THE HUMAN HISTAMINE H_3-RECEPTOR: A MOLECULAR MODELLING STUDY OF A G-PROTEIN COUPLED RECEPTOR

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Birgit Schlegel

aus Linz, Österreich

Düsseldorf 2005

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent:	Prof. Dr. Dr. h.c. Hans-Dieter Höltje
1. Korreferent:	Prof. Dr. Wolfgang Sippl
2. Korreferent:	Prof. Dr. Thierry Langer
Tag der mündlichen Prüfung:	18. Juli 2005

Part I regarding the molecular dynamics simulation of bovine rhodopsin has been published in the Journal of Molecular Modeling. The original publication of this part is available at www.springerlink.com.

Die vorliegende Arbeit wurde von November 2001 bis Mai 2005 am Institut für Pharmazeutische Chemie der Heinrich-Heine-Universität Düsseldorf unter der Betreuung von Prof. Dr. W. Sippl und Prof. Dr. Dr. h.c. H.-D. Höltje angefertigt.

Mein besonderer Dank gilt Prof. Dr. Dr. h.c. Hans-Dieter Höltje für die Aufnahme in seinen Arbeitskreis, die sehr guten Rahmenbedingungen, den Ansporn eigene wissenschaftliche Ideen zu verteidigen, sowie eine drei-dimensionale Sicht der Welt der Arzneistoffe, die ich während des Besuchs seiner Vorlesung erlangen konnte.

Bei Prof. Dr. Wolfgang Sippl möchte ich mich für die sehr gute Betreuung bedanken, die mir den nötigen Freiraum gelassen hat, eigene Ideen umzusetzen, aber gleichzeitig die Arbeit durch unzählige Diskussionen, Anregungen und Vorschläge in eine erfolgversprechende Richtung gelenkt hat. Neben seinem enormen fachlichen Wissen durfte ich auch von den sehr guten Ressourcen seines Arbeitskreises in Halle profitieren, was viele Aspekte der Arbeit erst ermöglicht hat.

Bei unserem Kooperationspartner Prof. Dr. Holger Stark möchte ich mich für die Überlassung des Datensatzes an H_3-Liganden bedanken, der essentiell für die Anfertigung dieser Arbeit war.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Introduction

More than 100 years have passed since Emil Fischer compared in 1890 the interaction between a ligand molecule and a protein binding site with a key that fits a lock. Since then, tremendous advances in X-ray crystallography and computer technology have been achieved and nowadays allow the complex drug-receptor interaction to be visualised at an atomic scale. The possibility to see and understand how drugs are accommodated within a protein binding site and how they might exert their effects opens a bright future for rational drug design.

In this work, the focus has been set on molecular modelling studies of the human histamine H_3-receptor. This receptor is involved in the regulation of important physiological processes and could be a potential target for the treatment of numerous diseases. The hH_3R is involved in cognition, food intake, sleep and pain perception, thus the list of diseases that might be influenced via this receptor is long: Alzheimer, schizophrenia, attention-deficit hyperactivity disorder (ADHD), obesity,. . .

The H_3R receptor belongs to the family of G-protein coupled receptors, on which a short introduction is given below.

1.2 G-Protein Coupled Receptors

1.2.1 Pharmaceutical Relevance and Classification

The human genome codes for approximately 30000 proteins, of which, however, only 10% are currently classified as druggable targets. The term *druggable target* thereby refers to

proteins of pharmaceutical interest that can be addressed by small drug-like molecules that bind to a defined cavity within the protein, thereby exerting their effect. Excluded from this definition are molecules targeting for example protein-protein interfaces, as they do not provide any well defined binding pocket. The pool of human druggable targets can be subdivided into target families of which kinases, proteases, ion-channels and GPCRs are the most important (see figure 1.1). Sequence analysis has led to the assumption that the human genome could code for approximately 1000 GPCRs, [1] a large number of which are however still orphan receptors (i.e. receptors for which no ligand is known and no function has been determined so far). Currently, more than 50% of all drugs on the market exert their function targeting a GPCR, which stresses the outstanding pharmaceutical importance of this protein family. [2] The large number of GPCR protein sequences has led to a complex classification scheme consisting of 6 families (A–F), each containing several subfamilies. Human GPCR sequences can only be found within families A–C, while families D–F contain sequences of yeast, amoeba and archea. Family A, which is also termed rhodopsin family, is the largest group, comprising receptors for most amine neurotransmitters, many neuropeptides, purines, prostaglandines, cannabinoids, ... Ligands targeting family A GPCRs either bind to the transmembrane helices (amines) or within the extracellular loop region (peptides). Family B, the secretin/glucagon family, contains receptors for peptide hormones, including secretin, glucagon, and calcitonin. The ligand binding domain is located within the *N*-terminal end of intermediate length. Family C, the metabotropic glutamate receptor family or calcium sensor family, is the smallest group comprising metabotropic glutamate receptors, $GABA_B$ receptors and calcium-sensing receptors. The ligands bind within the long extracellular tail. [1] All histamine receptors can be classified as biogen aminergic GPCRs, one of the several subfamilies within family A GPCRs. Other subfamilies in family A are the muscarinic acetylcholine, adrenergic, dopamine, serotonine, octopamine and trace amine-receptors.

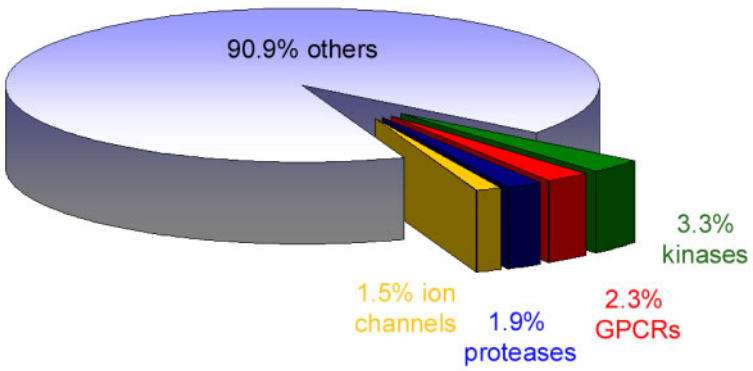


Figure 1.1: Important target families in the human genome.

1.2.2 Molecular Structure and Signal Transduction

GPCRs are heptahelical transmembrane proteins located in the cytoplasma membrane and have the common function of transducing extracellular signals into intracellular responses. Binding domains for the extracellular ligands thereby either lie within the helix bundle or within the *N*-terminal region of the receptor. A wide variety of ligands was reported to bind to GPCRs including biogen amines (*e.g. histamine, dopamine, acetylcholine*), amino acids (*e.g. glutamate*), ions (*e.g.* Ca^{2+}), lipids (*e.g. prostaglandines, leukotrienes*), peptides and proteins (*e.g. angiotensin, bradykinin, endorphins*). Agonist binding provokes a conformational change of the receptor that results in an alteration of the intracellular interaction with the heterotrimeric G-protein (see figure 1.2). [3] From coupling with the trimeric G-protein that assures a subsequent signal transduction in all GPCRs, the name *G-protein coupled receptor* originated. Following receptor activation, the heterotrimeric G-protein (composed of an α-, β- and γ-subunit) itself undergoes a conformational change that leads to the exchange of GDP for GTP bound to the α-subunit. The G$\alpha\beta\gamma$ complex breaks down into an α-subunit and the $\beta\gamma$-complex. The α-subunit can then stimulate effector molecules such as cyclases, phosphodiesterases and phospholipases, while for the $\beta\gamma$-complex in some cases a direct interaction with ion channels has been reported. As a result, the production of several second messengers such as cAMP (cyclic 3',5'-adenosine monophosphate), diacylglycerol or inositol(1,4,5)-trisphosphate is modulated. Finally, these second messengers can either prompt a fast cellular response (e.g. a modulation of intracellular ion concentration or regulation of enzyme activity) or cause a long-term biological effect by influencing transcription factors thereby regulating gene expression. The process is terminated when the catalytic conversion of GTP to GDP occurs in the α-subunit. αGDP dissociates from the effector and reunites with the $\beta\gamma$ complex.

A sufficient signal amplification is guaranteed, as each agonist/receptor complex can activate several G-proteins, which in turn can produce a large number of second messengers. The specificity of the biological response depends mainly on the nature of the α-subunit, of which more than 20 subtypes are known. [1]

Recently, in some cases, also G-protein independent signalling by GPCRs has been observed, thus further broadening the molecular mechanisms by which these receptors transduce extracellular signals. An example is the coupling of the G-protein coupled receptor kinase (GRK) to the phosphorylated *C*-terminal ending of GPCRs, thus labelling this receptor for endocytosis. [1, 3]

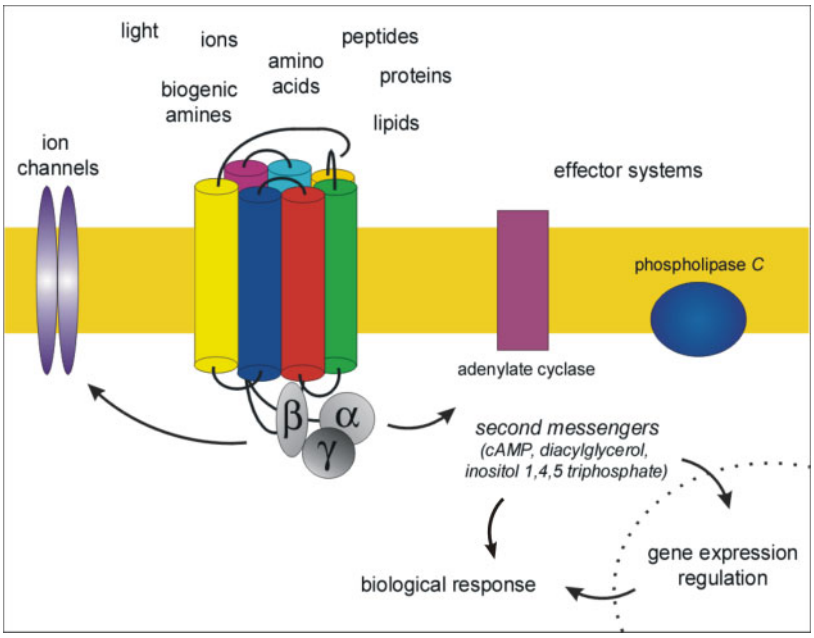


Figure 1.2: Diagram of GPCR function.

1.2.3 The Structure of Bovine Rhodopsin

A major breakthrough in the understanding of the GPCR receptor family was achieved in 2000 when the crystal structure of bovine rhodopsin was resolved [4] and for the first time detailed structural insights were gained (see figure 1.3). The crystallisation of other members of this protein family is hampered by usually low expression levels, which leads to problems in generating sufficient protein material for a crystallisation procedure as well as difficulties in the crystallisation process itself. [5]

Bovine rhodopsin represents a unique member among GPCRs as its intrinsic ligand *11-cis*-retinal is covalently bound to a lysine residue via a protonated Schiff-base linkage. Photon absorption results in retinal isomerisation to the *all-trans*-configuration triggering the activation of the protein. After photon capture, rhodopsin relaxes through a series of photoproducts until the active conformation, meta II, is formed which is capable of interacting with the G-protein. Hydrolysis of the Schiff-base results in the decay of meta II into the apoprotein opsin and *all-trans*-retinal. In vertebrates, *all-trans*-retinal is then regenerated in retinal pigment epithelial cells. [6]

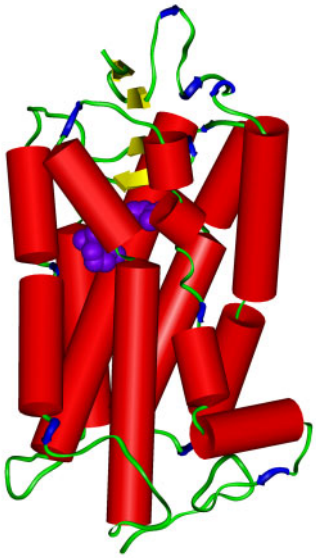


Figure 1.3: Crystal structure of bovine rhodopsin. Helical segments are depicted as red columns, beta-sheets as yellow arrows, turns as violet coil. The intrinsic ligand *11-cis*-retinal is displayed in its van der Waals representation.

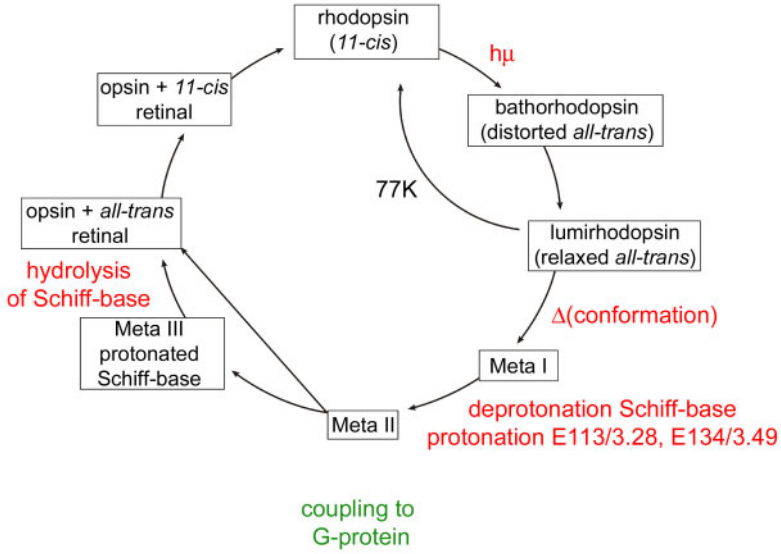


Figure 1.4: Intermediates in the photocycle: absorption of a photon results in the isomerisation of *11-cis*-retinal to a distorted *all-trans*-intermediate. The twisted double bond in bathorhodopsin relaxes to give lumirhodopsin. In lumirhodopsin the β-ionone moiety flips from its original position towards TM3 and TM4. Then, significant conformational changes take place and result in the meta I intermediate. Several de/protonation events produce meta II that represents the agonist-activated rhodopsin capable of interaction with the G-protein. Meta II finally decays to opsin and *all-trans*-retinal that is regenerated to *11-cis*-retinal outside the photoreceptor cells. The fact that lumirhodopsin can be reconverted to rhodopsin at 77 K where conformational changes are mainly impeded implies that no significant structural rearrangement takes place during the initial relaxation steps. [7]

1.2.4 Agonists, Antagonists and Inverse Agonists

Based on the biological response they provoke, ligands can be classified into agonists, partial agonists, antagonists, and inverse agonists.

Historically, a ligand that binds to a receptor and subsequently governs receptor activation has been referred to as an *agonist*; while a ligand that binds to a receptor without causing activation but impeding agonist binding, has been termed *antagonist*. Thus, although both molecules bind to the receptor (governed by their affinity), in the most simple cases only agonists will also possess an efficacy that results in receptor activation. *Partial agonists* are compounds that have a sub maximal tissue response even if they fully occupy the receptor.

The discovery of constitutive receptor activity has resulted in a reassessment of these terms. Constitutive activity describes the effect that receptors can exhibit an appreciable level of activity even in absence of any agonist. *Inverse agonists* are ligands that reduce this constitutive activation. *Neutral antagonists* restore the system towards the constitutive level of activity while agonists further activate the receptor. Figure 1.2 shows the classical two-state model that is nowadays however discussed controversially, as there is evidence for different coexisting receptor conformations.

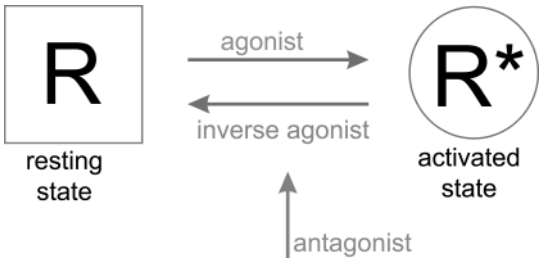


Figure 1.5: The two state model of receptor activity. The receptor exists in an equilibrium of activated and resting state. Antagonists preserve the level of constitutive activity, inverse agonists reduce activity and agonists enhance it.

1.3 Histamine

Histamine (2-[4-imidazolyl]ethyl amine) is a basic amine formed from L-histidine by histidine decarboxylase or an ubiquitous L-amino acid decarboxylase. High concentrations of histamine are found in the lungs, the skin and the gastrointestinal tract, where histamine functions as a mediator. In these tissues, histamine is predominantly stored in mast cells. During inflammatory or allergic reactions, histamine is liberated from mast cells by exocytosis. Alternatively, histamine is liberated upon destruction of these cells or by chemical substances (histamine liberators). When functioning as a mediator, histamine

exerts its effect predominantly via the H_1R, H_2R and the H_4R. Histamine additionally acts as a neurotransmitter in the central nervous system (CNS), where it interacts with the postsynaptically expressed H_1R and H_2R and praesynaptically expressed H_3R.

Two tautomeric forms of histamine exist, termed NπH and NτH (see figure 1.6). The ratio between the NτH and the NπH in an aqueous environment amounts to approximately 4:1. The pK_a of the α-amino group is 9.73 while the pK_a of the imidazole moiety is 5.91. [8] Histamine can thus exist as a dication whereby the protonated imidazole moiety is stabilised through delocalisation of the positive charge.

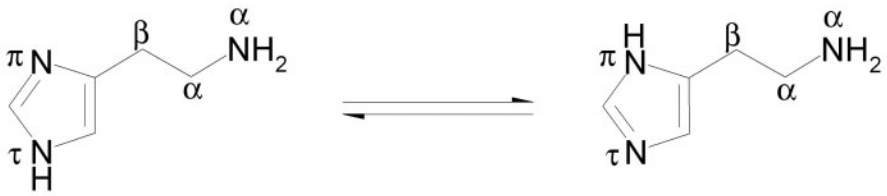


Figure 1.6: Tautomeric forms of histamine.

Released histamine is rapidly inactivated by histamine *N*-methyltransferase (HNMT, EC 2.1.1.8) or diamine oxidase (DAO, EC 1.4.3.6). The HNMT plays the dominant role in histamine metabolism within the human airways and gut, and is the only enzyme responsible for termination of the neurotransmitter action. [9] The HNMT inactivates histamine by transferring a methyl group from S-adenosyl-L-methionine to the Nτ-atom of the imidazole ring, yielding methylhistamine and S-adenosyl-L-homocysteine. The inactive methylhistamine is excreted in the urine or can be further metabolised by DAO or MAO into Nτ-methyl-imidazole-acetaldehyde, which can in turn be further oxidised by aldehyde dehydrogenase into Nτ-acetylimidazole acetic acid (see figure 1.7). The pathway of histamine metabolism starting with DAO is only relevant in the periphery.

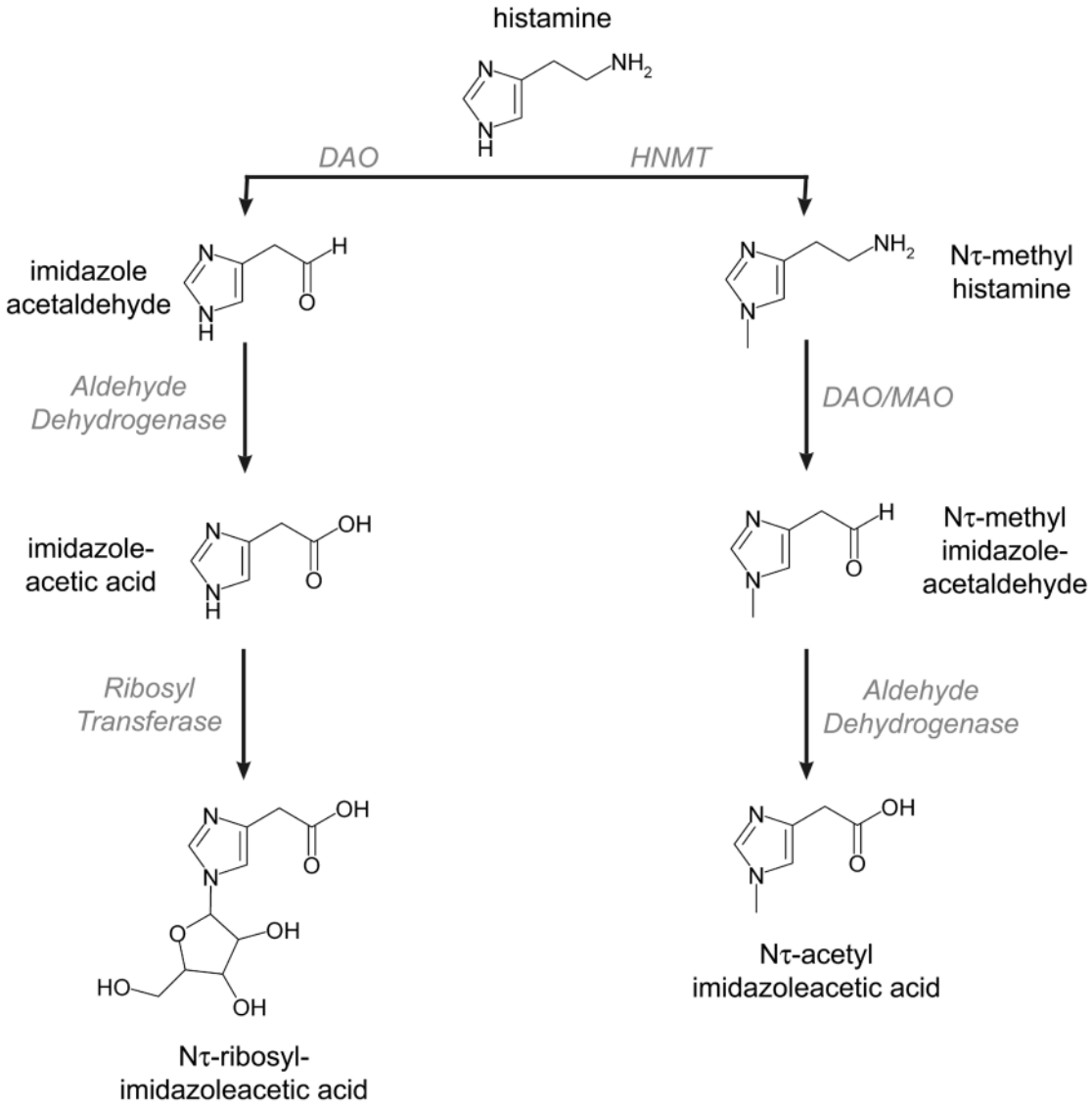


Figure 1.7: Metabolism of histamine.

1.4 The Histamine Receptor Family

The family of histamine receptors comprise to current knowledge four members, termed H_1-H_4. Compounds targeting the peripheral H_1-receptor are used in the therapy of allergic asthma and allergies; while compounds targeting the H_1R in the CNS can be used as sedatives or antiemetics. Antagonists addressing the H_2-receptor can down regulate gastric secretion and are widespread drugs for the treatment of peptic ulcer. To date no drugs targeting the H_3- or H_4-receptor are on the market, although some hH_3R ligands have by now reached clinic phase II. [10] It is expected that these agents have utility in the treatment of obesity, pain, cognitive disorders and allergic rhinitis. Finally, the most recently discovered H_4-receptor has been reported to play an important role in

inflammatory processes, therefore a broad spectrum of therapeutic application is expected also for this newest member of the histamine receptor family. Table 1.1 gives an overview on the main actions produced by histamine on the four receptors.

Table 1.1: Main actions of histamine on histamine receptors.

receptor	main action
hH_1R	• contraction of most smooth muscles in ileum, bronchi, uterus • increased vascular permeability • drop of blood pressure by dilatation of blood vessels • itching if injected to skin by stimulation of nerve endings • CNS: regulation of wakefulness
hH_2R	• stimulation of gastric secretion • cardiac stimulation (increase of rate and output of heart action)
hH_3R	• modulation of histamine release in histaminergic neurons • modulation of histamine synthesis in histaminergic neurons • modulation of the release of acetylcholine, dopamine, noradrenaline, serotonine, GABA, glutamate
hH_4R	• chemotaxis of eosinophils and mast cells to histamine • release of IL-16 from $CD8^+$ T cells

The first human histamine receptor cloned was the hH_2R, in 1991, [11] followed by the hH_1R in 1994, [12] the hH_3R in 1999 [13] and finally the hH_4R in 2000 [14–18]. Sequence analysis of the histaminergic receptor family revealed a low conservation within the family (20%) whereby the hH_3R and hH_4R share with approximately 40% the highest homology. Also the effector systems that are stimulated upon receptor activation are different within the receptor family: the hH_1R activates phospholipase C, the hH_2R stimulates the adenylyl cyclase, while the first cloned hH_3R subtype and the hH_4R inhibit cAMP production. As can be expected from the low degree of conservation, also the pharmacology of histamine receptor agonists and inverse agonists is clearly differentiable.

1.4.1 The Human H_3-Receptor

The histamine H_3-receptor was discovered in 1983 by Arrang and coworkers [19] and has been the focus of intense research over more than 20 years since then. Recently, several review articles have been published on the histamine receptor, [10] H_3R isoforms, [20, 21] on H_3R inverse agonists [22, 23] and agonists [24] that summarise the current knowledge on this receptor. Here, only a brief introduction on topics concerning this work can be given. Table 1.2 gives a short survey on interesting characteristics of the hH_3R.

