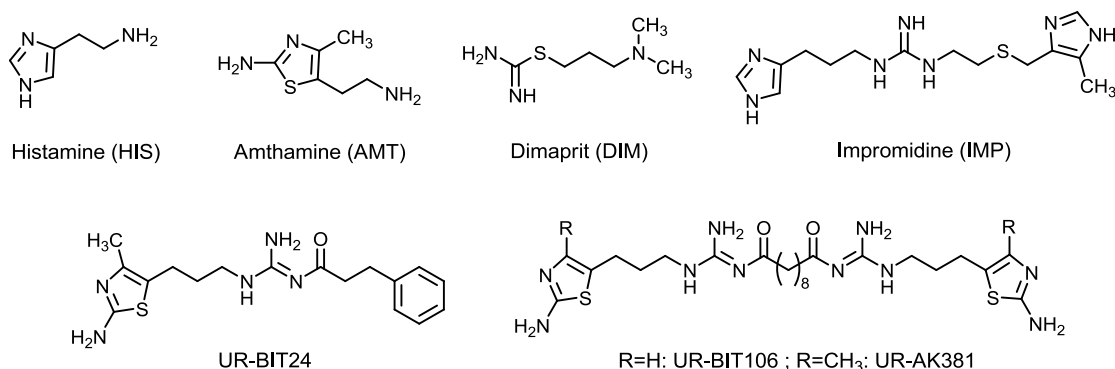


Figure 7.2: Structures of H₂R agonists

7.2 Materials and methods

7.2.1 Materials

The construction of the gpH₂R-G_{saS} plasmid was described previously (Kelley *et al.*, 2001). Phusion High-Fidelity DNA Polymerase was from New England Biolabs (Ipswich, MA, USA). Histamine, amthamine and dimaprit were purchased from Tocris (Avonmouth, Bristol, UK). UR-AK381, UR-BIT24 and UR-BIT106 were synthesized as described (Birnkammer *et al.*,

2012; Kraus *et al.*, 2009). Impromidine (Durant *et al.*, 1978) was synthesized as described (Durant *et al.*, 1985). UR-AK381 and UR-BIT106 (1 mM) were dissolved in 15% (v/v) dimethyl sulfoxide (DMSO) and dilutions were prepared in DMSO in order to attain a final DMSO concentration of 1.5% (v/v) in each assay tube. All other stock solutions were prepared with Millipore water. Other materials are described in section 6.2.

7.2.2 Construction of the cDNA for gpH₂R-D262S-D263S-E267S-E270S-G_{saS}

The cDNA for gpH₂R-D262S-D263S-E267S-E270S-G_{saS} (gpH₂R-ECL3-4Ser-G_{saS}) was generated by sequential overlap-extension PCR, using pVL1392-SF-gpH₂R-His₆-G_{saS} as template. In PCR 1A, the DNA region encoding the cleavable signal peptide from influenza hemagglutinin (S), the FLAG epitope (F) recognized by the M1 monoclonal antibody, and the N-terminal portion of the gpH₂R were amplified. The sense primer annealed with 27 base pairs (bp) of pVL1392-SF-gpH₂R-His₆-G_{saS} in front of the 5'-end of SF, containing a *Sac* I restriction site. The antisense primer encoded the sequence 5'-TGAAAACACCGAATTCACAGCACTACTCCCTTTCAGCCCACGGTAAACAAAAAC (mismatching base pairs underlined) to generate the Asp262→Ser262, Asp263→Ser263, Glu267→Ser267 and Glu270→Ser270 exchange, and a new *Eco* RI restriction site (CTTAAG). In PCR 1B, the DNA sequence of the remaining C-terminal part of the gpH₂R, a hexahistidine tag, and the entire sequence of G_{saS} were amplified using pVL1392-SF-gpH₂R-His₆-G_{saS} as template. The sense primer encoded the sequence 5'-AGTAGTGCTGTGAATTCGGTGTTTTCAGATGTTGTTCTGTGGCTGGGC-TATGCC to generate the Asp262→Ser262, Asp263→Ser263, Glu267→Ser267 and Glu270→Ser270 exchange, and a new *Eco* RI restriction site. The antisense primer annealed with 27 bp of the cDNA behind the C-terminus of G_{saS} and the stop codon, containing the *Xba* I restriction site. In PCR 2 the products of PCR 1A and 1B annealed in the region encoding the newly created mutations and the new *Eco* RI restriction site. Here, the sense primer of PCR 1A and the antisense primer of PCR 1B were used. In that way, the complete cDNA for the gpH₂R-D262S-D263S-E267S-E270S-G_{saS} fusion protein was amplified. The product of PCR 2 was double-digested with *Sac* I and *Xba* I and cloned into the pVL1392-SF-gpH₂R-His₆-G_{saS} plasmid digested with the same enzymes. The PCR-generated DNA sequence was verified by agarose gel electrophoresis, restriction enzyme analysis and sequencing (Entelechon, Regensburg, Germany).

7.2.3 Construction of the cDNA for gpH₂R-D262S-D263S-E267S-G_{saS} and gpH₂R-E270S-G_{saS}

To generate the cDNA for gpH₂R-D262S-D263S-E267S-G_{saS} (gpH₂R-ECL3-3Ser-G_{saS}) and gpH₂R-E270S-G_{saS}, the pVL1392 vector containing gpH₂R-D262S-D263S-E267S-E270S-G_{saS} and pVL1392-SF-gpH₂R-His₆-G_{saS}, respectively, was used as template. As in both cases a point mutation of one amino acid was sufficient, the constructs were prepared by analogy with a protocol known from the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene Cloning Systems, La Jolla, CA, USA). Two complementary primers, 5'-GCTGTGAATTCGGTGTTCGAGATGTTGTTCTGTGGCTG and 5'-CAGCCACAGAACAACATCTTCAAACACCGAATTCACAGC, both containing two mismatches for the Ser270→Glu270 exchange, respectively, annealed with the template DNA encoding the gpH₂R-D262S-D263S-E267S-E270S-G_{saS} receptor mutant. In the same way, two complementary primers, 5'-GCTGTCAATGAGGTGTTTCAGATGTT-GTTCTGTGGCTG and 5'-CAGCCACAGAACAACATCTGAAAACAC CTCATTGACAGC, both containing two mismatches for the Glu270→Ser270 exchange, respectively, annealed with the template pVL1392-SF-gpH₂R-His₆-G_{saS}. Both products were treated with *Dpn I* endonuclease (target sequence 5'-Gm6ATC), respectively, which is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template, originating from *E. coli* and therefore susceptible to *Dpn I* digestion. Transformation into competent *E. coli Top10* cells, single clone selection and plasmid preparation led to the desired gpH₂R-D262S-D263S-E267S-G_{saS} and gpH₂R-E270S-G_{saS} sequences in the pVL1392 vector, respectively. Both DNA sequences were confirmed by agarose gel electrophoresis and sequencing (Entelechon, Regensburg, Germany).

7.2.4 Cell culture, generation of recombinant baculoviruses and membrane preparation

Described in section 6.2.

7.2.5 Immunoblot analysis

Described in section 6.2.

7.2.6 Steady-state GTPase activity assay

Described in section 6.2.

7.2.7 [³⁵S]GTPγS binding assay

Described in section 6.2.

7.2.8 Multiple sequence alignment

Sequences of the aminergic class A rhodopsin-like GPCRs as listed in the GPCRDB database (Vroeling *et al.*, 2011), i.e. muscarinic acetylcholine, adrenergic, dopamine, histamine, serotonin, octopamine and trace amine receptors, which were classified as reviewed in the UniProt Knowledgebase (UniProtKB; Consortium, 2012), were taken from the latter database and aligned with ClustalW2 (Goujon *et al.*, 2010; Larkin *et al.*, 2007).

7.2.9 Homology model of the gpH₂R

The homology model of the gpH₂R was constructed using hH₂R_{a1} as template (chapter 3), which is based on the crystal structures of the β₁-adrenergic receptor (PDB ID 2VT4, chain B; Warne *et al.*, 2008) and opsin in its G-protein-interacting conformation (PDB ID 3DQB; Scheerer *et al.*, 2008). Side chain conformations of those amino acids which are identical in both H₂R species were retained unchanged. Deviating amino acids were mutated to the correct residue in the gpH₂R and conformations were adjusted as described in section 3.2.2.3. Further structural refinements were performed as described in section 3.2.2.3.

7.2.10 Miscellaneous

See section 6.2.

7.3 Results

7.3.1 Selection of a putative accessory binding site for bivalent H₂R agonists

The H₂R agonistic effects of bivalent *N*^G-acylated hetarylpropylguanidines synthesized in our group (Birnkammer *et al.*, 2012; Kraus *et al.*, 2009) were shown to be inhibited by H₂R antagonists like famotidine. The resulting K_B values for famotidine were comparable to those obtained when histamine was used as the agonist (Birnkammer *et al.*, 2012). This suggests that the bivalent compounds interact with the orthosteric binding site of the H₂R within TM3 and TM5. As a common binding mode of imidazole- and aminothiazole-type H₂R agonists, the heterocycle was suggested to interact with Asp186^{5.42} (Kraus *et al.*, 2009). At physiological pH (7.4) the acylguanidine moiety (pK_a = 7-8; Ghorai *et al.*, 2008; Rewinkel and

Adang, 1999; Schmuck and Lex, 1999) occurs in its protonated state predestinated for interaction with the negatively charged Asp98^{3.32}. This highly conserved residue was suggested to interact with the positively charged side chain nitrogen of histamine in the H₂R (Gantz *et al.*, 1992), by analogy with other biogenic amines and their corresponding GPCR, e.g. in the β_2 -adrenergic receptor (Strader *et al.*, 1988), the dopamine D₂ receptor (Mansour *et al.*, 1992) and the muscarinic acetylcholine receptor M₂ (Schwarz *et al.*, 1995). This strong interaction of the acylguanidine moiety of the first pharmacophore with Asp98^{3.32} reduces the degrees of freedom for interactions of the second pharmacophore, which also contains an acylguanidine. A simultaneous occupation of two orthosteric binding sites in H₂R dimers was excluded because of insufficient spacer lengths in the most potent acylguanidines. Bivalent compounds theoretically capable of bridging dimerizing H₂R protomers (necessary spacer length of about 20 CH₂ groups) are nearly inactive (Birnkammer, 2011; Birnkammer *et al.*, 2012).

To test the hypothesis that the second acylguanidine moiety interacts with acidic amino acids like in the orthosteric binding site, the receptor surface was screened for appropriate residues. Seven and ten acidic amino acids are located in the extracellular region of the hH₂R and the gpH₂R, respectively (Figure 7.3 and Table 7.1). These amino acids were considered potential candidates for a contribution to a second interaction site at the H₂R, explaining higher potencies of bivalent agonists compared to their monovalent congeners. Furthermore, this recognition site has to keep a suitable distance from the anchoring point Asp98^{3.32} related to spacer-dependent potencies of bivalent agonists. In the most potent bivalent acylguanidines, containing eight methylene groups as spacer, both pharmacophores are separated by 16 Å (distance between both carbon atoms of the guanidine moieties, extended conformation). In addition, Asp98^{3.32} is partially masked by other residues. For most residues at the extracellular surface it is impossible to bridge Asp98^{3.32} via the direct and shortest path, denoting that a distance of 16 Å between Asp98^{3.32} and a putative second interaction site is in most cases too large.

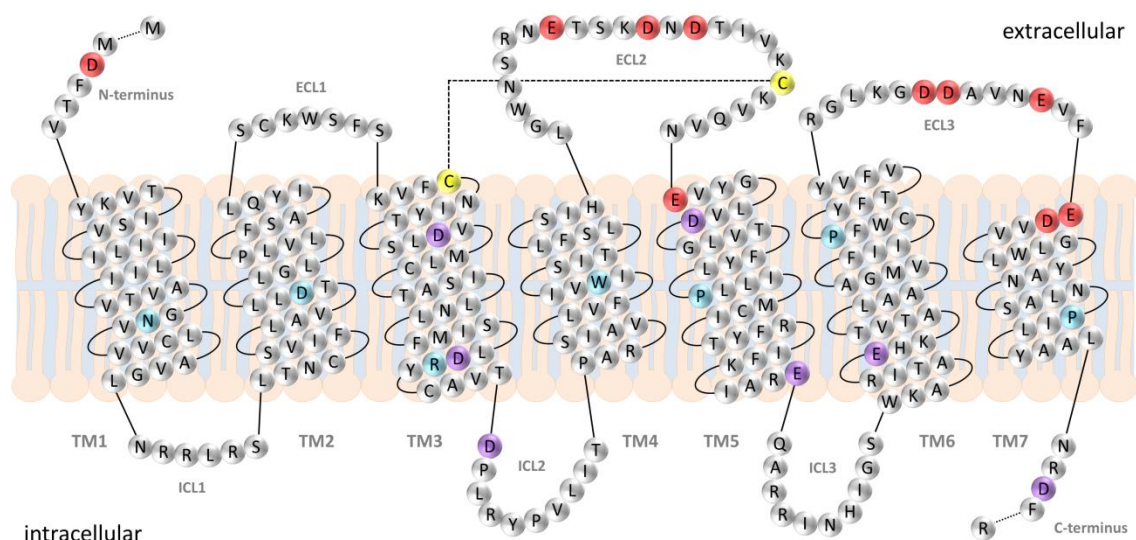


Figure 7.3: Snake representation of the guinea pig H₂R

Acidic residues at the extracellular part of the receptor are indicated with red balls, remaining ones with purple balls. Most conserved amino acids in each TM are shown in cyan. The N- and C-terminus (including one aspartic acid and four glutamic acids) are not shown for reasons of clarity. The disulfide bond between Cys91^{3,25} and Cys174^{5,28} in ECL2 is highlighted with dashed lines.

Table 7.1: Acidic amino acids in the human and guinea pig H₂ receptor

Amino acid	hH ₂ R	gpH ₂ R	Position	Amino acid	hH ₂ R	gpH ₂ R	Position
Asp13	x	x	N-term	Asp262	x	x	ECL3
Asp64 ^{2,50}	x	x	core	Asp263	x	x	ECL3
Asp98 ^{3,32}	x	x	OBS	Glu267	x	x	ECL3
Asp115 ^{3,49}	x	x	core/intrac	Glu270 ^{7,35}	x	x	ECL3/TM7
Asp122	x	x	ICL2	Asp271 ^{7,36}		x	OBS
Glu163	x	x	ECL2	Asp294	x	x	TM7/H8
Asp167		x	ECL2	Glu314		x	H8
Asp169		x	ECL2	Glu331	x	x	C-term
Glu180 ^{5,36}	x	x	ECL2/TM5	Glu336	x	x	C-term
Asp186 ^{5,42}	x	x	OBS	Asp337	x	x	C-term
Glu211 ^{5,67}	x	x	intrac	Glu349	x	x	C-term
Glu229 ^{6,30}	x	x	intrac	Asp358	x		C-term

An x indicates the presence of the respective residue in the hH₂R and the gpH₂R, respectively. Amino acids at the extracellular surface of the receptor are highlighted with a grey background. C-term, C-terminus; core, core of the receptor; N-term, N-terminus; intrac, intracellular side of the receptor; OBS, orthosteric binding site.

Asp13 resides in the N-terminus of both the human and guinea pig H₂R. In a recent study the N-terminus of the hH₂R was replaced by the corresponding amino acids in the gpH₂R (Birnkammer *et al.*, 2012; Brunskole, 2011). Only for one of six bivalent *N*^G-acylguanidines investigated at the hybrid, H₂R potency and intrinsic activity were shifted towards the gpH₂R. It was speculated about an interaction between the N-terminus and two or more extracellular loops. Asp13 of the H₂R is located in close proximity to the top of TM1 (Cys17^{1.31} in the hH₂R, Tyr17^{1.31} in the gpH₂R). The direct distance between Asp98^{3.32} and Cys/Tyr17^{1.31} (~ 26 Å, distance between C_α atoms) is too far for bridging Asp98^{3.32} in TM3 and Asp13 in the N-terminus with the most potent bivalent ligands.

The extracellular loop 2 of the human and guinea pig H₂R differ by four amino acids (Figure 7.4), including Asp167 and Asp169, which are only present in ECL2 of the gpH₂R. To explore the influence of ECL2 on the receptor orthologue selectivity of guanidine-type agonists, in a previous study (Preuss *et al.*, 2007c) the respective four amino acids in the hH₂R were replaced by the corresponding four amino acids of the gpH₂R (hH₂R_{gpE2}-G_{saS}) and *vice-versa* (gpH₂R_{hE2}-G_{saS}). Functional analysis of N-[3-(1*H*-imidazol-4-yl)propyl]guanidines and *N*^G-acylated analogues indicated that ECL2 does not contribute to preference of monovalent H₂R agonists for the guinea pig H₂R orthologue. Recently, selected bivalent acylguanidines were investigated at these reciprocal mutants as well (Birnkammer *et al.*, 2012). All compounds showed similar potencies and intrinsic activities at the hH₂R_{gpE2}-G_{saS} and the wild-type hH₂R-G_{saS} fusion proteins, indicating that the inserted acidic amino acids Asp167 and Asp169 are not involved in bivalent ligand binding. At the gpH₂R_{hE2}-G_{saS} mutant the intrinsic activities of the compounds were comparable to those at the wild-type gpH₂R-G_{saS}. However, three of eight investigated ligands exhibited significantly reduced pEC₅₀ values at the mutant gpH₂R (by 0.5–0.9 logarithmic units).

```

                160           170
                |           |
hH2R, ECL2  LGWNSRNETSKGNHTTSKCKVQVN
gpH2R, ECL2 LGWNSRNETSKDNDTIVKCKVQVN
***** * * *****

```

Figure 7.4: Sequence alignment of ECL2 of the human and guinea pig H₂R

An asterisk indicates corresponding amino acids in both species. Negatively charged amino acids are highlighted with a grey background

Although these inconsistent results do not suggest the mutated residues to be directly involved in ligand-receptor interactions, the integrity of ECL2 in the gpH₂R seems to be necessary for high-affinity binding of some bivalent acylguanidines (Birnkammer *et al.*, 2012). Hence, Asp167 and Asp169 in the gpH₂R were not considered promising candidates for an accessory binding site.

Glu163, present in both H₂R orthologues, is situated in the part of ECL2 farthest from the orthosteric binding site. The distance to the anchoring residue Asp98^{3,32} is about 24 Å (distance between C_α atoms). However, loops are among the most flexible regions in GPCRs, and there is some uncertainty about the orientation of ECL2 in the H₂R model. The outmost position of that part of ECL2 is comparable to the location in crystal structures of the tβ₁AR (Warne *et al.*, 2008), the hβ₂AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011a), the hA_{2A}AR (Jaakola *et al.*, 2008), the hH₁R (Shimamura *et al.*, 2011), the hD₃R (Chien *et al.*, 2010), the hmAChRM₂ (Haga *et al.*, 2012), the rmAChRM₃ (Kruse *et al.*, 2012) and the hCXCR4 (Wu *et al.*, 2010), suggesting that it is a general characteristic of some GPCRs. All in all, the contribution of Glu163 cannot be excluded but seems unlikely.

Table 7.2: Distance between C_α atoms of amino acids 3.32 and 5.36 in selected crystal structures of GPCRs

GPCR	PDB ID	Resolution [Å]	Amino acid 3.32	Amino acid 5.36	Distance [Å]
gpH ₂ R model	-	-	Asp98	Glu180	19.3
tβ ₁ AR ¹	2VT4	2.7	Asp121	Arg205	19.3
tβ ₁ AR ²	2Y00	2.5	Asp121	Arg205	18.8
hβ ₂ AR ³	2RH1	2.4	Asp113	Gln197	19.1
hβ ₂ AR ⁴	3P0G	3.5	Asp113	Gln197	19.0
hD ₃ R ⁵	3PBL	2.9	Asp110	Pro186	19.0
hH ₁ R ⁶	3RZE	3.1	Asp107	Thr188	19.1
Opsin ⁷	3DQB	3.2	Ala117	Glu201	21.1
bovine Rhodopsin ⁸	1U19	2.2	Ala117	Glu201	20.6

¹ Warne *et al.*, 2008 , chain B; ² Warne *et al.*, 2011 , chain B; ³ Cherezov *et al.*, 2007 ;

⁴ Rasmussen *et al.*, 2011a ; ⁵ Chien *et al.*, 2010 ; ⁶ Shimamura *et al.*, 2011 ; ⁷ Scheerer *et al.*, 2008 ; ⁸ Okada *et al.*, 2004

Glu180^{5.36}, present in the human and guinea pig H₂R, is a further candidate potentially interacting with the acylguanidine moiety of the second pharmacophore since it is located at the extracellular end of TM5. In the homology model of the gpH₂R the distance between Asp98^{3.32} and Glu180^{5.36} (about 20 Å; measured between C_α atoms) is equal to the distance of the corresponding amino acids in GPCR crystal structures (Table 7.2) and thus too large to enable an interaction of the second acylguanidine group with Glu180^{5.36}. However, the heterocycle connected to this acylguanidine moiety can definitely adopt a position in close proximity to Glu180^{5.36} (Figure 7.5).

A further amino acid which was previously investigated in more detail is Asp/Ala271^{7.36} of the H₂R (Ala in the hH₂R, Asp in the gpH₂R). To study the species selectivity of guanidine-type agonists, a single and a double hH₂R mutant (Ala271->Asp271, Cys17/Ala271->Tyr17/Asp271, respectively) was generated (Kelley *et al.*, 2001; Preuss *et al.*, 2007b). The insertion of the gpH₂R amino acids Asp and Tyr in positions 1.31 and 7.35, respectively, of the human receptor resulted in a change of the pEC₅₀ values towards those determined at the gpH₂R. The same tendency was observed for bivalent compounds tested recently at this hH₂R mutant (Birnkammer *et al.*, 2012). It was speculated about a direct or water-mediated hydrogen bond between Tyr17^{1.31} and Asp271^{7.35} in the gpH₂R, stabilizing the receptor in an active conformation. The absence of amino acids capable of forming an H-bond in the wild-type hH₂R could explain higher potencies at the gpH₂R and the hH₂R Ala271->Asp271 mutant. Nevertheless, Asp271^{7.35} is in close vicinity to the orthosteric binding site. Hence, a direct or at least an indirect interaction with ligands cannot be excluded, but an interaction with the second pharmacophore of bivalent agonists is unlikely.

Remaining acidic amino acids at the extracellular surface of the H₂R are located in ECL3 and the junction of ECL3 and TM7. Asp262, Asp263, Glu267 and Glu270 are present in both, the human and guinea pig H₂R. These amino acids are in a suitable distance to the orthosteric binding site to potentially explain the dependency of pEC₅₀ values of bivalent ligands on spacer length. Consequently, as a working hypothesis, these four acidic residues in ECL3 were considered the most promising candidates for an accessory binding site. Figure 7.5 shows a possible docking mode of UR-AK381 in the gpH₂R homology model. The first pharmacophore is hydrogen bonded to Asp98^{3.32} and Asp186^{5.42} in TM3 and TM5, respectively. The second acylguanidine moiety forms hydrogen bonds with Asp262, Asp263 and Glu267 in ECL3. Additionally, the heterocycle of the second pharmacophoric unit is able to interact with Val178^{ECL2} and Glu180^{5.36}.

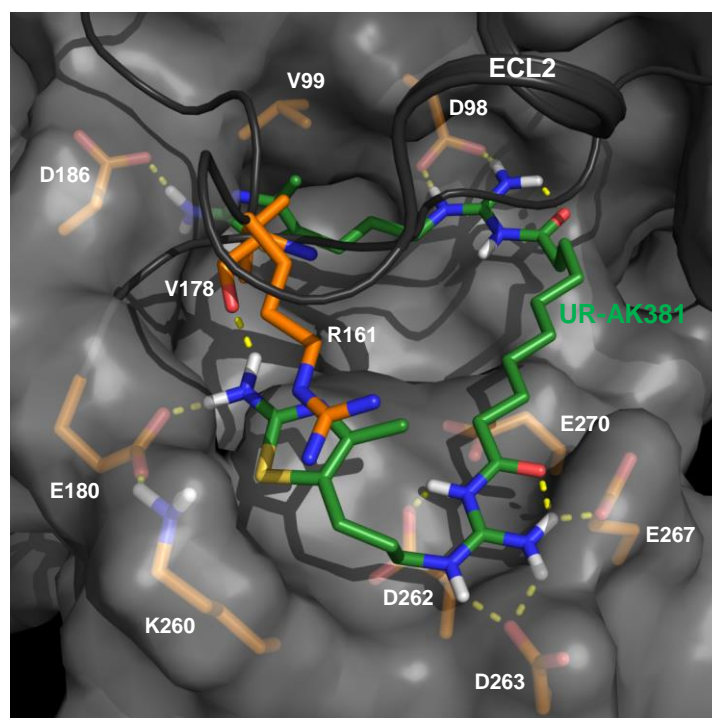


Figure 7.5: Possible binding mode of UR-AK381 in the gpH₂R homology model

View from the extracellular side at the gpH₂R. The surface of the receptor except ECL2 (ribbon and tube in dark grey) is shown with a grey transparent surface. Amino acids (carbon atoms in orange) close to the ligand UR-AK381 (carbon atoms in green) are shown. Hydrogen bonds are indicated with yellow dashed lines. The ligand was manually docked into the binding site. The binding mode of the second pharmacophore is based on physicochemical properties as well as the certain distance between the aspartate/glutamate residues of ECL3 and the anchor residue Asp98^{3,32}.

7.3.2 Multiple sequence alignment analysis of the third extracellular loop of aminergic GPCRs

The analysis of 340 aminergic GPCRs revealed a similar length of ECL3 with almost 80% of the receptors comprising 34 to 36 amino acids between the highly conserved prolines in TM6 and TM7. However, due to high sequence diversity, a common structural motif could not be identified (Table 7.3). Most conserved are residues Cys^{6.61} (46%), Ser^{6.64} (31%), Cys^{6.65} (80%), Pro^{7.31} (43%), Leu^{7.34} (46%) and Phe^{7.35} (40%). Sorting amino acids according to structural and physicochemical properties allowed the detection of further conserved sites. Charged amino acids are preferred in position 6.62, 6.63, 7.32 and 7.36 (38%, 45%, 32% and 48%, respectively). Furthermore, in position 6.64 tiny residues (49%), in position 7.34 hydrophobic amino acids (87%), especially aliphatic ones (77%) and in position 7.35 aromatic residues (59%) are favored.

Table 7.3: Sequence analysis of the extracellular loop 3 of 340 aminergic GPCRs¹

Position ²	6.61	6.62	6.63	6.64	6.65	7.27	7.28	7.29	7.30	7.31	7.32	7.33	7.34	7.35	7.36
hH ₂ R	R	G	D	-	-	-	D	A	I	N	E	V	L	E	A
gpH ₂ R	K	G	D	-	-	-	D	A	V	N	E	V	F	E	D
Amino acids ³	97	90	62	46	44	47	90	94	95	79	80	96	96	94	81
hydrophobic ⁴	39	31	23	30	4	40	52	53	57	66	47	51	87	67	24
polar ⁴	61	69	77	70	96	60	48	47	43	34	53	49	13	33	76
small ⁴	54	67	58	70	89	59	62	46	57	69	46	52	34	23	61
proline	0	11	7	4	1	4	16	15	13	43	10	4	0	0	0
tiny ⁴	6	23	22	49	6	8	7	8	5	12	18	25	5	15	22
aliphatic ⁴	12	0	2	8	2	24	21	28	34	19	7	17	77	12	15
aromatic ⁴	22	4	2	13	1	12	19	12	12	13	18	15	6	59	0
positive ⁴	6	19	13	8	1	12	2	12	3	1	6	8	2	2	15
negative ⁴	1	19	32	6	7	9	9	10	5	7	27	9	2	5	33
charged ⁴	7	38	45	13	8	21	11	22	8	8	32	17	3	8	48
Ala	1	3	8	15	1	3	0	2	1	1	7	13	3	1	7
Arg	4	9	2	4	1	2	0	1	3	1	0	5	2	0	3
Asn	1	4	1	4	1	19	6	3	0	3	1	3	0	2	11
Asp	1	12	20	5	1	8	5	4	2	5	9	1	0	0	22
Cys	46	15	2	3	80	9	15	7	0	0	0	0	1	0	0
Gln	2	3	13	1	1	3	3	1	1	0	3	2	1	1	2
Glu	0	8	11	1	5	1	4	6	3	2	18	8	1	5	11
Gly	5	17	4	3	0	1	0	1	0	0	6	9	0	13	1
His	3	4	0	10	1	4	1	4	6	8	4	2	0	1	0
Ile	3	0	1	3	0	12	5	11	16	5	1	5	10	2	5
Leu	8	0	0	4	2	1	5	12	3	7	6	5	46	8	4
Lys	2	9	11	4	0	10	2	11	0	0	6	3	0	2	12
Met	4	0	0	0	1	0	1	0	4	2	2	5	1	0	1
Phe	18	0	1	0	0	8	13	0	5	1	0	0	4	40	0
Pro	0	11	7	4	1	4	16	15	13	43	10	4	0	0	0
Ser	1	4	10	31	5	4	6	5	4	11	5	3	2	0	14
Thr	0	1	5	4	1	1	1	4	22	0	7	12	6	4	1
Trp	0	0	0	0	0	0	1	7	0	0	15	2	2	1	0
Tyr	1	0	0	3	0	0	4	0	0	3	0	10	0	18	0
Val	1	0	0	0	0	11	12	4	15	6	1	7	22	3	5

¹ Aminergic receptors of class A rhodopsin-like GPCRs (muscarinic acetylcholine, adrenergic, dopamine, histamine, serotonin, octopamine and trace amine receptors).