Table 1.2: A short overview on the hH_3R.

classification	biogen aminergic G-protein coupled receptor.
noteworthiness	constitutive activity (CA) *in vitro* and *in vivo* (might be due to octa-to-dodecapeptide motif in IC_3 loop) [21]; CA certainly involved in the modulation of histamine liberation, importance for the modulation of the release of other neurotransmitters not clear.
location of gene	chromosome 20
gene organisation	most likely 3 exons and 2 introns; splice variants observed in human, rat and guinea pig. [21]
protein sequence	445 amino acids (full length human H_3R)
homology to other receptors	hH_1R: 20%, hH_2R: 22%, hH_4R: 37%, other biogen aminergic receptors: 20-27%
affinity for histamine	pK_i = 7.8
expression loci	CNS-regions (e.g. cerebral cortex, hippocampus, hypothalamus,...), low expression in nerve terminals of heart, lung and intestine.
molecular function	*autoreceptor*: feedback inhibition of histamine release inhibition of histamine synthesis *heteroreceptor*: CNS: modulation of release of dopamine, noradrenaline, serotonine, GABA, glutamate; periphery: acetylcholine, neuropeptides (e.g. substance P).
potential therapeutic application	*agonists*: insomnia, antinociceptives, myocardial ischaemic arrhythmias (via modulation of noradrenaline liberation) *inverse agonists*: treatment of obesity, narcolepsy, attention-deficit hyperactivity disorder (ADHD), schizophrenia, Alzheimer's disease, nasal congestion (in periphery and in combination with H_1R antagonists)
isoforms	$\geq$ 20 isoforms due to alternative splicing resulting in receptors **(a)** with an truncated *N*-terminal end, **(b)** with a partial deletion of TM2 and E_1, **(c)** with E_3 loops of different length, **(d)** lacking TM5-7 or **(e)** with an elongated *C*-terminal end. Not all isoforms have been yet fully characterised. Differences in expression pattern and pharmacological profile are observed. [21]
species differences	H_3Rs of human, mouse, rat, guinea-pig, and monkey show >92% sequence conservation; different pharmacological profiles for inverse agonists; comparable profiles for agonists.
signal transduction	coupling to $G_{i/o}$ proteins and inhibition of adenylyl cyclase $\rightarrow$ [cAMP]$\downarrow$; other pathways probable (MAP kinase pathway via $\beta\gamma$ subunits, modulation of Ca^{2+} levels (agonists $\rightarrow$ $Ca^{2+}\downarrow$ $\rightarrow$ neurotransmitter-release$\downarrow$.))

hH_3R Function

As indicated in table 1.2, the hH_3R is mainly expressed in the CNS, but also in the peripheral nervous system. In the CNS, the hH_3R is expressed praesynaptically on histaminergic neurons and regulates the synthesis of histamine and its liberation from storing vesicles into the synaptic cleft. In this feedback mechanism, the hH_3R functions as an autoreceptor. Histamine can then postsynaptically stimulate the activation of hH_1R and hH_2R until it is inactivated by HNMT located in glia-cells (see figure 1.8). hH_3R expression is however not confined to histaminergic neurons. The hH_3R has been detected praesynaptically on several other neurons, where its action modulates the release of other neurotransmitters, such as dopamine, noradrenaline, serotonine, GABA and glutamate. Here, the constitutive activity of the hH_3R plays an important role. There is evidence that the $\beta\gamma$ complex can directly inhibit Ca-channels. The drop in Ca^{2+} concentration could then inhibit the fusion of the neurotransmitter storage vesicles with the praesynaptic membrane during the process of exocytosis. [25]

Regarding the peripheral nervous system, hH_3R expression has been described within the cardiovascular, gastrointestinal and bronchial system. Here, the hH_3R can modulate the release of noradrenaline (cardiovascular system) or acetylcholine and neuropeptides (gastrointestinal system, bronchial system).

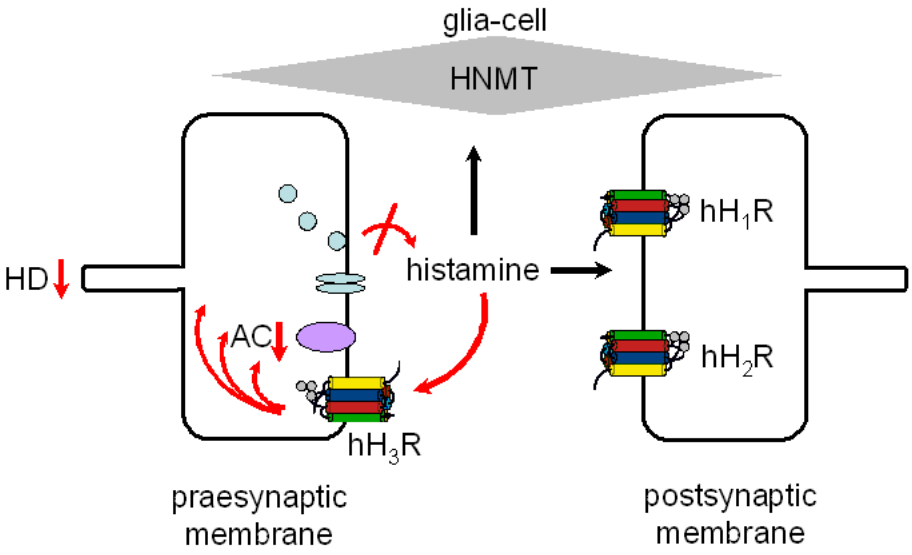


Figure 1.8: Function of hH_3R as an autoreceptor: Activation of the hH_3R by histamine results in an inhibition of adenylyl cyclase. As a consequence, the concentration of cAMP drops, which in turn reduces the activity of protein kinase A (PKA). cAMP-responsive element-binding proteins (CREB) thereupon are less active, which finally explains a reduced gene transcription. This likely represents the mechanism by which the conversion of L-histidine to histamine by the L-histidine decarboxylase (HD) is reduced. Through further pathways also the liberation of histamine into the synaptic cleft is inhibited.

hH_3R Ligands

The classification of hH_3R ligands into agonists, neutral antagonists and inverse agonists is extremely difficult. Depending on the signalling assay used, compounds can either appear to be agonists, antagonists or inverse agonists, as described for the compound proxyfan. [26] The reproducibility of *in vivo* tests is additionally hampered by the presence of various isoforms that have been described for the hH_3R and different expression patterns of these isoforms in different species. [21, 24]

As indicated in table 1.2, hH_3R agonists might be of therapeutic use for the treatment of insomnia or myocardial ischaemic arrhythmias, or could be antinociceptives. H_3R agonists have been developed starting from the natural ligand histamine that binds to human recombinant hH_3R with a binding affinity of $pK_i = 7.8$. The relatedness of most H_3R agonists to histamine consists mainly in the preservation of the 4(5)-substituted imidazole moiety, which so far has been neither replaced nor substituted without complete loss of affinity, prompting speculations that the capacity to tautomerise might be of importance in receptor activation. Yet, there is also evidence that histamine could bind to the hH_3R (and hH_4R) in its double protonated form: The hH_3R (and hH_4R) bind histamine with much higher affinity than the hH_1R and hH_2R [27] and the glutamate residue known to interact with the imidazole moiety could not be replaced by a glutamine residue indicating that a negative charge is essential at that site. [28] Larger structural variations in hH_3R agonists are allowed for the histamine side chain. In α and β position (see figure 1.6), methyl substituents are allowed, whereas a double methylation at the β position results in a significant affinity loss. One or two methyl groups can be added to the amino group increasing affinity and activity. Larger substituents such as an ethyl- or propyl-substituent result in a significant loss in activity, probably indicating that the space in the agonist binding site is very limited. [24] In several potent agonists the flexible side chain has been incorporated in a ring structure. The permitted size of the ring system is however limited. In figure 1.9 some structurally diverse hH_3R agonists are depicted.

For hH_3R inverse agonists, an even broader spectrum of potential applications has been reported due to their implication in the regulation of body-weight, arousal and sleep/wakefulness. A short summary is given in table 1.2. Known inverse agonists and antagonists can be classified into imidazole-containing compounds and non-imidazole ligands, of which the imidazole compounds have been discovered earlier.

Discovery of the constitutive activity of H_3-receptors, the cloning of the H_3R of different species, and knowledge of the histamine H_4-receptor has resulted in several reassessments regarding antagonism versus inverse agonism, affinity on the human H_3R and specificity. Although differences in species do not affect agonist binding, [29] they can influence inverse agonist binding. Significant differences in binding affinities of some

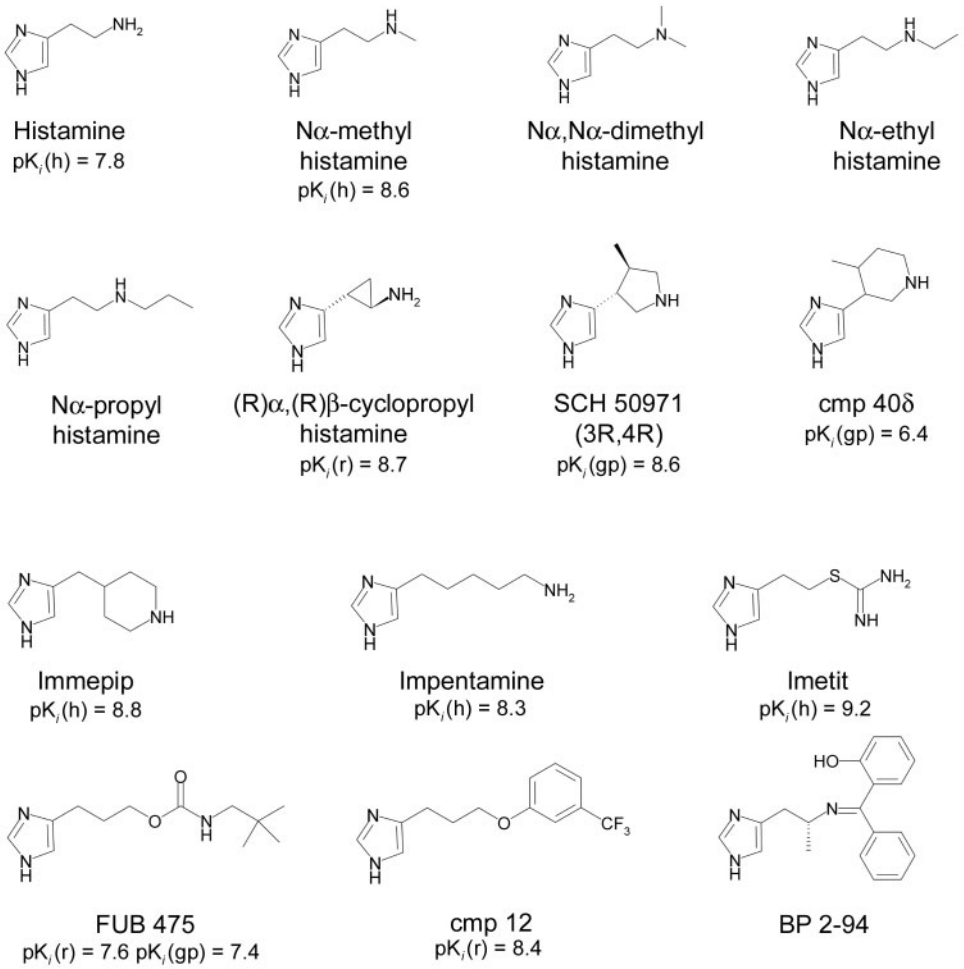


Figure 1.9: Structurally diverse H_3R agonists. pK_i values were measured on human (h), rat (r) or guinea pig (gp) receptors as indicated in brackets. Species differences were however reported not to influence agonist binding significantly. [29] For the depicted Nα-ethyl- and -propyl-histamine and compound 40δ, significant lower affinities and/or activities are observed, most probably, due to the larger substituents on the protonated moiety.

ligands (e.g. ciproxifan) have been observed, especially comparing rat and human H_3R test systems, while for other compounds (e.g. proxyfan, clobenpropit) comparable binding affinities on the hH_3R and rH_3R were observed. [29, 30]

The developing state of knowledge is best demonstrated for the imidazole-containing thiourea derivative thioperamide (see figure 1.10). Initially considered a highly selective antagonist with high affinity (pK_i(ratH_3R) = 8.4), thioperamide is nowadays classified as an inverse agonist with good selectivity over the H_1- and H_2-receptor but low selectivity over the H_4R (pK_i(hH_4R) = 7.3). On the human H_3R only a moderate affinity (pK_i = 7.2) is observed.

Some structurally diverse imidazole-containing inverse agonists are shown in figure 1.10.

Imidazole-containing compounds have a number of drawbacks that have resulted in

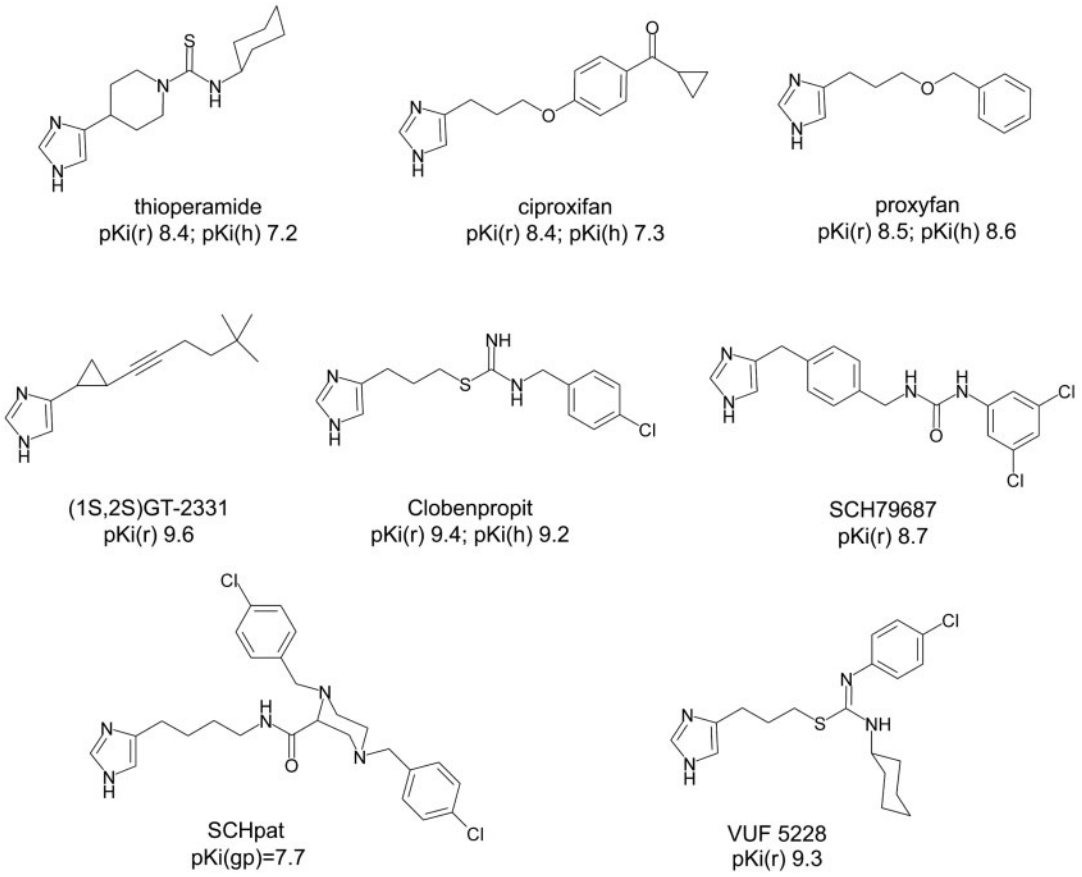


Figure 1.10: Structurally diverse imidazole-containing H_3R inverse agonists. For compound GT2331 (cipralisant) affinity for the (1S,2S) and (1R,2R) enantiomer are 9.6 and 8.5, respectively. [31, 32] As indicated in brackets, pK_i values were measured either on the human (h), rat (r) or guinea pig (gp) receptors.

considerable effort to substitute the imidazole moiety in hH_3R inverse agonists. Potential drawbacks of imidazole-containing compounds are listed below.

- Imidazole-containing ligands can interact with cytochrome P450 enzymes, which can result in a significant reduction of half-life or unpredictable side effects if other drugs have been administered.

- Imidazole compounds might be rapidly inactivated by the HNMT.

- The imidazole moiety is *per se* already rather polar due to its strong hydrogen donor and acceptor functionalities. Further polar groups could result in low CNS penetration. This may be desirable in drugs targeted to treat peripheral diseases such as SCH-79876 (see figure 1.10) that acts as a nasal decongestant in combination with H_1 antagonists; for CNS diseases it will however represent a limitation.

- Imidazole-containing compounds are usually less selective for the human H_3R compared to the hH_4R.

A replacement of the imidazole moiety was thus highly desirable. A comprehensive study on how the replacement of the imidazole moiety for a piperidine moiety affects binding affinity of structurally diverse H_3R inverse agonists has been carried out by Meier and coworkers. [33] Replacement of the imidazole moiety in thioperamide resulted for example in complete loss of affinity; for the clobenpropit analogue a pK_i = 6.3 was observed, which was significantly lower than the affinity observed for the imidazole compound (pK_i = 9.9). Similarly, a significant affinity loss was observed for replacements within the group of carbamate-containing structures, whereas inverse agonists containing an aliphatic or aromatic ether structure (e.g. ciproxifan) tolerated a replacement. Since then, many high-affinity compounds — mainly containing a piperidine-, piperazine-, or pyrrolidine moiety — were published. Some structurally diverse non-imidazole H_3R compounds are depicted in figure 1.11.

The actual state of knowledge regarding structure affinity relationship (SAR) of hH_3R inverse agonists can be summarised as followed:

- One strategy of inverse agonist design consists in increasing the distance between the basic moieties observed in agonists.
- In non/imidazole-containing inverse agonists, the basic moiety in the imidazole side chain can be omitted (see figures 1.9 and 1.10). In non-imidazole-containing compounds, incorporation of a second basic moiety in the side chain can however increase affinity significantly (e.g. JNJ-5207852 and JNJ-10181457 (see figure 1.11)).
- Attachment of large lipophilic groups to the linker moiety can increase inverse agonist affinity, selectivity and specificity. Aliphatic as well as aromatic lipophilic groups are tolerated.
- The linker moiety can cover a great structural variety: ethers and thioethers (e.g. proxyfan, ciproxifan), ketones, carbamates, esters, urea, thiourea (e.g. thioperamide), isothiourea (e.g. clobenpropit), guanidine, amidine, amine, amides, methylene units, phenyl rings (e.g. SCH79687), sulfoxides, sulphonamides, sulphamide, unsaturated hydrocarbon atoms, heterocyclic ring systems, etc.
- Many H_3R inverse agonists share a pharmacophore defined by a tertiary basic amine connected through an alkyloxy (often propyloxy) chain to a lipophilic moiety (usually an aromatic ring) [22] (see ciproxifan, FUB833, FUB836, etc.)

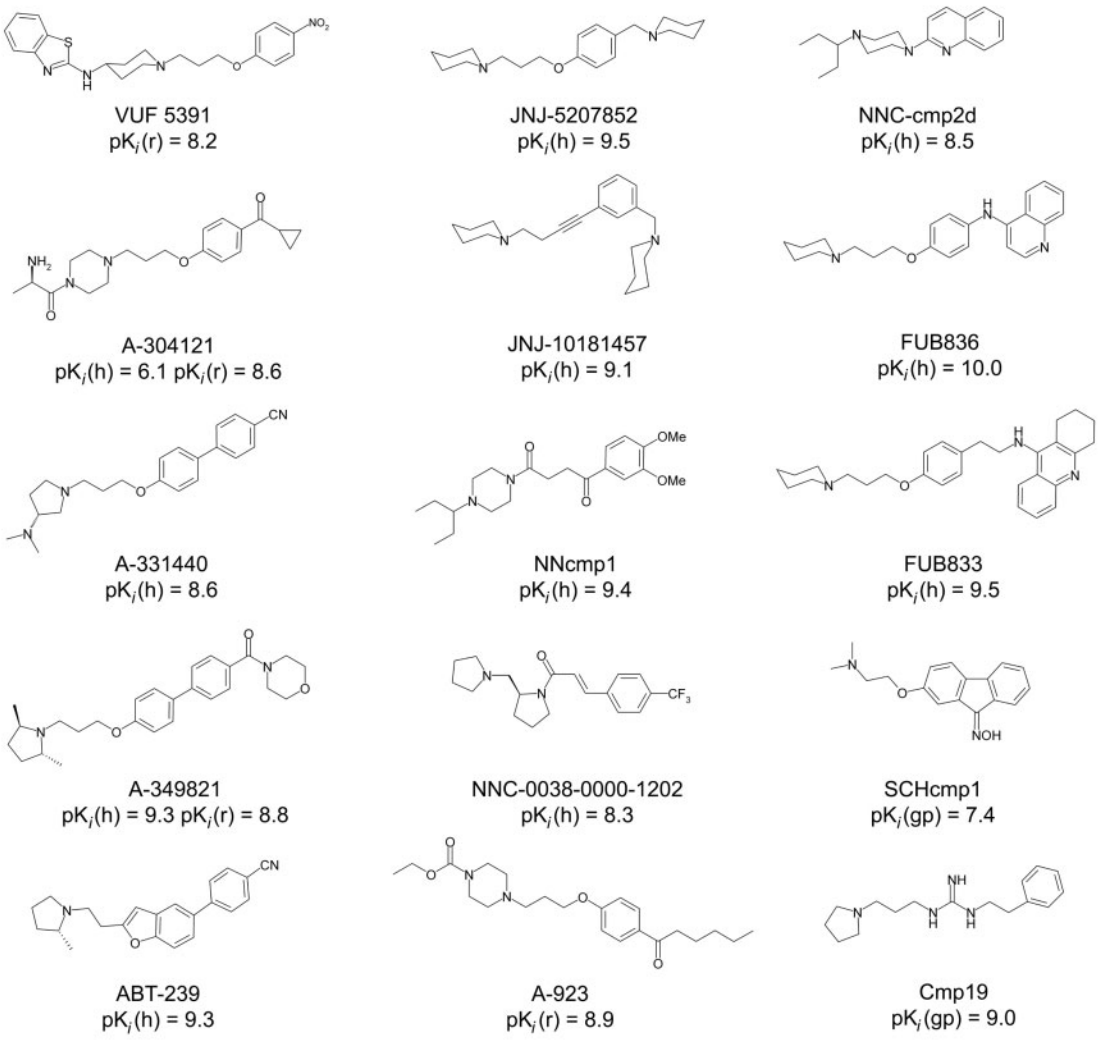


Figure 1.11: Structurally diverse non-imidazole-containing H_3R inverse agonists.

- Small structural changes can substantially influence the pharmacological profile. In a series of ether-containing H_3R ligands, only minor structural changes were required to transform (partial) agonists to inverse agonists (see figure 1.12). [34] The full range from full agonism to inverse agonism was observed also for a series of impentamine derived compounds (see figure 1.13). [35]

- In 2000, a H_3R pharmacophore model was described by De Esch and coworkers [36] proposing the existence of **two** lipophilic subpockets. At the same time, branched compounds were published from the Schering research group (see figure 1.10; SCHpat) affirming this pharmacophore model. [37] Compounds such as thioperamide and burimamide were described to interact with subpocket 2, while clobenpropit, iodoproxyfan and most other compounds were interacting with subpocket 1. Later, the pharmacophore model was extended for four hydrogen bond interaction sites. Only clobenpropit-derived branched ligands were predicted

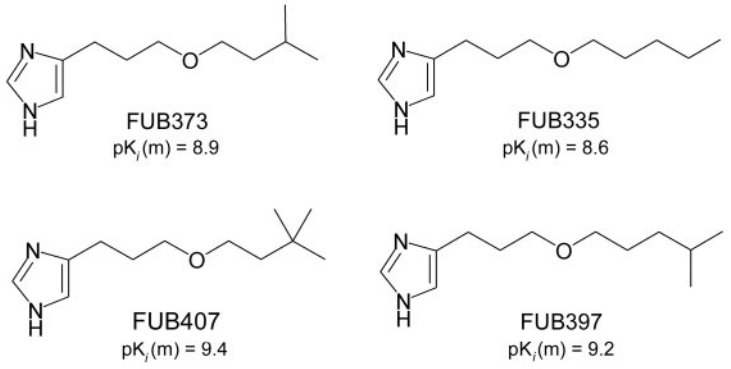


Figure 1.12: Slight modifications can determine if a compound is a (partial) agonist or inverse agonist. Compounds FUB373 and FUB407 were reported to be partial agonists, while FUB335 and FUB397 are inverse agonists. [34] pK$_i$ values were determined on mouse H_3R.

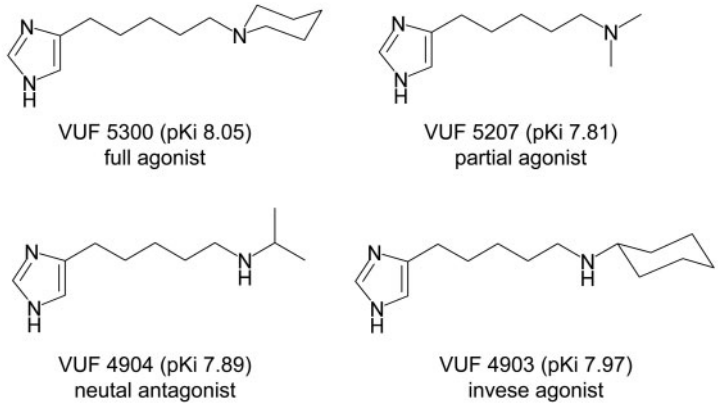


Figure 1.13: Structurally similar compounds covering the entire range from full agonism to inverse agonism. Binding affinities were determined on human H_3R.

to interact with all proposed sites simultaneously (see figure 1.10; VUF 5228). [38]

- Another SAR study was published by Abbott and was based upon compound A-923 (see figure 1.11) which showed high affinity towards the rat H_3R (pK$_i$ = 8.86), however lacked selectivity and oral bioavailability. [39] In this series, the piperazine moiety was most adequately replaced by a pyrrolidine group and the *n*-pentyl group by a cyclopropyl-group.
  A-923 was also the starting point for developing aminoalkoxybiphenylnitriles such as A-331440 (see figure 1.11). [40] From this series also D-alanine piperazine amides such as A-304121 were derived. Compound A-304121 is noteworthy as it is reported to be the most species-selective H_3R inverse agonist, with an affinity for the rat cortical H_3R of pK$_i$ = 8.6, 30-fold lower potency at the guinea pig ileum and 300-fold less potent in binding studies in human cortex (pK$_i$ = 6.1). [41] Sequence

heterogeneity is thereby mainly due to the point mutations at site 119 (human, guinea pig: T, rat: A) and 122 (human: A, guinea pig, rat: V). The importance of these residues was shown by the hH_3R double mutant (T119A, A121V) in which the high affinity of A-304121 observed for the rat H_3R was restored.

- In a series of aminoalkoxy-biaryl-4-carboxamides published by Abbott, [42] wide structural alterations were tolerated for the amide portion while the basic site seemed to be more restrictive. The most active compound A-349821 is depicted in figure 1.11.

- Compounds with a combined HNMT and hH_3R activity were synthesised by Apelt and coworkers [43] and resulted in ligands with subnanomolar affinity (see FUB833 and FUB836 in figure 1.11), however, lacking *in vivo* activity. Especially FUB833 gives a good idea on the volume available for ligand binding as it is one of the sterically most demanding compounds so far described.

- VUF 5391 is an interesting compound as a large functional moiety is attached to the 4-position of the piperidine group. Although so far only tested on the rat H_3R, the accommodation of sterically demanding substituents might be also possible in the human H_3R.

- Compounds from Johnson&Johnson were already mentioned with respect to the two basic moieties present in JNJ-5207852 and JNJ-10181457. Additionally noteworthy is the low conformational freedom of compound JNJ-10181457.

- Also ABT-239 (see figure 1.11) from Abbott was developed in order to reduce the conformational freedom of the linker moiety and has been reported to show cognition and attention enhancing properties in animal models. [44]

- Starting from the H_2R antagonist dimaprit, Linney and coworkers synthesised a series of clobenpropit-derived compounds (see compound 19 in figure 1.11). [45] The pyrrolidine and guanidine moieties were ideally separated by a propyl or butyl moiety. Variations in the chain length between the aromatic ring and the guanidine had less pronounced effects. No significant sensitivity to the nature of the aromatic substitution was observed. Increasing the size of the cyclic amine moiety from a 5- to 6-membered ring did not increase affinity; further enlargement however decreased affinity. Linney *et al.* speculated that the guanidine group would occupy the same site as the imidazole moiety.

- NNcmp1, NNcmp2d and NNC-0038-0000-1202 are structures published by Novo Nordisk (see figure 1.11). While the 1-alkyl-4-acylpiperazine compound NNcmp1 was reported to have a high polar surface area, 2-(1-piperazinyl)quinoline derivative

such as NNcmp2d were less polar. The bulky *N*-alkyl group was reported to be required for high affinity.

- Schering described a series of fluorenes of which SCHcmp1 (Z-isomer) is the most active (see figure 1.11). The SAR for this series revealed that both a two carbon linker and a three carbon linker maintained high affinity. The dimethylamino group could be replaced by piperidine or pyrrolidine moieties preserving high H_3R affinity. On the other hand, replacement of the oxime moiety by an hydroxy-group or a dimethylamine moiety significantly lowered binding affinity (pK_i = 6.72 or pK_i = 6.19, respectively). Acetates, carbamates and other nitrogen analogues were not tolerated as substituents for the oxime group. Similarly, affinity was decreased for ketones and substituted oximes. Also a methyloxime substitution resulted in inactives indicating that the hydrogen bond group of the oxime moiety could be important.

Mutational Studies on the hH_3R

Mutational analysis on the hH_3R have been reported by Uveges *et al.* who carried out alanine scanning mutagenesis in TM5 and evaluated the effect upon agonist binding, [27] Jacobsen *et al.* who mutated residues D114/3.32, E175/4.65 and E206/5.46, [28] and Yao *et al.* who mutated T119/3.37A and A122/7.40V. [46] Uveges and coworkers observed that mutation of residue D114/3.32 to either asparagine or glutamate resulted in no detectable cAMP response indicating that the mutant receptor could not interact with the G-protein. The point mutation D114/3.32N was also carried out by Jacobsen and coworkers who obtained similar results.

Uveges and coworkers additionally carried out alanine-scanning mutagenesis of a consecutive stretch of 14 residues in helix 5 starting from W196/5.36. A202/5.42 was mutated to glutamine. Potency was determined by the adenylyl cyclase assay where the forskolin induced inhibition of cAMP production by histamine was measured. While histamine could reduce forskolin-induced cAMP production to 91.6%, 2- to 5-fold increases in potency were observed on the mutant receptors W196/5.36A and T204/5.44A, while the point mutations L119/5.39A, A202/5.42Q, E206/5.46A and F207/5.47A significantly reduced histamine potency (e.g. A202Q: 72% inhibition). For other agonists similar effects were observed when tested on the mutant receptors. Binding affinity measurements were carried out using the radio ligand [^{125}I]-iodoproxyfan. [47] For this purpose, cells expressing the mutant receptors, were incubated with 25 pM [^{125}I]-iodoproxyfan until equilibrium was reached. The membranes were then filtered and washed and radioactivity was determined. Then, the competitive ligand was incubated and the decrease of radioactivity measured. The resulting IC_{50} could be converted to K_i values applying the Cheng-Prusoff

equation. The point mutation E206A had the greatest effect on the affinity of agonists, with histamine and R-α-methylhistamine being affected the most. Impentamine was relatively unaffected.

Jacobsen and coworkers described the mutations D114/3.32N, E175/4.65Q and E206/5.46Q. As already mentioned, the point mutation D114N resulted in loss of binding of [^{125}I]-iodoproxyfan. The pharmacological profile of the E175/4.65Q receptor was reported to be identical to the wild type receptor. The mutation E206/5.46Q resulted in a constitutively active receptor that could not bind R-α-methylhistamine, while binding of other agonists (e.g. iodoproxyfan) was only decreased 10-fold. Inverse agonists, such as ciproxifan could inhibit this constitutive activity.

Yao and coworkers mutated the amino acids T119/3.37 and A122/7.40 in the human H_3R to the corresponding residues in the rat H_3R (A3.37, V7.40) and could thereby restore the high binding affinity of A-304121 observed at the rat receptor for the human receptor.

1.4.2 The Human H_4-Receptor

The histamine H_4-receptor was discovered in 2000 simultaneously by several groups. [14–18] In retrospect, first hints for the existence of a fourth histamine receptor can be found already in 1975 when Clark *et al.* showed that histamine could induce eosinophil chemotaxis. [48] Today it is known that the H_4R is mainly expressed on hematopoietic cell types such as eosinophils, mast cells, basophils, ... The fact that these cells are involved in the development and symptomatology of asthma and allergies has led to the expectation that compounds targeting the H_4R alone or in combination with H_1R antihistamines could play an important role in the treatment of these diseases. Recently, a review has been published [49] that summarises the current knowledge on this receptor of which a short survey is given in table 1.3.

In the search for specific H_4R ligands, firstly, compounds of other histamine receptors were evaluated for their affinity towards the H_4R. As might be however expected due to the low sequence conservation of the H_4R to the H_1R and H_2R, pharmacological profiles are quite dissimilar for these receptors. The H_4R does not bind classical H_1R antihistamines and only for a few H_2R ligands such as burimamide (H_2/H_3 antagonist) or impromidine (H_2R agonist/H_3R inverse agonist) significant affinity has been observed (see figure 1.14). Classical H_2R ligands (cimetidine, ranitidine) lack affinity. In contrast, for many imidazole-based H_3R ligands, significant affinity has been observed. Most of these compounds are partial or full agonists such as R-α-methylhistamine, N-methylhistamine, clobenpropit, iodophenpropit or imetit. Of the evaluated inverse agonists, only thioperamide showed affinity for the H_4R. Non-imidazole based inverse agonists were usually devoid of H_4R affinity as shown in figure 1.14.

Table 1.3: A short overview on the hH_4R.

classification	biogen aminergic G-protein coupled receptor
noteworthiness	constitutive activity reported by [16]
location of gene	chromosome 18
gene organisation	3 exons, 2 introns; no splice variants observed so far
protein sequence	390 amino acids
homology to other receptors	hH_1R, hH_2R < 25%; hH_3R: 35% overall; 58% within TM
affinity for histamine	pK_i = 7.79-8.36
expression loci	cells of hematopoietic lineage (eosinophils, mast cells, basophils, dendritic cells, T-cells); expression could be regulated by interferon, TNFα, IL-6, IL-10 and/or IL-13
molecular function	chemotaxis of eosinophils and mast cells to histamine; control of IL-16 release from $CD8^+$ T-cells
potential therapeutic application	antiinflammatories, allergic rhinitis (via mast cells), asthma (via mast cells, eosinophils, T-cells), itching associated with atopic dermatitis, autoimmune diseases such as rheumatoid arthritis, multiple sclerosis, type I diabetes, systemic lupus erythematosus (via dendritic cells, T-cells)
isoforms	so far not reported
species differences	H_4R of rat, mouse, guinea pig and pig share only 65-73% homology; although expression patterns are similar, pharmacological profiles and signal transduction responses are dissimilar
signal transduction	coupling to $G_{i/o}$ proteins that are linked to several second messenger pathways (inhibition of adenylyl cyclase $\rightarrow$ [cAMP]$\downarrow$; in mast-cells probably link to PLC $\rightarrow$ release of Ca^{2+} from ER; induce of MAP kinase signalling cascades)

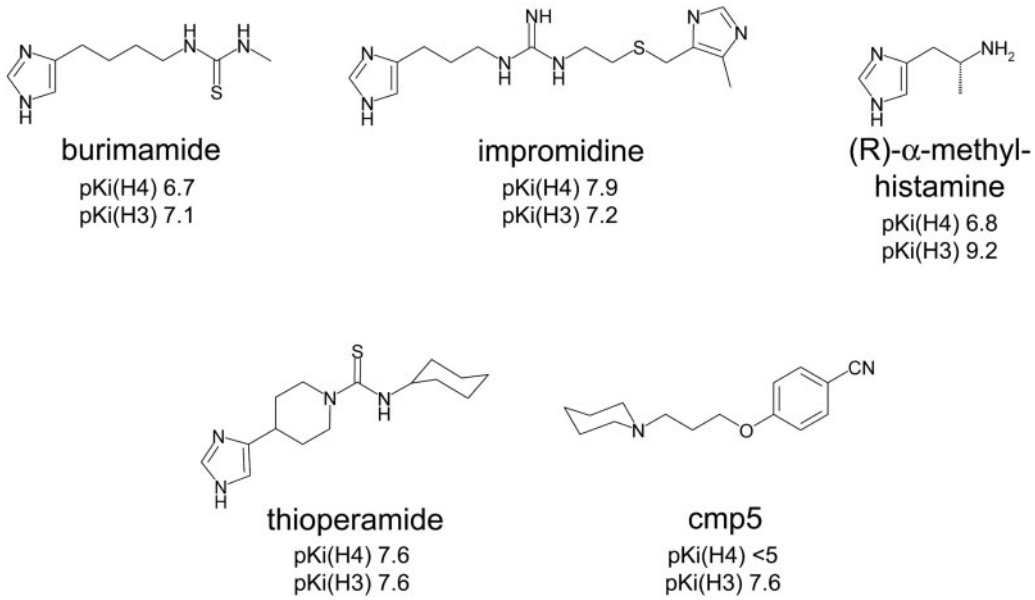


Figure 1.14: Unselective H_4R ligands. (R)-α-methylhistamine acts as a H_4R agonist, burimamide and impromidine as partial agonists, thioperamide as an antagonist (inverse agonist). Cmp5, a non-imidazole H_3R inverse agonist containing the commonly observed (3-piperidine-1-ylpropoxy)benzo-motif, is inactive at the H_4R.

The first selective H_4R antagonist was JNJ7777120 (see figure 1.15) published by Johnson&Johnson. [50] Various analogues of these indolylpiperazines were prepared. Regarding the *Y*-substitution, an increase in affinity was observed for a methyl-substituent while larger substituents (e.g. ethyl, phenethyl) lowered affinity significantly. The amide linkage seemed to be important for high affinity. Regarding the indole moiety, a free hydrogen bond donor function was essential. Diverse substituents R_4-R_7 were generally well tolerated except for a -OCH_3 group in R_5. Two compounds of this series are depicted in figure 1.15. They were more than 10000-fold selective for the H_4R. In figure 1.15 also a selective H_4R agonist (OUP-16) is depicted. [51]

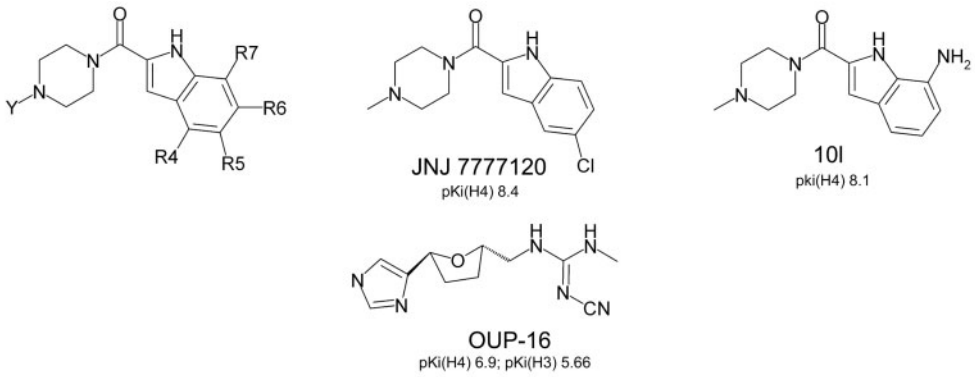


Figure 1.15: **(up)** Selective H_4R antagonist JNJ7777120 and compound 10l from the same series. **(down)** Selective H_4R agonist OUP-16.

Mutational analysis were reported by Shin and coworkers. [52] In this study, D94/3.32 and E182/5.46 were identified as main interaction sites for histamine binding. Site directed mutation of D94/3.32 to alanine, asparagine or glutamate resulted in complete loss of histamine binding. Likewise, the mutations E182A and E182Q abolished histamine binding while the mutation E182D partially preserved histamine affinity, which was however reduced 10-fold. Following the modelling studies carried out by Shin and coworkers, it might be expected that D94/3.32 forms a salt-bridge interaction with the amine moiety of histamine, while E182/5.46 forms a hydrogen bond or ion pair to the Nτ-atom of the imidazole moiety. The existence of a saltbridge interaction between E182 and the imidazole moiety is supported by the fact that E182 could not be replaced by a glutamine residue that would be capable of substituting the hydrogen bond acceptor function of E182. T178/5.42 and S179/5.43 were neither significantly involved in histamine binding nor in receptor activation. The point mutations T178A and S179A reduced affinity for histamine only 2-4-fold. Although T178/5.42 was interacting with the Nπ-atom of the imidazole moiety in the proposed model, the mutations T178A and S179A did not dramatically influence histamine affinity or signalling.

In the generated model, N147/4.57 and S320/6.52 were located in proximity to the putative histamine binding site. Yet, alanine substitution of both residues reduced histamine binding only slightly. Both residues seem to be however important in receptor activation. The point mutation N147Y (in order to mimic the H_3R) was detrimental to histamine signalling. The point mutation S320/6.52 to phenylalanine (in order to mimic the H_1R, H_2R or other biogen aminergic GPCRs) greatly reduced the potency of histamine binding, at the same time however doubling the maximal response of the receptor to histamine binding. Analogous results were also observed for other agonists.

Chapter 2

Methods

2.1 Generation of Homology Models

2.1.1 Introduction

In recent years, the number of protein structures determined experimentally and deposited in structural databases such as the Protein Data Base (PDB) [53] has been increasing exponentially. Still, the acquisition of structural information is a slow and expensive process. Especially membrane proteins have been proven difficult to crystallise due to low expression levels and difficulties regarding the crystallisation process itself. [5]
If the 3D structure of a protein is resolved and the sequence of a related protein of interest is known, the approach of comparative (homology) modelling becomes applicable. The structural information of the template protein can then be used as a scaffold for the generation of a model of the protein of interest (target protein). Since its release in 2000, bovine rhodopsin has been successfully used as such a scaffold for the generation of various GPCR homology models. [54–56] The resulting models can then effectively assist a drug discovery process as they offer the possibility to understand receptor-ligand interaction on an atomic level and can bring new impetus in lead-finding via database-screening methods or *de novo* design strategies (see scheme 2.1). Figure 2.2 shows a flowchart of homology model generation.

2.1.2 Sequence Analysis Tools

Prediction of Transmembrane Regions

Prediction of transmembrane regions relies on the distinctive patterns of hydrophobic (intramembraneous) and polar (loop) regions. [57] Prediction algorithms can be classified

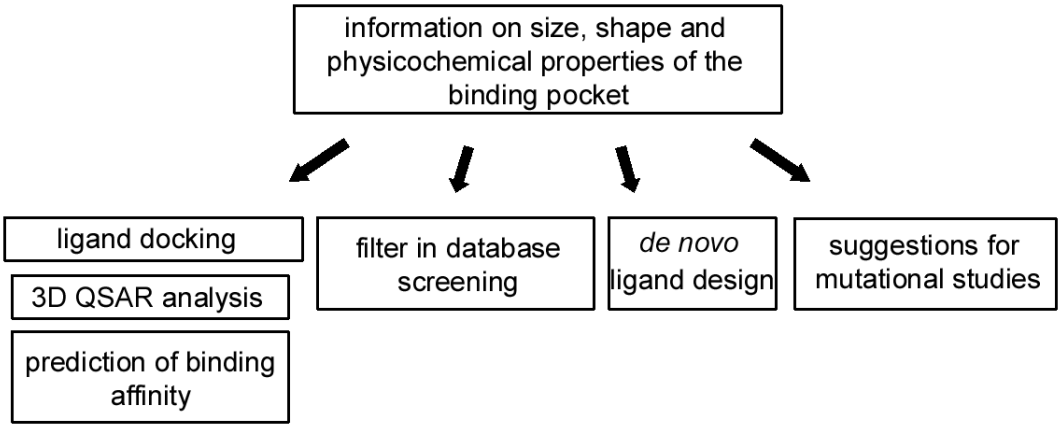


Figure 2.1: Possible applications of protein homology models in the drug discovery process.

into local and global approaches. *Local approaches* depend on local properties of the amino acid sequence. A conceptually easy strategy of this historically older class is the "sliding window" algorithm presented in 1982 by Kyte/Doolittle. [58] Each amino acid is assigned a specific hydropathy score that describes its tendency to be preferentially located either in a polar (negative score, e.g. R: -4.5) or an apolar (positive score, e.g. I: 4.5) environment. The protein sequence is then scanned with a window of defined size while the sum of hydropathy values is calculated within this window. The resulting value is compared with a cutoff defined and the residue consequently assigned to be transmembraneous or not. Combination of this simple hydrophobicity analysis with the "positive inside rule", which is based on the observation that the intracellular end of a helix is often marked by the presence of positively charged residues, gives the TopPred algorithm. [59] Another local approach is the DAS (Dense Alignment Surface) method that was originally introduced to improve sequence alignments in the GPCR family and was then generalised to any integral membrane protein. [60] TM Finder is based on a dual prediction algorithm incorporating segment hydrophobicity and non-polar phase helicity. To each amino acid of a candidate TM segment a hydrophobicity and helicity value is assigned, then a sliding window is applied to calculate the moving average. [61] *Global approaches* have in common that they try to predict the statistically most probable topology of a protein. A representative of this class is the program TMHMM that relies on a Hidden Markov Model (HMM). For each amino acid the probability that it belongs to one of several predefined states (e.g. inside loop, transmembrane region, outside loop) is calculated. TMHMM is thus a typical representative of the class of global approaches, as also sequence information outside of the transmembrane areas is incorporated in order to identify the protein's topology. [62] SPLIT4.0 is another global algorithm for TM prediction based on the observation that clusters of positively charged amino acids are observed

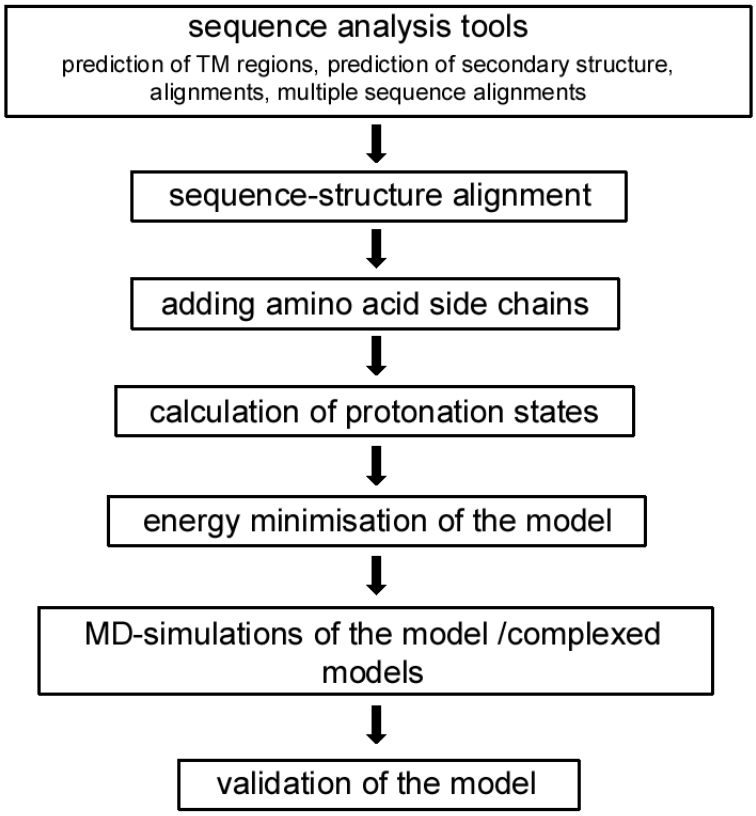


Figure 2.2: Flowchart of homology model generation

more frequently near the cytoplasmatic surface. [63]

Prediction of Secondary Structure

Prediction of secondary structure is more complex and no simple local approaches exist, as this would limit prediction accuracy to levels only slightly above 60%. [64] Currently, the best programs (PROF, PSIPRED, PHD, SSpro, . . .) have about 76% prediction accuracy, which required incorporation of advanced statistical methods (e.g. neural networks or HMM), sequence pattern matching and evolutionary conservation information. The background for incorporating sequence pattern matching is that the secondary structure of an unknown protein can often be predicted based on local sequence similarities to fragments of known structures.

Sequence Alignments

During evolution, sequences accumulate insertions and deletions as well as substitutions. In order to evaluate if two protein sequences are related and which amino acids correspond to each other, the sequences must be aligned in a way that an optimal score is obtained.

An alignment is scored by adding up match scores and penalties for gaps. In figure 2.3 the alignment procedure is explained aligning two short stretches of protein sequence (for a more detailed description see [65]). In order to find the optimal alignment score, one starts in the top-left corner. A step to the right would represent a deletion, a movement down an insertion while a diagonal move is equivalent to a match or a substitution. The simple scoring function gives no penalty for a match, each substitution costs 1 point. For each insertion or deletion a penalty of 1 point is given (see gap function). Starting from the top-left corner possible movements are a deletion (-/C), an insertion (A/-) or a substitution (A/C). The matrix is now filled up with values trying to reach each field from the three possible directions with minimal costs. The optimal alignment score for this scheme is four. Once, the optimal score has been evaluated, a bootstrapping process is carried out in order to find all alignments that have resulted in this optimal score. This process is depicted in figure 2.3 by red circles and small arrows. The two corresponding alignments are depicted on the right.