² Residues are named according to the Ballesteros/Weinstein nomenclature (Ballesteros and Weinstein, 1995). Between position 6.65 and 7.27, seven positions of the alignment are missing.

They are only present in 2-8% of the analyzed GPCRs, having a longer e3 loop. The borders of ECL3 were defined according to the transmembrane domains 6 and 7 in GPCR crystal structures and the presence of hydrophobic amino acids.

³ Percentage of amino acids in the respective position of the alignment.

⁴ Classification of amino acids according to Livingstone *et al.* (1993), i.e. small (Ala, Asn, Asp, Cys, Gly, Pro, Ser, Thr, Val), tiny (Ala, Gly, Ser), aliphatic (Ile, Leu, Val), aromatic (His, Phe, Trp, Tyr) and negatively charged (Asp, Glu) amino acids. Additionally, residues were classified as hydrophobic (Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val), polar (Arg, Asn, Asp, Cys, Gln, Glu, His, Lys, Ser, Thr, Tyr), positively charged (Arg, Lys) and charged (Arg, Asp, Glu, Lys).

Most GPCRs contain 40% to 60% polar residues in ECL3 (Figure 7.6 A). Less than 20% of this loop are charged residues in the majority of GPCRs. However, there are extreme examples at both sides. Whereas in some aminergic GPCRs ECL3 consists only of hydrophobic amino acids, there are GPCRs containing a high number of charged residues.

Considering that aminergic GPCRs bind to ligands which are positively charged at physiological pH, such as catecholamines, histamine and 5-hydroxytryptamine, it is remarkable that the third extracellular loops differ strongly in the content of negatively charged amino acids (Figure 7.6 B). The gpH₂R, which is among the GPCRs with the highest amount (42%) of acidic residues in ECL3, was selected to explore the importance of these negatively charged amino acids for ligand attraction, ligand binding and receptor activation.

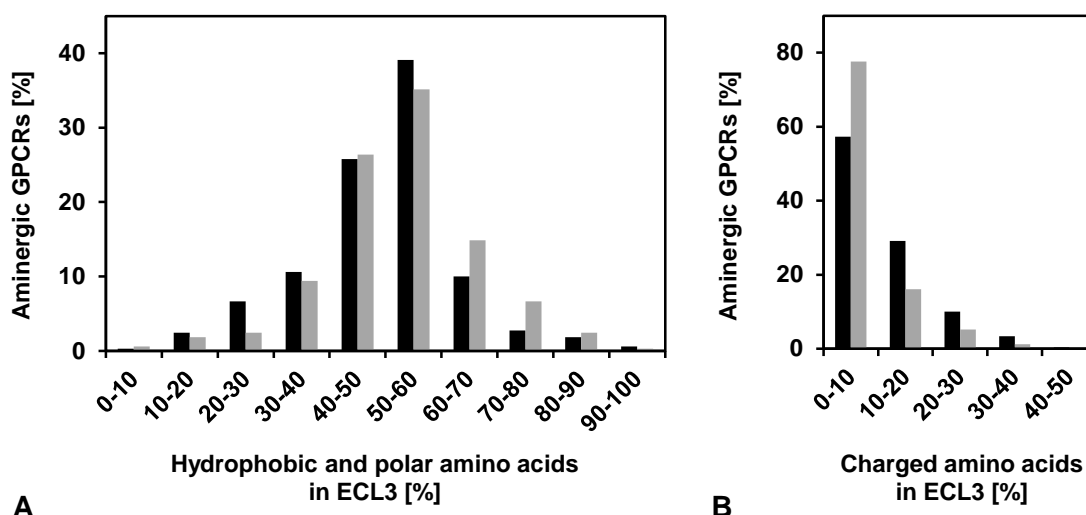


Figure 7.6: Hydrophobic, polar and charged amino acids in ECL3 of aminergic GPCRs

Shown is the percentage of 340 GPCRs that possess 0-10% to 90-100% (A) hydrophobic (black) and polar (grey) amino acids, or (B) negatively (black) and positively (grey) charged amino acids. The classification of amino acids is described in Table 7.3.

7.3.3 Immunological detection of recombinant proteins in Sf9 cell membranes

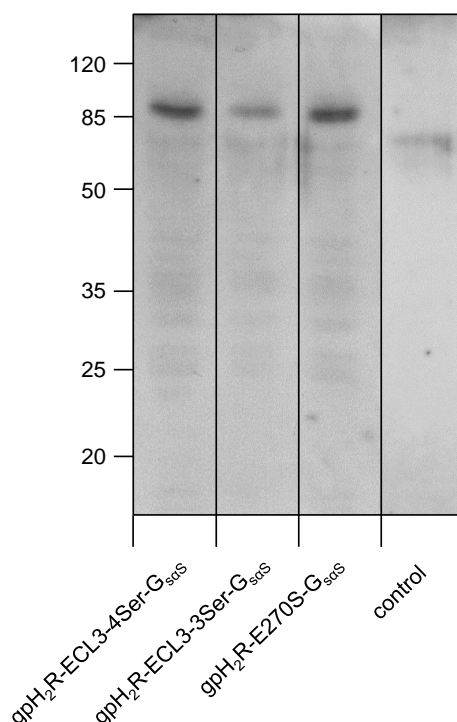


Figure 7.7: Immunological detection of the expression of recombinant proteins in Sf9 cells

Membranes of Sf9 cells expressing the three gpH₂R-G_{saS} mutants and membranes of uninfected Sf9 cells (control) were prepared, separated by SDS-PAGE on gels containing 12% (w/v) acrylamide, transferred onto nitrocellulose membranes and probed with the anti-FLAG (M1) antibody. In each lane, 7.5 µg of membrane protein was loaded onto the gel. Labels on the left designate masses of marker proteins in kDa.

The mutant receptors gpH₂R-ECL3-4Ser-G_{saS} and gpH₂R-E270S-G_{saS} were well expressed in Sf9 cells (Figure 7.7). The band for gpH₂R-ECL3-3Ser-G_{saS} appears weaker, possibly because of a lower amount of receptor protein in the membranes. The calculated molecular mass of the gpH₂R fusion protein is 88.2 kDa, neglecting the glycosylation of the protein. Bands for fusion proteins separated by SDS-PAGE on gels containing 12% (w/v) acrylamide were detected at about 80kDa (Houston *et al.*, 2002; Kelley *et al.*, 2001; Preuss *et al.*, 2007b; c). SDS-PAGE analysis of membranes expressing the gpH₂R-G_{saS} mutants yielded intense bands at 85-90 kDa, recognized by the anti-FLAG antibody. This is in good agreement with the result obtained previously. A slight band observed in uninfected Sf9 membranes at 80-85 kDa is also present in membranes containing the recombinant gpH₂R-G_{saS} proteins. In the control there is no protein detectable at 85-90 kDa, which could pretend the presence of the fusion proteins.

7.3.4 Intrinsic activities and potencies at the wild-type and mutant gpH₂Rs

Standard H₂R ligands (HIS, AMT, DIM, IMP), mono- (BIT24) and bivalent (UR-AK381, UR-BIT106) *N*⁶-acylated 3-(2-aminothiazol-5-yl)propylguanidines were tested in the [³⁵S]GTPγS binding assay at the wild-type and the three mutated gpH₂Rs (Table 7.4 and Figure 7.8).

Table 7.4: Intrinsic activities (*E*_{max}) and potencies at wild-type and three mutated gpH₂Rs in the [³⁵S]GTPγS assay

Cpd.	gpH ₂ R-G _{saS}		gpH ₂ R-ECL3-4Ser-G _{saS} ^b		gpH ₂ R-ECL3-3Ser-G _{saS} ^b		gpH ₂ R-E270S-G _{saS} ^b	
	pEC ₅₀	<i>E</i> _{max}	pEC ₅₀	<i>E</i> _{max}	pEC ₅₀	<i>E</i> _{max}	pEC ₅₀	<i>E</i> _{max}
	± SEM	± SEM ^a	± SEM	± SEM ^a	± SEM	± SEM ^a	± SEM	± SEM ^a
HIS	6.12 ± 0.01	1.00	4.14 ± 0.06 **	1.00	5.37 ± 0.08 **	1.00	6.03 ± 0.02 *	1.00
AMT	6.84 ± 0.06	0.99 ± 0.03	4.76 ± 0.12 **	0.98 ± 0.04	5.78 ± 0.09 **	1.02 ± 0.04	6.61 ± 0.05 *	1.00 ± 0.01
DIM	6.25 ± 0.02	0.94 ± 0.03	4.76 ± 0.12 **	0.45 ± 0.02 **	5.40 ± 0.05 **	0.76 ± 0.02 **	5.68 ± 0.02 **	0.87 ± 0.02
IMP	7.35 ± 0.07	0.93 ± 0.00	6.04 ± 0.03 **	0.93 ± 0.01	6.54 ± 0.05 **	0.96 ± 0.05	7.73 ± 0.05 *	1.03 ± 0.03 *
UR-BIT24	8.33 ± 0.12	0.83 ± 0.02	7.14 ± 0.06 **	0.78 ± 0.02	7.98 ± 0.06	0.79 ± 0.03	8.45 ± 0.05	0.94 ± 0.03 *
UR-AK381	9.29 ± 0.10	0.84 ± 0.05	7.84 ± 0.03 **	1.05 ± 0.02 *	8.86 ± 0.03 *	0.93 ± 0.02	9.15 ± 0.06	1.06 ± 0.02 *
UR-BIT106	8.90 ± 0.04	0.89 ± 0.03	7.87 ± 0.08 **	1.09 ± 0.02 **	8.78 ± 0.03	0.94 ± 0.01	9.24 ± 0.11 *	1.03 ± 0.04 *

[³⁵S]GTPγS binding was determined as described in section 6.2. Data shown are the means ± SEM of three to five experiments performed in triplicate. Intrinsic activities and potencies, respectively, of ligands at the mutated gpH₂Rs were compared to the corresponding parameters at gpH₂R-G_{saS}, using the t-test.

^a Intrinsic activity relative to the maximal response of histamine (*E*_{max} = 1.00).

^b Comparison with pEC₅₀/*E*_{max} at gpH₂R-G_{saS}; * *p* < 0.05; ** *p* < 0.01.

Compared to the wild-type gpH₂R, exchanging the four acidic amino acids Asp262, Asp263, Glu267 and Glu270 by serine, resulted in a strong decrease in potencies of all ligands tested ($p < 0.01$). Most pronounced decreases were observed for histamine and amthamine (96- and 120-fold, respectively), whereas the remaining ligands exhibited a 10- to 30-fold decrease.

To confirm the results obtained in the [³⁵S]GTPγS binding assay, histamine, amthamine and dimaprit were also investigated in the steady state GTPase assay at the mutant receptor (Table 7.5). No difference between both assays was observed for wild-type and mutant receptors, respectively. Compared to the wild-type gpH₂R-G_{saS}, pEC₅₀ values at the gpH₂R-ECL3-3Ser-G_{saS} were significantly reduced for standard H₂R ligands ($p < 0.01$) and UR-AK381 ($p < 0.05$). Potencies for UR-BIT24 and UR-BIT106 were reduced, though not significantly. Strikingly, EC₅₀ ratios of mutant to wild-type receptor were lowest for acylguanidines (1.3 to 2.7 compared to 5.6 to 11.4 for the remaining ligands). The mutation Glu270→Ser270 had only marginal effects on pEC₅₀ values (potency ratios 0.5 to 1.7) for all ligands except in the case of dimaprit (3.8-fold decrease, $p < 0.01$). IMP ($p < 0.05$), UR-BIT24 and UR-BIT106 ($p < 0.05$) showed slightly increased pEC₅₀ values compared to the wild-type receptor.

Table 7.5: Intrinsic activities (E_{\max}) and potencies at gpH₂R-G_{saS} and gpH₂R-ECL3-4Ser-G_{saS} in the GTPase assay and comparison to the [³⁵S]GTPγS assay

Cpd.	GTPase activity assay				[³⁵ S]GTPγS assay			
	gpH ₂ R-G _{saS} ^a		gpH ₂ R-ECL3-4Ser-G _{saS} ^c		gpH ₂ R-G _{saS} ^c		gpH ₂ R-ECL3-4Ser-G _{saS} ^{d,e}	
	pEC ₅₀	E_{\max}	pEC ₅₀	E_{\max}	pEC ₅₀	E_{\max}	pEC ₅₀	E_{\max}
	± SEM	± SEM ^b	± SEM	± SEM ^b	± SEM	± SEM ^b	± SEM	± SEM ^b
HIS	6.10 ± 0.08	1.00	4.26 ± 0.04 **	1.00	6.12 ± 0.01	1.00	4.14 ± 0.06 ##	1.00
AMT	6.74 ± 0.06	1.04 ± 0.00	4.71 ± 0.05 **	1.07 ± 0.05	6.84 ± 0.06	0.99 ± 0.03	4.76 ± 0.12 ##	0.98 ± 0.04
DIM	6.16 ± 0.10	0.94 ± 0.03	4.92 ± 0.05 **	0.44 ± 0.02 **	6.25 ± 0.02	0.94 ± 0.03	4.76 ± 0.12 ##	0.45 ± 0.02 ##

GTPase assay: GTPase activity in Sf9 membranes was determined as described in section 6.2. Data shown are the means ± SEM of three to five experiments performed in duplicate. GTPγS assay: see Table 7.4. Comparisons of pEC₅₀ and E_{\max} , respectively, were performed by t-tests.

^a Data at gpH₂R-G_{saS} in the GTPase assay were taken from Preuss *et al.* (2007a).

^b Intrinsic activity relative to the maximal response of histamine ($E_{\max} = 1.00$).

^c Comparison with pEC₅₀/ E_{\max} at gpH₂R-G_{saS} (GTPase assay); * $p < 0.05$; ** $p < 0.01$.

^d Comparison with pEC₅₀/ E_{\max} at gpH₂R-ECL3-4Ser-G_{saS} (GTPase assay); + $p < 0.05$; ++ $p < 0.01$.

^e Comparison with pEC₅₀/ E_{\max} at gpH₂R-G_{saS} (GTPγS assay); # $p < 0.05$; ## $p < 0.01$.

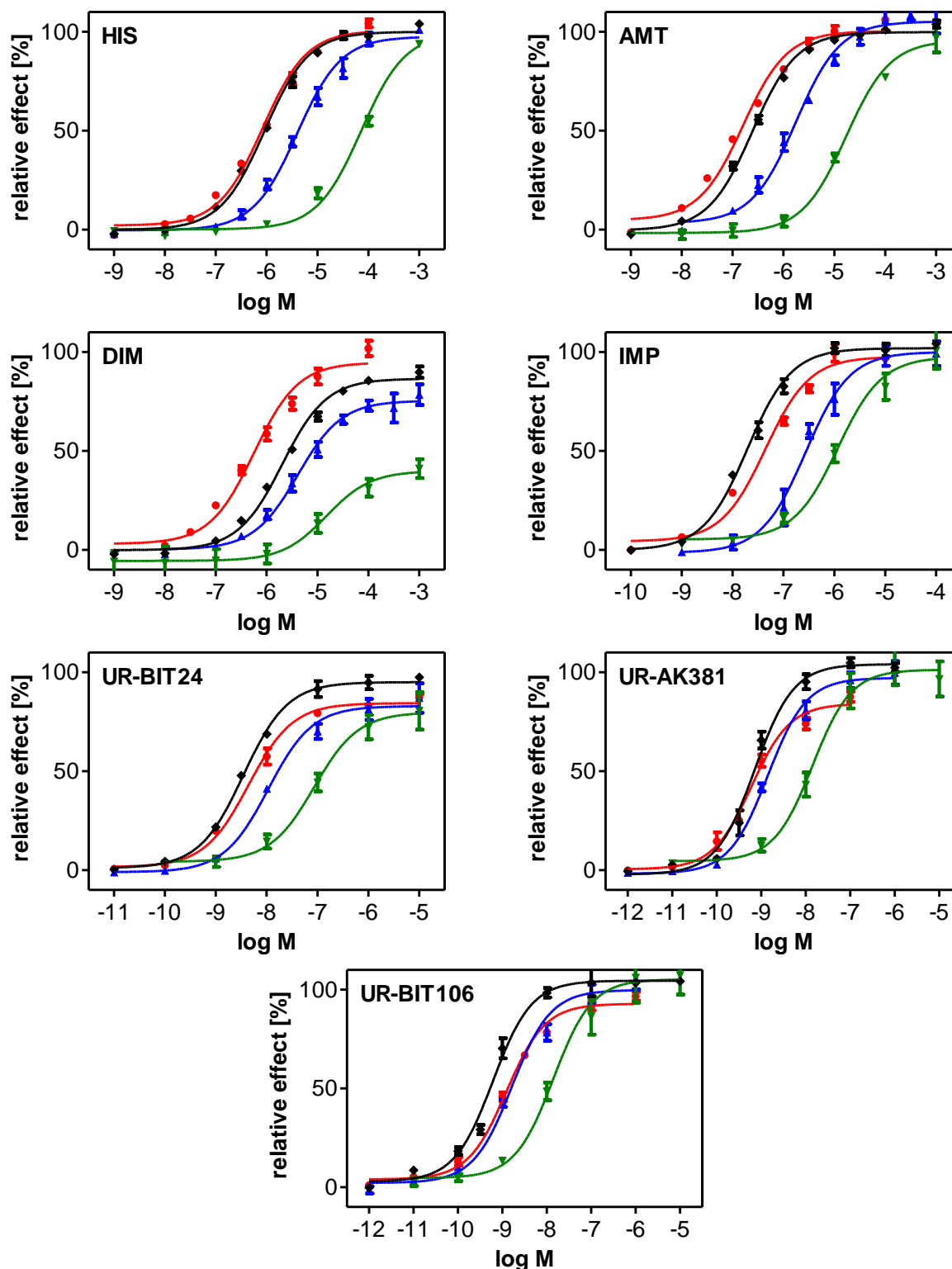


Figure 7.8: Concentration-response curves obtained in $[^{35}\text{S}]\text{GTP}\gamma\text{S}$ binding assays

$[^{35}\text{S}]\text{GTP}\gamma\text{S}$ binding in membranes of Sf9 cells expressing gpH₂R-G_{sa}S (●), gpH₂R-ECL3-4Ser-G_{sa}S (▼), gpH₂R-ECL3-3Ser-G_{sa}S (▲) and gpH₂R-E270S-G_{sa}S (◆) was determined as described in section 6.2. Reaction mixtures contained membranes (10 μg of protein/well) of Sf9 cells expressing fusion proteins and the respective ligand at concentrations indicated on the abscissa. Data shown are the means \pm SEM of three to five independent experiments performed in triplicates, expressed as percentage of $[^{35}\text{S}]\text{GTP}\gamma\text{S}$ binding relative to the maximum effect induced by histamine (set to 100%). Data were analyzed by nonlinear regression and were best fitted to sigmoidal concentration-response curves.

Significant ($p < 0.01$) decreases in intrinsic activity were only observed for dimaprit at gpH₂R-ECL3-4Ser-G_{saS} and gpH₂R-ECL3-3Ser-G_{saS}. At the gpH₂R-E270S-G_{saS} the intrinsic activities of all (acyl-)guanidines were increased ($p < 0.05$). The bivalent agonists UR-AK381 and UR-BIT106 showed also significantly increased E_{\max} values at gpH₂R-ECL3-4Ser-G_{saS} ($p < 0.05$ and $p < 0.01$, respectively).

7.4 Discussion

7.4.1 Second binding site for bivalent acylguanidine-type agonists and the role of Glu270 in the gpH₂R

Replacing four acidic amino acids by serine residues in ECL3 of the gpH₂R-G_{saS} resulted in reduced pEC₅₀ values of all investigated agonists. The decrease was most pronounced for HIS and AMT. No significant differences were observed between mono- and bivalent compounds. Molecular modeling studies revealed that Glu270^{7.35}, which was previously supposed to be part of the extracellular surface of ECL3 (Figure 7.5), is able to rotate towards the orthosteric binding site. The closest distance from the heteroatoms in the Glu270^{7.35} side chain to the acylguanidine moiety of the first pharmacophore and to Asp98^{3.32} is reduced from 6.7 Å to 3.8 Å, and from 10 Å to 8 Å, respectively. Hence, a contribution of Glu270^{7.35} to ligand binding within the orthosteric binding site cannot be excluded. In case of the β_2 AR, mutation of Tyr^{7.35} to alanine revealed a contribution of this residue to high affinity binding of agonists (Isogaya *et al.*, 1998; Kikkawa *et al.*, 1998). Furthermore, in crystal structures of the β_2 AR (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011b) the complexed ligands are in proximity to Tyr^{7.35}. To explore if the reduced potencies of standard H₂R agonists were caused by the missing glutamate in position 7.35, and if the accessory binding site of bivalent acylguanidines is composed only of Asp262, Asp263 and Glu267, two further gpH₂R-G_{saS} mutants were generated. The bivalent agonists as well as the monovalent ligand UR-BIT24 were similarly potent at the gpH₂R-ECL3-3Ser-G_{saS} and the wild-type receptor. Again the potency of the standard H₂R agonists histamine, dimaprit, amthamine and impromidine was significantly lower than at the wild-type gpH₂R-G_{saS}. However, the pEC₅₀ values decreased only about midway compared to the reductions in the case of the 4-Ser mutant. The results do not support the hypothesis that negatively charged residues in ECL3 form an accessory binding site for bivalent acylguanidine-type agonists. Alternatively, investigating Glu163 in ECL2 may be worthwhile. However, the prediction of a second binding site is complicated by the uncertain impact of the highly flexible N-terminus as part of the extracellular surface. In addition, the second pharmacophore of various bivalent

acylguanidines possibly stabilizes different conformations of the extracellular surface by interacting with distinct regions. This ligand specific behavior was partly observed at gpH₂R mutants with an altered N-terminus or ECL2 (Birnkammer *et al.*, 2012; Brunskole, 2011). Last but not least, supposing that depot formation of a lipophilic ligand in the lipid bilayer occurs as suggested by Portoghese (2001), the agonist concentration close to the extracellular surface of the receptor and therefore near the entrance to the binding site could be raised compared to the aqueous phase. For long acting β_2 -adrenoceptor agonists including a lipophilic moiety such as salmeterol, the existence of a specific 'exosite' or a membrane deposition (micro-kinetic theory) of the ligands is discussed (Coleman *et al.*, 1996; Patel *et al.*, 2011).

Contrary to the initial assumption, the replacement of Glu270^{7.35} did not affect receptor activation of the gpH₂R in general. Only the potency of DIM was sensitive to this mutation. DIM is an outlier in the present series since about 87% of the molecules are twofold positively charged at pH 7.4 (Durant *et al.*, 1977). Possibly, the high affinity of DIM at the wild-type gpH₂R-G_{saS} depends on the negatively charged residue in position 7.35, situated at the entrance of the binding pocket and close to the highly conserved Asp98^{3.32}, well known to interact with positively charged moieties of ligands.

7.4.2 Integrity of the orthosteric binding site

Amino acids in ECL3 are able to interact with the facing e2 loop. In the β_2 AR crystal structures, close contacts between Lys305^{7.32} in ECL3 and Asp192^{ECL2} or Phe193^{ECL2}, respectively, were observed (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2011a). Likewise, in the β_1 AR Arg317^{ECL3} interacts with Thr203^{ECL2} (Warne *et al.*, 2011; Warne *et al.*, 2008). Replacement of charged amino acids in ECL3, changing its electrostatic potential (Figure 7.9), could affect the contact to ECL2 and therefore alter the ECL2 conformation. ECL2 is in close contact to the orthosteric binding site and part of the binding pocket in several GPCRs (Peeters *et al.*, 2011). Furthermore, ECL3 constrains the extracellular parts of TM6 and 7. In a previous study on the delta opioid receptor, amino acids in ECL3 were suggested to restrain motions of TM6 and 7 on the extracellular side. Mutations may weaken this structural constraint and thereby affect the activation mechanism (Decaillot *et al.*, 2003). Moreover, amino acids of TM6 are part of the ligand binding site of several GPCRs, presumably including the H₂R (chapter 3). Therefore, the decreased potency of the ligands at the 4-Ser and 3-Ser mutants could result from a distorted orthosteric binding pocket, in particular from an altered distance between the aspartates in TM3 and TM5. The contraction of the binding site is discussed to be one of the most important roles of agonists in the activation process, observed in the β_1 - and β_2 -adrenoceptor (Rasmussen *et al.*, 2011a; Warne *et al.*, 2011). In

an extended conformation, the distance (6.1 Å and 7.7 Å) between the two polar centers in HIS and AMT, i.e. the heterocyclic nitrogen and the positively charged side chain amino group, is the smallest among all ligands. Strikingly, the greatest decrease in potency at the mutant compared to the wild-type H₂R was observed for histamine and amthamine. The distance between the basic centers is larger for IMP and the three acylguanidines UR-AK381, UR-BIT24 and UR-BIT106 (~ 10.0 Å). Possibly, these compounds tolerate a certain degree of 'deformation' of the binding site. In the case of DIM, the maximal distance between the polar heteroatoms (about 7.8 Å) is similar to HIS and AMT. However, DIM is the ligand that carries the highest charge per molecule. At pH 7.4, 87% of DIM is present as the dication and 5% as the monocation (Durant *et al.*, 1977). Possibly, the ability to form ionic interactions with both Asp98^{3,32} and Asp186^{5,42} allows a specific binding mode between TM3 and TM5. In line with this idea, acidic amino acids in ECL3 of the gpH₂R are necessary for a proper folding of the binding pocket. The decrease in potency at the gpH₂R-ECL3-3Ser-G_{saS} and the gpH₂R-ECL3-4Ser-G_{saS} mutants compared to the wild-type receptor-G_{saS} fusion protein appears to reflect the deformation of the binding site, depending on the number of replaced negatively charged residues.

7.4.3 Path of the ligand into the binding site

A further possible reason for reduced potencies at the modified gpH₂R_s is an altered recognition process of the ligands. Acidic amino acids in ECL3 potentially line the path of the ligand into the binding pocket. Asp300 in the e3 loop of the β₂AR, which corresponds to Asp262 in the gpH₂R, was shown to form a salt bridge to the ammonium group of alprenolol during its route into the binding pocket of the β₂AR in an unbiased molecular dynamic simulation on a microsecond timescale (Dror *et al.*, 2011). Glu448 in ECL3 of the gpH₁R, equivalent to Ala264 in the gpH₂R next to the mutated aspartates 262 and 263, was suggested to be relevant as a structure-recognition system for histaprodifen (Straßer and Wittmann, 2007). Acylguanidines, exhibiting a pK_a value of 7 to 8 (Ghorai *et al.*, 2008; Rewinkel and Adang, 1999; Schmuck and Lex, 1999), are only partly protonated at physiological pH. Remarkably, these compounds revealed the least decrease in potency at gpH₂R-ECL3-4Ser-G_{saS} and gpH₂R-ECL3-3Ser-G_{saS}. Obviously, these H₂R agonists tolerate the replacement of acidic amino acids better than the other ligands, which are charged at a higher percentage at pH 7.4. DIM is mainly twofold positively charged (87%). The predominant ionic species of IMP is likely to be the monocation (65%). Also the dication (33%) and trication (2%) are present in equilibrium (Durant *et al.*, 1985). At pH 7.4 the predominant form (almost 100%) of HIS and AMT is the monocation (Eriks *et al.*, 1993). Possibly, the route of these ligands into the binding pocket of the wild-type gpH₂R-G_{saS} is

more strongly dependent on ionic interactions with the extracellular receptor surface (Figure 7.9). In contrast, the dependence on negatively charged residues at the extracellular surface is less pronounced for acylguanidines. Furthermore, larger molecules such as acylguanidine-type H₂R agonists can form additional contacts apart from ionic interactions.