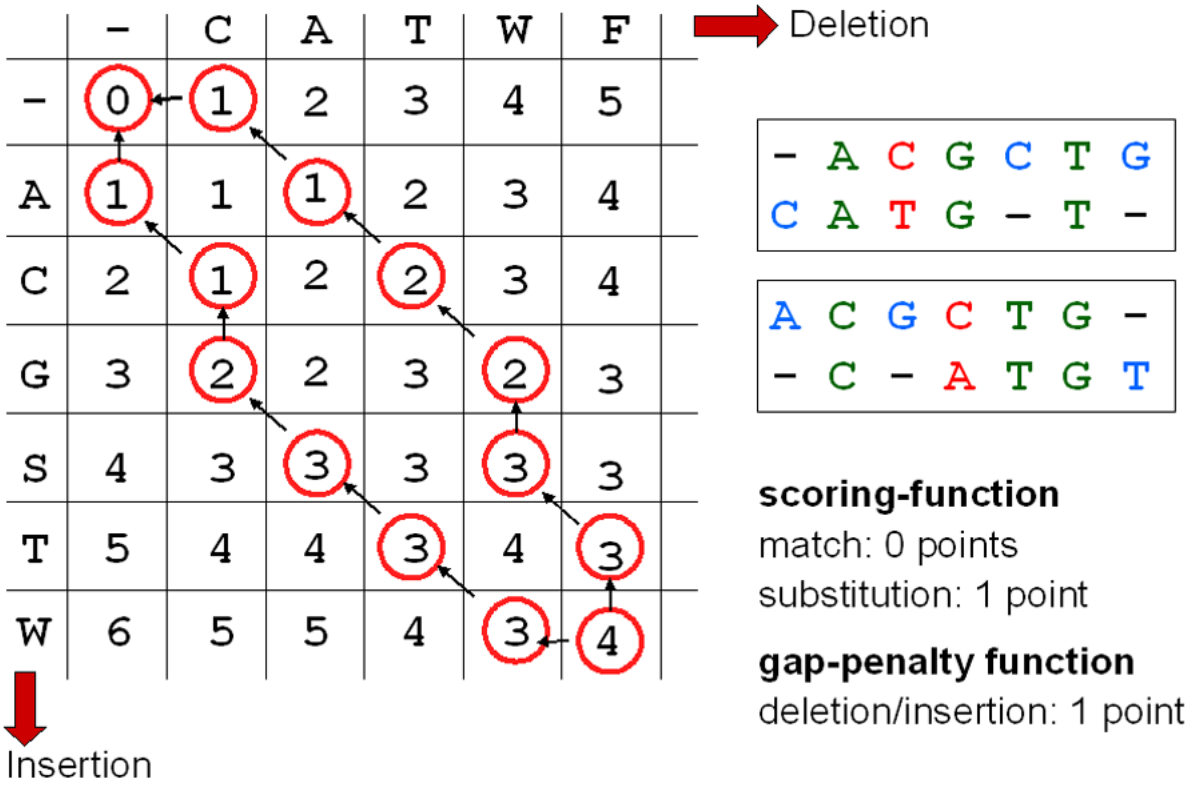


Figure 2.3: Alignment procedure (see text for explanation). In the two possible alignments obtained, matches are depicted in green, substitutions in red and insertion/deletions in blue.

In reality, scoring functions are more complex, as the score must account for the fact that certain amino acids can substitute for others in proteins without altering the protein's function or structure. Thus, substitution of a leucine for an isoleucine will result in a better

score than the pair leucine–arginine due to the more similar physicochemical properties of the Leu/Ile pair. Scoring matrices are 20 × 20 number schemes that assign a score for each possible substitution of amino acids. They have been empirically derived and can for example account for amino acid similarity (identity matrix) or evolutionary relatedness (PAM (percent accepted mutation)-matrix). Also gap-penalty functions are slightly more complicated in order to account for the fact that it is evolutionary more probable that once one big insertion/deletion occurred rather than several small ones. Thus, in most alignment programs, a gap-open and a (less costly) gap-extension penalty can be adjusted.

Multiple Sequence Alignments

Multiple sequence alignments of protein sequences can be used to find characteristic motifs and conserved residues in protein families, determine evolutionary relatedness and improve prediction of secondary structure. A multiple alignment is built up gradually by aligning the closest sequences first and successively adding more distant ones. [66] Thus, in a first step all pair wise alignments are generated and a guide-tree is built that reflects the relatedness of the sequences studied. Following the guide-tree, first pair wise alignments are generated and then combined to a complete alignment. This strategy has the advantage, that it leads to reproducible results and does not depend on the order in which the individual sequences are aligned (see figure 2.4).

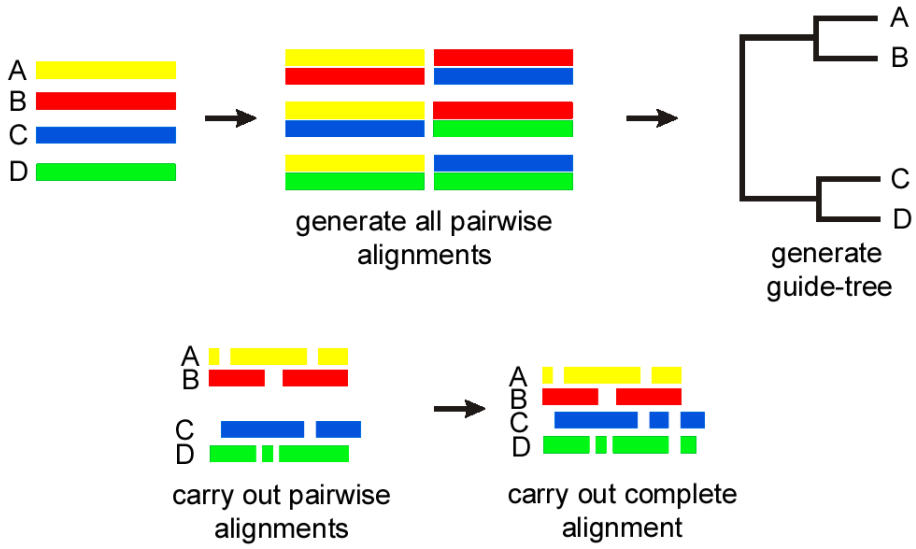


Figure 2.4: Steps in carrying out a multiple sequence alignment.

The information contained in a multiple sequence alignment can be displayed in form of a phylogenetic tree where related sequences are intuitively clustered. The closer a branch point lies to the center of the tree, the more divergent are the sequences on these

branches. If only related sequences are studied and no divergent sequence is included for comparison, one will obtain an unrooted tree. The distance to the center of the tree is then not related to the time in evolution when a common ancestor sequence started to diverge. For a detailed description of multiple sequence alignment methods and phylogenetic trees the reader is referred to [65].

2.1.3 Sequence Structure Alignment

Once a reasonable alignment has been found between the template and target protein, a sequence structure alignment can be carried out in which the backbone atoms of the target protein are arranged identically to that of the template protein. This will however only be possible in gap-free regions of the alignment, so called conserved regions. Much more difficult is the generation of non-conserved loop regions that often show little sequence conservation and may diverge in length from the template protein. A common method to obtain coordinates for these regions is carrying out a loop search. Resolved protein structures are searched for a peptide of identical length that can fill the gap in the protein model without introducing large distortions. Alternatively, loops can be generated using *de novo* strategies (protein threading), where by means of for example simulated annealing energetically favourable loop conformations are generated.

2.1.4 Adding Amino Acid Side Chains

During a sequence structure alignment, structural information is only obtained for the backbone region of the target protein. From statistical analysis of known protein structures is has been observed that amino acid side chains tend to exist in certain energetically favoured conformations (rotamers). The available conformational space for each side chain is further reduced by the dependency of the side chain conformation on the coordinates of the backbone. Programs such as `SCWRL` (**s**ide**c**hain **p**lacement **w**ith a **r**otamer **l**ibrary) incorporate this statistical information in the side chain prediction. [67] Additionally, the program allows to adopt side chain placements from the template structure, which is a useful strategy for conserved residues.

2.1.5 Prediction of Protonation States

At physiological pH of 7.4, acidic amino acids such as aspartate and glutamate (pK_a = 4.4) are deprotonated/negatively charged, while basic amino acids such as lysine (pK_a = 10.0) and arginine (pK_a = 12.0) are protonated, thus positively charged. Histidine (pK_a =

6.5) will be present in an equilibrium of protonated and deprotonated state while cysteins (pK_a = 8.5) and tyrosines (pK_a = 10.0) are to a large extent neutral. These pK_a values are however only true for an aqueous environment with a dielectric constant of $\varepsilon_r \approx 80$. In a protein environment with an estimated dielectric constant of 2-20, [68] the pK_a values can considerably shift from the values given. This shift is due to mainly two factors:

- A desolvation of the residue by nearby residues, which will favour the corresponding neutral forms of all titrable sites.
- Electrostatic interactions with nearby residues, which can result in stabilisation of a charge even in an apolar environment.

The `UHBD` program [69] calculates pK_a-shifts in proteins by determination of the differences in electrostatic work of altering the charge of a titrable group from the neutral to the charged state in the protein and the work of making the same alteration for the residue in aqueous solution. In solution, the ΔG_w value for the ionisation equilibrium is known from the pK_a value in solution:

$$\Delta G^0 = 2.303 \cdot RT \cdot pK_a \tag{2.1}$$

In order to calculate the corresponding ΔG_p value for the ionisation equilibrium in the protein environment, the free energy change for transferring each species of equation 2.2 into the protein environment is calculated:

$$HA + H_2O = A^- + H_3O^+ \tag{2.2}$$

However, this calculation merely accounts for the desolvation of the species. Additionally, the effect of electrostatic interactions with other residues has to be considered. Simultaneous consideration of multiple titrable sites is computationally quite demanding as for N ionisable residues 2^N possible ionisation states result, which is impracticable for moderately sized proteins. In order to obtain the electrostatic interaction energy, one titrable site is charged at each time and the electrostatic potential generated by this site at the other titrable residues is calculated. The calculation is repeated for all sites. Finally, a Monte Carlo simulation is carried out in order to determine a reasonable distribution of charged states within the protein (depending on the external pH, T and ionic strength).

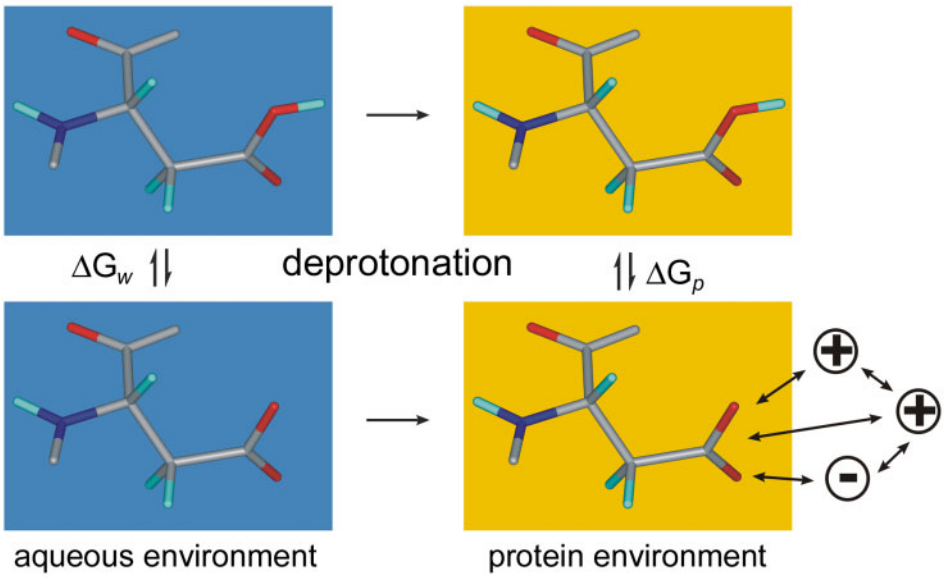


Figure 2.5: Calculation of pK$_a$-shifts as carried out in the UHBD program. The ΔG_w value for the aqueous environment is known from the pK$_a$ value of the residue. The ΔG_p value for the protein environment is calculated from the desolvation of de/protonated species and electrostatic interaction within the protein.

2.1.6 Force Field Methods: Energy Minimisation and MD Simulations

After the sequence structure alignment, the insertion of loop regions and addition of amino acid side chains, there will often remain steric clashes and distorted bonds in the resulting model. The goal of an energy minimisation is to relax the resulting structure and find an energetically favourable conformation of the protein. In order to carry out an energy minimisation, the coordinates of the protein model are required. Based on this information, the potential energy of the system can be calculated.

Theoretically, quantum mechanical methods would be required in order to solve this task as they account for both the position of nuclei and electrons of each atom considered. If the complexity of the molecular system studied increases, these methods will become impracticable due to the large time-consumption of the calculations required. Semiempirical methods represent an intermediate between quantum mechanical and force field methods that are usually applied for the simulation of larger molecular systems such as protein structures. Using force fields, the electrons are no longer explicitly considered and the energy of the molecule only depends on the position of the nuclei; a simplification that is permitted through the Born-Oppenheimer approximation.

From a mathematical point of view a force field is a function of potential energy that exclusively depends on the position of the nuclei. The contributions to the potential energy of the molecular system can be subdivided into bonded and non-bonded interactions (see

table 2.1 and figure 2.6). Bonded interactions can be further subdivided with regard to the number of particles involved resulting in a term describing bond stretching (two-body interaction), angle bending (three-body interaction) and a term describing bond rotation (torsion) (four-body interaction). Non-bonded interactions are calculated between all pairs of atoms (*i* and *j*) that are in different molecules or that are in the same molecule but separated by at least three bonds. Non-bonded interactions comprise electrostatic interactions and van der Waals interactions.

Table 2.1: Types of interactions considered in a force field.

bonded interactions	**non-bonded interactions**
bond stretching	electrostatic interactions
angle bending	van der Waals interactions
torsion angle rotation	

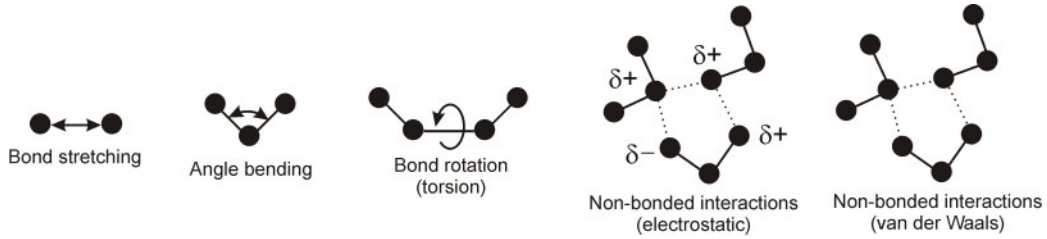


Figure 2.6: Bonded and non-bonded interactions included in a force field potential.

One functional form for a simple force field is given in equation 2.3. [70] In this example, the terms for bond stretching and angle bending are modelled by a harmonic potential and the torsion potential is described by a periodic cosine function. For calculating the non-bonded interactions the Coulomb potential is used for the electrostatic interactions and a Lennard Jones potential for the van der Waals interactions.

$$\begin{aligned} \mathrm{E}\left(r^N\right) = & \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right) \\ & + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \end{aligned} \tag{2.3}$$

An equation describing the contributions to potential energy represents the core of each force field and differences in force fields (such as GROMACS, CVFF, AMBER, ...) arise from variations in the terms implemented. A more thorough description of force field terms can be found in appendix 10.1.

Energy minimisation represents a “walk” on the multidimensional energy-hyperface into the valley of potential energy that is located closest to the starting point (see figure 2.7). This also represents the limitation of energy minimisation protocols as only a small percentage of the energy surface is explored during an energy minimisation and only the closest local minimum — and unlikely the global minimum — is encountered. Several algorithms are available to carry out energy minimisation that can be grouped into solely energy based methods (e.g. the Simplex method), gradient based algorithms (e.g. steepest descent, conjugate gradient) and procedures taking into account also the second derivative of the energy hyperface (Newton Raphson). [70, 71] Gradient based algorithms thereby are the most commonly used strategy. Here, the first derivative of potential energy determines the direction for the next move on the energy landscape.

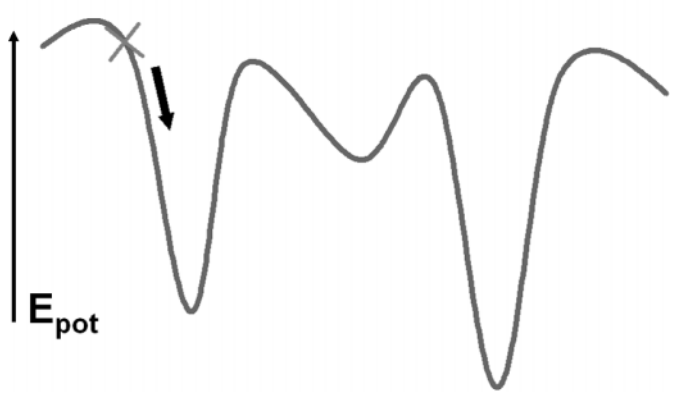


Figure 2.7: Energy minimisation represents a “walk” on the multidimensional energy-hyperface that is here simplified to two dimensions. Starting from a point of high potential energy, minimisation will result in a relaxed structure located in the closest local minimum. There is however no guarantee to encounter the global minimum.

Energy minimisation is a prerequisite to later study the protein structure by means of molecular dynamics simulations. When carrying out an MD simulation, in an initial step velocities are randomly generated acting on all atoms depending on the adjusted system temperature. The system then evolves from this starting point solving Newton’s laws of motion where the velocity v at time $t = 0$ will result in a set of coordinates x at time $t = i$. From these coordinates the potential energy E_{pot} can be calculated, the first derivation of which gives the force F acting on each atom resulting in a new velocity v at time $t = i$. When keeping the analogy to the “walk” on the energy hyperface represented by an energy minimisation, an MD simulation would correspond to floating over the energy landscape. Kinetic and potential energy give the total energy of the system that must be conserved during MD simulations. The importance of starting an MD simulation with an energetically favourable structure arises from the energy conservation within the system. If, during the course of the simulation, regions of low potential energy are sampled, the part of kinetic energy increases significantly, which can result in large distortions of the structure.

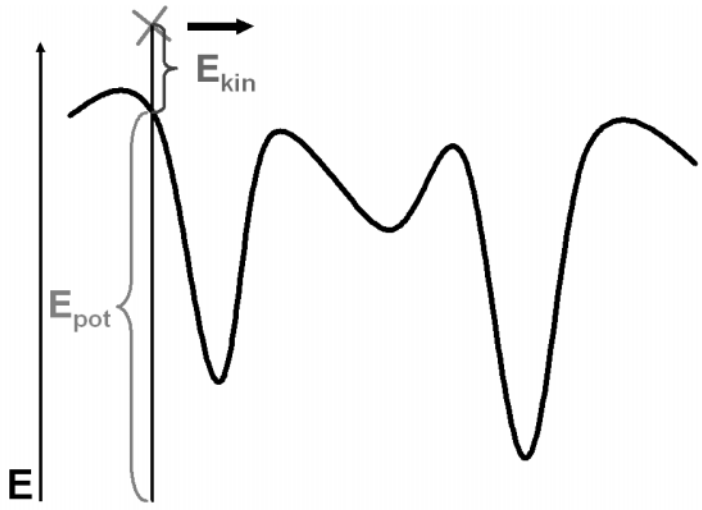


Figure 2.8: During an MD simulation one floats over the multidimensional energy-hyperface thereby exploring a greater fraction of the energy landscape. Due to the kinetic energy, peaks of potential energy can be overcome. During a sampling period thus several low energy conformations will be explored.

An input file for an MD simulation in GROMACS together with a short description of parameters that have to be adjusted is given in the appendix section 10.2.

MD simulations are a valuable tool for obtaining energetically favourable protein models and allow for structural adaptations within the protein model generated. This technique has however as well a number of limitations that must be considered. Some limitations arise simply from the neglection of electrons; in this regard, the disruption, generation or isomerisation of bonds cannot be studied by means of MD techniques, neither can charge-transfer complexes be reproduced. Theoretically, also π-π-interactions would be beyond the scope of MD techniques. This type of interaction can however be modelled by an atomic point charge model, where a C-H bond dipole is used to reproduce aromatic-aromatic interactions. [72] Yet, polarisation effects of π-systems by heteroatoms are beyond the scope of simple dipole models. Apart from these intrinsic limitations, restrictions result from approximations that must be introduced in order to obtain feasible computations: the system size is limited to small dimensions, cutoffs for van der Waals and electrostatic interactions must be accepted and the simulation period is limited to a nanosecond range. Thus, the amount of the energy landscape sampled during an MD simulation might only represent a small fraction of the conformational space available for a protein under study, and no information is available which part of the conformational space of a protein is sampled per time unit and if the entire conformational space will be actually accessible during an MD simulation or if the simulation is restricted to an isolated valley of potential energy.

2.1.7 Model Evaluation

During the generation of homology models, errors will inevitably occur that reduce the applicability of the model for later purposes. Commonly encountered errors are listed below:

- Errors in the target-template alignment in structurally conserved regions:
  When generating a family A GPCR homology model, this point is usually not a source of model incorrectness as several highly conserved residues and motifs help as pinpoints in order to obtain a correct alignment.

- Structural deviations of the template protein despite correct alignment:
  In order to cope with this problem, unconstrained MD simulation of the generated protein model can help to allow for the required adjustments.

- Errors in loop regions due to lack of structural information:
  The generation of reasonable conformations for divergent loop regions is one of the most difficult tasks to be accomplished. Both commonly applied methods, *de novo* design strategies and loop search by distance matrix approaches, are only feasible for short peptide stretches. In case of longer loop regions (> 10 amino acids) truncation of the corresponding region is often the only practicable strategy although a lot of effort has been put into the development of efficient threading algorithms. In case a truncation is carried out, the effects on the binding site geometry will strongly depend on the localisation of the neglected loop.

- Wrong side chain placement:
  Advanced algorithms for prediction of amino acid side chains exist that incorporate also backbone information in order to find a suitable amino acid side chain placement. Structural deviations from the template protein will however inevitably result in inconsistent amino acid side chain placements. Additionally, as will be described later, structural water molecules can uphold otherwise suboptimal side chain rotamers.

- Incorrect protonation states:
  Programs for calculating protonation states exist; however, an interdependence between the pK_a values of amino acids and the adopted side chain rotamers as well as the presence of putative structural water molecules exists.

- Neglection of structurally important water molecules:
  Favourable placements for water molecules within a structure can be found by analysing interaction fields for H_2O probes using programs such as `GRID`. [73]

The method is however too inaccurate — especially when applied to a homology model — and additionally depends on side chain placements and protonation states.

- Low stereochemical quality:
  Stereochemical parameters (bond lengths, bond angles, peptide bond and side chain ring planarities, chirality, main chain and side chain torsion angles, and clashes between non-bonded pairs of atoms) can be checked by programs such as `PROCHECK`. [74] Although a good stereochemical quality does not guarantee model correctness, it is a prerequisite for a subsequent application of the model.

2.2 Conformational Analysis of Ligand Molecules

In order to bind to a receptor with high affinity, a ligand must electrostatically and sterically match the binding pocket. The steric match will thereby depend primarily on the ligand conformation. Within a binding pocket, the ligand will not be necessarily present in its lowest energy conformation, as the gain in interaction energy with the receptor can compensate for a conformation with higher energy. Still, it can be expected that for a high-affinity ligand, the bioactive conformation is at least energetically favourable, as otherwise the conformational energy cost would reduce binding affinity. The relation between a high energetic binding conformation and the loss of free binding energy ΔG is given by equation 2.4 [75]:

$$\Delta G = -2.303 RT \log K_i \tag{2.4}$$

Under physiological conditions (T = 310 K), the free energy (in kJmol^{-1}) and the binding affinity are related by

$$\Delta G = -5.85 \log K_i \tag{2.5}$$

Thus, if a compound binds in a conformation that deviates 5.85 kJmol^{-1} from the global minimum structure, its affinity will be decreased by one magnitude. High-affinity compounds can thus be expected to bind in an energetically favourable conformation. In order to find the putative bioactive conformation, conformational analysis has to be carried out. Commonly used methods for this purpose are listed below and described in depth in [70, 71]:

- Systematic search:
  Each bond is rotated incrementally and the resulting structures are minimised.

Systematic search algorithms have the advantage that they sample the conformational space very well, yet, in case of a high number of rotable bonds this method may be computationally impracticable.

- Random search:
  In a random search, one can move from one region of the energy-surface to a completely unconnected region in a single step. A commonly applied method is the Metropolis Monte Carlo scheme that starts with a minimised conformation A of a molecule. Then a random move on the energy-landscape is carried out (e.g. torsion angles are rotated by a random amount) and the structure is minimised. The potential energy of the output structure B is evaluated. If $E_{pot}(B) < E_{pot}(A)$, the new conformation is accepted. If $E_{pot}(B) > E_{pot}(A)$, the move may still be accepted depending on the transition probability that in turn depends on the temperature. Monte Carlo methods efficiently sample the conformational space, however, there is no guarantee — as with all random search tools — that the entire energetic-landscape will be sampled.