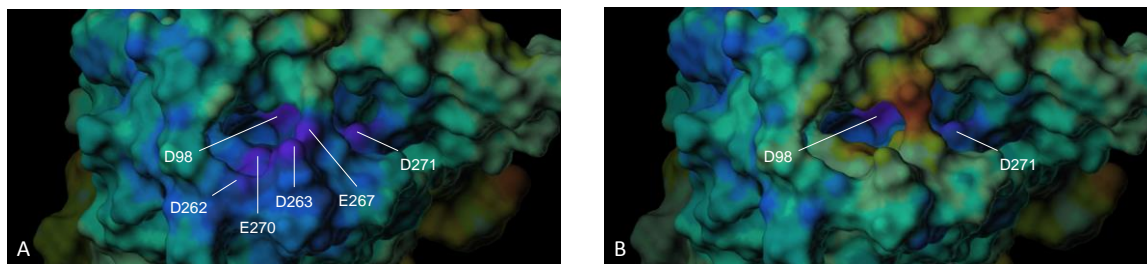


Figure 7.9 Surface of the gpH₂R and gpH₂R-ECL3-4Ser with the electrostatic potential

View from the extracellular side at the surfaces of the wild-type gpH₂R (A) and gpH₂R-ECL3-4Ser (B). The surface is colored according to the electrostatic potential, ranging from red (most positive) to purple (most negative). The electrostatic potential was calculated with the program MOLCAD (MOlecular Computer Aided Design), included in the modeling suite SYBYL X1.3 (Tripos, St. Louis, MO, USA).

7.4.4 Summary and conclusion

The aim of this study was to investigate the role of the negatively charged amino acids in the third extracellular loop of the gpH₂R regarding their general function in receptor activation and their potential to provide an accessory binding site for bivalent acylguanidine-type agonists. Three receptor mutants, gpH₂R-ECL3-4Ser-G_{sa}S, gpH₂R-ECL3-3Ser-G_{sa}S and gpH₂R-E270S-G_{sa}S, were generated, and standard H₂R agonists, mono- and bivalent (*N*⁶-acylated) hetarylpropylguanidines were characterized in the [³⁵S]GTPγS binding assay. The suggested second binding site located at ECL3 and composed of negatively charged residues was not confirmed. However, these residues are possibly responsible for the integrity of the orthosteric binding site and/or for the route of some ligands into the binding pocket. The particularly high percentage of negatively charged amino acids in ECL3 of the gpH₂R seems not to be generally needed for a proper function of the receptor. Some ligands (UR-BIT24, UR-BIT106) even tolerated the removal of three acidic residues. However, for all tested compounds the presence of at least one negatively charged amino acid in ECL3 proved to be essential to retain high agonistic potency. Further site directed mutagenesis studies on ECL3 of the gpH₂R may help to unveil the role of this loop in aminergic GPCRs.

7.5 References

- Ballesteros JA and Weinstein H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In *Methods Neurosci*, Stuart CS, Ed. Academic Press: **1995**; Vol. Volume 25, pp 366-428.
- Birnkammer T. Highly potent and selective acylguanidine-type histamine H₂ receptor agonists: synthesis and structure-activity relationships of mono- and bivalent ligands. Doctoral thesis, Regensburg, Regensburg, **2011**. <http://epub.uni-regensburg.de/22237/>
- Birnkammer T, Spickenreither A, Brunskole I, Lopuch M, Kagermeier N, Bernhardt G, Dove S, Seifert R, Elz S and Buschauer A. The Bivalent Ligand Approach Leads to Highly Potent and Selective Acylguanidine-Type Histamine H-2 Receptor Agonists. *J Med Chem* **2012**, 55, 1147-1160.
- Brunskole I. Molecular and Cellular Analysis of Aminergic G Protein-Coupled Receptors: Histamine H₂, H₄ and β 2-Adrenergic Receptors, a Scientific Paradigm. Doctoral thesis, Regensburg, Regensburg, **2011**. <http://epub.uni-regensburg.de/22963/>
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK and Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, 318, 1258-1265.
- Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V and Stevens RC. Structure of the human dopamine D₃ receptor in complex with a D₂/D₃ selective antagonist. *Science* **2010**, 330, 1091-1095.
- Claus M, Jaeschke H, Kleinau G, Neumann S, Krause G and Paschke R. A hydrophobic cluster in the center of the third extracellular loop is important for thyrotropin receptor signaling. *Endocrinology* **2005**, 146, 5197-5203.
- Coleman RA, Johnson M, Nials AT and Vardey CJ. Exosites: their current status, and their relevance to the duration of action of long-acting beta 2-adrenoceptor agonists. *Trends Pharmacol Sci* **1996**, 17, 324-330.
- Decaillot FM, Befort K, Filliol D, Yue S, Walker P and Kieffer BL. Opioid receptor random mutagenesis reveals a mechanism for G protein-coupled receptor activation. *Nat Struct Biol* **2003**, 10, 629-636.
- Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H and Shaw DE. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* **2011**.
- Durant GJ, Duncan WA, Ganellin CR, Parsons ME, Blakemore RC and Rasmussen AC. Impromidine (SK&F 92676) is a very potent and specific agonist for histamine H₂ receptors. *Nature* **1978**, 276, 403-405.
- Durant GJ, Ganellin CR, Hills DW, Miles PD, Parsons ME, Pepper ES and White GR. The histamine H₂-receptor agonist impromidine: synthesis and structure activity considerations. *J Med Chem* **1985**, 28, 1414-1422.
- Durant GJ, Ganellin CR and Parsons ME. Dimaprit, [S-[3-(N,N-dimethylamino)propyl]isothiourea]. A highly specific histamine H₂-receptor agonist. Part 2. Structure-activity considerations. *Inflamm Res* **1977**, 7, 39-43.
- Eriks JC, van der Goot H and Timmerman H. New activation model for the histamine H₂ receptor, explaining the activity of the different classes of histamine H₂ receptor agonists. *Mol Pharmacol* **1993**, 44, 886-894.
- Gantz I, DelValle J, Wang LD, Tashiro T, Munzert G, Guo YJ, Konda Y and Yamada T. Molecular basis for the interaction of histamine with the histamine H₂ receptor. *J Biol Chem* **1992**, 267, 20840-20843.
- Ghorai P, Kraus A, Keller M, Götte C, Igel P, Schneider E, Schnell D, Bernhardt Gn, Dove S, Zabel M, Elz S, Seifert R, et al. Acylguanidines as Bioisosteres of Guanidines: NG-

- Acylated Imidazolylpropylguanidines, a New Class of Histamine H₂ Receptor Agonists. *J Med Chem* **2008**, 51, 7193-7204.
- Gregory KJ, Hall NE, Tobin AB, Sexton PM and Christopoulos A. Identification of orthosteric and allosteric site mutations in M₂ muscarinic acetylcholine receptors that contribute to ligand-selective signaling bias. *J Biol Chem* **2010**, 285, 7459-7474.
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T and Kobayashi T. Structure of the human M₂ muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012**, 482, 547-551.
- Harterich S, Koschätzky S, Einsiedel J and Gmeiner P. Novel insights into GPCR-peptide interactions: mutations in extracellular loop 1, ligand backbone methylations and molecular modeling of neurotensin receptor 1. *Bioorg Med Chem* **2008**, 16, 9359-9368.
- Hawtin SR, Simms J, Conner M, Lawson Z, Parslow RA, Trim J, Sheppard A and Wheatley M. Charged extracellular residues, conserved throughout a G-protein-coupled receptor family, are required for ligand binding, receptor activation, and cell-surface expression. *J Biol Chem* **2006**, 281, 38478-38488.
- Houston C, Wenzel-Seifert K, Burckstummer T and Seifert R. The human histamine H₂-receptor couples more efficiently to Sf9 insect cell Gs-proteins than to insect cell Gq-proteins: limitations of Sf9 cells for the analysis of receptor/Gq-protein coupling. *J Neurochem* **2002**, 80, 678-696.
- Huang X-P, Williams FE, Peseckis SM and Messer WS. Differential Modulation of Agonist Potency and Receptor Coupling by Mutations of Ser388Tyr and Thr389Pro at the Junction of Transmembrane Domain VI and the Third Extracellular Loop of Human M₁ Muscarinic Acetylcholine Receptors. *Mol Pharmacol* **1999**, 56, 775-783.
- Isogaya M, Yamagiwa Y, Fujita S, Sugimoto Y, Nagao T and Kurose H. Identification of a key amino acid of the beta₂-adrenergic receptor for high affinity binding of salmeterol. *Mol Pharmacol* **1998**, 54, 616-622.
- Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP and Stevens RC. The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **2008**, 322, 1211-1217.
- Jager D, Schmalenbach C, Prilla S, Schrobang J, Kebig A, Sennwitz M, Heller E, Trankle C, Holzgrabe U, Holtje HD and Mohr K. Allosteric small molecules unveil a role of an extracellular E2/transmembrane helix 7 junction for G protein-coupled receptor activation. *J Biol Chem* **2007**, 282, 34968-34976.
- Kelley MT, Burckstummer T, Wenzel-Seifert K, Dove S, Buschauer A and Seifert R. Distinct interaction of human and guinea pig histamine H₂-receptor with guanidine-type agonists. *Mol Pharmacol* **2001**, 60, 1210-1225.
- Kikkawa H, Isogaya M, Nagao T and Kurose H. The role of the seventh transmembrane region in high affinity binding of a beta₂-selective agonist TA-2005. *Mol Pharmacol* **1998**, 53, 128-134.
- Kraus A, Ghorai P, Birnkammer T, Schnell D, Elz S, Seifert R, Dove S, Bernhardt G and Buschauer A. N-G-Acylated Aminothiazolylpropylguanidines as Potent and Selective Histamine H₂ Receptor Agonists. *ChemMedChem* **2009**, 4, 232-240.
- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, et al. Structure and dynamics of the M₃ muscarinic acetylcholine receptor. *Nature* **2012**, 482, 552-556.
- Lawson Z and Wheatley M. The third extracellular loop of G-protein-coupled receptors: more than just a linker between two important transmembrane helices. *Biochem Soc Trans* **2004**, 32, 1048-1050.
- Livingstone CD and Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* **1993**, 9, 745-756.
- Mansour A, Meng F, Meador-Woodruff JH, Taylor LP, Civelli O and Akil H. Site-directed mutagenesis of the human dopamine D₂ receptor. *Eur J Pharmacol* **1992**, 227, 205-214.

- Nawaratne V, Leach K, Felder CC, Sexton PM and Christopoulos A. Structural determinants of allosteric agonism and modulation at the M4 muscarinic acetylcholine receptor: identification of ligand-specific and global activation mechanisms. *J Biol Chem* **2010**, 285, 19012-19021.
- Okada T, Sugihara M, Bondar AN, Elstner M, Entel P and Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* **2004**, 342, 571-583.
- Patel S, Summerhill S, Stanley M, Perros-Huguet C and Trevethick MA. The reassertion profiles of long acting beta2-adrenoceptor agonists in the guinea pig isolated trachea and human recombinant beta2-adrenoceptor. *Pulm Pharmacol Ther* **2011**, 24, 247-255.
- Peeters MC, van Westen GJ, Li Q and AP IJ. Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol Sci* **2011**, 32, 35-42.
- Peeters MC, Wisse LE, Dinaj A, Vroling B, Vriend G and Ijzerman AP. The role of the second and third extracellular loops of the adenosine A1 receptor in activation and allosteric modulation. *Biochem Pharmacol* **2012**, 84, 76-87.
- Portoghese PS. 2000 Alfred Burger Award Address in Medicinal Chemistry. From Models to Molecules: Opioid Receptor Dimers, Bivalent Ligands, and Selective Opioid Receptor Probes. *J Med Chem* **2001**, 44, 3758-3758.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Constitutive activity and ligand selectivity of human, guinea pig, rat, and canine histamine H-2 receptors. *J Pharmacol Exp Ther* **2007a**, 321, 983-995.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Mutations of Cys-17 and Ala-271 in the human histamine H-2 receptor determine the species selectivity of guanidine-type agonists and increase constitutive activity. *J Pharmacol Exp Ther* **2007b**, 321, 975-982.
- Preuss H, Ghorai P, Kraus A, Dove S, Buschauer A and Seifert R. Point mutations in the second extracellular loop of the histamine H-2 receptor do the species-selective activity of not affect guanidine-type agonists. *Naunyn Schmiedebergs Arch Pharmacol* **2007c**, 376, 253-264.
- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011a**, 469, 175-180.
- Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **2011b**, 477, 549-555.
- Rewinkel JB and Adang AE. Strategies and progress towards the ideal orally active thrombin inhibitor. *Curr Pharm Des* **1999**, 5, 1043-1075.
- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP and Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, 455, 497-U430.
- Schmuck C and Lex J. Acetate Binding within a Supramolecular Network Formed by a Guanidiniocarbonyl Pyrrole Cation in the Solid State. *Org Lett* **1999**, 1, 1779-1781.
- Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Rompler H and Sangkuhl K. Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther* **2004**, 104, 173-206.
- Schwarz RD, Spencer CJ, Jaen JC, Mirzadegan T, Moreland D, Tecle H and Thomas AJ. Mutations of aspartate 103 in the Hm2 receptor and alterations in receptor binding properties of muscarinic agonists. *Life Sci* **1995**, 56, 923-929.
- Shi L and Javitch JA. The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. *Proc Natl Acad Sci U S A* **2004**, 101, 440-445.

- Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, et al. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **2011**, 475, 65-70.
- Strader CD, Sigal IS, Candelore MR, Rands E, Hill WS and Dixon RA. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *J Biol Chem* **1988**, 263, 10267-10271.
- Straßer A and Wittmann H-J. LigPath: a module for predictive calculation of a ligand's pathway into a receptor-application to the gpH1-receptor. *J Mol Model* **2007**, 13, 209-218.
- Thompson MD, Percy ME, McIntyre Burnham W and Cole DE. G protein-coupled receptors disrupted in human genetic disease. *Methods Mol Biol* **2008**, 448, 109-137.
- Tikhonova IG and Costanzi S. Unraveling the structure and function of G protein-coupled receptors through NMR spectroscopy. *Curr Pharm Des* **2009**, 15, 4003-4016.
- Warne T, Moukhametzianov R, Baker JG, Nehme R, Edwards PC, Leslie AGW, Schertler GFX and Tate CG. The structural basis for agonist and partial agonist action on a [bgr]1-adrenergic receptor. *Nature* **2011**, 469, 241-244.
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG and Schertler GFX. Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* **2008**, 454, 486-U482.
- Wheatley M, Simms J, Hawtin SR, Wesley VJ, Wootten D, Conner M, Lawson Z, Conner AC, Baker A, Cashmore Y, Kendrick R and Parslow RA. Extracellular loops and ligand binding to a subfamily of Family A G-protein-coupled receptors. *Biochem Soc Trans* **2007**, 35, 717-720.
- Wheatley M, Wootten D, Conner MT, Simms J, Kendrick R, Logan RT, Poyner DR and Barwell J. Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **2012**, 165, 1688-1703.
- Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, 330, 1066-1071.
- Zhao MM, Gaivin RJ and Perez DM. The third extracellular loop of the beta2-adrenergic receptor can modulate receptor/G protein affinity. *Mol Pharmacol* **1998**, 53, 524-529.

Chapter 8

Summary

G-protein coupled receptors (GPCR) are the most important drug target family in humans. In the last decades, our knowledge about GPCR structure, conformational states, activation and their role in signal transduction has substantially increased. A huge number of novel agonists, antagonists and inverse agonists, including bivalent and functionally selective ligands, has been developed. More recently, crystal structures of about 20 GPCRs have been resolved. Structures of inactive and active receptor states are available now which allow invaluable insights into GPCR activation. The histamine H₂ receptor (H₂R) belongs to the majority of GPCRs with unknown 3D structure. However, its relatedness to the β -adrenoceptors (β AR) with resolved structures of both states makes the H₂R an interesting target for structural analysis.

The aim of this thesis was to investigate differences between inactive and active human (h) H₂R states with in-silico methods like homology modeling and molecular dynamics (MD) simulations. To improve and extend tools provided by the MD package GROMACS for analyzing structures and interactions, novel routines were written. Furthermore, in-vitro site-directed mutagenesis studies should answer open questions on the role of specific amino acids in the H₂R. Tyr182^{5,38} residing in the orthosteric binding site of the hH₂R was replaced by Phe. In a second approach, guinea pig (gp) H₂R mutants were generated replacing acidic amino acids in the third extracellular loop (ECL3) in order to investigate the possibility that they belong to a second binding site for bivalent hetarylpropylguanidine-type agonists with higher potency than their monovalent analogs. Moreover, the question was whether the high number of acidic amino acids in ECL3 of the gpH₂R compared to other aminergic GPCRs is of importance at all.

Homology models of inactive and active hH₂R states were constructed using the structures of the β_1 AR (inactive) and β_2 AR (active) as templates. With the intention to compare different models of active hH₂R states with respect to reliability and to select the 'best' model, another active hH₂R state variant was generated, using the templates t β_1 AR and opsin. Both active

state models were complexed with a C-terminal G_{sc}-protein fragment. The resulting models were embedded in a natural environment consisting of a 1,2-dipalmitoyl-sn-glycero-3-phosphocholine bilayer and water molecules as solvent. The models were subjected to 80 ns MD simulations performed with GROMACS. The analysis of the results was improved by novel routines for the systematic calculation of direct and water mediated H-bonds and of van der Waals contacts. Additionally, a structure validation tool for proteins in MD simulation systems was written, probing the stereochemistry at C_α atoms and the planarity of aromatic side chains, delocalized π-electron systems and peptide bonds. Furthermore, backbone Φ/Ψ dihedral angles (Ramachandran analysis) as well as side chain torsion were measured and compared to experimental reference values.

The quality of the homology models before and during the MD simulations in terms of backbone and side chain conformations was in very good agreement to experimental protein data. Crucial rearrangements between the inactive and the active hH₂R state especially in the cytoplasmic domain were similar to changes observed at crystal structures. All molecular switches suggested to be essential for receptor activation were connected to the pronounced outward move of TM6 upon receptor activation: the parallel rearrangement of TM5 and TM6, the cleavage of the ionic lock, the disruption of a hydrophobic barrier in the center of the TM bundle, the inward move of TM7 including the highly conserved NPxxY(x)_{5,6}F motif, the rearrangement of the H-bond network around TM7, and alterations at the bottom of the binding site. As key amino acids for the hH₂R activation Arg116^{3,50}, Tyr202^{5,58}, Glu229^{6,30}, Phe243^{6,44}, Asn280^{7,45} and Tyr288^{7,53} were identified.

Site-directed mutagenesis experiments were based on the expression of mutant cDNA in Sf9 cells using baculoviruses. Membranes containing the recombinant proteins were prepared. The H₂R mutants were characterized by testing H₂R agonists and antagonists in the GTPase and GTPγS assay.

Pharmacological investigations with the Tyr182Phe mutant of the hH₂R revealed that, unlike previously expected, the hydroxyl group of Tyr182^{5,38} has only a minor effect on ligand binding. Most ligands tested exhibited a similar potency (pEC₅₀) and intrinsic activity as at the wild-type hH₂R. The suggested second binding site in ECL3 for bivalent agonists was not confirmed by results on three gpH₂R mutants in which acidic amino acids were replaced by serines. However, acidic residues in ECL3 are possibly responsible for the integrity of the orthosteric binding site and/or for the pathway of some ligands into the binding pocket. The exceptionally high number of negatively charged amino acids in ECL3 of the gpH₂R seems not to be generally necessary for proper receptor function.

The theoretical analysis of hH₂R states contributes to the understanding of GPCR states and their differences. In particular, results from MD simulations indicate that GPCR activation is based on a concerted interaction of molecular switches and rearrangements of transmembrane and cytoplasmic domains. Computational tools have been introduced which allow an improved and extended analysis of MD simulations as well as a validation of simulated proteins. The pharmacological investigations do not verify hypotheses about a possible contribution of Tyr182^{5,38} to hH₂R ligand binding and about the putative role of acidic amino acids in ECL3 of the gpH₂R.

Chapter 9

Appendix

9.1 Parameters for MD simulations in GROMACS

Parameter	Value	Description ¹
RUN CONTROL PARAMETERS:		
integrator	md	A leap-frog algorithm for integrating Newton's equations of motion.
tinit	0	Starting time of the simulation [ps].
dt	0.002	Time step for integration [ps].
nsteps	40000000	Maximum number of steps to integrate.
comm-mode	Linear	Remove center of mass translation.
nstcomm	1	Frequency for center of mass motion removal [steps].
comm-grps	Protein_DPPC SOL_CL	Groups for center of mass motion removal.
OUTPUT CONTROL:		
nstxout	500	Frequency to write coordinates to output trajectory file [steps].
nstvout	500	Frequency to write velocities to output trajectory [steps].
nstfout	500	Frequency to write forces to output trajectory [steps].
nstlog	500	Frequency to write energies to log file [steps].
nstenergy	500	Frequency to write energies to energy file [steps].
nstxtcout	500	Frequency to write coordinates to xtc trajectory [steps].
xtc-precision	1000	Precision to write to xtc trajectory.
energygrps	Protein_DPPC SOL_CL	Groups to write to energy file.

NEIGHBOR SEARCHING:

nstlist	10	Frequency to update the neighbor list [steps].
ns_type	grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every <i>nstlist</i> steps.
pbcb	xyz	Use periodic boundary conditions (pbcb) in all directions.
rlist	1.4	Cut-off distance for the short-range neighbor list [nm].

ELECTROSTATICS AND VAN DER WAALS:

coulombtype	PME	Use Particle-Mesh Ewald (PME) electrostatics to calculate long-range electrostatic interactions.
rcoulomb	1.4	Distance for the Coulomb cut-off [nm].
vdw-type	Cut-off	Apply twin-range cut-offs for van der Waals interactions.
rvdw	1.4	Distance for the Lennard-Jones cut-off [nm].
DispCorr	EnerPres	Apply long range dispersion corrections for energy and pressure.
fourierspacing	0.12	Maximum grid spacing for the Fast Fourier Transform (FFT) grid when using PME [nm].
pme_order	4	Interpolation order for PME. Four equals cubic interpolation.
ewald_rtol	1.00E-05	The relative strength of the Ewald-shifted direct potential.
ewald_geometry	3d	The Ewald sum is performed in all three dimensions.
optimize_fft	yes	Calculate the optimal FFT plan for the grid at startup. This saves a few percent for long simulations.

TEMPERATURE AND PRESSURE COUPLING:

tcoupl	berendsen	Temperature coupling with a Berendsen-thermostat.
tc-grps	Protein DPPC SOL_CL	Groups to couple separately to temperature bath.
tau-t	0.1 0.1 0.1	Time constant for coupling (one for each group in <i>tc_grps</i>) [ps].
ref-t	323 323 323	Reference temperature for coupling (one for each group in <i>tc_grps</i>) [K].

pcoupl	berendsen	Exponential relaxation pressure coupling.
pcoupltype	semiisotropic	Pressure coupling which is isotropic in the x and y direction, but different in the z direction. Useful for membrane simulations.
tau-p	2	Time constant for coupling [ps].
compressibility	4.5 E-05 4.5 E-05	Compressibility [1/bar]. For water at 1 atm and 300 K the compressibility is $4.5\text{e-}5 \text{ bar}^{-1}$.
ref-p	1 1	Reference pressure for coupling.

GENERATE VELOCITIES FOR STARTUP RUN:

gen-vel	no	Generate velocities in <i>grompp</i> according to a Maxwell distribution at temperature <i>gen_temp</i> , with random seed <i>gen_seed</i> . This is only used at the beginning of the equilibration.
gen-temp	323	Temperature for Maxwell distribution [K].
gen-seed	1982	Used to initialize random generator for random velocities.

BONDS:

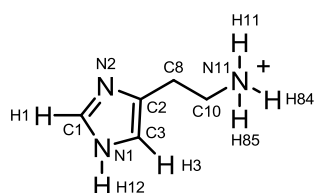
constraints	all-bonds	Convert all bonds to constraints.
constraint-algorithm	Lincs	LINear Constraint Solver. Allow the usage of a time step for integration of 2 fs.
lincs-order	4	Highest order in the expansion of the constraint coupling matrix. For 'normal' MD simulations an order of 4 usually suffices.
lincs-warnangle	30	Maximum angle that a bond can rotate before LINCS will complain [°].
morse	no	Bonds are represented by a harmonic potential.

NMR REFINEMENT:

disre	simple	Simple (per-molecule) distance restraints.
disre-weighting	equal	Divide the restraint force equally over all atom pairs in the restraint.
disre-fc	1000	Force constant for distance restraints [$\text{kJ mol}^{-1} \text{ nm}^{-2}$].

¹ The description of parameters was adapted from the GROMACS User Manual 4.5.6 (Hess *et al.*, 2008; Van Der Spoel *et al.*, 2005).

9.2 GROMOS96 53a6 force field parameters for histamine



Histamine with atom names

Atoms

Atom	Charge group	Charge	Mass
C3	1	0	12.011
H3	1	0.14	1.008
N1	1	-0.05	14.0067
H12	1	0.31	1.008
C1	1	0	12.011
H1	1	0.14	1.008
N2	1	-0.54	14.0067
C2	1	0	12.011
C8	2	0	14.027
C10	3	0.127	14.027
N11	3	0.129	14.0067
H84	3	0.248	1.008
H85	3	0.248	1.008
H11	3	0.248	1.008

Bonds

Atom 1	Atom 2	Parameter ^a
C3	H3	gb_3
C3	N1	gb_10
C2	C3	gb_10
N1	H12	gb_2
N1	C1	gb_10
C1	H1	gb_3
C1	N2	gb_10
C2	N2	gb_10
C2	C8	gb_27
C8	C10	gb_27
C10	N11	gb_21
N11	H84	gb_2
N11	H85	gb_2
N11	H11	gb_2

Angles

Atom 1	Atom 2	Atom 3	Parameter ^a	Function ^b
H3	C3	N1	ga_36	2
H3	C3	C2	ga_36	2
N1	C3	C2	ga_7	2
C3	N1	H12	ga_36	2
C3	N1	C1	ga_7	2
H12	N1	C1	ga_36	2
N1	C1	H1	ga_36	2
N1	C1	N2	ga_7	2
H1	C1	N2	ga_36	2
C1	N2	C2	ga_7	2
C3	C2	N2	ga_7	2
C3	C2	C8	ga_37	2
N2	C2	C8	ga_37	2
C2	C8	C10	ga_15	2
C8	C10	N11	ga_15	2
C10	N11	H84	ga_11	2
C10	N11	H85	ga_11	2
C10	N11	H11	ga_11	2
H84	N11	H85	ga_10	2
H84	N11	H11	ga_10	2
H85	N11	H11	ga_10	2

Dihedral angles

Atom 1	Atom 2	Atom 3	Atom 4	Parameter ^a	Function ^b
C10	C8	C2	C3	gd_40	1
N11	C10	C8	C2	gd_34	1
C8	C10	N11	H11	gd_29	1

Improper dihedral angles ^c

Atom 1	Atom 2	Atom 3	Atom 4	Parameter ^a	Function ^b
C3	C2	N1	H3	gi_1	2
N1	C1	H12	C3	gi_1	2
C1	N1	N2	H1	gi_1	2
C2	C8	N2	C3	gi_1	2
C3	N1	C1	N2	gi_1	2
N1	C1	N2	C2	gi_1	2
C1	N2	C2	C3	gi_1	2
N2	C2	C3	N1	gi_1	2
C2	C3	N1	C1	gi_1	2

^a Parameters as defined in the topology file ffG53a6bon.itp of the GROMOS 53a6 force field included in GROMACS 4.0.7.

^b Function used for the calculation.

^c Improper dihedrals are meant to keep planar groups (e.g. aromatic rings) planar, or to prevent molecules from flipping over to their mirror images.