- Simulated annealing or MD simulations:
  As previously described, MD simulations generate an ensemble of structures that does however not only represent minimum structures. In a simulated annealing protocol, the system temperature is periodically increased resulting in a significant rise of kinetic energy which makes it easier to overcome barriers of potential energy. Subsequently, the system is cooled down, thereby trapping the molecule in an energetically favourable conformation. MD simulation techniques for sampling the conformational space are quite time-consuming. Again, there is no guarantee of sampling the entire potential energy-surface.

2.3 Ligand Superposition Techniques

Ligand superposition techniques are frequently required in molecular modelling in order to accomplish for example one of the following tasks:

- In absence of structural information on a target protein, ligand superposition on an endogenous ligand in its bioactive conformation can help to deduce a pharmacophore model or generate a pseudoreceptor model.

- A set of superimposed structures is a prerequisite for a 3D QSAR analysis.

- Superposition methods can help to understand which structural features in a set of ligands correspond to one another.

Several algorithms for ligand superposition exist, including techniques that superimpose molecules by mapping and comparing shape and field properties of the structures (e.g. `Catalyst`) or by incrementally building up a test molecule upon a rigid reference molecule (e.g. `FLEXS`). When applying the `FLEXS` algorithm, [76] the flexible compound is first partitioned into fragments. In a first step, an anchor fragment is selected and placed onto the reference compound in a way that similar interactions can be established by both compounds. Then, the remaining fragments of the flexible molecule are incrementally added. Flexibility is considered by allowing each fragment to adopt a discrete set of energetically favourable conformations. Each superposition is then assigned a score that will be higher the better the match between the reference and target molecule in terms of highly directional (hydrogen bonds, hydrogen bond reinforced salt bridges) and less directional (lipophilic interactions) features. If, in the adopted conformation, both molecules are for example able to form a hydrogen bond to an identical countergroup site and the same lipophilic interactions can be established, a high score will be obtained. Additionally, a term considering the van der Waals volume overlap and a function for penalising deviations from ideal bond lengths and angles are included.

2.4 Analysis of Interaction Fields

One of the most commonly used programs in order to locate favourable interaction fields is the program `GRID`. [73] For this purpose, the molecule of interest is placed into a regularly shaped grid. A probe group is then placed on each grid vertex and the interaction energy with the molecule is evaluated at each point applying equation 2.6. The resulting energy on each `GRID` point (E_{GRID}) is composed of a term describing the van der Waals interaction (E_{LJ}), electrostatic interaction (E_{el}) and potential hydrogen bond interactions (E_{hb}):

$$E_{GRID} = \sum E_{LJ} + \sum E_{el} + \sum E_{hb} \tag{2.6}$$

Several probes have been developed that allow scanning a binding site or small ligand molecule for physicochemical properties (e.g. hydrophobic interactions, interactions with a protonated moiety, ...). The grid cage with energy values on each vertex can then be translated into interaction fields that highlights sites where a positive interaction with the grid probe can be expected. When scanning a protein binding site, this way, information can be obtained as to which chemical moieties a potential ligand molecule should obtain. Alternatively, scanning a ligand molecule, one can draw conclusions of how a ligand molecule might be oriented in the binding site.

2.5 Ligand Docking

In molecular docking, one attempts to generate and evaluate plausible structures of intermolecular complexes. [70] A vast number of docking algorithms has been developed in recent years and a detailed description of all algorithms is beyond the scope of this work (see [70, 71] for a more thourough discussion). Common to most docking procedures is that only ligand flexibility is considered while the protein is considered rigid; a consequence that arises from the vast conformational search problem. Three commonly encountered strategies of docking algorithms are mentioned below:

- Docking programs such as FLEXX incrementally build up a ligand molecule in a rigid protein binding pocket. [77] In a first step, a base fragment is identified and docked into the binding site. The rest of the ligand is then incrementally added to this base fragment placements. Ligand flexibility is considered by allowing each fragment to adopt a discrete set of energetically favourable conformations.

- Programs such as AUTODOCK use a Monte Carlo simulated annealing technique. [78] Starting with a ligand conformation in the protein binding pocket, at each iteration the ligand conformation and/or position is randomly changed (e.g. by varying a set of torsion angles or translating the ligand in the binding pocket). The energy of the new complex is then evaluated and the step is then accepted or rejected.

- The program GOLD (Genetic Optimisation for Ligand Docking) on the other hand uses a genetic algorithm. [79] The ligand conformation and orientation in the binding site is encoded by a chromosome. Genetic operations (crossover, mutation, migration) are then carried out resulting in random changes of conformation and position and the new complex is scored. The score thereby describes the fitness of the ligand and determines if the molecule is allowed to pass its genetic information to the offspring. The GOLD-score fitness function consists of a term for protein-ligand hydrogen bonding, protein-ligand van der Waals energy, ligand internal van der Waals energy and ligand torsional strain energy.

2.6 Pharmacophore Models for Screening Structural Databases

A pharmacophore can be defined as a 3D spatial arrangement of chemical and physical properties of a ligand molecule. The concept of pharmacophores shall be described at the

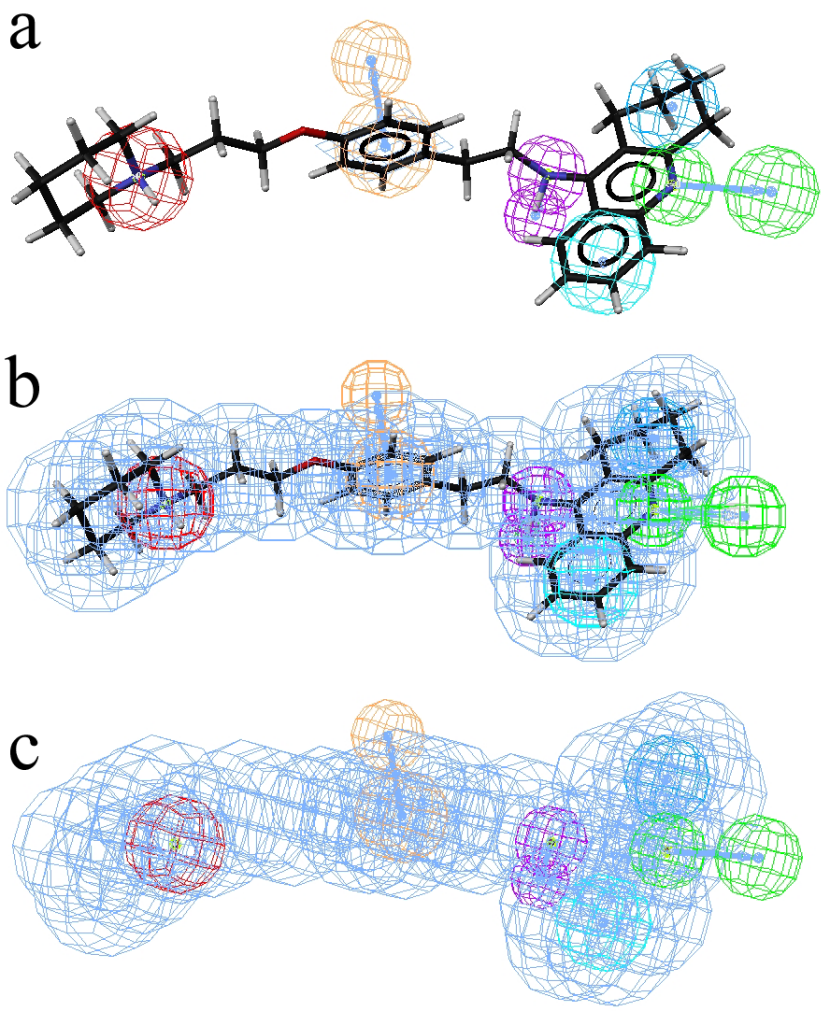


Figure 2.9: Concept of pharmacophores explained on hH_3R ligand FUB833. **(a)** Pharmacophoric features observed in molecule FUB833. **(b)** The molecule's shape can serve as an additional constraint in pharmacophore searches. **(c)** Possible pharmacophore representation of molecule FUB833.

example of ligand FUB833 (see figure 2.9). Table 2.2 lists the pharmacophoric features observed in molecule FUB833 and the corresponding colour code.

Thus, the protonated nitrogen atom of the piperidyl moiety can be translated into a center of a sphere with coordinates corresponding to the location of the nitrogen atom and a radius defining a volume around this atom that reflects the tolerances. If a molecule is compared to this pharmacophore model and its protonated nitrogen atom will lie within the sphere, this pharmacophoric feature will be said to be matched. The bigger the sphere, the easier it will be for a ligand conformation to match the pharmacophoric features. Similarly, a hydrophobic or a hydrophobic aliphatic moiety can be defined by a center of sphere and radius. Hydrogen bond acceptors and donors are represented by vectors in order to account for the directionality of h-bonds while aromatic rings can be either defined by spheres or the combination of center, plane and vector. If defined this way, also the

Table 2.2: Pharmacophoric features observed in ligand FUB833.

feature	colour	representation
positive charge	red	sphere
h-bond donor	magenta	sphere - vector - sphere
h-bond acceptor	green	sphere - vector - sphere
hydrophobic aliphatic	blue	sphere
aromatic ring	orange	plane, center of plane, vector, tolerance
hydrophobic	light blue	sphere

orientation of the aromatic plane in respect to the rest of the molecule is considered. Finally, also the shape of the molecule can be incorporated into a pharmacophore query by translating the van der Waals volume into an additional feature. Especially if the ligand is known to fill the binding pocket well, this way the available volume can be taken into account. The abstract definition of a molecule in form of a pharmacophore as defined in figure 2.9 facilitates comparison with other molecules.

In the previous example most features of FUB833 were considered for the generation of the pharmacophore model resulting in an almost unique fingerprint of FUB833. Depending on the number of features included in the model and the tolerances defined, it will be more or less difficult for other molecules to match the pharmacophore model. Matching a pharmacophore additionally strongly depends on the conformation that is adopted by the molecule that is compared to the pharmacophore model. Even a different conformation of FUB833 might not match the pharmacophore model defined above. Thus, in order to increase the likelihood of a specific molecule fitting a pharmacophore model, each molecule of interest is associated with a conformational model. When searching for similarities with the pharmacophore model, all conformations of a test molecule (below a reasonable energy cutoff) are tested on the pharmacophore before the best fit is evaluated.

The difficulty in defining a “useful” pharmacophore model lies in the restriction to only essential pharmacophoric features observed in most validated binders. A pharmacophore model capable of identifying validated binders is then a useful tool for screening large databases in order to retrieve new structures based on the similarity of pharmacophoric features. In order to screen commercial structure databases with a pharmacophore model, firstly, a set of conformations must be generated for all compounds deposited in the structure databases.

Part I

Molecular Modelling Studies of Bovine Rhodopsin

Chapter 3

Scope

G-protein coupled receptors (GPCRs) are a widespread family of membrane proteins with the common function of transducing extracellular signals into intracellular responses, prompting interest in these targets for several therapeutic applications. Unfortunately, so far only the crystal structure of bovine rhodopsin has been resolved. [4] Structural information on the binding crevice of pharmaceutically interesting GPCR targets would be of tremendous advantage for an efficient drug design with respect to an application in database screening or *de-novo* design strategies. Due to the conserved fold (7-helix bundle) and several highly conserved amino acids within the sequence of GPCRs, [80] the approach of homology modelling becomes applicable. Subsequent to an initial sequence structure alignment, energetically favourable conformations of the (ligand-complexed) model can be derived through molecular dynamics simulations.

In order to find optimal parameters for the simulation of GPRC homology models, calculations with the reference structure (bovine rhodopsin) will be carried out testing the influence of parameters such as the solvent environment used (CCl_4/H_2O, DPPC/H_2O), a potential *N*-terminal truncation, different protonation states for selected buried residues and a potential consideration of internal water molecules as resolved in [81]. In order to derive possible constraints for the simulation of GPCR homology models, the interhelical hydrogen bond pattern that seems responsible for upholding the inactive receptor conformation will be analysed. Finally, also conformational adaptations after the introduction of *all-trans*-retinal will be investigated by means of MD simulation, as an understanding of the initial amino acid rearrangements could potentially help in the design of antagonists or inverse agonists for other rhodopsin-like receptors that should impede these initial movements.

Chapter 4

Results

4.1 System Setups for the Simulation of Bovine Rhodopsin

4.1.1 Generation of Model Structures of Bovine Rhodopsin

The coordinates of the crystal structure 1HZX [82] of bovine rhodopsin with a resolution of 2.8 Å were used as a starting point for all herein described simulations. No *N*-terminal acetylation, glycosylations or a *C*-terminal palmitoyl moiety were considered as the presence of these structural features is *a priori* not known in other GPCR members. Thus, in order to preserve the transferability of results of these simulations to other GPCR homology models, these moieties were not included in the simulations. Missing residues (see figure 4.1) in the A-chain (236–240, 331–333) localised in the 3^{rd} intracellular loop and the *C*-terminal end respectively were added using the Loop Search and Splice Repair utilities of the `Insight 2000 Homology` module [83] where residues were added following a distance matrix approach. All further calculations were carried out with the `GROMACS` simulation package using the `ffG43a1` force field. In order to generate the protonated Schiff-base by which retinal is linked with lysine 296/7.43[1], the predefined retinal parameters (RTOL) in the `ffG43a1.rtp` file were adjusted and a new residue type (LYX) was defined in analogy to the predefined lysine (LYS) residue so that a covalent linkage could be set. Several models were generated varying in their sequence length, the protonation state of residues D83/2.50, E122/3.37 and E181/4.70 (E_2-loop) and a potential consideration of internal water molecules as resolved in the crystal structure 1L9H. [81] Table 4.1 summarises the models of bovine rhodopsin herein described.

[1] numbering scheme corresponding to Baldwin *et al.* [80]: the most conserved residue in each transmembrane segment is assigned position 50 (see figure 4.1). The first number refers to the helical segment.

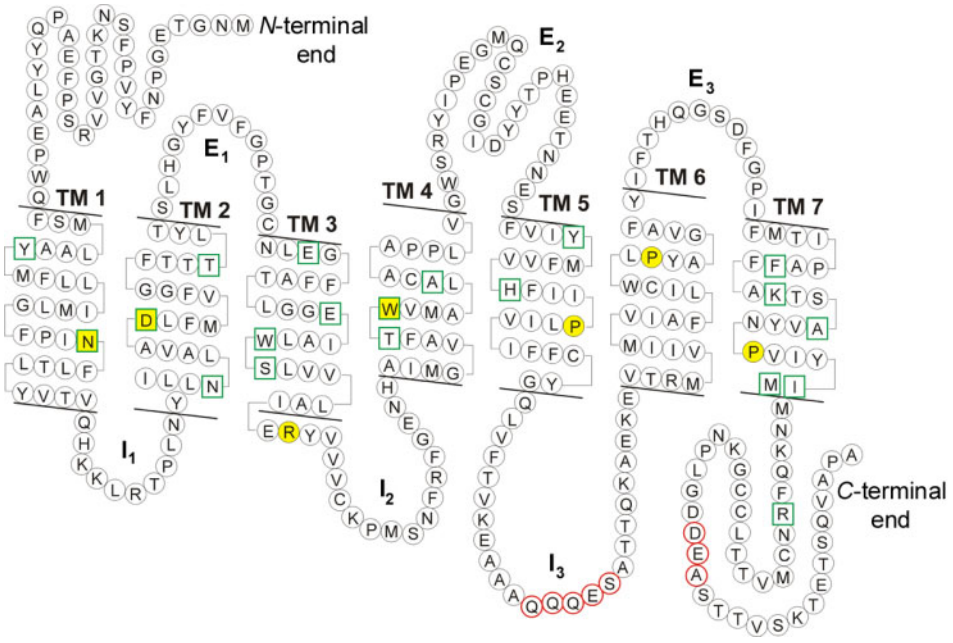


Figure 4.1: Snakeplot of bovine rhodopsin. The most conserved residue within each transmembrane region (TM) is highlighted in yellow (corresponding to position 50 in the Baldwin numbering scheme [80]). Residues in the 3^{rd} intracellular loop (I_1–I_3) and *C*-terminal end that could not be resolved in the crystal structure 1HZX are indicated by red circles. Residues that were found to be consistently involved in interhelical hydrogen bonds during MD simulations are written in green squares. The length of transmembrane segments (as described in Swissprot Entry PO2699) is indicated by parallel lines.

Table 4.1: Survey over models of bovine rhodopsin herein described.

abbreviation	properties
RT1	truncated model of bovine rhodopsin comprising the stretch 33-321 with D83/2.50 and E122/3.37 in the protonated neutral state. Internal water molecules were considered.
RT2	entire model of bovine rhodopsin with D83/2.50 and E122/3.37 in the protonated neutral state. Internal water molecules were not considered.
RT3	entire model of bovine rhodopsin with D83/2.50, E122/3.37 and E181/4.70 considered in their protonated neutral state. 11 internal water molecules were considered.
RT4	truncated model of bovine rhodopsin comprising the stretch 33-321 with D83/2.50, E122/3.37 and E181/4.70 considered in their deprotonated form. No internal water molecules were considered.

4.1.2 Bovine Rhodopsin in a CCl_4/Water Environment

Two CCl_4/solvent boxes with dimensions 8.56 $\times$ 6.45 $\times$ 8.09 nm and 8.56 $\times$ 6.45 $\times$ 9.01 nm were generated to simulate the truncated and entire rhodopsin models, respectively. Therefore, first a CCl_4-box of dimensions 8.56 $\times$ 6.45 $\times$ 3 nm was generated and filled with CCl_4 molecules up to the correct density of 1.596 gcm^{-3}. [84] Parameters for the CCl_4 molecule were predefined in the `ffG43a1.rtp` file. However, in order to preserve the tetrahedral structure of the CCl_4 molecules, an additional "virtual" bond was set between chlorine atoms 4 and 5 (see `ffG43a1.rtp` file for atom numbering scheme). Then water molecules were added to fill the box. During this procedure water molecules were also inserted into the apolar CCl_4 layer. After 5000 steps of steepest descent minimisation, an MD simulation was thus carried out for 1 ns using the parameters listed in table 4.2 in order to allow the misplaced water molecules to move out of this apolar environment. No pressure coupling was applied during the simulation. Instead, the density was adjusted to 1.596 gcm^{-3} and the volume kept fixed. This procedure gave rise to an extremely high pressure value of approximately 950 bar, as can be seen in figure 4.2 together with the course of potential energy during the equilibration period.

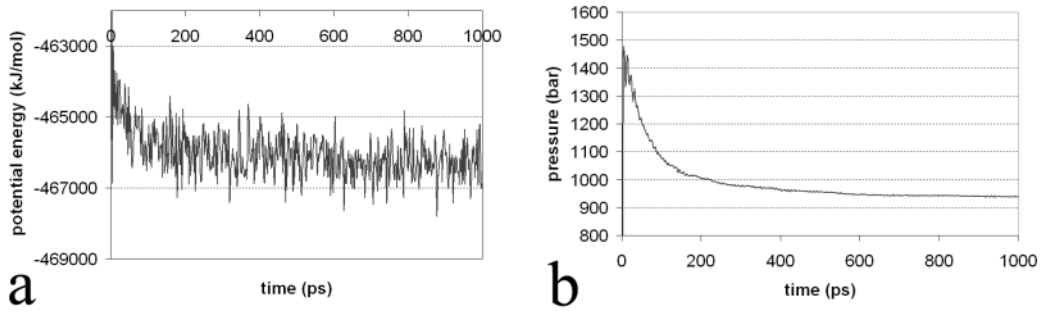


Figure 4.2: **(a)** Course of potential energy during the equilibration of a CCl_4/water box of dimension 8.56 $\times$ 6.45 $\times$ 9.01 nm with no pressure coupling applied. **(b)** Corresponding curve of system pressure.

Alternatively, the same system was equilibrated allowing for pressure coupling. In this case, the box dimensions increased to 108.2% of the original box size while the pressure value diminished to approximately 25 bar after a simulation period of 500 ps. The related decrease in density was thereby mainly due to a larger required volume of CCl_4 molecules. Simulation of a CCl_4-box with pressure coupling resulted in a 10.2% deviation of the correct density, whereas simulation of a water box gave rise to a minor density deviation of only 3.4%. This is most likely due to a better parameterization of the H_2O molecules compared to the CCl_4 molecules. As can be seen in figure 4.3, the apolar CCl_4 layer is slightly thicker after an equilibration with pressure coupling and at the same time the surface is more corrugated. Thus, even though artificially high values for pressure were introduced

in simulations without pressure coupling, this strategy was given preference as deviations of the density are expected to influence the simulation result in a more pronounced manner than deviations of pressure.

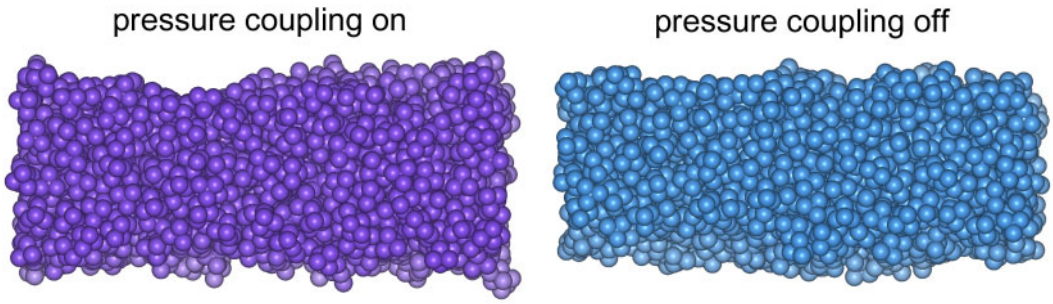


Figure 4.3: Comparison of the CCl_4 layer after equilibration of CCl_4/water boxes with the option of pressure coupling turned on (left) or off (right).

Coordinates of the retinal/opsin complex and the equilibrated box were then merged with the command `genbox` that automatically removed solvent molecules whose van der Waals radius would overlap with the protein structure. As an application of the PME (Particle Mesh Ewald) method [85] for calculation of Coulomb interactions requires the system to be neutral, sodium ions were added to yield an uncharged system. Additionally, sodium and chlorine ions were added to yield a final concentration of 156 mVal l^{-1}, as an explicit consideration of ions has been shown to have beneficial effects on secondary structure stability. [86] In all calculations the retinal/opsin complex was initially tethered with a force of 1000 kJmol^{-1}nm^{-2} in xyz directions. This tether was reduced stepwise (each 100 ps) from 1000 to 500 to 200 to 100 kJmol^{-1}nm^{-2} (resulting in characteristic plateaus of the RMSD (root mean square deviation) value at the beginning of each simulation). This proceeding led to superior results in terms of lower values of RMSD compared to a procedure where an unconstrained MD simulation was immediately carried out following the initial minimisation. Subsequently, an unconstrained dynamics simulation was carried out. No pressure coupling was applied in simulations using the CCl_4/solvent boxes; instead, the right density was adjusted and the volume kept fixed. This strategy, as described above, resulted in artificially high pressure values, seemed however justified due to smaller deviations of density. As deviations of the density are expected to influence the simulation result in a more pronounced manner than deviations of pressure, this strategy was given preference. Table 4.2 gives an overview of the parameters used for the simulations.

Figure 4.4 shows the resulting placement of bovine rhodopsin in the CCl_4/water box that was carried out in such way, that transmembrane regions came to lie within the membrane mimic and the amphiphilic helix 8 run parallel to the CCl_4 layer.

Table 4.2: GROMACS parameters for the simulation setup of a model of bovine rhodopsin in a CCl_4/water environment.

dt = 2 fs / 4 fs
Neighbour searching parameters: nstlist: 5, nstype: grid, pbc: xyz, rlist: 0.9 nm
Parameters for calculation of electrostatics and VdW interactions: Coulomb-type: PME, VdW-type: cutoff (r: 0.9 nm)
T-coupling: Berendsen (310K, τ: 0.1), p-coupling: no
constraints: h-bonds (LINCS)

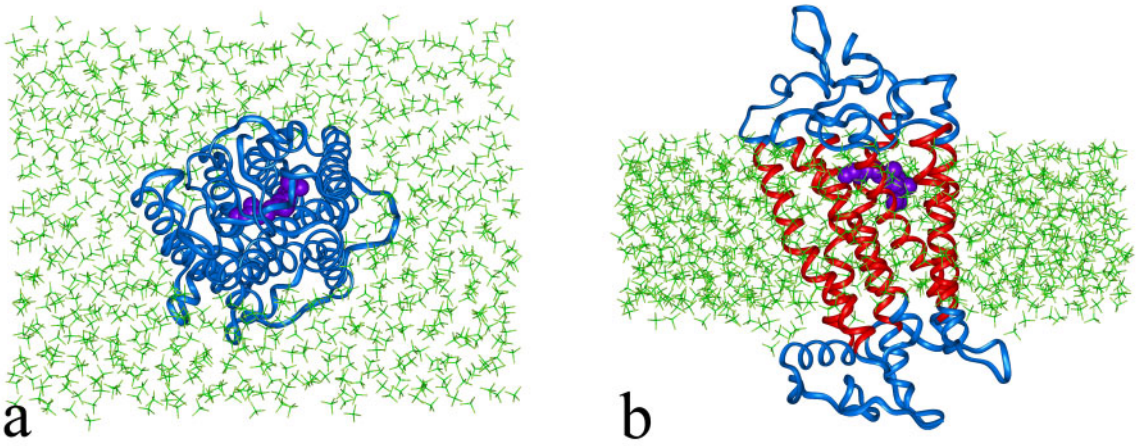


Figure 4.4: Model of bovine rhodopsin in a CCl_4/water box. **(a)** Viewed from the extracellular space. **(b)** Side view: transmembrane regions are coloured red, *11-cis*-retinal is shown in its van der Waals representation. Water molecules and ions are neglected for reasons of clarity. Helix 8 runs parallel to the membrane plane.

4.1.3 Bovine Rhodopsin in a DPPC/Water Environment

Membrane models are characterised by the set of force field parameters adjusted in order to reproduce realistic physical properties. In order to preserve the comparability of MD simulations, the ffgmx-DPPC parameters used by Tieleman *et al.* [87] were adapted following the procedure described in [88, 89] resulting in lipid parameters optimised for application of the ffG43a1 force field that had been previously used for simulations in the CCl_4/water environment. Starting from a pre-equilibrated box of 128 DPPC and 3655 water molecules, [90] the box was enlarged in the z-direction to yield the dimensions 6.56 $\times$ 6.59 $\times$ 9.21 nm resulting in a system of 128 DPPC and 7637 water molecules. Again, water molecules were placed inside the membrane layer during this procedure. Thus, the box was equilibrated for 500 ps without and 1000 ps with pressure coupling until all water molecules had moved outside the apolar core region and reasonable pressure values were obtained.

For the insertion of the retinal/opsin complex, a modified version of the `mdrun` program [91] was used to introduce a cylindrical hole of radius 2 nm (force constant for repulsive force: 50 kJmol^{-1}nm^{-1}; hx/hy/hz = 3.3/3.3/4.6). The approach of simply merging the DPPC box with the protein structure was not possible in this case, as too many DPPC molecules overlapped with the protein structure and would thus have been removed. Instead, DPPC molecules were restrained to movements within the membrane-plane and pulled out of the cylindrical hole defined. [91] Figure 4.5 shows several snapshots during the molecular dynamics simulation that resulted in the introduction of the hole in the DPPC/water box.

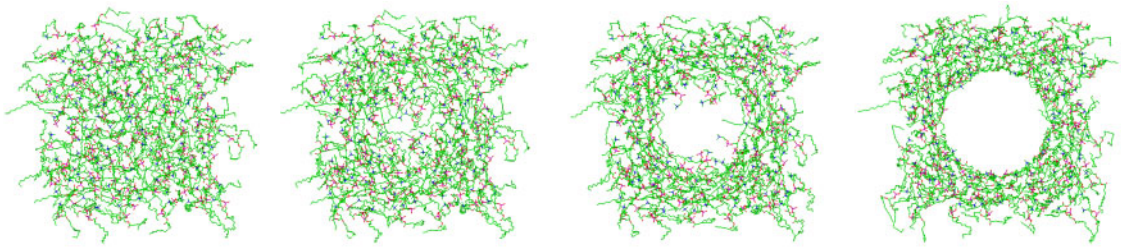


Figure 4.5: Snapshots during the introduction of a hole into the lipid bilayer. Only the DPPC layer is shown after 0, 0.8, 1.6 and 4 ns of simulation period.

The lipid bilayer with the introduced hole and the model of bovine rhodopsin were then merged and the resulting system equilibrated for 1 ns in order to allow for an adjustment of the box size to yield correct values for density and pressure. In analogy to [91], the introduction of the hole as well as the initial MD simulation to adjust the box size were carried out using the `ffgmx` force field. Then the system was transferred to the `ffG43a1` force field and further equilibrated. During this equilibration period the bovine rhodopsin model was highly restrained in order to avoid structural modifications during the setup of the simulation system. Sodium ions were again added to obtain a neutral system and, additionally, sodium and chlorine ions were added to yield a final concentration of 156 mVal l^{-1}. In all calculations the same equilibration procedure as described in section 4.1.2 was used including the stepwise removal of tethers (see table 4.3 for parameters).

Table 4.3: `GROMACS` parameters for the simulation setup of a model of bovine rhodopsin in a DPPC/water environment.

dt = 2 fs / 4 fs
Neighbour searching parameters: nstlist : 5, nstype: grid, pbc: xyz, rlist: 0.9 nm
Parameters for calculation of electrostatics and VdW interactions: Coulomb-type: PME, VdW-type: cutoff (r: 0.9 nm)
T-coupling: Berendsen (310K, τ: 0.1), p-coupling: Berendsen (isotropic, 1bar, τ: 0.5)
constraints: h-bonds (LINCS)

The placement of the protein within the lipid bilayer is depicted in figure 4.6. The amphiphilic helix 8 runs parallel to the membrane layer and lies at the interface of the polar DPPC-headgroups and the apolar core of the bilayer.

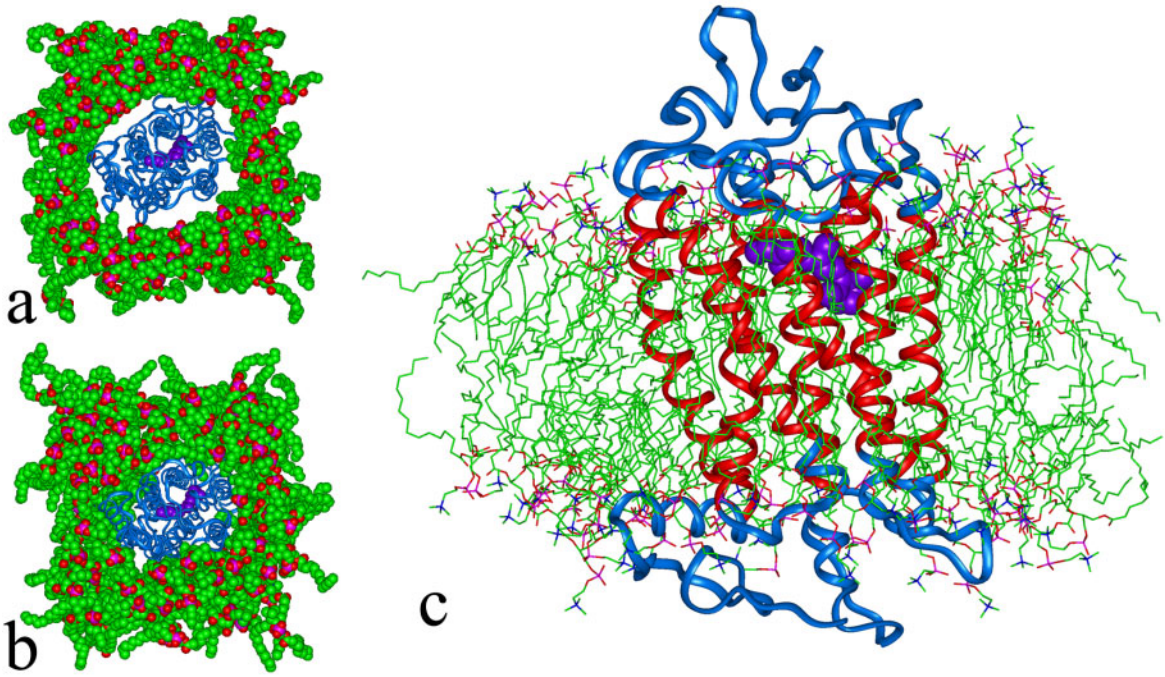


Figure 4.6: **(a)** Model of bovine rhodopsin inserted into the DPPC bilayer after merging the protein structure and the box and **(b)** after 1 ns of equilibration period. **(c)** Side view of the placement of bovine rhodopsin in the lipid bilayer. Transmembrane helices are coloured red, the chromophore is depicted in its van der Waals representation.

4.1.4 Calculation of pK_a-shifts of Titrable Amino Acid Residues

pK_a-shift calculations were carried out for the complete bovine rhodopsin models in absence and presence of internal solvent molecules resolved by Okada *et al.* [81] using the program `UHBD`. [69] Additional residue types for lysine 296/7.43 and retinal were defined and added to the `pkaS.dat` database. All histidines were assigned type `HisA`. Calculations were carried out at 310 K and with an ionic strength of 150 mM, setting the dielectric constants to 80 for solvent and 20 for the protein interior. [68] All cysteines (except those involved in the disulfide linkage) were included in the calculation. Four grids (2.5, 65/65/65; 1.2, 40/40/40; 0.75, 40/40/40; 0.25, 40/40/40) were used and the maximum number of iterations was set to 300.

4.2 Truncated versus Entire Protein Models

In order to test the influence of an explicit presence of the *N*-terminal ending on the overall protein integrity, a model of bovine rhodopsin lacking the initial stretch of 32 residues was simulated. In bovine rhodopsin, helix 1 interacts via two hydrogen bonds with helix 7, the highly conserved interaction of N55/1.50 – A299/7.46 and the hydrogen bond formed between Y43/1.38 and the backbone oxygen of F293/7.40 (see figure 4.7). Thus it seems that the *N*-terminal end, folding over the entire protein, is not necessary to anchor helix 1 to the remaining helix bundle. Yet, *N*-terminal truncation introduced undesirable side-effects as residues originally buried were brought to the protein surface facing the aqueous environment.

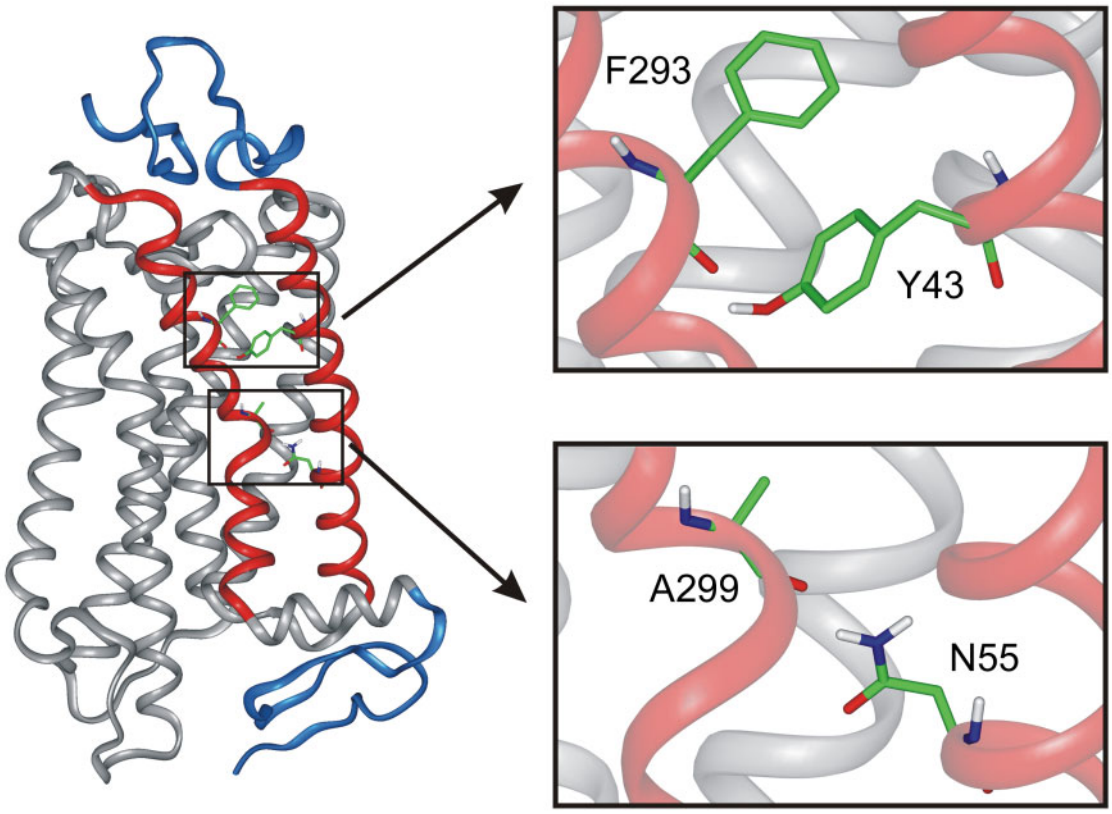


Figure 4.7: Crystal structure 1U19 of bovine rhodopsin: The *N*- and *C*-terminal ends are coloured blue, helices 1 and 7 red. The two sites where helix 1 is anchored to helix 7 by hydrogen bond interactions are magnified on the right.

During the simulation of the truncated rhodopsin model RT1 in a DPPC/water environment deviations of the second extracellular loop and of the position of helix 1 were observed. While distortions of the E_2-loop were due to the direct contact with the aqueous environment, artifacts in helix 1 were caused by the intrusion of extracellular water molecules and disruption of the interaction Y43/1.38 – F293/7.40 after 3500 ps of unconstrained simulation (see figure 4.8b). As a consequence, helix 1 separated with

its extracellular part from helix 7, leaving however the interaction N55/1.50 – A299/7.46 unaffected. At height of L1.38 the shift of helix 1 amounted to 1.83 Å (measured from backbone L1.38:C to L1.38:C').

When the simulation was repeated in a CCl_4/water environment the shift increased to 3.78 Å (see figure 4.8c). The interaction N55/1.50 – A299/7.46 was again unaffected, whereas the interaction Y43/1.38 – F293/7.40 was lost after 2700 ps. In this simulation setup the disruption of this hydrogen bond was however not caused by the intrusion of water molecules but rather due to the increased flexibility of the *N*-terminal end of helix 1 starting at residue G1.46.

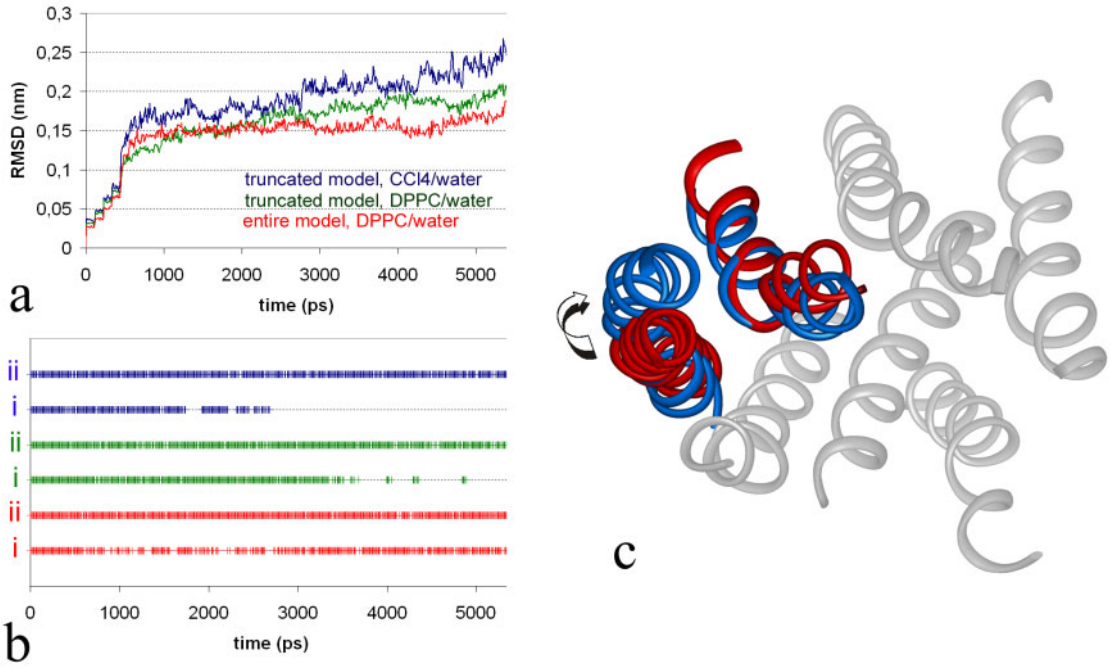


Figure 4.8: **(a)** Course of RMSD observed during the simulation of model RT1 in a DPPC/water and CCl_4/water environment compared to the RMSD plot obtained for the simulation of the entire model RT2 in a DPPC/water environment. **(b)** Plot of hydrogen bonds between residues Y43 – F293 (i) and N55 – A299 (ii) during the course of the simulations (colour code as in RMSD plot). **(c)** Transmembrane region of bovine rhodopsin viewed from the extracellular side. The reference crystal structure is coloured red (helices 1 and 7) and grey (helices 2-6). Helices 1 and 7 of model RT1 after 5 ns simulation in a CCl_4/water environment are depicted in blue.

4.3 Consideration of Internal Water Molecules

In the crystal structure 1L9H of bovine rhodopsin (resolution 2.6 Å) 11 internal water molecules have been described by Okada *et al.*, [81] of which seven are found within the transmembrane region. In order to analyse the dynamic behaviour of these water

molecules, a 25 ns MD simulation of an entire bovine rhodopsin model in a DPPC/water environment with D83/2.50, E122/3.37 and E181/4.70 considered in their protonated form (model RT3) was set up using a time-step of 2 fs. The 11 water molecules were positioned corresponding to the coordinates resolved in 1L9H.

In the herein described simulations the explicit consideration or neglection of certain water molecules did have pronounced effects on the hydrogen bond pattern. The time at which changes in the interhelical hydrogen bond pattern occurred can be seen in a plot of the RMSD as a function of time. In the RMSD plot of figure 4.9a three zones can be roughly discriminated. An initial short lasting plateau at 0.15 nm was formed during the first 2.5 ns, then the RMSD remained stable at 0.17 nm within the next 10 ns and eventually rose to a value of 0.2 nm, which could be observed until the end of this simulation. When the RMSD curve is related to the interhelical hydrogen pattern (see figure 4.9a), a potential reason for the increase of the value of RMSD from 0.15 to 0.17 nm could be the loss of the interaction between Q64/1.59 and T320/7.67. More obvious is the coincidence of the increase of RMSD after 12 ns with the switch from hydrogen bond N55/1.50 – A299/7.46 to the linkage between D83/2.50 – A299/7.46.

The consideration of internal water molecules did thus not necessarily improve the quality of the simulation. Internal water molecules were able to disrupt conserved hydrogen bonds due to their high flexibility in regions rich in polar residues, where changes in the hydrogen bond pattern could occur easily. Rather than stabilising the existing hydrogen bond pattern, internal water molecules tended to catalyse the switch to alternatively possible hydrogen bonds. In figure 4.10 the disruption of the highly conserved interaction between N55/1.50 and A299/7.46 is shown. Although the interaction was not observed in all frames during the course of the simulation, it persisted for about 12 ns of simulation period. Then a temporary loss of the hydrogen bond coincided with the insertion of two water molecules. In order to accommodate the solvent molecules, the distance between these two residues increased, which gave rise to the increase in RMSD observable after 12 ns. Once the interaction was lost, an alternative interaction between the backbone oxygen of A299/7.46 and D83/2.50 could be established that persisted up to the end of the simulation (see figure 4.11). The switch between alternative interactions is, of course, quite likely the intended natural function, however, the original hydrogen bond interaction pattern as observed in the crystal structure was preserved better when the disruption of a hydrogen bond could not be compensated by a temporary interaction with a flexible water molecule before a new interaction was established.

This holds however true only for water molecules located in protein-regions that were rich in polar residues. Water molecule 2021 at site 2b (numbering scheme referring to [81]) formed a stable interaction with E113/3.28 maintaining the exact side chain conformation

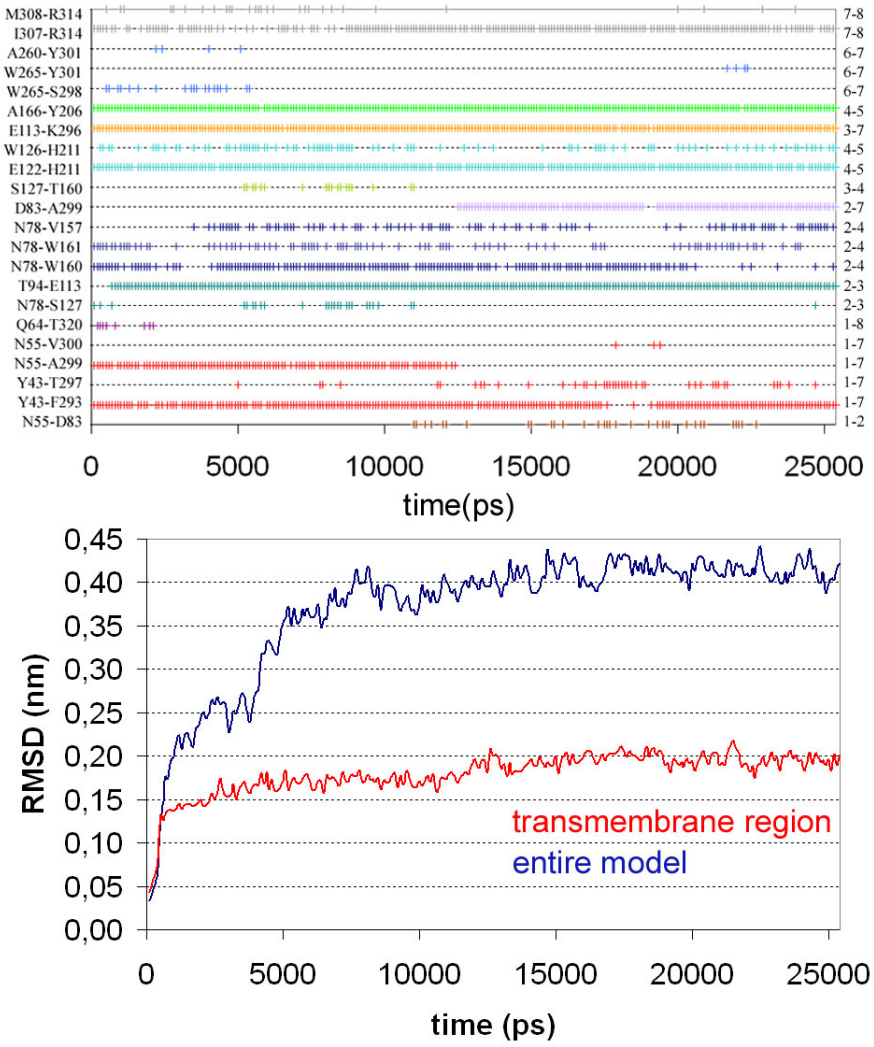


Figure 4.9: **(a)** Interhelical hydrogen bond pattern during the simulation of model RT3 in a DPPC/water environment. Helices involved in the contacts are written on the right side. **(b)** Corresponding plot of RMSD for the entire model (blue curve) and the transmembrane region (red curve).

for this residue as observed in the crystal structure during the entire simulation. Molecule 964 (site 3) likewise did not substantially deviate from its original position. The reason for this persistence was that these water molecules were trapped with their interaction partner in an otherwise apolar environment.

On the other hand, a great flexibility was observed for water molecules 2015, 2017 and 2020 at sites 1a, 1b and 1c respectively (crosslinking helices 2, 3 and 7) and 2024 at site 4 (linking helices 2, 3 and 7). The best example for this behaviour was molecule 2020, which hopped during the course of the simulation from its original position 1c to the proximal position 1a on to position 4 that was more than 12 Å away from the original site. Thus, there seems to exist a passage from the water cluster near D83/2.50 to the intracellular space.

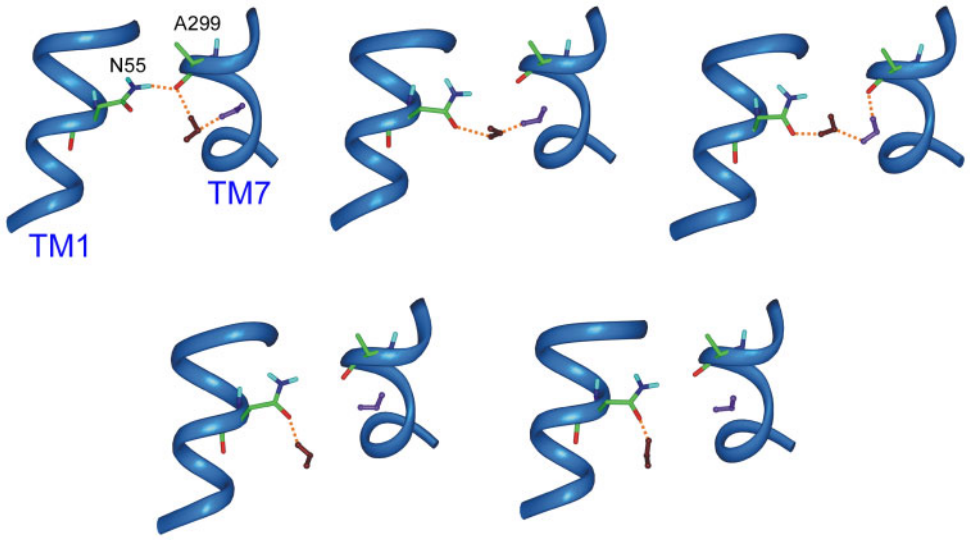


Figure 4.10: Disruption of the highly conserved interaction between N55/1.50 and A299/7.46 by flexible water molecules. Insertion of the water molecules increased the distance between the interaction partners so that the hydrogen bond could not be reestablished at a later point of the simulation.