9.3 Systematic calculation of hydrogen bonds: *gro_hbonds*

9.3.1 Parameter file *gro_hbonds-para.txt*

```

1  # PLEASE ENTER: name of the simulation
2  var-sim=inact-md1-2
3
4  # PLEASE ENTER: name of the gro-file (also xtc and tpr files must have this
5  # name; otherwise change respective lines in gro_hbonds.sh)
6  var-filename=../../hH2R_inactive_2VT4_md1_2
7
8  # PLEASE ENTER: duration of the simulation in picoseconds
9  var-duration=80000
10
11 # PLEASE ENTER: time step for analysis
12 var-timestep=10
13
14 # PLEASE ENTER: the number of amino acids, lipids and ligands in the coordinate
15 # file (has to be in one sequence)
16 var-residues=412
17
18 # PLEASE ENTER: cut-off for heteroatom1-heteroatom2 distance for H-bonds in
19 # picometer; natural number
20 var-cutoffdist=350
21
22 # PLEASE ENTER: cut-off for heteroatom1-H-heteroatom2 angle for H-bonds in degree;
23 # natural number
24 var-cutoffang=120
25
26 # PLEASE ENTER: number of data points the average in the output files should be
27 # calculated (division of duration/timestep/histo has to be possible without a
28 # rest)
29 var-histo=50

```

9.3.2 Shell script *gro_hbonds.sh*

```

1  #!/bin/bash
2
3  #####
4  ### Content:
5  #   Systematic calculation of direct and water mediated H-bonds
6  #   Helix analysis
7  #   Analysis of buried unsatisfied hydrogen bond donors and acceptors
8  #   Detailed information (distance and angle) of a specific H-bond during the
9  #       entire simulation
10 #   Picture (.jpg) of the existence of an H-bond over time
11
12 #####
13
14 ### Section 1
15 #####
16
17 # Compilation of the C program gro_hbonds-calc
18 gcc gro_hbonds-calc.c /usr/lib64/libm.a -o gro_hbonds-calc
19
20 # Storage of the result files
21 mkdir output_detail
22 mkdir output_sum
23

```

```

24 # Extract variables from gro_hbonds-para.txt
25 descr=gro_hbonds-para.txt
26
27 sim=`sed -n /var-sim=/p $descr | cut -d= -f2`
28 filename=`sed -n /var-filename=/p $descr | cut -d= -f2`
29 duration=`sed -n /var-duration=/p $descr | cut -d= -f2`
30 timestep=`sed -n /var-timestep=/p $descr | cut -d= -f2`
31 residues=`sed -n /var-residues=/p $descr | cut -d= -f2`
32 cutoffdist=`sed -n /var-cutoffdist=/p $descr | cut -d= -f2`
33 cutofffang=`sed -n /var-cutofffang=/p $descr | cut -d= -f2`
34 histo=`sed -n /var-histo=/p $descr | cut -d= -f2`
35
36 # Extract the number of atoms/sites from the coordinate file (.gro)
37 coordline=`sed -n '2p' $filename.gro`
38 # Create a list with residue numbers and names
39 cut -c1-9 $filename.gro | sed -n '1,2!p' | sed 's/d' | sort -n -u | cut -c6-9 | sed
's/ //g' > aa-name.txt
40 # Determine number of residues
41 resline=`wc -l aa-name.txt | cut -d \ -f 1`
42 # Total number of time points to analyze
43 frames=`expr $duration / $timestep`
44
45
46 #####
47 #### Section 2: Calculate H-bonds
48 #####
49
50 # The program works with slices of 1000 frames
51 slice=1000
52 # Number of trjconv runs
53 parts=0
54 # Beginning and end of one slice (in ps)
55 begin=$(echo "scale=2; $timestep-0.5" | bc)
56 end=$(echo "scale=2; $slice*$timestep+0.5" | bc)
57 start=$timestep
58
59 while [ $start -le $duration ]; do
60
61 # Create coordinate file with trjconv
62 echo 0 | trjconv -f $filename.xtc -s $filename.tpr -o frame_$begin-$end.gro -pbc
mol -dt $timestep -b $begin -e $end >> trjconv-result.txt 2>&1
63
64 # Combine sequence and coordinates
65 cat aa-name.txt frame_$begin-$end.gro > input_frame_$begin-$end.txt
66 rm -f frame_$begin-$end.gro
67
68 ./gro_hbonds-calc 1 $resline $coordline $slice $residues $start $cutoffdist
$cutofffang $timestep < input_frame_$begin-$end.txt >> out_h-bonds_int_all.txt
69
70 rm -f input_frame_$begin-$end.txt
71
72 parts=`expr $parts + 1`
73 begin=$(echo "scale=2; $parts*$slice*$timestep+$timestep-0.5" | bc)
74 end=$(echo "scale=2; ($parts+1)*$timestep*$slice+0.5" | bc)
75 start=`echo $(echo "scale=2; $begin+0.5" | bc) | cut -d\ -f1`
76 done
77
78 rm -f aa-name.txt
79 rm -f trjconv-result.txt
80
81 # Sort the output; necessary for the subprogram arg1=2
82 sort -n -t- -k 1,1 -k 2,2 -k 5,5 -k 6,6 -k 9,9 out_h-bonds_int_all.txt > out_h-
bonds_int_all_sort.txt
83 rm -f out_h-bonds_int_all.txt
84 # Summarize all single H-bond files; the data for a heteroatom1-heteroatom2
85 # interaction is combined over the whole simulation
86 minus=`grep -o - out_h-bonds_int_all_sort.txt | wc -l`
87 ./gro_hbonds-calc 2 $minus 0 0 0 0 0 0 < out_h-bonds_int_all_sort.txt >
output_detail/$sim-h-bond-all-int-Het_.txt
88 rm -f out_h-bonds_int_all_sort.txt
89 sort -n -t- -k 1,1 -k 2,2 -k 3,3 -k 4,4 -k 5,5 -k 6,6 output_detail/$sim-h-bond-
all-int-Het_.txt > output_detail/$sim-hbond-Het_int.txt
90 rm -f output_detail/$sim-h-bond-all-int-Het_.txt
91 # Create summarized output: Heteroatom-based interactions without water
92 # interactions
93 sed -n '/^[0-9]*-[0-9]*-[0-9]*-[012356]-/p' output_detail/$sim-hbond-Het_int.txt >

```

```

output_detail/$sim-hbond-Het_int_-SOL.txt
94 minus=`grep -o - output_detail/$sim-hbond-Het_int_-SOL.txt | wc -l`
95 ./gro_hbonds-calc 7 $minus 8 $frames $timestep $histo 0 0 0 < output_detail/$sim-
    hbond-Het_int_-SOL.txt > output_sum/$sim-hbond-Het_-SOL.txt
96
97 # Interactions: type-based (e.g. bb, sc, terminus) without water interactions
98 minus=`grep -o - output_detail/$sim-hbond-Het_int.txt | wc -l`
99 ./gro_hbonds-calc 4 $minus 1 0 0 0 0 0 < output_detail/$sim-hbond-Het_int.txt >
    output_detail/$sim-hbond-type_int.txt
100 minus=`grep -o - output_detail/$sim-hbond-type_int.txt | wc -l`
101 ./gro_hbonds-calc 7 $minus 4 $frames $timestep $histo 0 0 0 < output_detail/$sim-
    hbond-type_int.txt > output_sum/$sim-hbond-type.txt
102 sed -n 1p output_sum/$sim-hbond-type.txt > output_sum/$sim-hbond-type_-SOL.txt
103 sed -n '/^[0-9]*\t[0-9]*\t[0-9]*\t[012356]\t/p' output_sum/$sim-hbond-type.txt >>
    output_sum/$sim-hbond-type_-SOL.txt
104
105 # Interactions: residue-based without water interactions
106 minus=`grep -o - output_detail/$sim-hbond-Het_int.txt | wc -l`
107 ./gro_hbonds-calc 4 $minus 2 0 0 0 0 0 < output_detail/$sim-hbond-Het_int.txt >
    output_detail/$sim-hbond-res_int.txt
108 sed -n '/^[0-9]*-[0-9]*-[0-9]*-[012356]/-p' output_detail/$sim-hbond-Het_int.txt >
    output_detail/$sim-hbond-Het_int_-SOL.txt
109 minus=`grep -o - output_detail/$sim-hbond-Het_int_-SOL.txt | wc -l`
110 ./gro_hbonds-calc 4 $minus 2 0 0 0 0 0 < output_detail/$sim-hbond-Het_int_-
    SOL.txt > output_detail/$sim-hbond-res_int_-SOL.txt
111 minus=`grep -o - output_detail/$sim-hbond-res_int_-SOL.txt | wc -l`
112 ./gro_hbonds-calc 7 $minus 2 $frames $timestep $histo 0 0 0 < output_detail/$sim-
    hbond-res_int_-SOL.txt > output_sum/$sim-hbond-res_-SOL.txt
113
114 rm -f output_sum/$sim-hbond-type.txt
115 rm -f output_detail/$sim-hbond-Het_int_-SOL.txt
116 rm -f output_detail/$sim-hbond-res_int_-SOL.txt
117
118 #####
119 #### Section 3: Calculate water mediated H-bonds
120 #####
121
122 # Extract all lines where a water molecule is present: code X-X-X-4-X-X-X-X
123 sed -n '/^[0-9]*-[0-9]*-[0-9]*-4-*/p' output_detail/$sim-hbond-Het_int.txt >
    water_hbonds.txt
124
125 # Sort water molecules
126 sort -n -t- -k 2,2 -k 1,1 -k 5,5 water_hbonds.txt > water_hbonds_sort.txt
127 rm -f water_hbonds.txt
128 minus=`grep -o - water_hbonds_sort.txt | wc -l`
129
130 ./gro_hbonds-calc 3 $minus 1 0 0 0 0 0 < water_hbonds_sort.txt > $sim-SOL-
    mediated-hbond-int_.txt
131 rm -f water_hbonds_sort.txt
132
133 # Sort the output according to the residue numbers
134 sort -t- -k 1,1 -k 2,2 -k 3,3 -k 4,4 -k 5,5 -k 6,6 -n $sim-SOL-mediated-hbond-
    int_.txt > $sim-SOL-mediated-hbond-int-sort.txt
135 rm -f $sim-SOL-mediated-hbond-int_.txt
136
137 # Generate a detailed file (each atom separate): Het-based
138 minus=`grep -o - $sim-SOL-mediated-hbond-int-sort.txt | wc -l`
139 ./gro_hbonds-calc 3 $minus 2 0 0 0 0 0 < $sim-SOL-mediated-hbond-int-sort.txt >
    output_detail/$sim-SOL-mediated-hbond-int-Het_.txt
140 rm -f $sim-SOL-mediated-hbond-int-sort.txt
141 sort -n -t- -k 1,1 -k 2,2 -k 3,3 -k 4,4 -k 5,5 -k 6,6 output_detail/$sim-SOL-
    mediated-hbond-int-Het_.txt > output_detail/$sim-hbond_SOL-Het_int.txt
142 rm -f output_detail/$sim-SOL-mediated-hbond-int-Het_.txt
143 minus=`grep -o - output_detail/$sim-hbond_SOL-Het_int.txt | wc -l`
144 ./gro_hbonds-calc 7 $minus 8 $frames $timestep $histo 0 0 0 < output_detail/$sim-
    hbond_SOL-Het_int.txt > output_sum/$sim-hbond_SOL-Het.txt
145
146 # Generate a summarized file: type-based (e.g. bb, sc, terminus)
147 minus=`grep -o - output_detail/$sim-hbond_SOL-Het_int.txt | wc -l`
148 ./gro_hbonds-calc 4 $minus 1 0 0 0 0 0 < output_detail/$sim-hbond_SOL-
    Het_int.txt > output_detail/$sim-hbond_SOL-type_int.txt
149 minus=`grep -o - output_detail/$sim-hbond_SOL-type_int.txt | wc -l`
150 ./gro_hbonds-calc 7 $minus 4 $frames $timestep $histo 0 0 0 < output_detail/$sim-
    hbond_SOL-type_int.txt > output_sum/$sim-hbond_SOL-type.txt
151
152 # Generate a summarized file: residue-based

```

```

153 minus=`grep -o - output_detail/$sim-hbond_SOL-Het_int.txt | wc -l`
154 ./gro_hbonds-calc 4 $minus 2 0 0 0 0 0 < output_detail/$sim-hbond_SOL-
  Het_int.txt > output_detail/$sim-hbond_SOL-res_int.txt
155 minus=`grep -o - output_detail/$sim-hbond_SOL-res_int.txt | wc -l`
156 ./gro_hbonds-calc 7 $minus 2 $frames $timestep $histo 0 0 0 < output_detail/$sim-
  hbond_SOL-res_int.txt > output_sum/$sim-hbond_SOL-res.txt
157
158 #####
159 #### Section 4: Calculate bb1-3, bb1-4 and bb1-5 interactions for helicity-ckeck
160 #####
161
162 # bb-bb interactions: Het-based
163 sed -n /^[0-9]*-[0-9]*-0-0-[0-9]*-[0-9]*-*/p output_detail/$sim-hbond-Het_int.txt
  > output_detail/$sim-hbond-bb-int-Het.txt
165 minus1=`grep -o - output_detail/$sim-hbond-bb-int-Het.txt | wc -l`
166
167 # bb-bb-interactions-1-3
168 ./gro_hbonds-calc 5 3 $minus1 $frames $residues 2 0 0 0 < output_detail/$sim-
  hbond-bb-int-Het.txt > output_detail/$sim-hbond-helix-1-3-res_int.txt
169 minus=`grep -o - output_detail/$sim-hbond-helix-1-3-res_int.txt | wc -l`
170 ./gro_hbonds-calc 7 $minus 2 $frames $timestep $histo 0 0 0 < output_detail/$sim-
  hbond-helix-1-3-res_int.txt > output_sum/$sim-hbond-helix-1-3-res.txt
171
172 # bb-bb-interactions-1-4
173 ./gro_hbonds-calc 5 4 $minus1 $frames $residues 2 0 0 0 < output_detail/$sim-
  hbond-bb-int-Het.txt > output_detail/$sim-hbond-helix-1-4-res_int.txt
174 minus=`grep -o - output_detail/$sim-hbond-helix-1-4-res_int.txt | wc -l`
175 ./gro_hbonds-calc 7 $minus 2 $frames $timestep $histo 0 0 0 < output_detail/$sim-
  hbond-helix-1-4-res_int.txt > output_sum/$sim-hbond-helix-1-4-res.txt
176
177 # bb-bb-interactions-1-5
178 ./gro_hbonds-calc 5 5 $minus1 $frames $residues 2 0 0 0 < output_detail/$sim-
  hbond-bb-int-Het.txt > output_detail/$sim-hbond-helix-1-5-res_int.txt
179 minus=`grep -o - output_detail/$sim-hbond-helix-1-5-res_int.txt | wc -l`
180 ./gro_hbonds-calc 7 $minus 2 $frames $timestep $histo 0 0 0 < output_detail/$sim-
  hbond-helix-1-5-res_int.txt > output_sum/$sim-hbond-helix-1-5-res.txt
181
182 rm -f output_detail/$sim-hbond-bb-int-Het.txt
183
184 #####
185 #### Section 5: Calculate percentage of frames where heteroatoms are hydrogen
186 #### bonded (structure validation)
187 #####
188
189 # Get a list of the amino acid sequence
190 cut -c1-9 $filename.gro | sed -n 1,2!'p | sort -u -n | sed -n 1,"$residues"p |
  cut -c 6-9 | sed 's/ \+//g' > aa-list.txt
192 minus=`grep -o - output_detail/$sim-hbond-Het_int.txt | wc -l`
193 cat aa-list.txt output_detail/$sim-hbond-Het_int.txt > $sim-input-str-val-h-
  bonds.txt
194
195 # Output with all frames where H-bonds occur
196 ./gro_hbonds-calc 6 0 $minus $frames $residues 0 0 0 0 < $sim-input-str-val-h-
  bonds.txt > $sim-str-val-h-bonds-sum-int.txt
197
198 cut -f1,3,4,6- $sim-str-val-h-bonds-sum-int.txt | sed -n 1!'p | sed 's/\t/-/g' >
  output_detail/$sim-hbonds-amount_int.txt
199
200 minus=`grep -o - output_detail/$sim-hbonds-amount_int.txt | wc -l`
201 ./gro_hbonds-calc 7 $minus 3 $frames $timestep $histo 0 0 0 < output_detail/$sim-
  hbonds-amount_int.txt > output_sum/$sim-hbonds-amount.txt
202
203 rm -f $sim-input-str-val-h-bonds.txt
204 rm -f $sim-str-val-h-bonds-sum-int.txt
205 rm -f aa-list.txt
206
207 #####
208 #### Section 6: detailed output for one special H-bond (heteroatom1 - hydrogen -
209 #### heteroatom2)
210 #####
211
212 # Atom numbers in the gro-file (example)
213 het1=228
214

```

```

215 het2=212
216 H=229
217
218 echo \[ hbond\ \] > index-$het1-$het2-$H.ndx
219 echo $het1 $het2 $H >> index-$het1-$het2-$H.ndx
220
221 begin=$(echo "scale=2; $timestep-0.5" |bc)
222 end=$(echo "scale=2; $duration+0.5" |bc)
223 trjconv -f $filename.xtc -s $filename.tpr -o hbond-$het1-$het2-$H.gro -n index-
    $het1-$het2-$H.ndx -b $begin -e $end -dt $timestep >> trjconv-result.txt 2>&1
224 rm -f trjconv-result.txt
225 rm -f index-$het1-$het2-$H.ndx
226
227 sed -n '/[0-9]\.[0-9][0-9][0-9]$/p' hbond-$het1-$het2-$H.gro > tmp1.txt
228 rm -f hbond-$het1-$het2-$H.gro
229 cut -c21-44 tmp1.txt > tmp2.txt
230 rm -f tmp1.txt
231
232 lines=`wc -l tmp2.txt | cut -d\ -f1`
233
234 echo -e time\[ps\]\t distance\[nm\]\t angle\[°\]\t hbond > hbond-dist-ang-$het1-
    $het2-$H.txt
235
236 ./gro_hbonds-calc 8 $lines $timestep 0 0 0 $cutoffdist $cutoffang 0 < tmp2.txt >>
    hbond-dist-ang-$het1-$het2-$H.txt
237
238 rm -f tmp2.txt
239
240
241 #####
242 #### Section 7: jpg-picture for one H-bond (for line 1 in output_detail/$sim-
243 #### hbond-Het_int.txt)
244 #####
245
246 sed -n 1p output_detail/$sim-hbond-Het_int.txt > interaction.txt
247
248 ./gro_hbonds-calc 9 8 $frames $timestep 0 0 0 0 0 < interaction.txt > interaction-
    pic.xpm
249
250 convert -density 300 interaction-pic.xpm hbond-existance_example.jpg
251
252 rm -f interaction.txt
253 rm -f interaction-pic.xpm
254 rm -f gro_hbonds-calc

```

9.3.3 C program `gro_hbonds-calc.c`

```

1  #include <stdio.h>
2  #include <stdlib.h>
3  #include <string.h>
4  #include <math.h>
5
6  // Compile with:
7  // gcc gro_hbonds-calc.c /usr/lib64/libm.a -o gro_hbonds-calc
8
9  // Number of residues which are defined by the AAcode in this program
10 // (charge_termini)
11 #define no_def_res 40
12
13 int main(int argc, char *argv[]) {
14
15     int arg1, arg2, arg3, arg4, arg5, arg6, arg7, arg8, arg9;
16
17     // With parameters in the shell script the following variables are defined
18     arg1 = strtol(argv[1], NULL, 10);
19     arg2 = strtol(argv[2], NULL, 10);

```

```

20  arg3 = strtol(argv[3], NULL, 10);
21  arg4 = strtol(argv[4], NULL, 10);
22  arg5 = strtol(argv[5], NULL, 10);
23  arg6 = strtol(argv[6], NULL, 10);
24  arg7 = strtol(argv[7], NULL, 10);
25  arg8 = strtol(argv[8], NULL, 10);
26  arg9 = strtol(argv[9], NULL, 10);
27
28  float cutoffdist = ((float) arg7)/1000;
29  float cutofffang = ((float) arg8)/1;
30
31  // This number (> duration of the MD) is always at the end of a line/section;
32  int endl ine=9999999;
33
34  // (charge_termini) for every charged terminal residue; also change "no_def_res"
35  // if necessary
36  // int terminuslaa=319; // residue number
37  // int terminuslcode=39; // position in AAc ode
38
39  // In AAn ame the amino acid name is stored and compared to the amino acid sequence
40  // which is read in; all listed standard amino acids of GROMACS (ffG53a6.rtp) are
41  // included (+ ACE and NH2), except HYP (hydroxyproline); additionally: DPPC, SOL,
42  // SOLc, CL-, HAH
43  char AAn ame[no_def_res][5]={
44      {'A','C','E'}, // 0
45      {'N','H','2'}, // 1
46      {'A','L','A'}, // 2
47      {'A','R','G'}, // 3
48      {'A','R','G','N'}, // 4
49      {'A','S','N'}, // 5
50      {'A','S','N','1'}, // 6
51      {'A','S','P'}, // 7
52      {'A','S','P','H'}, // 8
53      {'C','Y','S'}, // 9
54      {'C','Y','S','H'}, // 10
55      {'C','Y','S','1'}, // 11
56      {'C','Y','S','2'}, // 12
57      {'G','L','N'}, // 13
58      {'G','L','U'}, // 14
59      {'G','L','U','H'}, // 15
60      {'G','L','Y'}, // 16
61      {'H','I','S','A'}, // 17
62      {'H','I','S','B'}, // 18
63      {'H','I','S','H'}, // 19
64      {'H','I','S','1'}, // 20
65      {'H','I','S','2'}, // 21
66      {'I','L','E'}, // 22
67      {'L','E','U'}, // 23
68      {'L','Y','S'}, // 24
69      {'L','Y','S','H'}, // 25
70      {'M','E','T'}, // 26
71      {'P','H','E'}, // 27
72      {'P','R','O'}, // 28
73      {'S','E','R'}, // 29
74      {'T','H','R'}, // 30
75      {'T','R','P'}, // 31
76      {'T','Y','R'}, // 32
77      {'V','A','L'}, // 33
78      {'D','P','P','C'}, // 34
79      {'S','O','L'}, // 35
80      {'S','O','L','c'}, // 36
81      {'C','L','-'}, // 37
82      {'H','A','H'} // 38
83  };
84
85  // For each amino acid a code necessary for calculating the H-bonds is stored; the
86  // sequence of the code has to be the same as the amino acid sequence in AAn ame
87  int AAc ode[no_def_res][18]={
88      {3,1,2,0}, // 0-ACE
89      {3,1,0,2}, // 1-NH2
90      {6,2,0,1,5,0}, // 2-ALA
91      {17,5,0,1,6,1,9,2,12,2,16,0}, // 3-ARG
92      {16,5,0,1,6,1,9,1,11,2,15,0}, // 4-ARGN
93      {11,4,0,1,5,0,6,2,10,0}, // 5-ASN
94      {11,4,0,1,5,0,6,2,10,0}, // 6-ASN1
95      {9,4,0,1,5,0,6,0,8,0}, // 7-ASP

```

```

96         {10,4,0,1,5,0,6,1,9,0}, // 8-ASPH
97         {7,3,0,1,4,0,6,0}, // 9-CYS
98         {8,3,0,1,4,1,7,0}, // 10-CYSH
99         {7,3,0,1,4,0,6,0}, // 11-CYS1
100        {7,3,0,1,4,0,6,0}, // 12-CYS2
101        {12,4,0,1,6,0,7,2,11,0}, // 13-GLN
102        {10,4,0,1,6,0,7,0,9,0}, // 14-GLU
103        {11,4,0,1,6,0,7,1,10,0}, // 15-GLUH
104        {5,2,0,1,4,0}, // 16-GLY
105        {14,4,0,1,5,1,11,0,13,0}, // 17-HISA
106        {14,4,0,1,5,0,10,1,13,0}, // 18-HISB
107        {15,4,0,1,5,1,11,1,14,0}, // 19-HISH
108        {14,4,0,1,5,1,11,0,13,0}, // 20-HIS1
109        {14,4,0,1,5,1,11,0,13,0}, // 21-HIS2
110        {9,2,0,1,8,0}, // 22-ILE
111        {9,2,0,1,8,0}, // 23-LEU
112        {12,3,0,1,7,2,11,0}, // 24-LYS
113        {13,3,0,1,7,3,12,0}, // 25-LYSH
114        {9,3,0,1,5,0,8,0}, // 26-MET
115        {17,2,0,1,16,0}, // 27-PHE
116        {7,2,0,0,6,0}, // 28-PRO
117        {8,3,0,1,4,1,7,0}, // 29-SER
118        {9,3,0,1,4,1,8,0}, // 30-THR
119        {21,3,0,1,8,1,20,0}, // 31-TRP
120        {18,3,0,1,14,1,17,0}, // 32-TYR
121        {8,2,0,1,7,0}, // 33-VAL
122        {50,8,6,0,8,0,9,0,10,0,13,0,15,0,32,0,34,0}, // 34-DPPC
123        {3,1,0,2}, // 35-SOL
124        {3,1,0,2}, // 36-SOLc
125        {1,1,0,0}, // 37-CL-
126        {14,3,2,1,6,0,10,3}, // 38-HAH
127        {10,3,0,1,8,0,9,0} // (charge_termini) 39-terminal aa of GaCT is LEU
128                           // and charged (ends with -COO)
129    };
130
131
132    //////////////////////////////////////
133    // Section 1: Calculate H-bonds: need arg1, arg2, arg3, arg4, arg5, arg6, arg7,
134    // arg8, arg9
135    //////////////////////////////////////
136
137    if(arg1==1)
138    {
139
140        int a,b,e,f,k,l,i,j,t,u,v,x,y,z,cmp;
141        int lineaa1, lineaa2, lineH, begaa1, begaa2, aalfr, aa2fr, aa1, aa2, type1, type2;
142        char w;
143        float d;
144        float alphab, alphaw, cosalphab;
145        float pi=3.14159;
146
147        // In matrixres the amino acid sequence is stored
148        char matrixres[arg2][5];
149
150        // Initialize matrixres with NULL
151        for (i=0; i<arg2; i++) {for (j=0; j<5; j++) matrixres[i][j]=0;}
152
153        // In arrayres the amino acid sequence as integer code is saved
154        int arrayres[arg2];
155
156        // Get memory from RAM for matrixcoord where all coordinates are stored
157        float ** matrixcoord;
158        matrixcoord = malloc((arg4*arg3) * sizeof(float *));
159        if(NULL == matrixcoord)
160        {
161            printf("NO RAM for matrixcoord 1!");
162            return EXIT_FAILURE;
163        }
164        for(i=0; i<(arg4*arg3); i++)
165        {
166            matrixcoord[i] = malloc(3 * sizeof(float));
167            if(NULL == matrixcoord[i])
168            {
169                printf("NO RAM for matrixcoord 2!");
170                return EXIT_FAILURE;
171            }

```

```

172     }
173
174     // Interactions per residue
175     int resultlines=100000;
176     // In intresulttable 8 integers describing both residues are stored
177     int intresulttable[resultlines][8];
178
179     // Get memory from RAM for resulttable -> here it is stored if there is an
180     // interaction at a specific time point
181     char ** resulttable;
182     resulttable = malloc((resultlines) * sizeof(char *));
183     if(NULL == resulttable)
184     {
185         printf("NO RAM for resulttable 1!");
186         return EXIT_FAILURE;
187     }
188     for(i=0; i<(resultlines); i++)
189     {
190         resulttable[i] = malloc((arg4) * sizeof(char));
191         if(NULL == resulttable[i])
192         {
193             printf("NO RAM for resulttable 2!");
194             return EXIT_FAILURE;
195         }
196     }
197
198     // Read amino acid sequence
199     for(i=0; i<arg2; i++)
200     {
201         for(j=0; ; j++)
202         {
203             scanf("%c", &matrixres[i][j]);
204             if(matrixres[i][j]=='\n'){matrixres[i][j]=0; break;}
205         }
206     }
207
208     // Convert amino acids to numbers
209     for(i=0; i<arg2; i++)
210     {
211         for(j=0; j<no_def_res; j++)
212         {
213             if((cmp=strcmp(matrixres[i], AName[j]))==0) {arrayres[i]=j; break;}
214         }
215     }
216
217     // Read coordinates
218     for(i=0; i<(arg4*arg3); )
219     {
220         for(j=0; ; j++) {scanf("%c", &w); if(w=='\n') break;}
221         scanf("%d",&a);
222
223         for(a=0; a<arg3; a++)
224         {
225             for(j=0; j<21; j++) scanf("%c", &w);
226             scanf("%f", &matrixcoord[i][0]);
227             scanf("%f", &matrixcoord[i][1]);
228             scanf("%f", &matrixcoord[i][2]);
229             i++;
230         }
231         scanf("%c", &w);
232         for(j=0; ; j++) {scanf("%c", &w); if(w=='\n') break;}
233     }
234
235     /* control output
236     for (i=0; i<arg2; i++)
237     printf("%s - %d\n", matrixres[i], arrayres[i]);
238     for (i=0; i<100; i++)
239     printf("%f %f %f\n", matrixcoord[i][0], matrixcoord[i][1], matrixcoord[i][2]);
240     */
241
242     // Start the H-bond calculation
243     // x is the first amino acid, arg5 the number of residues which should be analyzed
244     for(x=0; x<arg5; x++)
245     {
246
247

```

```

248 // Initialize intresulttable with 0
249 for (i=0; i<resultlines; i++) {for(j=0; j<8; j++) intresulttable[i][j]=0;}
250
251 // Initialize resulttable with SPACE
252 for (i=0; i<resultlines; i++) {for(j=0; j<arg4; j++) resulttable[i][j]=' ';}
253
254 // Count the number of result lines of amino acid no x (in intresulttable and
255 // resulttable)
256 u=0;
257
258 // First line of the first amino acid (x)
259 begaal=0;
260 for(i=0;i<x;i++)
261 {j=arrayres[i]; begaal+=AAcode[j][0];}
262
263 // Amino acid type (as integer code) of the first amino acid (x)
264 aal=arrayres[x];
265
266 // (charge_termini) if there is more than one charged termini add corresponding
267 // lines
268 // if(x==terminuslaa-1) aal=terminuslcode;
269 // if(x>terminuslaa-1) begaal++;
270
271 // Only interactions to residues having a higher sequence number are calculated
272 for (y=x;y<arg2;y++)
273 {
274 // First line of the second amino acid (y)
275 begaa2=0;
276 for(i=0;i<y;i++)
277 {j=arrayres[i]; begaa2+=AAcode[j][0];}
278
279 // Amino acid type (as integer code) of the second amino acid (y)
280 aa2=arrayres[y];
281
282 // (charge_termini) if there is more than one charged termini add corresponding
283 // lines
284 // if(y==terminuslaa-1) aa2=terminuslcode;
285 // if(y>terminuslaa-1) begaa2++;
286
287 // Control output: printf("%d_%d_%d_%d ", x, begaal, y, begaa2);
288
289 for(z=0;z<arg4;z++)
290 {
291 aalfr=begaal+z*arg3;
292 aa2fr=begaa2+z*arg3;
293
294 // Check if the first atoms of the first and second amino acid has a greater
295 // distance than 3 nm (critical for lipids)
296 d=sqrt((matrixcoord[aalfr][0]-matrixcoord[aa2fr][0])*(matrixcoord[aalfr][0]-
matrixcoord[aa2fr][0])+(matrixcoord[aalfr][1]-
matrixcoord[aa2fr][1])*(matrixcoord[aalfr][1]-
matrixcoord[aa2fr][1])+(matrixcoord[aalfr][2]-
matrixcoord[aa2fr][2])*(matrixcoord[aalfr][2]-matrixcoord[aa2fr][2]));
297 if(d<=3.0)
298 {
299 for(a=2, b=0; b<AAcode[aal][1];a+=2,b++)
300 // heteroatom in the first amino acid/residue
301 {
302 for(e=2,f=0;f<AAcode[aa2][1];e+=2,f++)
303 // heteroatom of the second amino acid/residue
304 {
305
306 lineaal=aalfr+AAcode[aal][a];
307 lineaa2=aa2fr+AAcode[aa2][e];
308
309 if(lineaal<lineaa2)
310 {
311
312 d=sqrt((matrixcoord[lineaal][0]-matrixcoord[lineaa2][0])*(matrixcoord[lineaal][0]-
matrixcoord[lineaa2][0])+(matrixcoord[lineaal][1]-
matrixcoord[lineaa2][1])*(matrixcoord[lineaal][1]-
matrixcoord[lineaa2][1])+(matrixcoord[lineaal][2]-
matrixcoord[lineaa2][2])*(matrixcoord[lineaal][2]-matrixcoord[lineaa2][2]));
313
314 // check if the distance is below the cutoff and if there is at least one hydrogen
315 // connected to the heteroatoms

```

```

316 if(d<=cutoffdist && ( AAcode[aa1][a+1]!=0 || AAcode[aa2][e+1]!=0) )
317 {
318
319 // Code for interacting heteroatoms: 0 backbone, 1 side chain, 2 termini (ACE,
320 // NH2), 3 DPPC, 4 SOL und SOLc, 5 CL-, 6 ligand
321 if(aa1==0 || aa1==1) type1=2; //0: ACE; 1: NH2
322 else if(aa1==34) type1=3; // DPPC
323 else if(aa1==35 || aa1==36) type1=4; // SOL and SOLc
324 else if(aa1==37) type1=5; // chloride
325 else if(aa1==38) type1=6; // ligand
326 else if(AAcode[aa1][1]==2) type1=0; // if there are only two heteroatoms in the
327 // remaining residues than this must be a aa without heteroatoms in the side chain
328 else if( (AAcode[aa1][1]>2) && ((a==2) || (a==(2*AAcode[aa1][1])) ) ) type1=0;
329 // if there are more than 2 heteroatoms and it is the first (a=2) or the last in
330 // the AAcode definition
331 else type1=1; // remaining: side chain
332 if(aa2==0 || aa2==1) type2=2; //0: ACE; 1: NH2
333 else if(aa2==34) type2=3; // DPPC
334 else if(aa2==35 || aa2==36) type2=4; // SOL and SOLc
335 else if(aa2==37) type2=5; // chloride
336 else if(aa2==38) type2=6; // ligand
337 else if(AAcode[aa2][1]==2) type2=0; // if there are only two heteroatoms in the
338 // remaining residues than this must be a aa without heteroatoms in the side chain
339 else if( (AAcode[aa2][1]>2) && ((e==2) || (e==(2*AAcode[aa2][1])) ) ) type2=0;
340 // if there are more than 2 heteroatoms and it is the first (a=2) or the last in
341 // the AAcode definition
342 else type2=1; // remaining: side chain
343
344 for(k=1, l=0; l<AAcode[aa1][a+1]; k++, l++)
345 {
346 lineH=lineaa1+k;
347
348 cosalphab=((matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2]))/((sqrt((matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2]))*sqrt((matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2]))));
349 alphab=acos( cosalphab );
350 alphaw=alphab*180/pi;
351
352 if(alphaw>=cutoffang)
353 {
354 v=0;
355 for (t=0;t<u;t++)
356 {
357 // Check if these amino acids interacted over another hydrogen before or were
358 // present at an earlier time point
359 if( (intresulhtable[t][1]==x+1) && (intresulhtable[t][2]==b+1) &&
(intresulhtable[t][5]==y+1) && (intresulhtable[t][6]==f+1) )
360 {resulttable[t][z]='1'; v=1; break;}
361 }
362 if(v==0)
363 {
364 // 0:amino acid name 1; 1:amino acid number 1; 2: number of heteroatom in amino
365 // acid number 1; 3:type of heteroatom in amino acid 1 (backbone etc.); 4:amino
366 // acid name 1; 5:amino acid number 1; 6: number of heteroatom in amino acid
367 // number 1; 7:type of heteroatom in amino acid 1 (backbone etc.)
368 intresulhtable[u][0]=(lineaa1-z*arg3)+1;
369 intresulhtable[u][1]=x+1;
370 intresulhtable[u][2]=b+1;
371 intresulhtable[u][3]=type1;
372 intresulhtable[u][4]=(lineaa2-z*arg3)+1;
373 intresulhtable[u][5]=y+1;

```

```

374 intresulttable[u][6]=f+1;
375 intresulttable[u][7]=type2;
376 resulttable[u][z]='1';
377 u++;
378 }
379 }
380 }
381
382 for(k=1, l=0; l<Acode[aa2][e+1]; k++, l++)
383 {
384   lineH=lineaa2+k;
385
386   cosalphab=((matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2]))/(sqrt((matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa1][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa1][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa1][2]-
matrixcoord[lineH][2])))*sqrt((matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])*(matrixcoord[lineaa2][0]-
matrixcoord[lineH][0])+(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])*(matrixcoord[lineaa2][1]-
matrixcoord[lineH][1])+(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2])*(matrixcoord[lineaa2][2]-
matrixcoord[lineH][2]))));
387   alphab=acos( cosalphab );
388   alphaw=alphab*180/pi;
389
390   if(alphaw>=cutoffang)
391   {
392     v=0;
393     for(t=0;t<u;t++)
394     {
395       // Check if these amino acids interacted over another hydrogen before or were
396       // present at an earlier time point
397       if( (intresulttable[t][1]==x+1) && (intresulttable[t][2]==b+1) &&
(intresulttable[t][5]==y+1) && (intresulttable[t][6]==f+1) )
398       {resulttable[t][z]='1'; v=1; break;}
399     }
400     if(v==0)
401     {
402       // 0:amino acid name 1; 1:amino acid number 1; 2: number of heteroatom in amino
403       // acid number 1; 3:type of heteroatom in amino acid 1 (backbone etc.); 4:amino
404       // acid name 1; 5:amino acid number 1; 6: number of heteroatom in amino acid
405       // number 1; 7:type of heteroatom in amino acid 1 (backbone etc.)
406       intresulttable[u][0]=(lineaa1-z*arg3)+1;
407       intresulttable[u][1]=x+1;
408       intresulttable[u][2]=b+1;
409       intresulttable[u][3]=type1;
410       intresulttable[u][4]=(lineaa2-z*arg3)+1;
411       intresulttable[u][5]=y+1;
412       intresulttable[u][6]=f+1;
413       intresulttable[u][7]=type2;
414       resulttable[u][z]='1';
415       u++;
416     }
417   }
418 }
419 }
420 }
421 }
422 }
423 }
424 }
425 }
426
427 // Output: aa1-no, aa2-no, type1, type2, Het1, Het2, atomNo1, atomNo2
428 for(i=0;i<u;i++)
429 {
430   printf("%d-", intresulttable[i][1]);
431   printf("%d-", intresulttable[i][5]);

```



```

507 if(arg1==3)
508 {
509
510 int a,b,i,j,t,v;
511
512 // In intall all integer values of the input are stored
513 int *intall = malloc(arg2 * sizeof(int));
514 if(intall == NULL)
515 {printf("NO RAM\n");
516 return EXIT_FAILURE;
517 }
518
519 // In result summarized data is stored
520 int *result = malloc(arg2 * sizeof(int));
521 if(result == NULL)
522 {printf("NO RAM\n");
523 return EXIT_FAILURE;
524 }
525
526 // Read the data
527 for(i=0;i<arg2;i++)
528 scanf("%d-", &intall[i]);
529
530 //////////////////////////////////////
531
532 if(arg3==1)
533 {
534
535 t=0;
536 for(i=0;i<arg2;i++)
537 {
538
539 j=i;
540 for( ; ; )
541 {
542
543 for( ;j<arg2;j++)
544 {
545 if(intall[j]==newline) {j++; break;}
546 }
547
548 if(j>=arg2) break;
549
550 if(intall[i+1]==intall[j+1])
551 {
552
553 v=0;
554 for(a=0;intall[i+8+a]!=newline; a++)
555 {
556 for(b=0;intall[j+8+b]!=newline;b++)
557 {
558 if(intall[i+8+a]==intall[j+8+b])
559 {
560 if(v==0)
561 {
562 result[t]=intall[i];t++;
563 result[t]=intall[j];t++;
564 result[t]=intall[i+2]; t++;
565 result[t]=intall[j+2]; t++;
566 result[t]=intall[i+4]; t++;
567 result[t]=intall[j+4]; t++;
568 result[t]=intall[i+6]; t++;
569 result[t]=intall[j+6]; t++;
570 }
571 result[t]=intall[j+8+b]; t++;
572 v++;
573 }
574 }
575 }
576 if(v!=0) {result[t]=newline; t++;}
577 }
578
579 else // for if(intall[i+1]==intall[j+1])
580 {
581 for( ; ;i++)
582 {

```

```

583     if(intall[i]==endline) break;
584     }
585     break;
586     }
587     }
588     }
589
590     // Output
591     for(i=0;i<t;i++)
592     {
593         printf("%d-", result[i]);
594         if(result[i]==endline) printf("\n");
595     }
596
597     }
598
599     //////////////////////////////////////
600
601     if(arg3==2)
602     {
603
604         for(i=0;i<arg2; )
605         {
606
607             printf("%d-%d-%d-%d-%d-%d-%d-", intall[i], intall[i+1], intall[i+2],
608                 intall[i+3], intall[i+4], intall[i+5], intall[i+6], intall[i+7]);
609             t=0;
610             for(j=i+8;j<arg2;j++,t++)
611             {
612                 if(intall[j]==endline) {break;}
613                 result[t]=intall[j];
614             }
615             j=i;
616             for( ;j<arg2 ; )
617             {
618
619                 for( ;j<arg2;j++)
620                 {
621                     if(intall[j]==endline) {j++; break;}
622                 }
623
624                 if(intall[i]==intall[j] && intall[i+1]==intall[j+1] && intall[i+4]==intall[j+4] &&
625                     intall[i+5]==intall[j+5])
626                 {
627                     for( ;intall[j+8]!=endline;t++,j++) result[t]=intall[j+8];
628                 }
629                 else
630                 {
631                     result[t]=endline;
632
633                     for(a=0;result[a]!=endline;a++)
634                     {
635                         for(b=a+1;result[b]!=endline;b++)
636                         {
637
638                             for( ;result[b]!=endline; )
639                             {
640                                 if(result[a]==result[b]) { for(v=b+1; result[v-1]!=endline; v++) result[v-
641                                     1]=result[v]; }
642                                 else break;
643                             }
644                             if(result[b]<result[a])
645                             {
646                                 v=result[a];
647                                 result[a]=result[b];
648                                 result[b]=v;
649                             }
650                         }
651                     }
652
653                     for(v=0;result[v]!=endline;v++) printf("%d-", result[v]);
654                     printf("%d-\n",endline);
655

```

```

656     i=j;
657     break;
658 }
659 }
660 }
661 }
662
663 // Free memory
664 free(intall);
665 free(result);
666 }
667
668
669 ///////////////////////////////////////////////////
670 // Section 4: Summarize code- and residue-based interactions from Het-based input
671 // file: need arg1, arg2, arg3
672 ///////////////////////////////////////////////////
673
674 if(arg1==4)
675 {
676
677     int a,b,i,j,t,v;
678
679     // In intall all integer values of the input are stored
680     int *intall = malloc(arg2 * sizeof(int));
681     if(intall == NULL)
682     {printf("NO RAM\n");
683     return EXIT_FAILURE;
684     }
685
686     // In result summarized data is stored
687     int *result = malloc(arg2 * sizeof(int));
688     if(result == NULL)
689     {printf("NO RAM\n");
690     return EXIT_FAILURE;
691     }
692
693     // Read the data
694     for(i=0;i<arg2;i++)
695     scanf("%d-", &intall[i]);
696
697     ///////////////////////////////////////////////////
698
699     if(arg3==1)
700     {
701
702         for(i=0;i<arg2; )
703         {
704
705             printf("%d-%d-%d-%d-", intall[i], intall[i+1], intall[i+2], intall[i+3]);
706             t=0;
707             for(j=i+8;j<arg2;j++,t++)
708             {
709                 if(intall[j]==newline) {break;}
710                 result[t]=intall[j];
711             }
712
713             j=i;
714             for( ;j<arg2 ; )
715             {
716
717                 for( ;j<arg2;j++)
718                 {
719                     if(intall[j]==newline) {j++; break;}
720                 }
721
722                 if((intall[i]==intall[j] && intall[i+1]==intall[j+1] && intall[i+2]==intall[j+2]
723                 && intall[i+3]==intall[j+3]))
724                 {
725                     for( ;intall[j+8]!=newline;t++,j++) result[t]=intall[j+8];
726                 }
727
728                 else
729                 {
730                     result[t]=newline;

```

```

731 for(a=0;result[a]!=endl; a++)
732 {
733     for(b=a+1;result[b]!=endl; b++)
734     {
735
736         for( ;result[b]!=endl; )
737         {
738             if(result[a]==result[b]) { for(v=b+1; result[v-1]!=endl; v++) result[v-
739             1]=result[v]; }
740             else break;
741         }
742         if(result[b]<result[a])
743         {
744             v=result[a];
745             result[a]=result[b];
746             result[b]=v;
747         }
748     }
749 }
750
751 for(v=0;result[v]!=endl;v++) printf("%d-", result[v]);
752 printf("%d-\n",endl);
753
754 i=j;
755 break;
756 } // for else ...
757 } // for for( ;j<arg2 ; )
758 } // for for(i=0;i<arg2; )
759 } // for if(arg3==1)
760
761 //////////////////////////////////////
762
763 if(arg3==2)
764 {
765
766     for(i=0;i<arg2; )
767     {
768
769         printf("%d-%d-", intall[i], intall[i+1]);
770         t=0;
771         for(j=i+8;j<arg2;j++,t++)
772         {
773             if(intall[j]==endl) {break;}
774             result[t]=intall[j];
775         }
776
777         j=i;
778         for( ;j<arg2 ; )
779         {
780
781             for( ;j<arg2;j++)
782             {
783                 if(intall[j]==endl) {j++; break;}
784             }
785
786             if(intall[i]==intall[j] && intall[i+1]==intall[j+1])
787             {
788                 for( ;intall[j+8]!=endl;t++,j++) result[t]=intall[j+8];
789             }
790
791             else
792             {
793                 result[t]=endl;
794
795                 for(a=0;result[a]!=endl; a++)
796                 {
797                     for(b=a+1;result[b]!=endl; b++)
798                     {
799
800                         for( ;result[b]!=endl; )
801                         {
802                             if(result[a]==result[b]) { for(v=b+1; result[v-1]!=endl; v++) result[v-
803                             1]=result[v]; }
804                             else break;

```

```

805
806 if(result[b]<result[a])
807 {
808 v=result[a];
809 result[a]=result[b];
810 result[b]=v;
811 }
812 }
813 }
814
815 for(v=0;result[v]!=endline;v++) printf("%d-", result[v]);
816 printf("%d-\n",endline);
817
818 i=j;
819 break;
820 } // for else ...
821 } // for for( ;j<arg2 ; )
822 } // for for(i=0;i<arg2; )
823 } // for if(arg3==2)
824
825 //////////////////////////////////////
826
827 // Free memory
828 free(intall);
829 free(result);
830 }
831
832
833 //////////////////////////////////////
834 // Section 5: Calculate bb-interactions for helicity-check: need arg1, arg2,
835 // arg3, arg4, arg5, arg6
836 //////////////////////////////////////
837
838 if(arg1==5)
839 {
840
841 int a,b,i,j;
842 float contact;
843
844 // In alldata the input is stored
845 int *alldata = malloc(arg3 * sizeof(int));
846 if(alldata == NULL)
847 {printf("NO RAM\n");
848 return EXIT_FAILURE;
849 }
850
851 // Read the data
852 for(i=0;i<arg3;i++) scanf("%d",&alldata[i]);
853
854 // Check if there are 1-arg2 interactions (alpha-helix) and if so count the number
855 // of interactions
856
857 for(i=0, a=1 ;i<arg3; i++)
858 {
859
860 if(a==arg5-arg2) break;
861
862 if(alldata[i]==a)
863 {
864
865 // check if there are 1-4 interactions and if the first heteroatom (oxygen) has a
866 // bigger Het-no than the second (nitrogen)
867 if( (alldata[i]==(alldata[i+1]-arg2)) && (alldata[i+4]>alldata[i+5]) )
868 {
869
870 for(j=i+8,b=0; ;j++,b++)
871 {
872 if(alldata[j]==endline)
873 {
874 contact=1.0*(float)b/(float)arg4;
875 if(arg6==1) printf("%d\t%d\t%.3f\n", alldata[i], alldata[i+1], contact);
876 if(arg6==2) printf("%d-%d-", alldata[i], alldata[i+1]);
877 break;
878 }
879 }
880 if(arg6==2)

```

```

881 {
882 for(j=i+8; ;j++)
883 {
884 printf("%d-", alldata[j]);
885 if(alldata[j]==endl) {printf("\n"); break;}
886 }
887 }
888
889 a++;
890 }
891
892 for( ; ; i++) {if(alldata[i]==endl) break;}
893
894 }
895
896 else if(alldata[i]<a)
897 {
898 for( ; ; i++) {if(alldata[i]==endl) break;}
899 }
900
901 else
902 {
903 if(arg6==1) printf("%d\t%d\t0.000\n", a, a+arg2);
904 if(arg6==2) printf("%d-%d-%d-\n", a, a+arg2, endl);
905 a++;
906 i--;
907 }
908 }
909 }
910
911 // Free memory
912 free(alldata);
913 }
914
915
916 ///////////////////////////////////////////////////////////////////
917 // Section 6: Calculate percentage of frames where heteroatoms are hydrogen
918 // bonded: need arg1, arg3, arg4, arg5
919 ///////////////////////////////////////////////////////////////////
920
921 if(arg1==6)
922 {
923
924 int a,b,i,j,k,t,u,v,y,z;
925 int cmp, code;
926
927 // In matrixres the amino acid sequence is stored
928 char matrixres[arg5][5];
929 // Initialize matrixres with NULL
930 for (i=0; i<arg5; i++) {for (j=0; j<5; j++) matrixres[i][j]=0;}
931
932 // In arrayres the amino acid sequence as integer code is saved
933 int arrayres[arg5];
934
935 // In intall all values are stored
936 int *intall = malloc(arg3 * sizeof(int));
937 if(intall == NULL)
938 {printf("NO RAM\n");
939 return EXIT_FAILURE;
940 }
941
942 // In result summarized data is stored; *2 if there are double interactions at the
943 // same time point analyzed
944 int *result = malloc((arg4*2) * sizeof(int));
945 if(result == NULL)
946 {printf("NO RAM\n");
947 return EXIT_FAILURE;
948 }
949
950 // Read amino acid sequence
951 for(i=0; i<arg5; i++)
952 {
953 for(j=0; ; j++)
954 {
955 scanf("%c", &matrixres[i][j]);
956 if(matrixres[i][j]=='\n') {matrixres[i][j]=0; break;}

```

```

957     }
958     }
959
960     // Convert amino acids to numbers
961     for(i=0; i<arg5; i++)
962     {
963         for(j=0;j<no_def_res;j++)
964         {
965             if((cmp=strcmp(matrixres[i], AAname[j]))==0) {arrayres[i]=j; break;}
966         }
967     }
968
969     // Read the data
970     for(i=0;i<arg3;i++)
971         scanf("%d-", &intall[i]);
972
973     printf("aa-no\ttaa-name\tHet-no\tttype\tth-bonded\n");
974
975     for(i=0;i<arg5;i++)
976     {
977
978         j=arrayres[i];
979         // (charge_termini) if there is more than one charged termini add corresponding
980         // lines
981         // if(i==terminuslaa-1) j=arrayres[terminuslcode];
982
983         // Check all heteroatoms
984         for(k=0; k<AAcode[j][1]; k++)
985         {
986
987             for(z=0;z<2*arg4;z++) result[z]=0;
988
989             u=0;
990             for(y=0; y<arg3; y++)
991             {
992                 if( (intall[y]==i+1 && intall[y+4]==k+1) || (intall[y+1]==i+1 && intall[y+5]==k+1)
993                 )
994                 {
995                     for(t=y+8; ;u++, t++)
996                     {
997                         if(intall[t]==endline) break;
998                         result[u]=intall[t];
999                     }
1000                 } //if( (intall[y]==j &
1001
1002                 for( ; ; y++)
1003                 {
1004                     if(intall[y]==endline) break;
1005                 }
1006             } //for(y=0; y<arg3; y++)
1007
1008             if(j==0 || j==1) code=2; //0: ACE; 1: NH2
1009             else if(j==34) code=3; // DPPC
1010             else if(j==35 || j==36) code=4; // SOL and SOLc
1011             else if(j==37) code=5; // chloride
1012             else if(j==38) code=6; // ligand
1013             else if(AAcode[j][1]==2) code=0;
1014             else if( (AAcode[j][1]>2) && ((k==0) || (k==AAcode[j][1]-1)) ) code=0;
1015             else code=1; // remaining: side chain
1016
1017             if(u==0) printf("%d\t%s\t%d\t%d\t0.000\t%d-\n", i+1, matrixres[i],k+1, code,
1018                 endline);
1019             else
1020             {
1021                 result[u]=endline;
1022                 for(t=0;t<u-1;t++)
1023                 {
1024                     for(b=t+1;b<u;b++)
1025                     {
1026
1027                         if(result[b]==result[t]) { for(v=b+1; result[v-1]!=endline; v++) {result[v-1]=result[v];} u--;}
1028                     }
1029                 }

```

```

1030     if(result[b]<result[t])
1031     {
1032         v=result[t];
1033         result[t]=result[b];
1034         result[b]=v;
1035     }
1036 }
1037 } // else
1038
1039 printf("%d\t%s\t%d\t%d\t%.3f\t", i+1, matrixres[i],k+1, code, (float)(u-
1040 1)/(float)arg4); // (float)(u-1)/(float)arg4
1041 for(z=0;z<u;z++) printf("%d-", result[z]);
1042 printf("%d-\n", endl);
1043 } //else
1044 } //for(k=0; k<Acode[j][1]; k++) // check all heteroatoms
1045 } //for(i=0;i<arg5; )
1046 }
1047
1048
1049 ///////////////////////////////////////////////////////////////////
1050 // Section 7: Analyze H-bond data: need arg1, arg2, arg3, arg4, arg5
1051 ///////////////////////////////////////////////////////////////////
1052
1053 if(arg1==7)
1054 {
1055
1056     // Time step (in picoseconds) used for H-bond calculation for the final output
1057     int timestep = arg5;
1058     int a,b,c,i,j;
1059     int histo[arg4/arg6];
1060
1061     // In intall all values are stored
1062     int *intall = malloc(arg2 * sizeof(int));
1063     if(intall == NULL)
1064     {printf("NO RAM\n");
1065     return EXIT_FAILURE;
1066     }
1067
1068     // Read the data
1069     for(i=0;i<arg2;i++)
1070     scanf("%d-", &intall[i]);
1071
1072     if(arg3==2) printf("a1\taa2\tinteractions\t");
1073     if(arg3==3) printf("aa\tHet\tcode\tH-bonded\t");
1074     if(arg3==4) printf("a1\taa2\tc1\tc2\tinteractions\t");
1075     if(arg3==8)
1076     printf("a1\taa2\tc1\tc2\tHet1\tHet2\tatomNo1\tatomNo2\tinteractions\t");
1077
1078     for(i=0;i<(arg4/arg6);i++) printf("%d\t", timestep*((i+1)*arg6-(arg6/2)));
1079     printf("\n");
1080
1081     for(i=0;i<arg2;i++)
1082     {
1083         for(j=0;j<(arg4/arg6);j++) histo[j]=0;
1084
1085         for(j=i;j<i+arg3;j++) printf("%d\t", intall[j]);
1086         a=0;
1087         for(j=i+arg3; j++,a++)
1088         {
1089             for(b=timestep,c=0; b<arg4*timestep; b+=arg6*timestep,c++)
1090             {
1091                 if(intall[j]>=b && intall[j]<(b+arg6*timestep)) {histo[c]++; break;}
1092             }
1093             if(intall[j]==endl) break;
1094         }
1095         printf("%.3f\t", (float)a/(float)arg4);
1096         for(j=0;j<(arg4/arg6);j++) printf("%.3f\t", (float)histo[j]/(float)arg6);
1097         printf("\n");
1098         for(;i++) if(intall[i]==endl) break;
1099     }
1100
1101     free(intall);
1102 }
1103