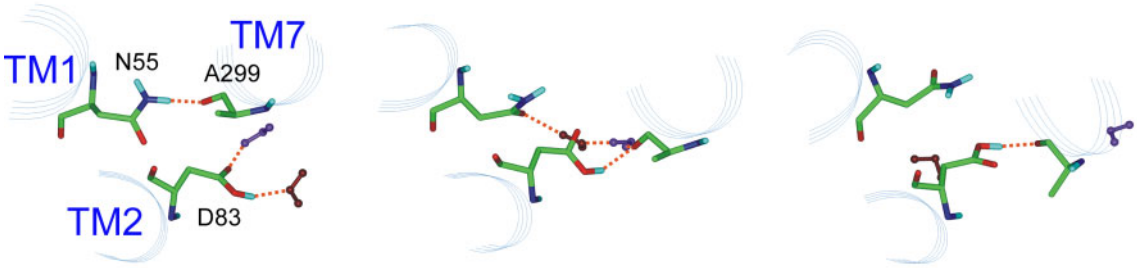


Figure 4.11: Change in the hydrogen bond pattern observed after 12 ns in a simulation of model RT3 catalysed by highly mobile water molecules. The original interaction between N55/1.50 and A299/7.46 was disrupted after 12 ns and a new contact between A299/7.46 and D83/2.50 was established.

Comparison of simulations in presence and absence of internal water molecules showed that neglection of water molecules 2027, 2028, 2000 and 2014 did not alter side chain conformations at the corresponding sites substantially. When these water molecules were, however, considered, they moved away from their original position interfering with existing hydrogen bonds located nearby.

As a subsumption, only two (2021, 964) of the eleven water molecules resolved in reference [81] remained stable at their original position hydrogen bonding to E113/3.28 and Y268/6.51 – C264/2.47 respectively. Three water molecules (2015, 2017, 2020) are described to form a cluster in proximity to the highly conserved residues D83/2.50 and N302/7.49. These solvent molecules as well as molecule 2024 close to the intracellular loop were highly flexible indicating a possible passage to the intracellular space. Four

water molecules (2027, 2000, 2014, 2028) did not seem to be essential.

4.4 CCl_4/Water versus DPPC/Water Environment

A typical phospholipid membrane such as DPPC/water spans about 4.5 nm, [92] of which the hydrocarbon interior accounts for approximately 3.0 nm. Although the lecithin headgroup is of zwitterionic nature, the resulting charge distribution is almost completely cancelled, since the charge distribution of choline and the phosphate atom overlap to a large extent and the remaining charge is neutralised further by the distribution of water dipoles. [93]

In a CCl_4/water simulation environment, the CCl_4 layer is adjusted to 3 nm thus mimicking only the hydrophobic core of the membrane. This simulation system obviously lacks the DPPC-headgroup interface; however, since this region is rich in water molecules, a replacement by an aqueous environment should not introduce large errors.

In order to compare the suitability of the CCl_4/water- with the DPPC/water-environment for a GPCR simulation, the simulation of model RT2 in a DPPC/water mimic was compared to a simulation in a CCl_4/water environment. Figure 4.12 shows the RMSD curve within the transmembrane region for both the DPPC/water (red curve) and a CCl_4/water (blue curve) environment. During the simulation in the DPPC/water environment, a lower RMSD value was maintained for about 4800 ps until a hydrogen bond between helices 5 (Y223/5.58) and 6 (R252/6.35) was established that gave rise to an increase of RMSD up to a value of 0.2 nm. In a CCl_4/water environment, formation of a hydrogen bond between helices 5 and 6 took place already at an earlier stage (after 500 ps), resulting in a 0.02 nm upward shift of the RMSD curve during the first 4800 ps of simulation until the two curves eventually converged.

Apart from the interaction between helix 5 and 6, the interhelical hydrogen bond pattern was highly comparable within the two simulation environments (see figure 4.13). An exception was the interaction between helix 2 and 7. In the CCl_4/water environment, a hydrogen bond interaction was established between residues D83/2.50 and N302/7.49 that did not emerge in the DPPC/water environment. This hydrogen bond can be however regarded as an artifact of simulation setups lacking internal water molecules that would otherwise impede this interaction as D83/2.50 was in contact with water molecules in the crystal structure (see 4.11).

Thus, the only relevant differences in the hydrogen bonding pattern rather took place due to local metastabilities resulting from the neglection of the G-protein (that could be a potential reason for the formation of the hydrogen bond between residue Y223/5.58

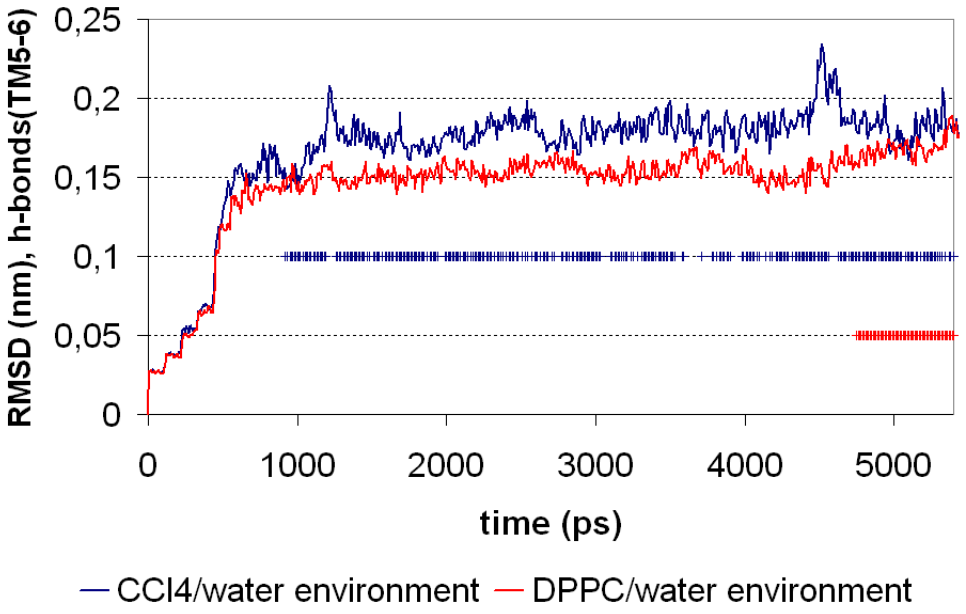


Figure 4.12: Comparison of RMSD plots of the simulations of model RT2 in either a DPPC/water or a CCl_4/water environment. Additionally, the frames in which a hydrogen bond interactions between helix 5 and 6 occurred are indicated by + signs.

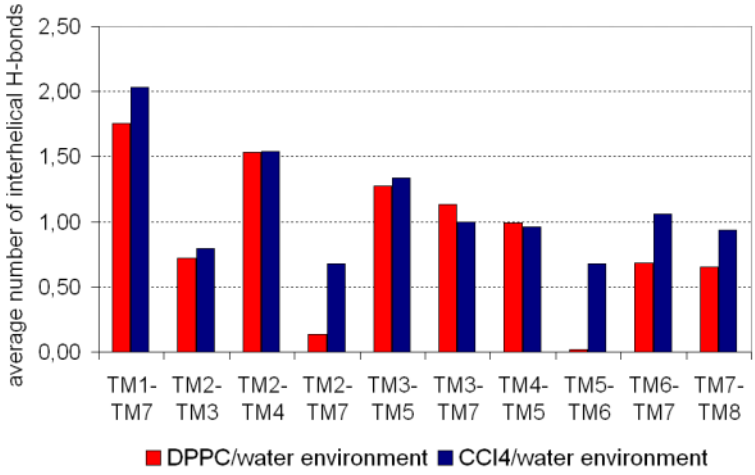


Figure 4.13: Comparison of the average number of interhelical hydrogen bond interactions when simulating model RT2 in a DPPC/water or CCl_4/water environment.

and R252/6.35) or internal water molecules than to differences of the solvent environment used. In both cases a stable simulation (resulting in a plateau of RMSD value) was possible and eventually the same errors occurred. The earlier onset of deviations from the original structure in the CCl_4/water environment lets the DPPC/water environment appear more suitable to maintain the model close to the experimental structure at first sight. This result is however consistent with two possible interpretations: a DPPC/water simulation system temporarily maintains smaller deviations due either to the more natural environment or to

a generally decreased flexibility in a more viscous medium. In order to test this hypothesis, the simulation in different membrane mimics was repeated with a truncated model of bovine rhodopsin with D83/2.50 and E122/3.37 in the deprotonated state — for which large deviations from the starting structure were observed when simulated in a CCl_4/water environment (see figure 4.14, blue curve) — in the DPPC/water environment (figure 4.14, red curve). Apparently, a DPPC/water environment seems to slow down adaptations of the protein structure even when they seem justified. This must be considered as a drawback of this simulation system when applied in the simulation of homology models where structural adaptations are expected to take place during the course of a simulation (e.g. in helix 1). Visual inspection of the resulting protein structures revealed that local distortions of the backbone regions occurred in both simulation setups, however these modifications were accompanied by translocations of entire helices in the CCl_4/water environment that seem to be impeded in the DPPC/water environment.

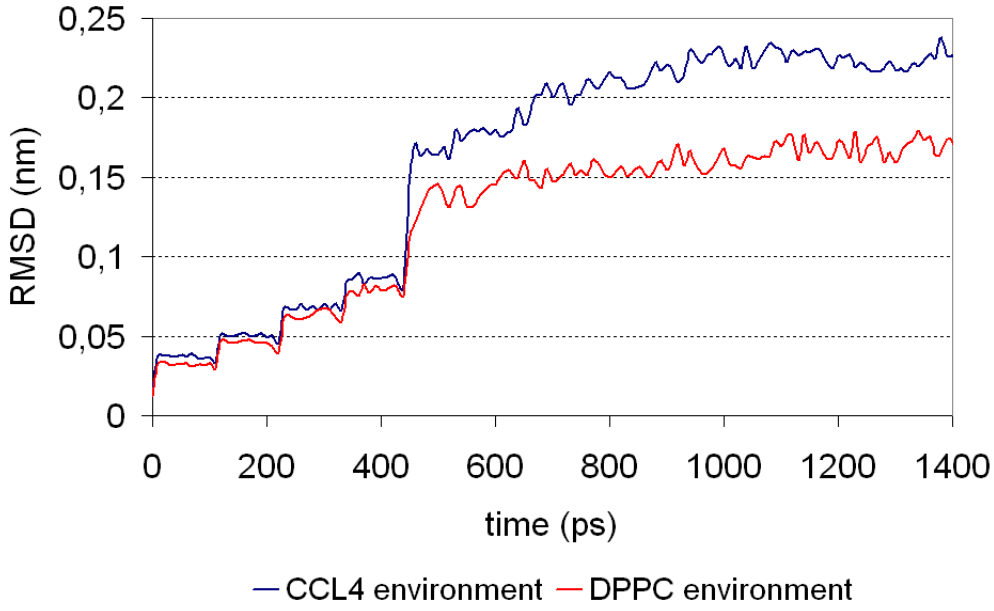


Figure 4.14: Comparison of RMSD plots of the imperfect model RT4 in a CCl_4/water and DPPC/water environment. The expected increase of RMSD is significantly larger in the CCl_4/water environment where structural changes are less impeded.

4.5 Choice of the Correct State of Protonation

4.5.1 Calculation of pK_a-shifts Using the `UHBD` Program

In table 4.4 the state of protonation and the calculated pK_a-shifts, as suggested by the `UHBD` program, are listed for those residues which significantly deviated from the pK_a

in aqueous solution. The calculations were carried out twice, once for the retinal/opsin complex, then also taking into account water molecules as resolved in the crystal structure 1L9H. [81] However, pK_a-shifts were identical, so only values of one calculation are given. The most striking pK_a-shifts were observed for the acidic residues D83/2.50, E122/3.37 and E181/4.70 in the E_2 loop that were all predicted to be in their neutral state.

Table 4.4: pK_a-shifts for titrable sites in the retinal/opsin complex showing significant deviations from the expected protonation states in aqueous solution.

residue	aqueous pK_a	intrinsic pK_a	charge at pH=7.4
D83/2.50	4.0	9.8	0.0
E122/3.37	4.4	12.2	0.0
E181/4.70	4.4	12.1	0.0

While D83/2.50 is a highly conserved residue within biogen aminergic GPCRs, E122/3.37 aligns to different residues as shown in table 4.5. In most sequences, the corresponding position is occupied by a threonine, serine or asparagine, thus the hydrogen donor ability seems to be the key function of residues at this location rather than the introduction of a charge. Residue E181/3.28 is located in the E_2 loop at the end of helix 4. A multiple sequence alignment for this residue is — due to the low sequence conservation of loop regions — however meaningless.

Table 4.5: Multiple sequence alignment of representatives of the family of human biogene aminergic GPCRs and bovine rhodopsin starting with the highly conserved cysteine residue in the E_1 loop involved in the disulfide linkage. The corresponding position to E122 in bovine rhodopsin is highlighted and aligns with residues characterised by their hydrogen bond donor functions. The alignment was carried out using the program `ClustalW` with default settings and the identity scoring matrix.

α_1-AA	C N I W A A V D V L C C T A S
β_1-A	C E L W T S V D V L C V T A S
D_1	C N I W V A F D I M C S T A S
M_2	C D L W L A L D Y V V S N A S
α_2-AA	C E I Y L A L D V L F C T S S
D_2	C D I F V T L D V M M C T A S
5-HT_{2A}	C A V W I Y L D V L F S T A S
5-HT_{1A}	C D L F I A L D V L C C T S S
5-HT_4	C L V R T S L D V L L T T A S
H_3	C K L W L V V D Y L L C T S S
Rhodopsin	C N L E G F F A T L G G E I A

4.5.2 MD Simulations Comparing Different States of Protonation for Residues D83/2.50, E122/3.37, and E181/4.70

In order to analyse the effect of different states of protonation of residues D83/2.50, E122/3.37, and E181/4.70 on the overall protein stability, MD simulations with models varying in these parameters were carried out. If residue D83/2.50 was considered negatively charged and no internal water molecules were considered (model RT4), this residue would disrupt the highly conserved hydrogen bond interaction between N55/1.50 and A299/7.46 (see figure 4.15b). Residue D83/2.50 was — apart from interactions with water molecules — not involved in any stabilising interaction in the crystal structure (see figure 4.15a). If considered in its deprotonated state D83/2.50 formed a hydrogen bond with the backbone nitrogen of residue V300/7.47 which was accompanied by a pronounced relocation of the backbone region of helix 7, resulting in deviations from the ideal helix structure (see figure 4.15b). Whereas consideration of D83 in its protonated state (model RT2) resulted in a conformation analogue to the one observed in the crystal structure (see figure 4.15c).

In one simulation setup with D83 considered in its charged state, an intracellular water molecule intruded and moved up to the location of the water cluster described in [81] during the equilibration period. Under these conditions, residue D83/2.50 permanently interacted with G120/3.35 and S298/7.45 bridged by this water molecule that prevented the distortions previously described.

Of the four possible setups for residue D83/2.50, i.e. D83 un/protonated $\pm$ internal water molecules, the alternative D83(-) without any water molecules consistently resulted in distortion of the protein backbone. The variant D83H in absence of internal water molecules temporarily reproduced the circumstances in the crystal structure represented however a meta stable system as D83 was not involved in any hydrogen bond interaction. In the presence of internal water molecules, both protonation states of D83 were consistent with a structural preservation. Yet, due to the high flexibility of internal water molecules in an unconstrained dynamics simulation, the hydrogen bonding pattern that was observed in the crystal structure was only temporarily preserved (see figure 4.9).

The situation for residue E122/3.37 was rather straightforward. A multiple sequence alignment shown in table 4.5 already suggested that at this position a hydrogen donor function might be important. This was supported also by the MD simulations carried out. In the setup where E122 was considered negatively charged, a distortion of the backbone region of helix 4 could be observed, and residues in this region significantly deviated from the conformation observed in the crystal structure, whereas consideration of E122 in its protonated state led to negligible deviations from the crystal structure (see figure 4.16).

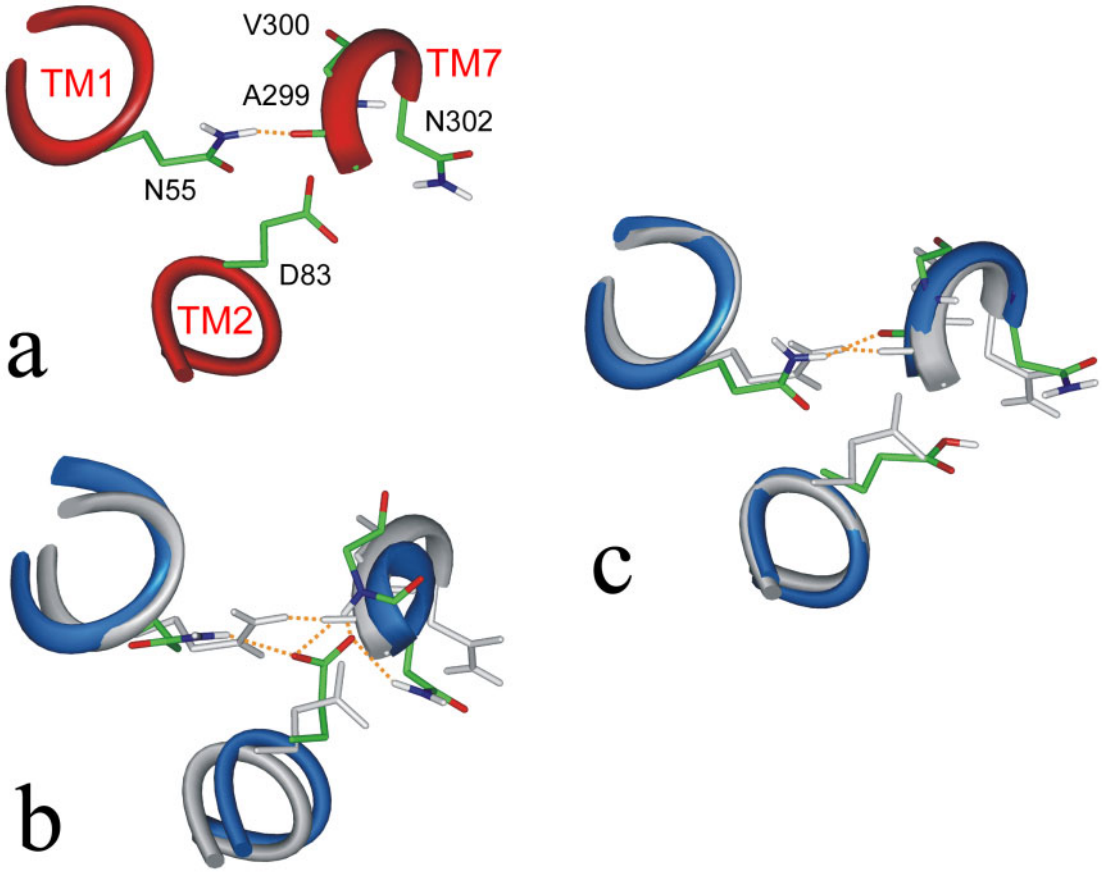


Figure 4.15: **(a)** Hydrogen bond pattern around residue D83/2.50 as observed in the crystal structure 1HZX. **(b)** Hydrogen bond interactions obtained after 1 ns unconstrained MD simulation of model RT4. **(c)** Hydrogen bond pattern observed after 5 ns unconstrained MD simulation of model RT2. In figures **(b)** and **(c)** the reference structure is superimposed in light grey.

Similarly, a protonation of residue E181/4.70 resulted in smaller deviation from the crystal structure (see figure 4.17).

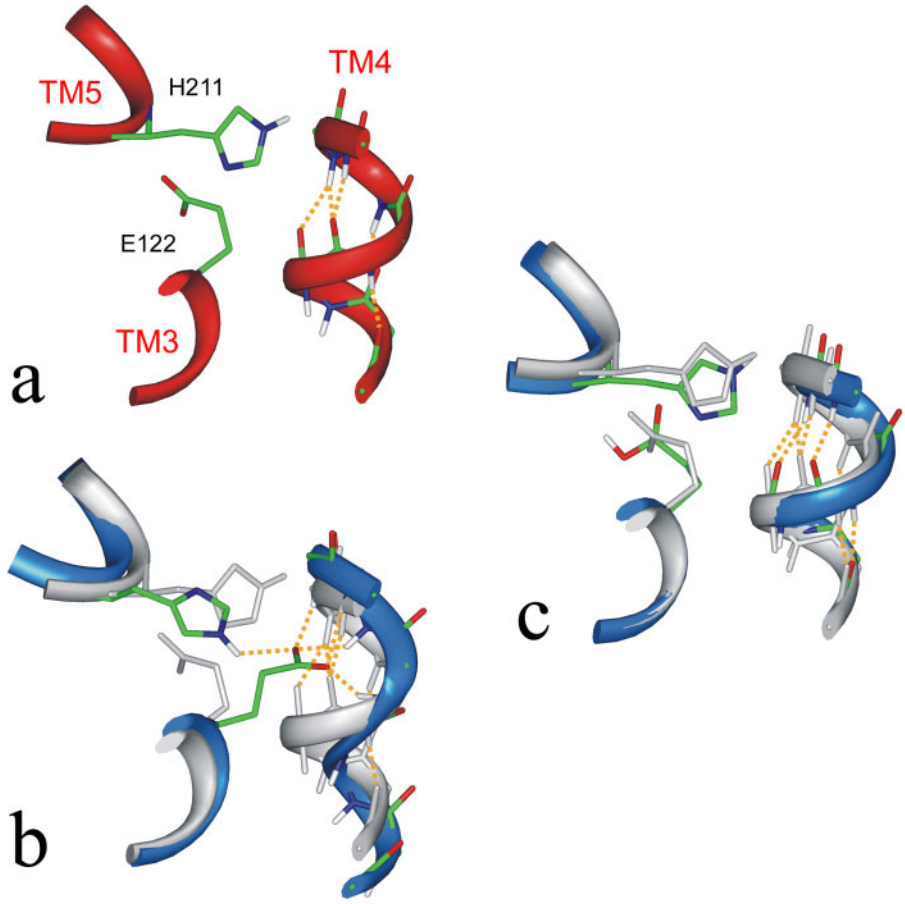


Figure 4.16: **(a)** Hydrogen bond pattern around residue E122/3.37 as observed in the crystal structure 1HZX. **(b)** Hydrogen bond interactions obtained after 1 ns unconstrained MD simulation of model RT4. **(c)** Hydrogen bond pattern observed after 5 ns unconstrained MD simulation of model RT2. In figures **(b)** and **(c)** the reference structure is superimposed in light grey.

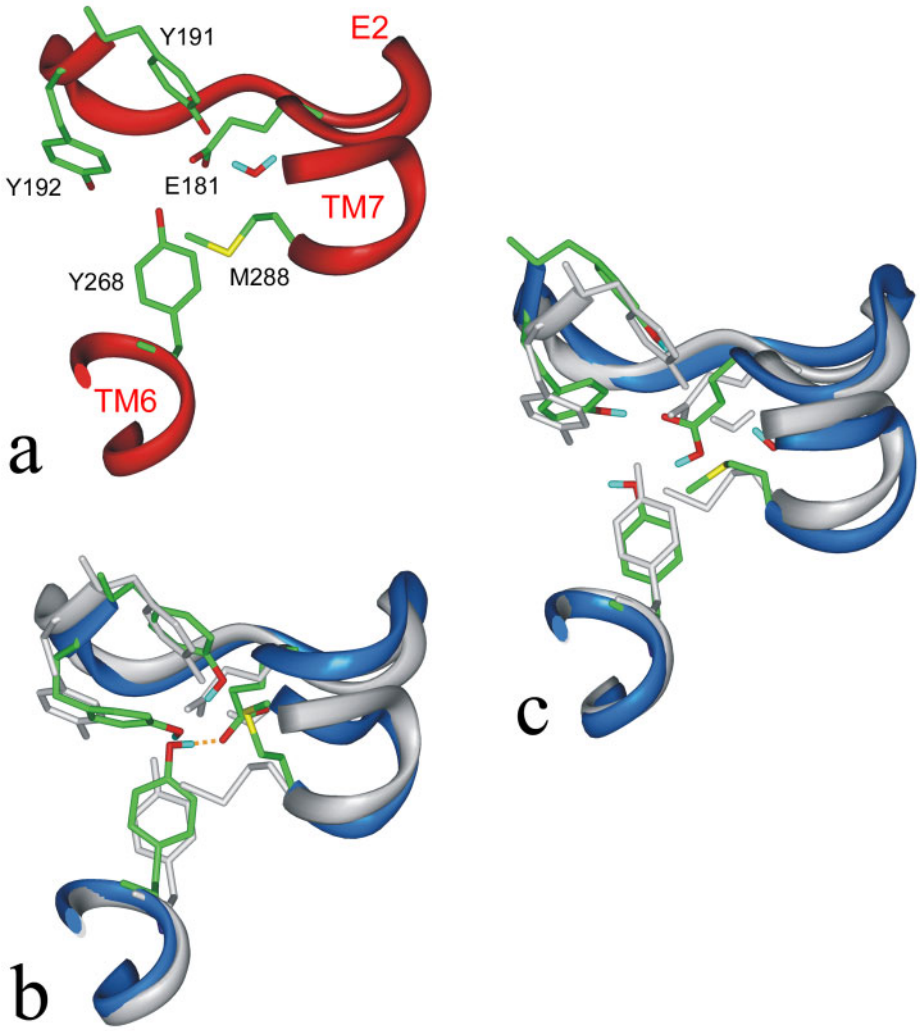


Figure 4.17: **(a)** Hydrogen bond pattern around residue E181/4.70 as observed in the crystal structure 1HZX. **(b)** Hydrogen bond interactions obtained after 1 ns unconstrained MD simulation of model RT4. **(c)** Hydrogen bond pattern observed after 5 ns unconstrained MD simulation of model RT2. In figures **(b)** and **(c)** the reference structure is superimposed in light grey.