```

```

1104
1105 ///////////////////////////////////////////////////
1106 // Section 8: Calculate a single H-bond: distance and angle values
1107 ///////////////////////////////////////////////////
1108
1109 if(arg1==8)
1110 {
1111
1112     int i,j;
1113
1114     float d;
1115     float alphab, alphaw, cosalphab;
1116     float pi=3.14159;
1117
1118     // Get memory from RAM for matrixcoord were all coordinates are stored
1119     float ** matrixcoord;
1120     matrixcoord = malloc(arg2 * sizeof(float *));
1121     if(NULL == matrixcoord)
1122     {
1123         printf("NO RAM for matrixcoord 1!");
1124         return EXIT_FAILURE;
1125     }
1126     for(i=0; i<arg2; i++)
1127     {
1128         matrixcoord[i] = malloc(3 * sizeof(float));
1129         if(NULL == matrixcoord[i])
1130         {
1131             printf("NO RAM for matrixcoord 2!");
1132             return EXIT_FAILURE;
1133         }
1134     }
1135
1136     // Read coordinates
1137     for(i=0; i<arg2; i++)
1138     {
1139         scanf("%f", &matrixcoord[i][0]);
1140         scanf("%f", &matrixcoord[i][1]);
1141         scanf("%f", &matrixcoord[i][2]);
1142     }
1143
1144     // Calculate distance and angle
1145     for(i=0; i<arg2; i+=3)
1146     {
1147         d=sqrt((matrixcoord[i][0]-matrixcoord[i+1][0])*(matrixcoord[i][0]-
matrixcoord[i+1][0])+(matrixcoord[i][1]-matrixcoord[i+1][1])*(matrixcoord[i][1]-
matrixcoord[i+1][1])+(matrixcoord[i][2]-matrixcoord[i+1][2])*(matrixcoord[i][2]-
matrixcoord[i+1][2]));
1148
1149         cosalphab=((matrixcoord[i][0]-matrixcoord[i+2][0])*(matrixcoord[i+1][0]-
matrixcoord[i+2][0])+(matrixcoord[i][1]-matrixcoord[i+2][1])*(matrixcoord[i+1][1]-
matrixcoord[i+2][1])+(matrixcoord[i][2]-matrixcoord[i+2][2])*(matrixcoord[i+1][2]-
matrixcoord[i+2][2]))/((sqrt((matrixcoord[i][0]-
matrixcoord[i+2][0])*(matrixcoord[i][0]-matrixcoord[i+2][0])+(matrixcoord[i][1]-
matrixcoord[i+2][1])*(matrixcoord[i][1]-matrixcoord[i+2][1])+(matrixcoord[i][2]-
matrixcoord[i+2][2])*(matrixcoord[i][2]-
matrixcoord[i+2][2]))*sqrt((matrixcoord[i+1][0]-
matrixcoord[i+2][0])*(matrixcoord[i+1][0]-
matrixcoord[i+2][0])+(matrixcoord[i+1][1]-
matrixcoord[i+2][1])*(matrixcoord[i+1][1]-
matrixcoord[i+2][1])+(matrixcoord[i+1][2]-
matrixcoord[i+2][2])*(matrixcoord[i+1][2]-matrixcoord[i+2][2]))));
1150         alphab=acos( cosalphab );
1151         alphaw=alphab*180/pi;
1152
1153         if(d<=cutoffdist && alphaw>=cutoffang) j=1; else j=0;
1154
1155         printf("%d\t%.3f\t%.3f\t%d\n", i/3*arg3+arg3, d, alphaw, j);
1156
1157     }
1158
1159     // Free memory
1160     for(i=0; i<arg2; i++)
1161         free(matrixcoord[i]);
1162     free(matrixcoord);
1163 }
1164

```

```

1165 ///////////////////////////////////////////////////
1166 // Section 9: Generate a xpm picture for one hydrogen bonding interaction
1167 ///////////////////////////////////////////////////
1168
1169 if(arg1==9)
1170 {
1171
1172
1173     int d,i,j;
1174
1175     // In alldata the input is stored
1176     int *alldata = malloc((arg2+arg3) * sizeof(int));
1177     if(alldata == NULL)
1178     {printf("NO RAM\n");
1179     return EXIT_FAILURE;
1180     }
1181
1182     // Read the data
1183     for(i=0;i<(arg2+arg3);i++)
1184     {
1185         scanf("%d-",&alldata[i]);
1186         if (alldata[i]==endline) break;
1187     }
1188
1189     // Print header
1190     printf("/ * XPM * /\n");
1191     printf("static char * pic_xpm[] = {\n");
1192     printf("\"\"%d %d 3 1\", \n", arg3, arg3/20);
1193     printf("\"\" \\tc None\", \n");
1194     printf("\"\"0\\tc #FFFFFF\", \n");
1195     printf("\"\"1\\tc #000000\", \n");
1196
1197     // Print the data
1198
1199     for(j=0; j<arg3/20; j++) // Height of the image is 1/20 of ist width
1200     {
1201
1202         printf("\"");
1203         for(i=arg2, d=1 ;d<arg3+1; d++)
1204         {
1205             if(alldata[i]/arg4==d) {printf("1"); i++;}
1206             else printf("0");
1207         }
1208         if(j==(arg3/20)-1) printf("\"}"); else printf "\",");
1209         printf("\n");
1210     }
1211
1212     // Free memory
1213     free(alldata);
1214 }
1215
1216 return EXIT_SUCCESS;
1217 }

```

9.4 Calculation of hydrophobic contacts: *gro_contacts*

9.4.1 Parameter file *gro_contacts-para.txt*

```

1  # PLEASE ENTER: name of the simulation
2  var-sim=inact-mdl-2
3
4  # PLEASE ENTER: name of the gro-file (also xtc and tpr files must have this
5  # name; otherwise change respective lines in gro_contacts.sh)
6  var-filename=../../hH2R_inactive_2VT4_md1_2
7
8  # PLEASE ENTER: duration of the simulation in picoseconds
9  var-duration=80000
10
11 # PLEASE ENTER: time step for analysis
12 var-timestep=10
13
14 # PLEASE ENTER: cut-off value for contacts in picometer (the number has to be
15 # divided by 10 without a rest)
16 var-cutoff=500
17
18 # PLEASE ENTER: the number of amino acids, lipids and ligands in the coordinate
19 # file (has to be in one sequence)
20 var-last=412

```

9.4.2 Reference file *contacts-atoms.txt*

```

1  ALA CB
2  ARG CB
3  ARG CG
4  ARG CD
5  ARGN CB
6  ARGN CG
7  ARGN CD
8  ASN CB
9  ASN1 CB
10 ASP CB
11 ASPH CB
12 CYS CB
13 CYSH CB
14 CYS1 CB
15 CYS1 SG
16 CYS2 CB
17 CYS2 SG
18 GLN CB
19 GLN CG
20 GLU CB
21 GLU CG
22 GLUH CB
23 GLUH CG
24 GLY CA
25 HISA CB
26 HISA CG
27 HISA CD2
28 HISA CE1
29 HISB CB
30 HISB CG
31 HISB CD2
32 HISB CE1
33 HISH CB

```

34	HISH CG
35	HISH CD2
36	HISH CE1
37	HIS1 CB
38	HIS1 CG
39	HIS1 CD2
40	HIS1 CE1
41	HIS2 CB
42	HIS2 CG
43	HIS2 CD2
44	HIS2 CE1
45	ILE CB
46	ILE CG1
47	ILE CG2
48	ILE CD
49	LEU CB
50	LEU CG
51	LEU CD1
52	LEU CD2
53	LYS CB
54	LYS CG
55	LYS CD
56	LYS CE
57	LYSH CB
58	LYSH CG
59	LYSH CD
60	LYSH CE
61	MET CB
62	MET CG
63	MET CE
64	PHE CB
65	PHE CG
66	PHE CD1
67	PHE HD1
68	PHE CD2
69	PHE HD2
70	PHE CE1
71	PHE HE1
72	PHE CE2
73	PHE HE2
74	PHE CZ
75	PHE HZ
76	PRO CB
77	PRO CG
78	PRO CD
79	SER CB
80	THR CB
81	THR CG2
82	TRP CB
83	TRP CG
84	TRP CD1
85	TRP HD1
86	TRP CD2
87	TRP CE2
88	TRP CE3
89	TRP HE3
90	TRP CZ2
91	TRP HZ2
92	TRP CZ3
93	TRP HZ3
94	TRP CH2
95	TRP HH2
96	TYR CB
97	TYR CG
98	TYR CD1
99	TYR HD1
100	TYR CD2
101	TYR HD2
102	TYR CE1
103	TYR HE1
104	TYR CE2
105	TYR HE2
106	TYR CZ
107	VAL CB
108	VAL CG1
109	VAL CG2

```

110 DPPC C22
111 DPPC C23
112 DPPC C24
113 DPPC C25
114 DPPC C26
115 DPPC C27
116 DPPC C28
117 DPPC C29
118 DPPC C210
119 DPPC C211
120 DPPC C212
121 DPPC C213
122 DPPC C214
123 DPPC C215
124 DPPC C216
125 DPPC C12
126 DPPC C13
127 DPPC C14
128 DPPC C15
129 DPPC C16
130 DPPC C17
131 DPPC C18
132 DPPC C19
133 DPPC C110
134 DPPC C111
135 DPPC C112
136 DPPC C113
137 DPPC C114
138 DPPC C115
139 DPPC C116
140 HAH C3
141 HAH H3
142 HAH C1
143 HAH H1
144 HAH C2
145 HAH C8
146 HAH C10

```

9.4.3 Shell script `gro_contacts.sh`

```

1  #!/bin/bash
2
3  #####
4  ### Content:
5  # Calculates contacts between hydrophobic sites
6  # A file with atoms should be provided: "contacts-atoms.txt"
7  #####
8
9  # Compilation of the C program gro_contacts-calc
10 gcc gro_contacts-calc.c -o gro_contacts-calc
11
12 # Extract variables from parameter.txt
13 descr=contacts-para.txt
14
15 sim=`sed -n /var-sim=/p $descr | cut -d= -f2`
16 filename=`sed -n /var-filename=/p $descr | cut -d= -f2`
17 duration=`sed -n /var-duration=/p $descr | cut -d= -f2`
18 timestep=`sed -n /var-timestep=/p $descr | cut -d= -f2`
19 cutoff=`sed -n /var-cutoff=/p $descr | cut -d= -f2`
20 last=`sed -n /var-last=/p $descr | cut -d= -f2`
21 histo=`sed -n /var-histo=/p $descr | cut -d= -f2`
22
23 # Extract residue name
24 cut -c1-9 $filename.gro | sed -n '1,2!p' | sed '$d' | sort -n -u | cut -c6-9 | sed
's/ //g' | sed -n 1,"$last"p > aa-name.txt
25 residues=`wc -l aa-name.txt | cut -d\  -f 1`
26
27 # Create index-file for g mdat

```

```

28 echo -e "del 10-20 \n \nq\n" | make_ndx -f $filename.gro -o index_contacts.ndx >>
make_index.txt 2>&1;
29 echo \[ contacts \] >> index_contacts.ndx;
30
31 # Prepare the gro-file for extracting data
32 sed 's/ / /g' $filename.gro > $filename.txt
33 delspace=1
34 while [ $delspace -le 10 ]; do
35 sed 's/ / /g' $filename.txt > $filename.txt1
36 rm -f $filename.txt
37 mv $filename.txt1 $filename.txt
38 delspace=`expr $delspace + 1`
39 done
40
41 # Extract the atom numbers of the atoms out of the modified gro-file
42 lineno=`/usr/bin/wc -l contacts-atoms.txt | cut -d\ -f 1`
43 linecount=1
44 while [ $linecount -le $lineno ]; do
45 var1=`sed -n '$linecount'p' contacts-atoms.txt`
46 sed -n "/$var1\ /p" $filename.txt | cut -d\ -f 4 >> index_contacts_tmp1.ndx
47 linecount=`expr $linecount + 1`
48 done
49
50 # Sort the atom numbers: necessary for g_mdmat
51 sort -n index_contacts_tmp1.ndx > index_contacts_tmp2.ndx
52 rm -f index_contacts_tmp1.ndx
53
54 # Add the atom numbers to the index file
55 cat index_contacts.ndx index_contacts_tmp2.ndx > index_contacts_.ndx
56 rm -f index_contacts.ndx
57 rm -f index_contacts_tmp2.ndx
58 mv index_contacts_.ndx index_contacts.ndx
59
60 # Resort the index-file (to GROMACS format)
61 echo -e "\nq\n" | make_ndx -f $filename.gro -n index_contacts.ndx -o
index_contacts_.ndx >> make_index.txt 2>&1;
62 rm -f index_contacts.ndx
63 mv index_contacts_.ndx index_contacts.ndx
64 rm -f make_index.txt
65
66 # g_mdmat
67 echo 10 | g_mdmat -f $filename.xtc -s $filename.tpr -frames all.xpm -mean
dm_all.xpm -n index_contacts.ndx -dt $timestep -t 1.0 -nlevels 51 -e $duration >>
g_mdmat-result_all.txt 2>&1
68
69 # Reduction of the output-file to important lines
70 sed -e '/yyy/!d' all.xpm > all_reorder.txt
71 sed -e 's/,//g' all_reorder.txt > all_reorder_.txt
72 rm -f all_reorder.txt
73 sed -e 's/"//g' all_reorder_.txt > all_reorder.txt
74 rm -f all_reorder_.txt
75
76 # Number of lines of the modified output file of g_mdmat
77 var2=`/usr/bin/wc -l all_reorder.txt | cut -d\ -f 1`
78
79 # File size of the modified output file of g_mdmat
80 bytes=`wc -c all_reorder.txt | cut -d\ -f 1`
81
82 # Determine number of frames and number of amino acids: input to C-program
83 frames=`sed -n /t=/p all.xpm | wc -l`
84 reshydr=`expr $var2 / $frames`
85
86 cat aa-name.txt all_reorder.txt > input.txt
87
88 ./gro_contacts-calc $bytes $frames $reshydr $residues $cutoff < input.txt > $sim-
contacts.txt
89 cut -f 1-3 $sim-contacts.txt > $sim-contacts_sum.txt
90
91 rm -f aa-name.txt
92 rm -f all.xpm
93 rm -f dm_all.xpm
94 rm -f g_mdmat-result_all.txt
95 rm -f all_reorder.txt
96 rm -f input.txt
97 rm -f index_contacts.ndx
98 rm -f gro_contacts-calc

```

9.4.4 C program `gro_contacts-calc.c`

```

1  #include <stdio.h>
2  #include <stdlib.h>
3  #include <string.h>
4
5  // Compile with:
6  // gcc gro_contacts-calc.c -o gro_contacts-calc
7
8  // Number of residues which are defined by the AAcode in this program
9  #define no_def_res 39
10
11 int main(int argc, char *argv[]) {
12
13     int bytes, frames, reshydr, residues, cutoff;
14
15     // With parameters in the shell script the following variables are defined
16     bytes = strtol(argv[1], NULL, 10);
17     frames = strtol(argv[2], NULL, 10);
18     reshydr = strtol(argv[3], NULL, 10);
19     residues = strtol(argv[4], NULL, 10);
20     cutoff = strtol(argv[5], NULL, 10);
21
22
23     int a,b,c,i,j,r,s,cmp,amount,count,far,aa1,aa2,aa1temp,aa2temp,aa1print,aa2print;
24     float d;
25     char *array;
26     char aatmp[frames];
27     float aares[frames];
28     float y = ((float) cutoff)/1000;
29
30     // In matrixres the amino acid sequence is stored
31     char matrixres1[residues][5];
32     // Initialize matrixres with NULL
33     for (i=0; i<residues; i++) {for (j=0; j<5; j++) matrixres1[i][j]=0;}
34
35     // In arrayres the amino acid sequence as integer code is saved
36     int arrayres[residues];
37
38     // In matrixres2 the amino acid sequence with hydrophobic atoms as integer code is
39     // saved
40     int matrixres2[residues];
41
42
43     // Memory for array
44     array = calloc(bytes,sizeof(char));
45     if(NULL == array)
46     {
47         printf("NO RAM!");
48         return EXIT_FAILURE;
49     }
50
51
52     char AAname[no_def_res][5]={
53         {'A','C','E'}, // 0
54         {'N','H','2'}, // 1
55         {'A','L','A'}, // 2
56         {'A','R','G'}, // 3
57         {'A','R','G','N'}, // 4
58         {'A','S','N'}, // 5
59         {'A','S','N','1'}, // 6
60         {'A','S','P'}, // 7
61         {'A','S','P','H'}, // 8
62         {'C','Y','S'}, // 9
63         {'C','Y','S','H'}, // 10
64         {'C','Y','S','1'}, // 11
65         {'C','Y','S','2'}, // 12
66         {'G','L','N'}, // 13
67         {'G','L','U'}, // 14
68         {'G','L','U','H'}, // 15
69         {'G','L','Y'}, // 16
70         {'H','I','S','A'}, // 17

```

```

71      {'H','I','S','B'}, // 18
72      {'H','I','S','H'}, // 19
73      {'H','I','S','L'}, // 20
74      {'H','I','S','2'}, // 21
75      {'I','L','E'}, // 22
76      {'L','E','U'}, // 23
77      {'L','Y','S'}, // 24
78      {'L','Y','S','H'}, // 25
79      {'M','E','T'}, // 26
80      {'P','H','E'}, // 27
81      {'P','R','O'}, // 28
82      {'S','E','R'}, // 29
83      {'T','H','R'}, // 30
84      {'T','R','P'}, // 31
85      {'T','Y','R'}, // 32
86      {'V','A','L'}, // 33
87      {'D','P','P','C'}, // 34
88      {'S','O','L'}, // 35
89      {'S','O','L','c'}, // 36
90      {'C','L','-'} // 37
91      {'H','A','H'} // 38
92    };
93
94    // 0: no hydrophobic atom; 1: hydrophobic atom in residue (sequence according to
95    // AName)
96    int AAcode[no_def_res]=
97    {0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,1};
98
99    // Read amino acid sequence
100   for(i=0; i<residues; i++)
101   {
102     for(j=0; j<jmax; j++)
103     {
104       scanf("%c", &matrixresl[i][j]);
105       if(matrixresl[i][j]=='\n'){matrixresl[i][j]=0; break;}
106     }
107   }
108
109   // Convert amino acids to numbers
110   for(i=0; i<residues; i++)
111   {
112     for(j=0; j<no_def_res; j++)
113     {
114       if((cmp=strncmp(matrixresl[i], AName[j]))==0) {arrayres[i]=j; break;}
115     }
116   }
117
118   // Create the sequence of residues with hydrophobic atoms
119   for(i=0, a=0; i<residues; i++)
120   {
121     j=arrayres[i];
122     if(AAcode[j]==1) {matrixres2[a]=i+1; a++;}
123   }
124
125   // Read g_mdmat data
126   for(i=0; i<bytes; i++)
127     scanf("%c", &array[i]);
128
129   // Header for output
130   printf("aal\ttaa2\tcontacts<cutoff\tcontacts>2.0 nm\t");
131   for(i=0; i<frames; i++) printf("%d\t", i*10);
132   printf("\n");
133
134   // Analyze
135   for(aal=(reshydr-1), s=1; aal>0; aal--, s++)
136   {
137     for(aa2=s; aa2<reshydr; aa2++)
138     {
139       for(i=(reshydr+1)*aal+aa2, c=0; c<frames; c++, i+=((reshydr+1)*reshydr))
140       {
141         aatmp[c]=array[i];
142
143         count=0;
144         for(b=0; b<frames; b++)

```

```

146     {
147         if( ((aatmp[b]>64 && aatmp[b]<91) || (aatmp[b]>96 && aatmp[b]<122)) &&
aatmp[b]<(cutoff/10/2)+65 ) count++;
148     }
149
150
151     if(count>0)
152     {
153         far=0;
154         for(r=0; r<frames; r++)
155         {
156             for(a = 65; a < 91; a++)
157             {
158                 if (aatmp[r] == a)
159                 {d = (a-64)*0.02; aares[r]=d;}
160             }
161             for(a = 97; a < 121; a++)
162             {
163                 if (aatmp[r] == a)
164                 {d = (a-70)*0.02; aares[r]=d;}
165             }
166             if (aatmp[r] == 121)
167             {d = 2.0; aares[r]=d;far++;}
168         }
169     }
170
171     aalprint=reshydr-aal-1;
172     aa2print=aa2;
173     if(aalprint>aa2print) {aaltemp=aa2print; aa2temp=aalprint;} else
174     {aaltemp=aalprint; aa2temp=aa2print;}
175     printf("%d\t%d\t%.4f\t%.4f\t", matrixres2[aaltemp], matrixres2[aa2temp],
count*1.0/frames, far*1.0/frames);
176     for(r=0; r<frames; r++)
177     printf("%.2f\t", aares[r]);
178     printf("\n");
179 }
180 }
181 }
182
183 free(array);
184
185 return EXIT_SUCCESS;
186
187 }

```

9.5 Structure validation of MD simulations: *gro_validation*

9.5.1 Parameter file *gro_validation-para.txt*

```

1  # PLEASE ENTER: name of the simulation
2  var-sim=inact-md1-2
3
4  # PLEASE ENTER: name of the gro file (also xtc and tpr files must have this
5  # name; otherwise change respective lines in gro_validation.sh)
6  var-filename=../../hH2R_inactive_2VT4_md1_2
7
8  # PLEASE ENTER: duration of the simulation in picoseconds
9  var-duration=80000
10
11 # PLEASE ENTER: time step for analysis

```

```

12 var-timestep=10
13
14 # PLEASE ENTER: the residue number in the gro file of the FIRST amino acid to
15 # analyze
16 var-first=1
17
18 # PLEASE ENTER: the number of the LAST amino acid to analyze
19 var-last=292

```

9.5.2 Shell script *gro_validation.sh*

```

1  #!/bin/bash
2
3  #####
4  ### Content:
5  #   Chirality check of amino acids
6  #   Check for planarity of aromatic side chains and of polar amino acid side
7  #   chains containing a delocalized pi-electron system
8  #   Check peptide dihedral angle (peptide bond) for planarity
9  #   Compare the side chain rotamers with experimental values of Lovell et al.
10 #   (2000)
11 #   Make a ramachandran analysis using the data provided by Lovell et al. (2000)
12
13 ### Reference files (folder: reference)
14 #   chirality-dihedral.txt -> contains a list of amino acid atoms which were used
15 #   to calculate the dihedral angle at C-alpha atoms to check for proper
16 #   chirality
17 #   improvers.txt -> contains a list of amino acid atoms to calculate the dihedral
18 #   angles for the planarity check
19 #   dihedrals.txt -> contains a list of amino acid atoms defining the dihedral
20 #   angles of the side chain rotamers
21 #   lovell-rotamer.data -> reordered reference data for rotamer analysis
22 #   from Lovell et al. (2000)
23 #   lovell-rama.data -> reordered reference data for ramachandran analysis from
24 #   Lovell et al. (2000)
25
26
27 #####
28 #### Section 1
29 #####
30
31 # Compilation of the C program gro_validation-calc
32 gcc gro_validation-calc.c /usr/lib64/libm.a -o gro_validation-calc
33
34 # In 'output' all result files are stored
35 mkdir output
36
37 # Extract variables from parameter.txt
38 descr=gro_validation-para.txt
39
40 sim=`sed -n /var-sim=/p $descr | cut -d= -f2`
41 filename=`sed -n /var-filename=/p $descr | cut -d= -f2`
42 duration=`sed -n /var-duration=/p $descr | cut -d= -f2`
43 timestep=`sed -n /var-timestep=/p $descr | cut -d= -f2`
44 first=`sed -n /var-first=/p $descr | cut -d= -f2`
45 last=`sed -n /var-last=/p $descr | cut -d= -f2`
46
47 # Number of amino acids
48 aminoacids=`expr $last - $first + 1`
49
50 # Number of frames to analyze
51 frames=`expr $duration / $timestep`
52 lastframe=`expr $duration + 1`
53
54 # Get a list of the amino acid numbering and sequence
55 cut -c1-9 $filename.gro | sed -n 1,2!p | sort -u -n | sed -n "$first","$last"p |

```

```

56 cut -c 1-5 | sed 's/ \+//g' > aa-no.txt
57 cut -c1-9 $filename.gro | sed -n 1,2!'p | sort -u -n | sed -n "$first","$last"p |
58 cut -c 6-9 | sed 's/ \+//g' > aa-list.txt
59
60 # Generate a modified gro file
61 paste -d- aa-no.txt aa-list.txt | sed -n 's/-//p' > aa-no_list.txt
62 number=1
63 while [ $number -le $aminoacids ]; do
64   var=`sed -n '$number'p' aa-no_list.txt`
65   sed -n '/[\ ^]"$var"/p $filename.gro | sed 's/ \+ /g' >> gro-file-mod_.txt
66   number=`expr $number + 1`
67 done
68 cut -d\ -f1-4 gro-file-mod_.txt > gro-file-mod.txt
69 rm -f gro-file-mod_.txt
70
71 #####
72 ### Section 2: Get the atom numbers for the chirality check
73 #####
74 end=`wc -l reference/chirality-dihedral.txt | cut -d\ -f 1`
75
76 l1=1
77 while [ $l1 -le $end ]; do
78
79   var=`sed -n "$l1"p reference/chirality-dihedral.txt`
80   sed -n "/$var\" /p gro-file-mod.txt | cut -d\ -f 4 > atom-numbers_0.txt
81   l1=`expr $l1 + 1`
82   var=`sed -n "$l1"p reference/chirality-dihedral.txt`
83   sed -n "/$var\" /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
84   l1=`expr $l1 + 1`
85   var=`sed -n "$l1"p reference/chirality-dihedral.txt`
86   sed -n "/$var\" /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
87   l1=`expr $l1 + 1`
88   var=`sed -n "$l1"p reference/chirality-dihedral.txt`
89   sed -n "/$var\" /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
90   l1=`expr $l1 + 1`
91
92   l2=`wc -l atom-numbers_0.txt | cut -d\ -f 1`
93   if [ $l2 -gt 1 ]; then
94
95     step=`expr $l2 / 4`
96     a1=1
97     while [ $a1 -le $step ]; do
98
99       a2=`expr $a1 + $step`
100      a3=`expr $a1 + $step + $step`
101      a4=`expr $a1 + $step + $step + $step`
102      sed -n "$a1"p atom-numbers_0.txt >> atom-numbers_1.txt
103      sed -n "$a2"p atom-numbers_0.txt >> atom-numbers_1.txt
104      sed -n "$a3"p atom-numbers_0.txt >> atom-numbers_1.txt
105      sed -n "$a4"p atom-numbers_0.txt >> atom-numbers_1.txt
106
107      a1=`expr $a1 + 1`
108
109     done
110
111     cat atom-numbers_1.txt | tr "\n" " " > atom-numbers_2.txt
112     rm -f atom-numbers_1.txt
113     l3=1
114     while [ $l3 -le $l2 ]; do
115
116       l4=`expr $l3 + 3`
117       cut -d\ -f "$l3"-"$l4" atom-numbers_2.txt >> atom-numbers_3.txt
118       l3=`expr $l3 + 4`
119
120     done
121
122     else
123       echo > noaa.txt
124       fi
125
126     done
127
128     sort -n atom-numbers_3.txt >> atom-numbers_final-chirality.txt
129     rm -f atom-numbers_0.txt

```

```

130 rm -f atom-numbers_1.txt
131 rm -f atom-numbers_2.txt
132 rm -f atom-numbers_3.txt
133 rm -f noaa.txt
134
135
136 #####
137 ### Section 3: Get the atom numbers for the planarity check
138 #####
139
140 end=`wc -l reference/impropers.txt | cut -d\ -f 1`
141 l1=1
142 while [ $l1 -le $end ]; do
143
144     var=`sed -n "$l1"p reference/impropers.txt`
145     sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 > atom-numbers_0.txt
146     l1=`expr $l1 + 1`
147     var=`sed -n "$l1"p reference/impropers.txt`
148     sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
149     l1=`expr $l1 + 1`
150     var=`sed -n "$l1"p reference/impropers.txt`
151     sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
152     l1=`expr $l1 + 1`
153     var=`sed -n "$l1"p reference/impropers.txt`
154     sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
155     l1=`expr $l1 + 1`
156
157     l2=`wc -l atom-numbers_0.txt | cut -d\ -f 1`
158     if [ $l2 -gt 1 ]; then
159
160         step=`expr $l2 / 4`
161         a1=1
162         while [ $a1 -le $step ]; do
163
164             a2=`expr $a1 + $step`
165             a3=`expr $a1 + $step + $step`
166             a4=`expr $a1 + $step + $step + $step`
167             sed -n "$a1"p atom-numbers_0.txt >> atom-numbers_1.txt
168             sed -n "$a2"p atom-numbers_0.txt >> atom-numbers_1.txt
169             sed -n "$a3"p atom-numbers_0.txt >> atom-numbers_1.txt
170             sed -n "$a4"p atom-numbers_0.txt >> atom-numbers_1.txt
171
172             a1=`expr $a1 + 1`
173         done
174
175         cat atom-numbers_1.txt | tr "\n" " " > atom-numbers_2.txt
176         rm -f atom-numbers_1.txt
177         l3=1
178         while [ $l3 -le $l2 ]; do
179
180             l4=`expr $l3 + 3`
181             cut -d\ -f "$l3"-"$l4" atom-numbers_2.txt >> atom-numbers_3.txt
182             l3=`expr $l3 + 4`
183
184         done
185
186     else
187         echo > noaa.txt
188     fi
189
190 done
191
192 sort -n atom-numbers_3.txt >> atom-numbers_final-impropers.txt
193 rm -f atom-numbers_0.txt
194 rm -f atom-numbers_1.txt
195 rm -f atom-numbers_2.txt
196 rm -f atom-numbers_3.txt
197 rm -f noaa.txt
198
199
200 #####
201 ### Section 4: Get the atom numbers for the side chain rotamers
202 #####
203
204 line=`wc -l reference/dihedrals.txt | cut -d\ -f 1`
205 l1=2

```

```

206 while [ $l1 -le $line ]; do
207
208   var=`sed -n "$l1"p reference/dihedrals.txt`
209   sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 > atom-numbers_0.txt
210   l1=`expr $l1 + 1`
211   var=`sed -n "$l1"p reference/dihedrals.txt`
212   sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
213   l1=`expr $l1 + 1`
214   var=`sed -n "$l1"p reference/dihedrals.txt`
215   sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
216   l1=`expr $l1 + 1`
217   var=`sed -n "$l1"p reference/dihedrals.txt`
218   sed -n /"$var"\ /p gro-file-mod.txt | cut -d\ -f 4 >> atom-numbers_0.txt
219   l1=`expr $l1 + 1`
220
221   sort -n atom-numbers_0.txt > atom-numbers_1.txt
222   end=`wc -l atom-numbers_1.txt | cut -d\ -f 1`
223   cat atom-numbers_1.txt | tr "\n" " " > atom-numbers_2.txt
224   l2=1
225   while [ $l2 -le $end ]; do
226
227     l3=`expr $l2 + 3`
228     cut -d\ -f "$l2"-"$l3" atom-numbers_2.txt >> atom-numbers_3.txt
229     l2=`expr $l2 + 4`
230
231   done
232
233 done
234
235 sort -n atom-numbers_3.txt >> atom-numbers_final-rotamer.txt
236 rm -f atom-numbers_0.txt
237 rm -f atom-numbers_1.txt
238 rm -f atom-numbers_2.txt
239 rm -f atom-numbers_3.txt
240
241
242 #####
243 ### Section 5: Measure dihedrals for chirality, planarity and rotamers
244 #####
245
246 echo \[ dihedrals \] > atom-numbers_final-chirality-planarity-rotamer.ndx
247 cat atom-numbers_final-chirality.txt atom-numbers_final-impropers.txt atom-
  numbers_final-rotamer.txt >> atom-numbers_final-chirality-planarity-rotamer.ndx
248
249 # Measure all dihedrals in protein side chain
250 g_angle -f $filename.xtc -b 1 -e $lastframe -dt $timestep -all -n atom-
  numbers_final-chirality-planarity-rotamer.ndx -type dihedral -od $sim-dihedrals-
  angdist.xvg -ov $sim-dihedrals-angaver.xvg > g_angle-result.txt 2>&1
251
252 # Extract important columns from the output of g_angle
253 sed -n '/[#@]!/p' $sim-dihedrals-angaver.xvg | sed 's/^ \+//g' | sed 's/ \+ /g'
  | cut -d\ -f 3- > $sim-dihedrals-angaver-imp.txt
254
255
256 #####
257 ### Section 6: Chirality check
258 #####
259
260 beg=1
261 end=`wc -l atom-numbers_final-chirality.txt | cut -d\ -f 1`
262 cut -d\ -f $beg-$end $sim-dihedrals-angaver-imp.txt > $sim-chirality-protein-
  angaver-imp.txt
263 # Generate input file from aa-no, aa-list and measured angles
264 cat aa-no.txt aa-list.txt $sim-chirality-protein-angaver-imp.txt > input-
  chirality.txt
265 # Number of impropers
266 intangles=`wc -l atom-numbers_final-chirality.txt | cut -d\ -f 1`
267
268 ./gro_validation-calc 1 $aminoacids $frames $intangles 0 0 < input-chirality.txt >
  output/$sim-chirality.txt
269
270
271 #####
272 ### Section 7: Planarity check
273 #####
274

```

```

275 beg=`expr $end + 1`
276 end=`wc -l atom-numbers_final-impropers.txt | cut -d\ -f 1`
277 end=`expr $beg + $end`
278 end=`expr $end - 1`
279 cut -d\ -f $beg-$end $sim-dihedrals-angaver-imp.txt > $sim-impropers-protein-
    angaver-imp.txt
280 # Generate input file from aa-no, aa-list and measured angles
281 cat aa-no.txt aa-list.txt $sim-impropers-protein-angaver-imp.txt > input-
    planarity.txt
282 # Number of impropers in protein side chain
283 intangles=`wc -l atom-numbers_final-impropers.txt | cut -d\ -f 1`
284 # Complete output with all single impropers (mean, SD, and values within 2*SD
285 # interval, histogram) incl. residue based mean and histogram
286 ./gro_validation-calc 2 $aminoacids $frames $intangles 0 0 < input-planarity.txt >
    output/$sim-planarity.txt
287
288
289 #####
290 ### Section 8: Check peptide dihedral angle
291 #####
292
293 # In g_chi the output files of g_chi are stored; in resid the modified files
294 mkdir g_chi-omega
295 mkdir resid-omega
296
297 cd g_chi-omega
298 g_chi -f ../$filename.xtc -s ../$filename.tpr -omega -all -dt $timestep -b 0 -e
    $lastframe >> ../g_chi.txt 2>&1
299 cd ..
300
301 # List of aa without ACE and NH2
302 paste -d- aa-list.txt aa-no.txt | sed -n /ACE/'!'p | sed -n /NH2/'!'p > aa-
    no_list-omega.txt
303 aminoacidsomega=`wc -l aa-no_list-omega.txt | cut -d\ -f 1`
304
305 cut -d- -f 1 aa-no_list-omega.txt > omega-aa-list.txt
306 cut -d- -f 2 aa-no_list-omega.txt > omega-aa-no.txt
307
308 # Print header
309 ./gro_validation-calc 3 1 $frames 0 0 0 > output/$sim-omega.txt
310
311 number=1
312 while [ $number -le $aminoacidsomega ]; do
313
314 no=`sed -n '$number'p' omega-aa-no.txt`
315 aa=`sed -n '$number'p' omega-aa-list.txt`
316 sed '/[#@]/d' g_chi-omega/omega$aa$no.xvg | cut -c13- | sed 's/ *//g' | sed -n
    1!'p > resid-omega/resid$no.txt
317 ./gro_validation-calc 3 $no $frames 0 1 0 < resid-omega/resid$no.txt >>
    output/$sim-omega.txt
318
319 number=`expr $number + 1`
320 done
321
322
323 #####
324 ### Section 9: Compare the rotamers with experimental values
325 #####
326
327 beg=`expr $end + 1`
328 end=`wc -l atom-numbers_final-rotamer.txt | cut -d\ -f 1`
329 end=`expr $beg + $end`
330 end=`expr $end - 1`
331 cut -d\ -f $beg-$end $sim-dihedrals-angaver-imp.txt > $sim-dihedrals-protein-
    angaver-imp.txt
332 # Generate input file from aa-no, aa-list, reference angles and measured angles
333 cat aa-no.txt aa-list.txt reference/lovel-rotamer.data $sim-dihedrals-protein-
    angaver-imp.txt > input-rotamer.txt
334 # Number of reference dihedrals
335 reflines=`wc -l reference/lovel-rotamer.data | cut -d\ -f 1`
336 # Number of dihedrals in protein side chain
337 intangles=`wc -l atom-numbers_final-rotamer.txt | cut -d\ -f 1`
338 # Rotamer location (favored, allowed, disallowed) for every amino acid
339 ./gro_validation-calc 4 $aminoacids $frames $intangles 1 $reflines < input-
    rotamer.txt > output/$sim-rotamer-regions.txt
340 # Matrices for generating a surface graph (Residues with one and two rotamers;

```

```

341 # more than 2 rotamers: only chi1 and chi2)
342 ./gro_validation-calc 4 $aminoacids $frames $intangles 2 $reflines < input-
    rotamer.txt > output/$sim-rotamer-graph.txt
343 # Time resolution of side chain rotamers
344 ./gro_validation-calc 4 $aminoacids $frames $intangles 3 $reflines < input-
    rotamer.txt > output/$sim-rotamer-time.txt
345
346 #####
347 ### Section 10: Backbone phi/psi dihedrals: ramachandran plot
348 #####
349 #####
350
351 # g_rama
352 g_rama -f $filename.xtc -s $filename.tpr -o phi-psi.xvg -b 1 -dt $timestep -e
    $lastframe >> rama-result.txt 2>&1
353
354 sed -e '/[#@]/d' phi-psi.xvg > phi-psi_header.txt
355 sed -n 's/ / /pg' phi-psi_header.txt > phi-psi_space.txt
356
357 # Number the lines of the file
358 nl phi-psi_space.txt | sed -n 's/\t/ /p' > phi-psi_space_no.txt
359
360 # List of aa without ACE and NH2
361 paste -d- aa-list.txt aa-no.txt | sed -n /ACE/'!'p | sed -n /NH2/'!'p > aa-
    no_list-rama.txt
362 aminoacidsrama=`wc -l aa-no_list-rama.txt | cut -d\ -f 1`
363
364 # Extract the right angles of the complete xvg-file
365 number=1
366 while [ $number -le $aminoacidsrama ]; do
367     var=`sed -n "$number"p aa-no_list-rama.txt`
368     sed -n /"$var"/p phi-psi_space_no.txt >> phi-psi_.txt
369     number=`expr $number + 1`
370 done
371
372 # Reduce the output of g_rama to couple of phi/psi values
373 sed 's/^[ ]\+//p' phi-psi_.txt | sort -n -u -t\ -k 1,1 | cut -d\ -f 2,3 > phi-
    psi_in.txt
374
375 intangle=`wc -l phi-psi_in.txt | cut -d\ -f 1`
376
377 cut -d- -f 1 aa-no_list-rama.txt > rama-aa-list.txt
378 cut -d- -f 2 aa-no_list-rama.txt > rama-aa-no.txt
379
380 # Combine the sequence, reference data, and the measured phi/psi values to one
381 # file
382 cat rama-aa-no.txt rama-aa-list.txt reference/lovell-rama.data phi-psi_in.txt >
    rama-input.txt
383
384 reflines=`wc -l reference/lovell-rama.data | cut -d\ -f 1`
385
386 # Ramachandran values for every amino acid
387 ./gro_validation-calc 5 $aminoacidsrama $frames $intangle 1 $reflines < rama-
    input.txt > output/$sim-rama-regions.txt
388 # List of phi and psi values, tab-separated
389 ./gro_validation-calc 5 $aminoacidsrama $frames $intangle 2 $reflines < rama-
    input.txt > output/$sim-rama-data.txt
390 # 180x180 matrix for generating a surface graph
391 ./gro_validation-calc 5 $aminoacidsrama $frames $intangle 3 $reflines < rama-
    input.txt > output/$sim-rama-graph.txt
392 # Time resolution of ramachandran values
393 ./gro_validation-calc 5 $aminoacidsrama $frames $intangle 4 $reflines < rama-
    input.txt > output/$sim-rama-time.txt
394
395 #####
396 ### Clean directory
397
398 rm -f aa-no.txt
399 rm -f aa-no_list.txt
400 rm -f aa-list.txt
401 rm -f gro-file-mod.txt
402 rm -f g_angle-result.txt
403 rm -f omega-aa-no.txt
404 rm -f omega-aa-list.txt
405 rm -f g_chi.txt
406

```

```

407 rm -rf g_chi-omega
408 rm -f aa-no_list-omega.txt
409 rm -rf resid-omega
410 rm -f rama-result.txt
411 rm -f phi-psi_space_no.txt
412 rm -f aa-no_list-rama.txt
413 rm -f phi-psi_.txt
414 rm -f rama-aa-no.txt
415 rm -f rama-aa-list.txt
416 rm -f phi-psi_in.txt
417 rm -f rama-input.txt
418 rm -f atom-numbers_final-chirality.txt
419 rm -f atom-numbers_final-impropers.txt
420 rm -f atom-numbers_final-rotamer.txt
421 rm -f atom-numbers_final-chirality-planarity-rotamer.ndx
422 rm -f input-chirality.txt
423 rm -f input-planarity.txt
424 rm -f input-rotamer.txt
425 rm -f $sim-dihedrals-angdist.xvg
426 rm -f $sim-dihedrals-angaver.xvg
427 rm -f $sim-chirality-protein-angaver-imp.txt
428 rm -f $sim-impropers-protein-angaver-imp.txt
429 rm -f $sim-dihedrals-protein-angaver-imp.txt
430 rm -f $sim-dihedrals-angaver-imp.txt
431 rm -f CA.txt
432 rm -f phi-psi.xvg
433 rm -f phi-psi_header.txt
434 rm -f phi-psi_space.txt
435 rm -f gro_dihedrals
436 rm -f gro_validation-calc

```

9.5.3 C program *gro_validation-calc.c*

```

1  #include <stdio.h>
2  #include <stdlib.h>
3  #include <string.h>
4  #include <math.h>
5
6  // Compile with:
7  // gcc gro_validation-calc.c /usr/lib64/libm.a -o gro_validation-calc
8
9  // Definition of constants
10 // Number of residues which are defined by the variable AAcodes in this program
11 #define no_def_res 34
12 // Number of bars in histogram for chirality (even number)
13 #define histono_chir 52
14 // Width of the bars for chirality
15 #define z_chir 1
16 // Mean dihedral angle for chirality
17 #define imprref 35
18 // Number of bars in histogram for planarity check (even number)
19 #define histono_imp 32
20 // Width of the bars for planarity
21 #define z_imp 2
22 // Number of bars in histogram for peptide dihedral angle distribution (even
23 // number)
24 #define histono_omega 34
25 // Width of the bars for peptide bond
26 #define z_omega 5
27
28 int main(int argc, char *argv[]) {
29
30 int angtype, aminoacids, frames, intangles, selection, reflines;
31
32 // With parameters in the shell script the following variables are defined
33 angtype = strtol(argv[1], NULL, 10);

```

```

34 aminoacids = strtol(argv[2], NULL, 10);
35 frames = strtol(argv[3], NULL, 10);
36 intangles = strtol(argv[4], NULL, 10);
37 selection = strtol(argv[5], NULL, 10);
38 reflines = strtol(argv[6], NULL, 10);
39 int datsize;
40 datsize=frames*intangles;
41
42 // In AName the amino acid name is stored and compared to the amino acid sequence
43 // which is read in; all listed standard amino acids of GROMACS (ffG53a6.rtp) are
44 // included, incl. ACE and NH2, except HYP (hydroxyproline)
45
46 char AName[no_def_res][5]=
47 {
48     {'A','C','E'}, // 0
49     {'N','H','2'}, // 1
50     {'A','L','A'}, // 2
51     {'A','R','G'}, // 3
52     {'A','R','G','N'}, // 4
53     {'A','S','N'}, // 5
54     {'A','S','N','1'}, // 6
55     {'A','S','P'}, // 7
56     {'A','S','P','H'}, // 8
57     {'C','Y','S'}, // 9
58     {'C','Y','S','H'}, // 10
59     {'C','Y','S','1'}, // 11
60     {'C','Y','S','2'}, // 12
61     {'G','L','N'}, // 13
62     {'G','L','U'}, // 14
63     {'G','L','U','H'}, // 15
64     {'G','L','Y'}, // 16
65     {'H','I','S','A'}, // 17
66     {'H','I','S','B'}, // 18
67     {'H','I','S','H'}, // 19
68     {'H','I','S','1'}, // 20
69     {'H','I','S','2'}, // 21
70     {'I','L','E'}, // 22
71     {'L','E','U'}, // 23
72     {'L','Y','S'}, // 24
73     {'L','Y','S','H'}, // 25
74     {'M','E','T'}, // 26
75     {'P','H','E'}, // 27
76     {'P','R','O'}, // 28
77     {'S','E','R'}, // 29
78     {'T','H','R'}, // 30
79     {'T','R','P'}, // 31
80     {'T','Y','R'}, // 32
81     {'V','A','L'} // 33
82 };
83
84 ////////////////////////////////////////////////////
85 // Chirality check: needs angtype, aminoacids, datsize (-> frames, intangles)
86 ////////////////////////////////////////////////////
87
88 if(angtype==1)
89 {
90
91     int a,d,i,j,k,x,y;
92     int cmp;
93     int SD2spann;
94     int check;
95     float ang;
96     float mean;
97     float SD;
98     float histofl[histono_chir];
99     float checkfl;
100
101     // In matrixres the amino acid sequence is stored
102     char matrixres[aminoacids][5];
103     // Initialize matrixres with NULL
104     for (i=0; i<aminoacids; i++)
105     {
106         for (j=0; j<5; j++)
107             matrixres[i][j]=0;
108     }
109     // In arrayres the amino acid sequence as integer code is saved
110     int arrayres[aminoacids];

```

```

110 // In matrixno the amino acid numbers are stored
111 int matrixno[aminoacids];
112 // Each amino with a chiral center at the C-alpha atom contains one
113 int AAcode[no_def_res][1]=
114 {
115     {0}, // 0
116     {0}, // 1
117     {1}, // 2
118     {1}, // 3
119     {1}, // 4
120     {1}, // 5
121     {1}, // 6
122     {1}, // 7
123     {1}, // 8
124     {1}, // 9
125     {1}, // 10
126     {1}, // 11
127     {1}, // 12
128     {1}, // 13
129     {1}, // 14
130     {1}, // 15
131     {0}, // 16
132     {1}, // 17
133     {1}, // 18
134     {1}, // 19
135     {1}, // 20
136     {1}, // 21
137     {1}, // 22
138     {1}, // 23
139     {1}, // 24
140     {1}, // 25
141     {1}, // 26
142     {1}, // 27
143     {1}, // 28
144     {1}, // 29
145     {1}, // 30
146     {1}, // 31
147     {1}, // 32
148     {1} // 33
149 };
150
151 // In dihedralsdata the angle data is stored
152 float *dihedralsdata = malloc((datsize) * sizeof(float));
153 if(dihedralsdata == NULL)
154 { printf("NO RAM\n");
155   return EXIT_FAILURE;
156 }
157
158 // Read amino acid numbers
159 for(i=0; i<aminoacids; i++)
160 scanf("%d", &matrixno[i]);
161
162 // Read amino acid sequence
163 scanf("%c", &matrixres[0][0]);
164 for(i=0; i<aminoacids; i++)
165 { for(j=0; ; j++)
166     { scanf("%c", &matrixres[i][j]);
167       if(matrixres[i][j]=='\n')
168       { matrixres[i][j]=0;
169         break;
170       }
171     }
172 }
173
174 // Convert amino acid sequence to numbers
175 for(i=0; i<aminoacids; i++)
176 { for(j=0; j<no_def_res; j++)
177     { if((cmp=strcmp(matrixres[i], AAcname[j]))==0)
178       { arrayres[i]=j;
179         break;
180       }
181     }
182 }
183
184 // Read the measured dihedrals
185 for(i=0; i<datsize; i++)

```

```

186 scanf("%f", &dihedraldata[i]);
187
188 // Print header for histogram
189 printf("AS-no\tAS\tmean\tSD\tchiralityOK\t");
190 printf("<%.1f\t", imprpref-(1.0)*((histono_chir-2)/2)*z_chir);
191 for(y=0, ang=imprpref-(1.0)*((histono_chir-2)/2)*z_chir; y<histono_chir-2; y++,
    ang+=z_chir)
192 printf("%.1f-%.1f\t", ang, ang+z_chir);
193 printf(">%.1f\n", imprpref+((histono_chir-2.0)/2)*z_chir);
194
195 // Calculate dihedral angles
196 for(i=0; i<aminoacids ; i++)
197 {
198
199 mean=0.0;
200 SD=0.0;
201
202 // Determine the starting point of the measured angles
203 k=0;
204 for(d=0;d<i;d++)
205 {
206     a=arrayres[d];
207     k+=AAcode[a][0];
208 }
209 a=arrayres[i];
210
211 if(AAcode[a][0]>0)
212 {
213
214 check=0;
215 for(d=0;d<histono_chir;d++)
216 histofl[d]=0.0;
217
218 // Calculate mean, histogram and perform chirality-check
219 x=k;
220 for(d=0;d<frames; d++, x+=intangles)
221 {
222     mean+=dihedraldata[x];
223     for(y=0, ang=imprpref-(1.0)*((histono_chir-2)/2)*z_chir; y<histono_chir-2; y++,
        ang+=z_chir)
224     {
225         if(dihedraldata[x]>=ang && dihedraldata[x]<ang+z_chir)
226             histofl[y+1]+=1.0/frames;
227
228         if(dihedraldata[x]<imprpref-(1.0)*((histono_chir-2)/2)*z_chir)
229             histofl[0]+=1.0/frames;
230         if(dihedraldata[x]>=imprpref+((histono_chir-2)/2)*z_chir)
231             histofl[histono_chir-1]+=1.0/frames;
232         if(dihedraldata[x] < 0 || dihedraldata[x] > 70)
233             check++;
234     }
235 }
236
237 // 0 and 70: +-35 degree of the reference of 35 degree
238 }
239
240 mean/=frames;
241 checkfl=(float)check/(float)frames;
242
243 // Calculate standard deviation SD
244 x=k;
245 for(d=0; d<frames; d++, x+=intangles)
246 SD+=(dihedraldata[x]-mean)*(dihedraldata[x]-mean);
247
248 SD=sqrt(SD/(frames-1));
249
250 // Calculate fraction of values within 2*SD
251 x=k;
252 SD2spann=0;
253 for(d=0; d<frames; d++, x+=intangles)
254 {
255     if(dihedraldata[x]>(mean-2*SD) && dihedraldata[x]<(mean+2*SD))
256         SD2spann++;
257 }
258
259 printf("%d\t%s\t%.3f\t%.3f\t%.3f\t", matrixno[i], matrixres[i], mean, SD, 1.0-
    checkfl);
260 for(d=0; d<histono_chir; d++) printf("%.3f\t", histofl[d]);
261 printf("\n");
262 }
263 }
264
265 // Free memory
266 free(dihedraldata);

```

```

259 }
260
261
262 ///////////////////////////////////////////////////////////////////
263 // Check for planarity: needs angtype, aminoacids, datsize (-> frames, intangles)
264 ///////////////////////////////////////////////////////////////////
265
266 if(angtype==2)
267 {
268
269     int a,d,i,j,k,x,y;
270     int cmp;
271     float ang;
272     float meansum;
273     float SDsum;
274     float SD2spannsum;
275     int SD2spann[20]; // Max 20 impropers per residue (TRP)
276     float mean[20];
277     float SD[20];
278     float histofl[20][histono_imp];
279     float histoflsum[histono_imp];
280
281     // In matrixres the amino acid sequence is stored
282     char matrixres[aminoacids][5];
283     // Initialize matrixres with NULL
284     for (i=0; i<aminoacids; i++)
285     {     for (j=0; j<5; j++)
286           matrixres[i][j]=0;
287     }
288     // In arrayres the amino acid sequence as integer code is saved
289     int arrayres[aminoacids];
290     // In matrixno the amino acid numbers are stored
291     int matrixno[aminoacids];
292     // In AAcode for each residue the amount of dihedral angles used for the planarity
293     // control is stored
294     int AAcode[no_def_res][1]=
295     {
296         {0}, // 0
297         {0}, // 1
298         {0}, // 2
299         {4}, // 3
300         {3}, // 4
301         {2}, // 5
302         {2}, // 6
303         {1}, // 7
304         {1}, // 8
305         {0}, // 9
306         {0}, // 10
307         {0}, // 11
308         {0}, // 12
309         {2}, // 13
310         {1}, // 14
311         {1}, // 15
312         {0}, // 16
313         {9}, // 17
314         {9}, // 18
315         {10}, // 19
316         {9}, // 20
317         {9}, // 21
318         {0}, // 22
319         {0}, // 23
320         {0}, // 24
321         {0}, // 25
322         {0}, // 26
323         {12}, // 27
324         {0}, // 28
325         {0}, // 29
326         {0}, // 30
327         {20}, // 31
328         {12}, // 32
329         {0} // 33
330     };
331
332     // In dihedralsdata the angle data is stored
333     float *dihedralsdata = malloc((datsize) * sizeof(float));
334     if(dihedralsdata == NULL)

```

```

335 { printf("NO RAM\n");
336     return EXIT_FAILURE;
337 }
338
339 // Read amino acid numbers
340 for(i=0; i<aminoacids; i++)
341     scanf("%d", &matrixno[i]);
342
343 // Read amino acid sequence
344 scanf("%c", &matrixres[0][0]);
345 for(i=0; i<aminoacids; i++)
346 { for(j=0; ; j++)
347     { scanf("%c", &matrixres[i][j]);
348         if(matrixres[i][j]=='\n'){matrixres[i][j]=0; break;}
349     }
350 }
351
352 // Convert amino acid sequence to numbers
353 for(i=0; i<aminoacids; i++)
354 { for(j=0; j<no_def_res; j++)
355     { if((cmp=strcmp(matrixres[i], AName[j]))==0)
356         { arrayres[i]=j;
357             break;
358         }
359     }
360 }
361
362 // Read the measured dihedrals
363 for(i=0; i<datsize; i++)
364     scanf("%f", &dihedraldata[i]);
365
366 // Print header for histogram
367 printf("AS-no\tAS\tmean-sum\tSD-sum\t2SD-interval\t");
368 printf("<%.1f\t", (-1.0)*((histono_imp-2)/2)*z_imp);
369 for(y=0, ang=(-1)*((histono_imp-2)/2)*z_imp; y<histono_imp-2; y++, ang+=z_imp)
370     printf("%.1f-%.1f\t", ang, ang+z_imp);
371 printf(">%.1f\n", ((histono_imp-2.0)/2)*z_imp);
372
373 // Calculate improvers
374 for(i=0; i<aminoacids ; i++)
375 {
376
377     for(j=0; j<20; j++)
378     { mean[j]=0.0;
379         SD[j]=0.0;
380     }
381
382     // Determine the starting point of the measured angles
383     k=0;
384     for(d=0; d<i; d++)
385     { a=arrayres[d];
386         k+=AAcode[a][0];
387     }
388
389     a=arrayres[i];
390
391     if(AAcode[a][0]>0)
392     {
393
394         for(j=0; j<20; j++)
395         { for(d=0; d<histono_imp; d++)
396             histofl[j][d]=0.0;
397         }
398
399         for(d=0; d<histono_imp; d++)
400             histoflsum[d]=0.0;
401         for(j=0; j<AAcode[a][0]; j++, k++)
402         {
403
404             // Calculate mean and histogram
405             x=k;
406             for(d=0; d<frames; d++, x+=intangles)
407             { mean[j]+=dihedraldata[x];
408                 for(y=0, ang=(-1)*((histono_imp-2)/2)*z_imp; y<histono_imp-2; y++, ang+=z_imp)
409                 { if(dihedraldata[x]>=ang && dihedraldata[x]<ang+z_imp)
410                     histofl[j][y+1]+=1.0/frames;

```

```

411     }
412     if(dihedraldata[x]<(-1)*((histono_imp-2)/2)*z_imp)
413         histofl[j][0]+=1.0/frames;
414     if(dihedraldata[x]>((histono_imp-2)/2)*z_imp)
415         histofl[j][histono_imp-1]+=1.0/frames;
416 }
417 mean[j]/=frames;
418
419 // Calculate SD
420 x=k;
421 for(d=0; d<frames; d++, x+=intangles)
422     SD[j]+=(dihedraldata[x]-mean[j])*(dihedraldata[x]-mean[j]);
423
424 SD[j]=sqrt(SD[j]/(frames-1));
425
426 // Calculate fraction of values within 2*SD
427 x=k;
428 SD2spann[j]=0;
429 for(d=0; d<frames; d++, x+=intangles)
430 {     if(dihedraldata[x]>(mean[j]-2*SD[j]) && dihedraldata[x]<(mean[j]+2*SD[j]))
431         SD2spann[j]++;
432 }
433 }
434
435 for(d=0;d<histono_imp;d++)
436 {     for(j=0; j<AAcode[a][0]; j++)
437         histoflsum[d]+=histofl[j][d];
438     histoflsum[d]/=AAcode[a][0];
439 }
440
441 meansum=0.0;
442 for(j=0; j<AAcode[a][0]; j++)
443     meansum+=mean[j];
444 meansum/=AAcode[a][0];
445 printf("%d\t%s\tsummary\t%.4f\t-\t-\t", matrixno[i], matrixres[i], meansum);
446 for(d=0; d<histono_imp; d++)
447     printf("%.4f\t", histoflsum[d]);
448 printf("\n");
449 for(j=0; j<AAcode[a][0]; j++)
450 {     printf("%d\t%s\t%d\t%.4f\t%.4f\t%.4f\t", matrixno[i],
451         matrixres[i],j+1,mean[j],SD[j], (float)SD2spann[j]/frames);
452     for(d=0; d<histono_imp; d++)
453         printf("%.4f\t", histofl[j][d]);
454     printf("\n");
455 }
456 }
457
458 // Free memory
459 free(dihedraldata);
460 }
461
462
463 ////////////////////////////////////////////////////
464 // Calculate omega backbone angle: needs angtype, aminoacids, frames, selection
465 ////////////////////////////////////////////////////
466
467 if(angtype==3)
468 {
469
470     int i,k,y;
471     float ang;
472     float mean, SD, k_SD;
473     float histofl[histono_omega];
474     float values[frames];
475
476     // Print header
477     if(selection==0)
478     {     printf("aa-no\tmean\tSD\t2SDinterval\t");
479         printf("<%.1f\t", 180-((histono_omega-2.0)/2)*z_omega);
480         for(y=0, ang=180.0-((histono_omega-2)/2)*z_omega; y<histono_omega-2; y++,
481             ang+=z_omega)
482             printf("%.1f-%.1f\t", ang, ang+z_omega);
483         printf(">%.1f\n", 180+((histono_omega-2.0)/2)*z_omega);
484     }

```

```

485 if(selection==1)
486 {
487
488 // Read data
489 for(i=0; i<frames; i++)
490 scanf("%f\n", &values[i]);
491
492 // Convert data
493 for(i=0; i<frames; i++)
494 { if(values[i]<0)
495     values[i]+=360.0;
496 }
497
498 // Calculation of the mean
499 mean=0.0;
500 for(i=0;i<frames;i++)
501 mean+=values[i];
502 mean=mean/frames;
503
504 // Standard deviation
505 SD=0.0;
506 for(i=0;i<frames;i++)
507 SD+=(values[i]-mean)*(values[i]-mean);
508 SD=sqrt(SD/(frames-1));
509
510 // Interval +- 2*SD
511 k=0;
512 for(i=0;i<frames;i++)
513 { if(values[i]>(mean - 2*SD) && values[i]<(mean + 2*SD))
514     k++;
515 }
516 k_SD = (float) k/ (float) (frames);
517
518 // Histogram
519 for(i=0;i<histono_omega;i++)
520 histofl[i]=0.0;
521 for(i=0;i<frames;i++)
522 { for(y=0, ang=180.0-(((histono_omega-2)/2)*z_omega); y<histono_omega-2; y++,
ang+=z_omega)
523 { if(values[i]>=ang && values[i]<ang+z_omega)
524     histofl[y+1]+=1.0/frames;
525 }
526 if(values[i]<180.0-(((histono_omega-2)/2)*z_omega))
527 histofl[0]+=1.0/frames;
528 if(values[i]>=180.0+((histono_omega-2)/2)*z_omega)
529 histofl[histono_omega-1]+=1.0/frames;
530 }
531
532 // Output
533 printf("%d\t%.3f\t%.3f\t%.3f\t",aminoacids,mean,SD,k_SD);
534 for(i=0;i<histono_omega;i++)
535 printf("%.3f\t", histofl[i]);
536 printf("\n");
537 }
538 }
539
540
541 ///////////////////////////////////////////////////////////////////
542 // Compare side chain dihedrals with Lovell values: needs angtype, aminoacids,
543 // datsize (-> frames, intangles), selection, reflines
544 ///////////////////////////////////////////////////////////////////
545
546 if(angtype==4)
547 {
548
549 // Cut-off values for favored, allowed, and disallowed
550 float cutoff_mf=0.02;
551 float cutoff_dis=0.002;
552
553 int a,d,i,j,k,x,y,z;
554 int cmp, place, tmp, mata, matb;
555 float mat[360][360];
556 float ang[4];
557 int angint[4];
558 int mf, al, dis;
559 float mf_f, al_f, dis_f;

```

```

560 char timeres[frames];
561
562 // In matrixres the amino acid sequence is stored
563 char matrixres[aminoacids][5];
564 // Initialize matrixres with NULL
565 for (i=0; i<aminoacids; i++)
566 {   for (j=0; j<5; j++)
567     matrixres[i][j]=0;
568 }
569 // In arrayres the amino acid sequence as integer code is saved
570 int arrayres[aminoacids];
571 // In matrixno the amino acid numbers are stored
572 int matrixno[aminoacids];
573 // In AAcode for each amino acid the data of dihedrals is stored; Position 1:
574 // number of dihedrals in the side chain; Position 2: start of the reference
575 // values in the respective reference file; Position 3 and further: step size for
576 // each dihedral
577 int AAcode[no_def_res][18]=
578 {
579     {0}, // 0
580     {0}, // 1
581     {0}, // 2
582     {4,0,10,360,10,360,10,360,10,360}, // 3
583     {4,0,10,360,10,360,10,360,10,360}, // 4
584     {2,1679616,5,360,5,360}, // 5
585     {2,1679616,5,360,5,360}, // 6
586     {2,1684800,5,360,5,180}, // 7
587     {2,1684800,5,360,5,180}, // 8
588     {1,1687392,1,360}, // 9
589     {1,1687392,1,360}, // 10
590     {1,1687392,1,360}, // 11
591     {1,1687392,1,360}, // 12
592     {3,1687752,8,360,8,360,8,360}, // 13
593     {3,1778877,8,360,8,360,8,180}, // 14
594     {3,1778877,8,360,8,360,8,180}, // 15
595     {0}, // 16
596     {2,1825452,5,360,5,360}, // 17
597     {2,1825452,5,360,5,360}, // 18
598     {2,1825452,5,360,5,360}, // 19
599     {2,1825452,5,360,5,360}, // 20
600     {2,1825452,5,360,5,360}, // 21
601     {2,1830636,5,360,5,360}, // 22
602     {2,1835820,5,360,5,360}, // 23
603     {4,1841004,10,360,10,360,10,360,10,360}, // 24
604     {4,1841004,10,360,10,360,10,360,10,360}, // 25
605     {3,3520620,8,360,8,360,8,360}, // 26
606     {2,3611745,5,360,5,180}, // 27
607     {0}, // 28
608     {1,3614337,1,360}, // 29
609     {1,3614697,1,360}, // 30
610     {2,3615057,5,360,5,360}, // 31
611     {2,3611745,5,360,5,180}, // 32
612     {1,3620241,1,360} // 33
613 };
614
615 // In angleref the angle data is stored
616 float *angleref = malloc((reflines) * sizeof(float));
617 if(angleref == NULL)
618 {   printf("NO RAM\n");
619     return EXIT_FAILURE;
620 }
621
622 // In dihedralsdata the angle data is stored
623 float *dihedralsdata = malloc((datsize) * sizeof(float));
624 if(dihedralsdata == NULL)
625 {   printf("NO RAM\n");
626     return EXIT_FAILURE;
627 }
628
629 // Read amino acid numbers
630 for(i=0; i<aminoacids; i++)
631     scanf("%d", &matrixno[i]);
632
633 // Read amino acid sequence
634 scanf("%c", &matrixres[0][0]);
635 // New line character after amino acid numbers

```

```

636 for(i=0; i<aminoacids; i++)
637 {   for(j=0; ; j++)
638     {   scanf("%c", &matrixres[i][j]);
639         if(matrixres[i][j]=='\n')
640         {   matrixres[i][j]=0;
641             break;
642         }
643     }
644 }
645
646 // Convert amino acid sequence to numbers
647 for(i=0; i<aminoacids; i++)
648 {   for(j=0; j<no_def_res; j++)
649     {   if((cmp=strcmp(matrixres[i], AName[j]))==0)
650         {   arrayres[i]=j;
651             break;
652         }
653     }
654 }
655
656 // Read the reference data in the order: arg-asn-asp-cys-gln-glu-his-ile-leu-lys-
657 // met-phetyr-ser-thr-trp-val
658 for(i=0; i<reflines; i++)
659 scanf("%f", &angleref[i]);
660
661 // Read the measured dihedrals
662 for(i=0; i<datsize; i++)
663 {   scanf("%f", &dihedraldata[i]);
664     if(dihedraldata[i]<0)
665         dihedraldata[i]=360.0+dihedraldata[i];
666 }
667
668 // Compare dihedrals
669 for(i=0; i<aminoacids ; i++)
670 {
671
672 if(selection==2)
673 {   for(mata=0; mata<360; mata++)
674     {   for(matb=0; matb<360; matb++)
675         {   mat[mata][matb]=0.0;
676         }
677     }
678
679 // Determine the starting point of the measured angles
680 k=0;
681 for(d=0; d<i; d++)
682 {   a=arrayres[d];
683     k+=AAcode[a][0];
684 }
685 a=arrayres[i];
686
687 if(AAcode[a][0]>0)
688 {
689
690 mf=0;
691 al=0;
692 dis=0;
693 x=k;
694
695 for(d=0; d<frames; d++, x=x-AAcode[a][0]+intangles)
696 {
697
698 place=0; mata=0; matb=0;
699 // Determine the reference value (the right line)
700 for(j=0, y=2 ; j<AAcode[a][0]; j++, y+=2)
701 {
702
703 if( (a==7 || a==8 || a==14 || a==15 || a==27 || a==32) && (j==(AAcode[a][0]-1))
704 && (dihedraldata[x]>180.0))
705     dihedraldata[x]-=180.0;
706 // The last angle for ASP, GLU, PHE and TYR has only a range of 180 degrees in the
707 // reference data
708 if(selection==0)
709     printf("%f\t", dihedraldata[x]);
710 // Special treatment of GLU and GLUH

```

```

711 if(a==14 || a==15)
712 {
713
714     if(selection==2)
715     {
716         if(j==0) {mata=(int) (dihedraldata[x]/8);}
717         if(j==1) {matb=(int) (dihedraldata[x]/8);}
718     }
719     if(j==0 ) {angint[j]=((int) (dihedraldata[x]/8)*45*23);}
720     else if(j==1) {angint[j]=((int) (dihedraldata[x]/8)*23);}
721     else if(j==2) {angint[j]=((int) (dihedraldata[x]/(180.0/23)))};
722     if(selection==0) printf("%d\t",angint[j]); // Control
723 }
724
725 else
726 {
727     // The angle is divided by the step size of the angle in the reference data
728     ang[j]=dihedraldata[x]/AAcode[a][y];
729     // The integer of this new angle is used
730     angint[j]=(int)ang[j];
731     if(selection==2)
732     {
733         if(j==0) {mata=(int) (dihedraldata[x]/AAcode[a][y]);}
734         if(j==1) {matb=(int) (dihedraldata[x]/AAcode[a][y]);}
735     }
736     tmp=1;
737     for(z=y+2; z<AAcode[a][0]*2+1; z+=2)
738     tmp*= (AAcode[a][z+1]/AAcode[a][z]);
739
740     angint[j]*=tmp;
741     if(selection==0) printf("%d\t",angint[j]);
742 }
743
744 place+=angint[j];
745 x++;
746 }
747
748 place+=AAcode[a][1];
749 if(selection==0)
750 printf("%d\t%d\t%f\t\n", place, place-AAcode[a][1], angleref[place]);
751 if(selection==2)
752 mat[mata][matb]+=1.0/frames;
753 if(angleref[place]>cutoff_mf)
754 {
755     mf++;
756     timeres[d]='f';
757 }
758 else if(angleref[place]<=cutoff_mf && angleref[place]>=cutoff_dis)
759 {
760     al++;
761     timeres[d]='a';
762 }
763 else if(angleref[place]<cutoff_dis)
764 {
765     dis++;
766     timeres[d]='d';
767 }
768 }
769
770 // Print the data
771 mf_f=(float)mf/frames;
772 al_f=(float)al/frames;
773 dis_f=(float)dis/frames;
774 if(selection==1)
775 printf("%d\t%s\t%.3f\t%.3f\t%.3f\n", matrixno[i], matrixres[i], mf_f, al_f,
776 dis_f);
777 if(selection==2)
778 {
779     j=(AAcode[a][3]/AAcode[a][2]);
780     if(AAcode[a][0]==1)
781     {
782         printf("%s-%d\n", matrixres[i], matrixno[i]);
783         for(mata=0; mata<j; mata++)
784             printf("%f\t", mat[mata][0]);
785         printf("\n");
786     }
787     else if (AAcode[a][0]>1)
788     {
789         d=(AAcode[a][5]/AAcode[a][4]);
790         printf("%s-%d\n", matrixres[i], matrixno[i]);
791         for(mata=0; mata<j; mata++)
792         {
793             for(matb=0; matb<d; matb++)
794                 printf("%f\t", mat[mata][matb]);
795             printf("\n");
796         }
797     }
798 }

```

```

786     }
787 }
788 }
789 if(selection==3)
790 {   printf("%d\t", matrixno[i]);
791     for(z=0;z<frames;z++)
792         printf("%c", timeres[z]);
793     printf("\n");
794 }
795 }
796 }
797
798 // Free memory
799 free(dihedraldata);
800 free(angleref);
801 }
802
803
804 ////////////////////////////////////////////////////
805 // Ramachandran plot: compares phi and psi combinations with experimental values
806 // from Lovell: needs angtype, aminoacids, frames, intangles, selection, reflines
807 ////////////////////////////////////////////////////
808
809 if(angtype==5)
810 {
811
812     // Cut-off values for favored, allowed, and disallowed
813     float cutoff_mf=0.02;
814     float cutoff_dis=0.002;
815
816     int mf, al, dis;
817     float mf_f, al_f, dis_f;
818     reflines=2*reflines;
819     int ramaref1[reflines]; // There are 180*180*2 phi and psi values
820     int ramaref2[reflines];
821     int ramaref3[reflines];
822     int ramaref4[reflines];
823     reflines=reflines/2;
824     float ramavalue1[reflines]; // There are 180*180 values
825     float ramavalue2[reflines];
826     float ramavalue3[reflines];
827     float ramavalue4[reflines];
828     int phi[frames];
829     int psi[frames];
830     int a,b,d,i,j,f,k;
831     int ang;
832     float matr1[359][359];
833     float matr2[359][359];
834     float matr3[359][359];
835     float matr4[359][359];
836     char timeres[frames];
837
838     // Initialize with zero
839     for(i=0;i<359;i++)
840     {   for(j=0;j<359;j++)
841         {   matr1[i][j]=0.0;
842             matr2[i][j]=0.0;
843             matr3[i][j]=0.0;
844             matr4[i][j]=0.0;
845         }
846     }
847
848     // Matrix where temporarily the reference data are stored depending on the current
849     // amino acid
850     float matref[359][359];
851     // Store fraction of angles
852     float mat[359][359];
853     // In ramadat the angle data is stored
854     float *ramadat = malloc((2*intangles) * sizeof(int));
855     if(ramadat == NULL)
856     {   printf("NO RAM\n");
857         return EXIT_FAILURE;
858     }
859     // In matrixres the amino acid sequence is stored
860     char matrixres[aminoacids][5];
861     // Initialize matrixres with NULL

```

```

862 for (i=0; i<aminoacids; i++)
863 {   for (j=0; j<5; j++)
864     matrixres[i][j]=0;
865 }
866 // In matrixno the amino acid numbers are stored
867 int matrixno[aminoacids];
868 // For GLY, PRO, and pre-PRO other reference values are used
869 char AName[2][4]= {   {'P','R','O','\0'},
870                     {'G','L','Y','\0'}}
871                     };
872
873 // Read amino acid numbers
874 for(i=0; i<aminoacids; i++)
875 scanf("%d", &matrixno[i]);
876
877 // Read amino acid sequence
878 scanf("%c", &matrixres[0][0]);
879 for(i=0; i<aminoacids; i++)
880 {   for(j=0; ; j++)
881     {   scanf("%c", &matrixres[i][j]);
882         if(matrixres[i][j]=='\n')
883         {   matrixres[i][j]=0;
884             break;
885         }
886     }
887 }
888
889 // Read the reference data in the order: general, glycine, pro, pre-pro
890 for(i=0, j=0; j<reflines/4; i+=2, j++)
891 {   scanf("%d", &ramaref1[i]);
892     scanf("%d", &ramaref1[i+1]);
893     scanf("%f", &ramavalue1[j]);
894 }
895
896 for(i=0, j=0; j<reflines/4; i+=2, j++)
897 {   scanf("%d", &ramaref2[i]);
898     scanf("%d", &ramaref2[i+1]);
899     scanf("%f", &ramavalue2[j]);
900 }
901
902 for(i=0, j=0; j<reflines/4; i+=2, j++)
903 {   scanf("%d", &ramaref3[i]);
904     scanf("%d", &ramaref3[i+1]);
905     scanf("%f", &ramavalue3[j]);
906 }
907
908 for(i=0, j=0; j<reflines/4; i+=2, j++)
909 {   scanf("%d", &ramaref4[i]);
910     scanf("%d", &ramaref4[i+1]);
911     scanf("%f", &ramavalue4[j]);
912 }
913
914 // Convert the reference data
915 for(i=0, j=0; j<reflines/4; i+=2, j++)
916 {   a=ramaref1[i]+179; b=ramaref1[i+1]+179;
917     matr1[a][b]=ramavalue1[j];
918     a=ramaref2[i]+179; b=ramaref2[i+1]+179;
919     matr2[a][b]=ramavalue2[j];
920     a=ramaref3[i]+179; b=ramaref3[i+1]+179;
921     matr3[a][b]=ramavalue3[j];
922     a=ramaref4[i]+179; b=ramaref4[i+1]+179;
923     matr4[a][b]=ramavalue4[j];
924 }
925
926 // Read the angle data
927 for(i=0; i<(2*intangles); i++)
928 scanf("%f", &ramadat[i]);
929
930 // Ramachandran plot
931 for(f=0, d=0; d<aminoacids; f+=2, d++)
932 {
933
934 if ((a=strcmp(matrixres[d], AName[0]))==0) // Check if PRO
935 {   for(i=0; i<359; i++)
936     {   for(j=0; j<359; j++)
937         matref[i][j]=matr3[i][j];

```

```

938     }
939 }
940 else if ((a=strcmp(matrixres[d+1],AAname[0]))==0 && d!=aminoacids-1)
941 // Check if pre-PRO until second to last amino acid
942 {
943     for(i=0;i<359;i++)
944     {
945         for(j=0;j<359;j++)
946             matref[i][j]= matr4[i][j];
947     }
948 }
949 else if ((a=strcmp(matrixres[d],AAname[1]))==0) // Check if GLY
950 {
951     for(i=0;i<359;i++)
952     {
953         for(j=0;j<359;j++)
954             matref[i][j]= matr2[i][j];
955     }
956 }
957 else
958 {
959     for(i=0;i<359;i++)
960     {
961         for(j=0;j<359;j++)
962             matref[i][j]= matr1[i][j];
963     } // Else: "general"
964 }
965
966 if(selection==2)
967 printf("%s-%d\n", matrixres[d], matrixno[d]);
968 if(selection==3)
969 {
970     printf("%s-%d\n", matrixres[d], matrixno[d]);
971     for(i=0;i<359;i++)
972     {
973         for(j=0;j<359;j++)
974             mat[i][j]=0.0;
975     }
976 }
977
978 for(i=f, j=0; j<frames; i+=(2*aminoacids), j++)
979 {
980     ang=(int) (ramadat[i]);
981     if ((ang%2) == 0)
982     {
983         if (ramadat[i]<0) ang--;
984         else ang++;
985     }
986     phi[j]=ang;
987     if(selection==2) printf("%.4f\t", ramadat[i]);
988     ang=(int) (ramadat[i+1]);
989     if ((ang%2) == 0)
990     {
991         if (ramadat[i+1]<0) ang--;
992         else ang++;
993     }
994     psi[j]=ang;
995     if(selection==2) printf("%.4f\n", ramadat[i+1]);
996     a=phi[j]+179; b=psi[j]+179;
997     mat[a][b]+=(100*(1/(float) frames));
998 }
999
1000 if(selection==3)
1001 {
1002     for(i=0;i<360;i+=2,k++)
1003     {
1004         for(j=0;j<360;j+=2)
1005             printf("%f\t", mat[i][j]);
1006         printf("\n");
1007     }
1008 }
1009
1010 if(selection==1 || selection==4)
1011 {
1012     mf=0;
1013     al=0;
1014     dis=0;
1015     for(j=0;j<frames;j++)
1016     {
1017         a=phi[j]+179; b=psi[j]+179;
1018         if(matref[a][b]>cutoff_mf)
1019         {
1020             mf++;
1021             timeres[j]='f';
1022         }
1023         else if(matref[a][b]<=cutoff_mf && matref[a][b]>=cutoff_dis)
1024         {
1025             al++;
1026             timeres[j]='a';
1027         }
1028         else if(matref[a][b]<cutoff_dis)
1029         {
1030             dis++;
1031             timeres[j]='d';
1032         }
1033     }
1034 }

```

```

1014 }
1015 mf_f=(float)mf/frames;
1016 al_f=(float)al/frames;
1017 dis_f=(float)dis/frames;
1018 if(selection==1)
1019 printf("%d\t%s\t%.3f\t%.3f\t%.3f\n", matrixno[d], matrixres[d], mf_f, al_f,
dis_f);
1020 if(selection==4)
1021 { printf("%d\t", matrixno[d]);
1022   for(j=0;j<frames;j++)
1023     printf("%c", timeres[j]);
1024   printf("\n");
1025 }
1026 }
1027 }
1028
1029 // Free memory
1030 free(ramadat);
1031 }
1032
1033 return EXIT_SUCCESS;
1034 }

```

9.6 Time resolved Ramachandran analysis: *rama_time.sh*

```

1  #!/bin/bash
2
3
4  ### PLEASE ENTER #####
5  file=inact-md1-2-rama-time #example
6  residue=74                 #example
7  frames=8000                #example
8  #####
9
10 sed -n '/^#C#/p' test.sh | cut -c4- > rama-xpm-file.c
11
12 gcc rama-xpm-file.c -o rama-xpm-file
13
14 sed -n '/^$residue\\t/p $file.txt | cut -f2 > $residue.txt
15 ./rama-xpm-file $frames < $residue.txt > $residue.xpm
16
17 convert -density 300 $residue.xpm $residue.jpg
18
19 rm -f $residue.txt
20 rm -f $residue.xpm
21 rm -f rama-xpm-file.c
22 rm -f rama-xpm-file
23
24 ### C program rama-xpm-file
25 #C#
26 #C# //////////////////////////////////////
27 #C# // Generate an xpm picture for one residue containing the time
28 #C# // resolved ramachandran analysis
29 #C# //////////////////////////////////////
30 #C#
31 #C# #include <stdio.h>
32 #C# #include <stdlib.h>
33 #C# #include <string.h>
34 #C#
35 #C# int main(int argc, char *argv[]) {
36 #C#
37 #C# int frames;
38 #C# frames = strtol(argv[1], NULL, 10);
39 #C#

```

```
40  #C# int i,j;
41  #C#
42  #C# // In alldata the input is stored
43  #C# char alldata[frames];
44  #C#
45  #C# // Read the data
46  #C# for(i=0;i<frames;i++)
47  #C# {
48  #C#   scanf("%c",&alldata[i]);
49  #C# }
50  #C#
51  #C# // Print header
52  #C# printf("/* XPM */\n");
53  #C# printf("static char * pic_xpm[] = {\n");
54  #C# printf("\t%d %d 5 1\\",\n", frames, frames/20);
55  #C# printf("\t\t\t\t\tNone\\",\n");
56  #C# printf("\t\t\t\t\t#FFFFFF\\",\n");
57  #C# printf("\t\t\t\t\t#FFFFFF\\",\n"); // blue: #0101DF
58  #C# printf("\t\t\t\t\t#FFFFFF\\",\n"); // green: #088A08; yellow: #FFFF00
59  #C# printf("\t\t\t\t\t#FF0000\\",\n");
60  #C#
61  #C# // Height of the picture is 1/20 of its width
62  #C# for(j=0; j<frames/20; j++)
63  #C# {
64  #C#   printf("\t\t\t\t\t");
65  #C#   for(i=0 ;i<frames; i++)
66  #C#   {
67  #C#     printf("%c", alldata[i]);
68  #C#   }
69  #C#   if(j==(frames/20)-1) printf("\t\t\t\t\t"); else printf("\t\t\t\t\t");
70  #C#   printf("\n");
71  #C# }
72  #C#
73  #C# return EXIT_SUCCESS;
74  #C# }
```

9.7 References

Hess B, Kutzner C, van der Spoel D and Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, 4, 435-447.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, 26, 1701-1718.

Publications, Posters and Professional Training

Publication (published results prior to the submission of this thesis):

Keller M, Kaske M, Holzammer T, Bernhardt G and Buschauer A. Dimeric argininamide-type neuropeptide Y receptor antagonists: chiral discrimination between Y1 and Y4 receptors. Bioorg Med Chem (submitted June 2013)

Appl H, Holzammer T, Dove S, Haen E, Strasser A and Seifert R. Interactions of recombinant human histamine H(1)R, H(2)R, H(3)R, and H(4)R receptors with 34 antidepressants and antipsychotics. Naunyn Schmiedebergs Arch Pharmacol 2012, 385, 145-170.

Poster Presentations:

Holzammer T, Dove S and Buschauer A. The influence of acidic amino acids in the third extracellular loop of the histamine H₂ receptor on ligand binding. Poster contribution, 6th Summer School Medicinal Chemistry, University of Regensburg, September 26-28, 2012.

Kagermeier N, Holzammer T, Bernhardt G, Dove S and Buschauer A. Thr190 in the human histamine H₂ receptor contributes to binding of imidazole-type ligands. Poster contribution, 6th Summer School Medicinal Chemistry, University of Regensburg, September 26-28, 2012.

Kaske M, Keller M, Bernhardt G, Berlicki L, Holzammer T, Cabrele C, Reiser O and Buschauer A. Bivalent Argininamide-Type Neuropeptide Y Y₁R Ligands as Lead Structure for the Development of Non-Peptide Y₄R Antagonists. Poster contribution, 6th Summer School Medicinal Chemistry, University of Regensburg, September 26-28, 2012.

Kaczor A, Dove S and Holzammer T. Molecular Dynamics Study of the D₂R-mGluR5 Heterodimer in the Inactive state. Poster contribution, XXIInd International Symposium on Medicinal Chemistry (ISMC) 2012, Berlin, Germany, September 2-6, 2012.

Keller M, Kaske M, Bernhardt G, Berlicki L, Holzammer T, Cabrele C, Reiser O and Buschauer A. Bivalent Argininamide-Type Neuropeptide YY₁R Ligands as Lead Structure for the Development of Nonpeptide Y₄R Antagonists. Poster contribution, XXIInd International Symposium on Medicinal Chemistry (ISMC) 2012, Berlin, Germany, September 2-6, 2012.

Holzammer T, Buschauer A and Dove S. Active state model of the human histamine H₂ receptor. Poster contribution, 5th Summer School Medicinal Chemistry, University of Regensburg, September 13-15, 2010.

Professional Training:

July 2009 to September 2011 member of the Research Training Group (Graduiertenkolleg 760) "Medicinal Chemistry: Molecular Recognition – Ligand Receptor Interactions"