4.6 Studying the Conformational Adaptations after Introduction of *all-trans*-Retinal

Activation of bovine rhodopsin takes place after absorption of a photon that leads to the isomerisation of *11-cis*- to *all-trans*-retinal. The conformational changes thereby occur in a stepwise manner, producing the intermediates photorhodopsin, bathorhodopsin, lumirhodopsin, metarhodopsin I and metarhodopsin II (see figure 1.4). [6] Isomerisation of the C_{11}-C_{12} double bond takes place during the transition from batho- to lumirhodopsin. Furutani *et al.* have demonstrated that *all-trans*-retinylidene can be accommodated without large structural rearrangements by reconverting lumi- to bathorhodopsin at 77 K. [7]

In model RT2 of bovine rhodopsin, an *all-trans*-retinylidene chromophore was introduced by changing the dihedral angle of the C_{11}-C_{12} double bond. The backbone was restrained with a tether of 9000 kJmol^{-1}nm^{-2} in order to avoid displacements of entire helices or local distortions of the backbone region. After a simulation period of 5 ns in a DPPC/water environment, only minor deviations could be observed (see figure 4.18). In order to accommodate the ionone ring in proximity to A169/4.58, the ring conformation of P171/4.60 changed and residues Y178/4.67 and I189/E_2 reoriented slightly in order to avoid steric clashes. The most pronounced change in an amino acid side chain conformation was observed for F212/5.47. Phenylalanine 212/5.47 previously interacting with residues L216/5.51 and F208/5.43 reoriented as to fill the volume previously occupied by the β-ionone ring. In succession also M207/5.42 reoriented. The switch of F212/5.47 thereby only depended on the empty space introduced by the isomerisation as could be demonstrated within a simulation where the retinylidene ligand was truncated and only the C_{15} atom was considered in order to preserve the Schiff-base moiety.

Accommodation of the *all-trans*-retinylidene chromophore required structural adaptation of P171/4.60. The resulting ring conformation was however quite likely introduced due to large constraints on the backbone region and the *all-trans* retinylidene ligand. In order to allow for slight backbone rearrangements, the simulation was repeated in absence of backbone tethers, only restraining the retinylidene ligand (9000 kJmol^{-1}nm^{-2}). Again, the most pronounced adaptation was the change in side chain conformation of residue F212/5.47. Successively, also M207/5.42 and F203/5.38 reoriented. The result of this cascade was the disruption of the interaction between W175/4.64 and S202/5.37 that had been highly stable in other simulation setups (see figure 4.19b). The loss of this interaction increased the flexibility of helix 4 that separated from helix 5 (2.67 Å measured from C167/C_α-C_α´) provoking also the loss of the interhelical contact between A166/4.55 and Y206/5.41. The inward movement of F212/5.47 also induced a reorientation of residue E122/3.37 that was previously interacting with the backbone-oxygen of H211/5.46. Now,

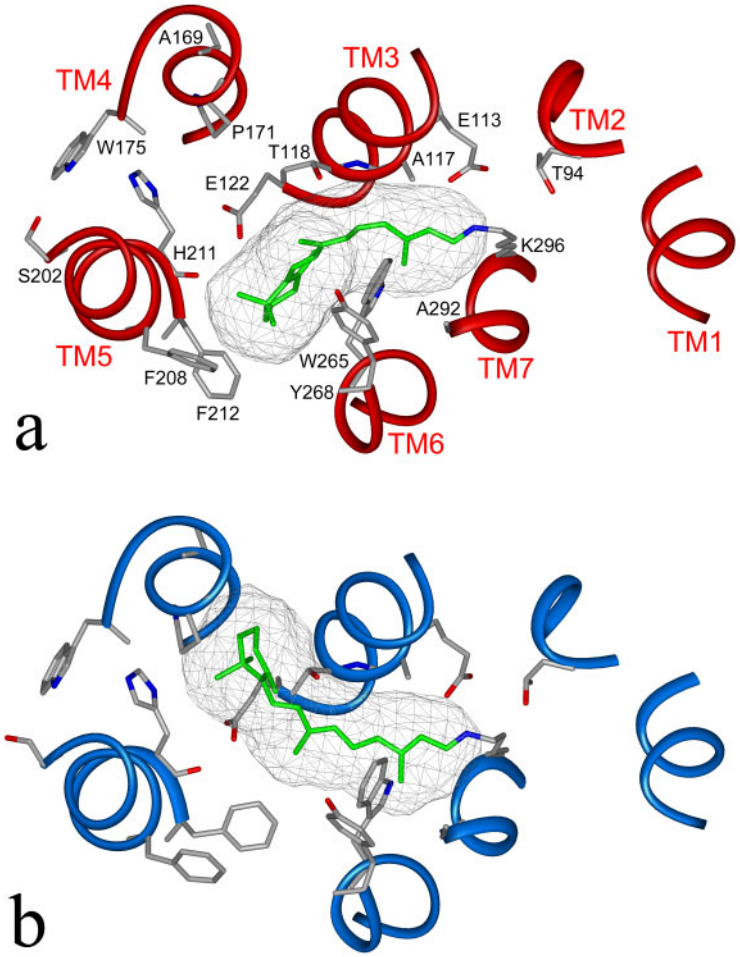


Figure 4.18: **(a)** Position of *11-cis*-retinal as observed in the crystal structure 1HZX. **(b)** Position of *all-trans*-retinal after 5 ns MD simulation with constraints on the backbone atoms.

an alternative interaction was established between these residues, involving the imidazole moiety of H211/5.46 and the carboxyl-group of E122/3.37. This modification led to a deviation of the intracellular part of helix 5 that approached helix 6 alongside slightly increasing the distance between helix 3 and 5.

Briefly, the change in the side chain conformation of residue F212/5.47 resulted in the loss of interhelical contacts between helices 4 and 5. Thereupon helix 4 shifted outwards with its *N*-terminal ending. The reorientation of residue E122/3.37 resulted in an approach of helix 5 towards helix 6. The described adaptations took place during the first 500 ps and resulted in a RMSD value of 0.2 nm within the transmembrane region. Thereupon no significant adaptations occurred until the end of the simulation.

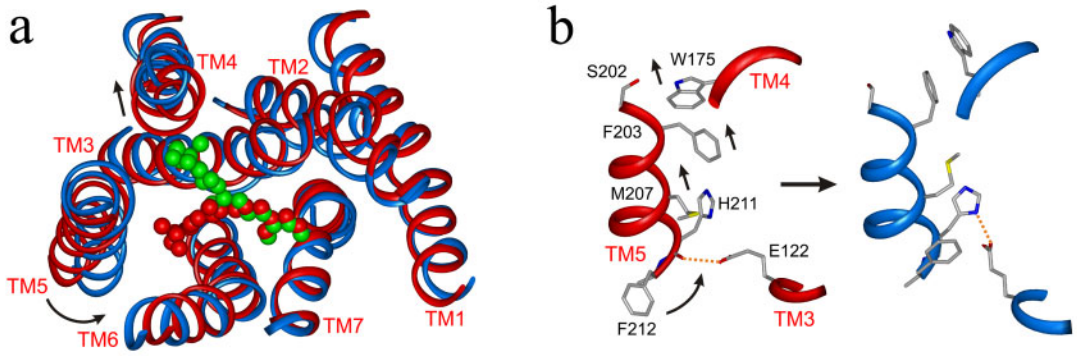


Figure 4.19: **(a)** Transmembrane helix bundle of bovine rhodopsin viewed from the extracellular side before (red) and after (blue) introduction of *all-trans*-retinal and an unconstrained MD simulation of 5 ns. Pronounced movements of helices are indicated by black arrows. **(b)** Cascade of amino acid side chain rearrangements provoked by the initial flip-over of residue F212/5.47 leading to the disruption of the interaction W175/4.46 – S202/5.37 and to the modification of the hydrogen bond established between E122/3.37 and H211/5.46.

4.7 Simulation Setup for the Analysis of Interhelical Contacts

The positional root mean square deviation from the experimental X-ray structure and analysis of the interhelical hydrogen bond pattern were used in order to assess the structural preservation. Based on these criteria, the simulation setup defined through the following parameters resulted in a minimal deviation from the original crystal structure during the first 5 ns:

- consideration of the entire protein sequence
- residues D83/2.50 and E122/3.37 in the protonated (neutral) state
- no consideration of internal water molecules
- DPPC/water environment
- 2 fs time-step

With the above setup, the RMSD within the helical backbone region remained stable for approximately 4200 ps after the initial equilibration phase (see figures 4.20, red curve and 4.21). The observed RMSD value of 0.15 nm, together with the high similarity of the side chain conformations compared to the crystal structure suggested this model be suitable for analysing the dynamic behaviour of interhelical hydrogen bonds.

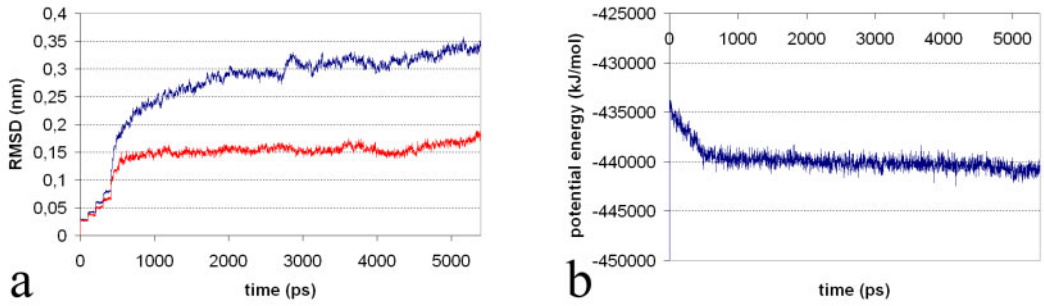


Figure 4.20: **(a)** RMSD-course within the helical region (red curve) and the entire backbone region (blue curve) during the simulation of model RT2 in a DPPC/water environment. **(b)** Corresponding plot of potential energy.

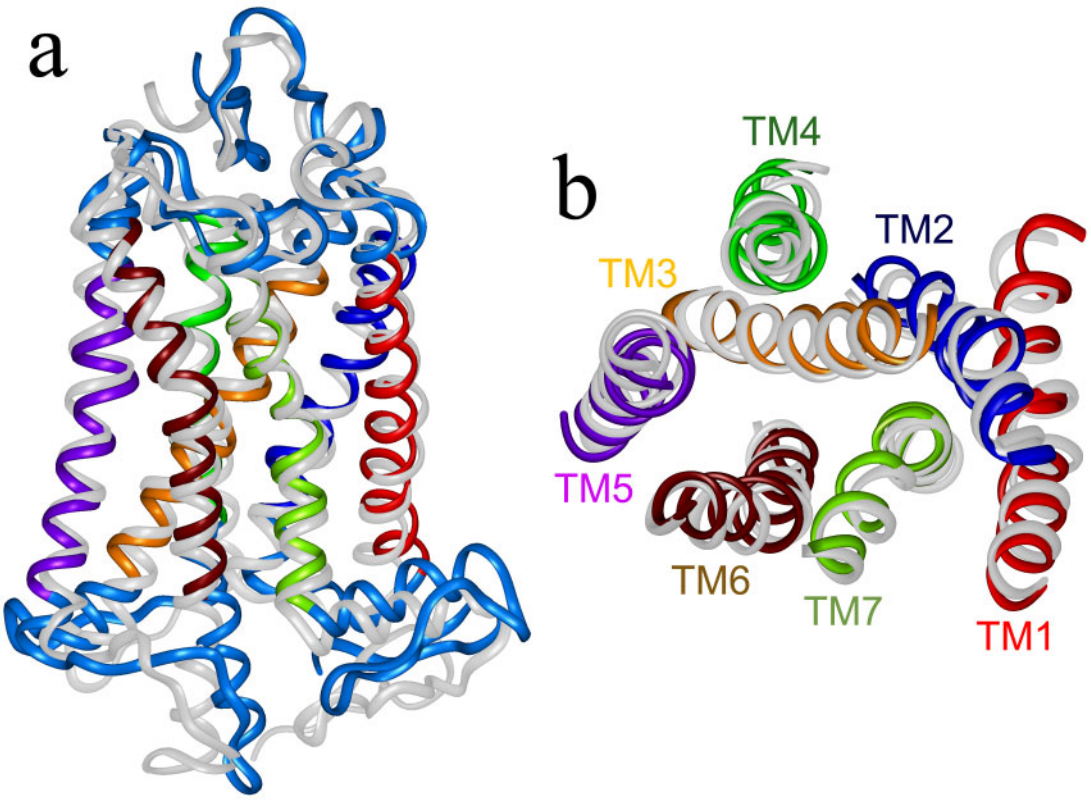


Figure 4.21: Comparison of the backbone region of the crystal structure of bovine rhodopsin 1HZX (light grey) and model RT2 that had been simulated for 5 ns without constraints: **(a)** side view **(b)** helix bundle viewed from the extracellular space (loops are neglected for reasons of clarity). Deviations from the crystal structure are apparently small and predominantly limited to the intracellular loop region.

The RMSD within the entire backbone region (see figures 4.20, blue curve) of the bovine rhodopsin model was considerably higher due to deviations that occurred especially in the intracellular loop region. This flexibility is in agreement with the experimental finding that this part of the protein has been difficult to crystallise due to its high flexibility. In contrast, the *N*-terminal end and the extracellular loop region including the 2^{nd} extracellular

loop remained highly stable during the simulation which is in agreement with [94]. The largest deviation in this region consisted in an inward shift of the loop connecting helix 6 and 7 with a maximal displacement of 5.4 Å. Minor changes occurred as well within the initial stretch of 10 amino acids of the *N*-terminal ending. After 4700 ps, a rise in RMSD can be observed which is related to the formation of a hydrogen bond between residues Y223/5.58 and R252/6.35. Formation of a hydrogen bond between the highly conserved residue Y223/5.58 [80] and different residues of helix 6 has been observed as well in other simulation setups. In the crystal structure, Y223/5.58 is not involved in any hydrogen bond interaction and is pointing towards residue E232 located in the 3^{rd} intracellular loop. When the MD simulation is prolonged up to 25 ns, Y223/5.58 gets permanently involved in hydrogen bond interactions with various residues from helix 6.

The temporary equilibration phase can be seen when the potential energy is plotted as a function of time (see figure 4.20b). Apparently, formation of the interaction between helix 5 and 6 is accompanied by a decrease in potential energy of the protein structure.

Figure 4.22 shows all interhelical hydrogen bonds described by Teller *et al.* [82] and additional ones that occurred at frequencies above 1% of the analysed frames that were written out every 2 ps. Table 4.6 shows the corresponding frequencies of occurrence.

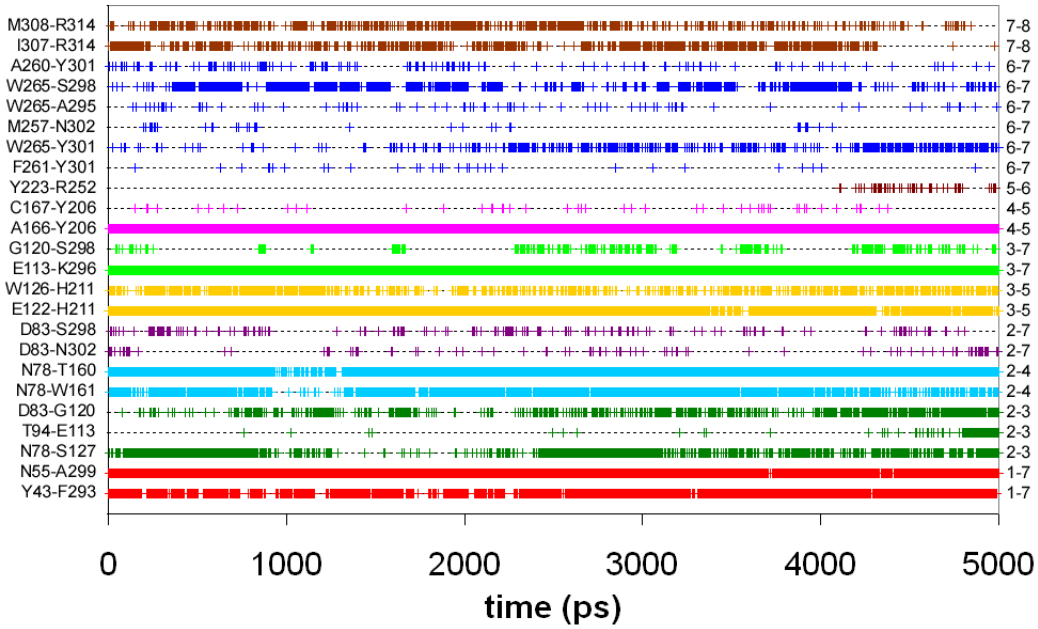


Figure 4.22: Interhelical hydrogen bonds observed during the simulation of model RT2 in a DPPC/water environment. Helices involved in the contact are written on the right side.

The calculations imply that strong interhelical contacts exist between helices TM3-TM7 that are on average linked by 1.9 hydrogen bonds. These interactions are produced mainly through the hydrogen bond reinforced ionic interaction of residue E113/3.28 and K296/7.43. Although in the crystal structure only a lateral interaction of one carbonyl

Table 4.6: Frequencies of interhelical hydrogen bonds observed during the simulation of a model of bovine rhodopsin. Frames were written out every 2 ps. Not all of the spatially possible interactions, as derived from the crystal structure by Teller *et al.*, [82] were observed during the simulation and seemed necessary for upholding the protein conformation. An asterisk was added to hydrogen bond interactions that occurred due to the neglection of internal water molecules that would otherwise have prevented these hydrogen bonds.

interhelical contact	residues involved	frequency (%)	interhelical contact	residues involved	frequency (%)
TM1-2	N55 – A80 1.50/2-47	<1	TM3-7	E113 – K296 3.28/7.43	99
TM1-2	N55 – D83 1.50/2.50	<1	TM3-7	G120 –S298 3.35/7.45	13∗
TM1-7	Y43 – F293 1.38/7.40	75	TM4-5	M163 – H211 4.52/5.46	<1
TM1-7	Y43 – F294 1.38/7.41	<1	TM4-5	A166 – H211 4.55/5.46	<1
TM1-7	N55 – A299 1.50/7.46	95	TM4-5	A166 – Y206 4.55/5.41	97
TM1-8	Q64 – T320 1.59/7.67	6	TM4-5	C167 – Y206 4.56/5.41	2
TM2-3	N78 – S127 2.45/3.42	39	TM4-5	C167 – H211 4.56/5.46	<1
TM2-3	T94 – E113 2.61/3.28	5	TM5-6	Y223 – R252 5.58/6.35	2
TM2-3	D83 – G120 2.50/3.35	27∗	TM6-7	E249 – M309 6.32/7.56	<1
TM2-4	N78 – W161 2.45/4.50	61	TM6-7	F261 – Y301 6.44/7.48	1
TM2-4	N78 – T160 2.45/4.49	92	TM6-7	W265 – Y301 6.48/7.48	18
TM2-7	D83 – S298 2.50/7.45	10∗	TM6-7	M257 – N302 6.40/7.49	1
TM2-7	D83 – A299 2.50/7.46	<1	TM6-7	W265 – A295 6.48/7.42	3
TM2-7	D83 – V300 2.50/7.47	<1	TM6-7	W265 – S298 6.48/7.45	40
TM2-7	D83 – N302 2.50/7.49	4∗	TM6-7	A260 – Y301 6.43/7.48	5
TM3-4	W126 – M163 3.41/4.52	<1	TM7-8	I307 – R314 7.54/7.61	40
TM3-5	E122 – H211 3.37/5.46	91	TM7-8	M308 – R314 7.55/7.61	26
TM3-5	W126 – H211 3.41/5.46	37			

oxygen atom with the nitrogen atom is observed, two simultaneous hydrogen bonds are found in most frames of the molecular dynamics simulation due to the fact that force field methods favour an chelate-type interaction that is, however, rarely found for this kind of interaction in natural protein-ligand complexes. TM1 – TM7 (1.7), TM2 – TM4 (1.5), TM3 – TM5 (1.3), TM4 – TM5 (1.0), TM2 – 3 (0.8), TM6 – TM7 (0.7) are linked by the average number of interactions indicated in brackets (see figure 4.23 for detailed contacts). In this simulation setup also an occasional hydrogen bond was observed between helices 2 and 7 that seemed however artificial due to the neglection of internal water molecules. Consideration of an internal water molecule at a position corresponding to site 1a [81] would especially affect the hydrogen bonding pattern around D83/2.50, as this water molecule forms contacts to D83/2.50, N302/7.49 and potentially G120/3.35, thus crosslinking helices 2, 3 and 7.

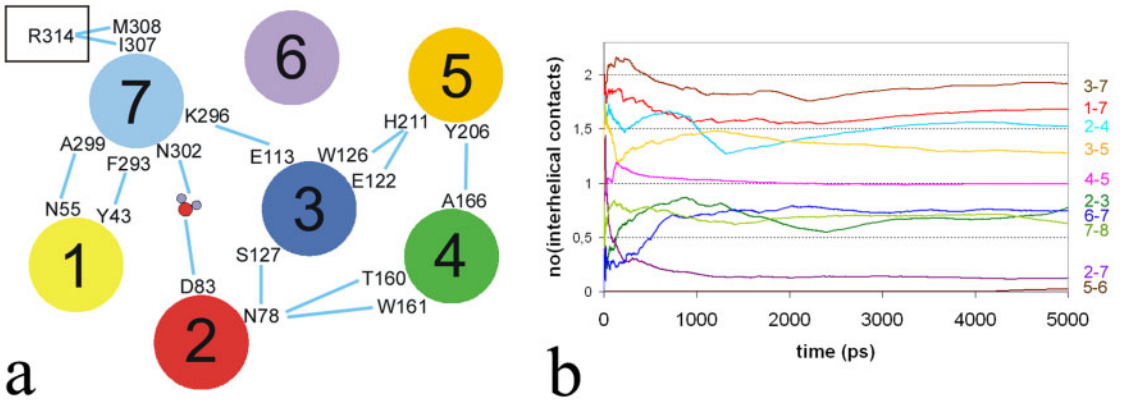


Figure 4.23: **(a)** Scheme of bovine rhodopsin depicting important interhelical contacts. **(b)** Course of the number of interhelical contacts during an unconstrained MD simulation of model RT2 in a DPPC/water environment. Helices involved into the contact are written on the right side.

Furthermore, important hydrogen bonds within the extracellular and intracellular loop regions were established between residues W175/4.64 – S202/5.37 (92%) and R135/3.50 – E247/6.30 (47%) at the frequencies given in brackets. Obviously, the ionic lock between helix 3 and 6, which had been previously described by Shapiro *et al.*, [95] was not present continuously. However, it was never lost during the course of the simulation.

Chapter 5

Discussion

In chapter 4 several long lasting simulations of models of bovine rhodopsin are described varying in several parameters that one frequently has to adjust when setting up simulation systems of GPCR homology models. Since the crystal structure of bovine rhodopsin has been released, several long lasting simulations of the structure of bovine rhodopsin have been published by several groups. [94, 96] In the work of Crozier *et al.* the main focus was set on a postulated correlation between conformational changes of retinal and surrounding amino acid residues to a larger scale conformational change of the entire protein structure. It was speculated that fluctuations observed during a simulation of the dark-adapted rhodopsin (with retinal in its *11-cis*-conformation) could be part of the pathway to the light-adapted state. Additionally, interactions of the protein structure with the surrounding lipid and water environment were analysed during the course of the simulation. Huber *et al.* investigated the complex interactions between the rhodopsin structure and the lipid environment using a POPC membrane mimic. Additionally, the dynamic behaviour of internal water molecules and structural fluctuations of the entire rhodopsin structure were analysed.

Although simulations of bovine rhodopsin already exist, the main focus is usually set on rhodopsin specific questions such as activation or interaction with the lipid environment. Here, the simulations are seen from the view-point of a homology modeller using the structure of bovine rhodopsin as a template structure. The main interest was to evaluate which circumstances lead to the most stable simulation and the consequence of simulation setups differing from this optimal setup.

5.1 System Setup

All models of bovine rhodopsin were based on the crystal structures 1HZX [82] that had been resolved with a resolution of 2.8 Å and 1L9H [81] with a resolution of 2.6 Å. Both models are quite similar regarding the placement of amino acid side chains, and both lack the stretches 236–240 and 331–333 located in the intracellular loop region. The main difference consists in the number of internal water molecules resolved, with significantly more H_2O molecules present in the structure 1L9H. As will be discussed in section 5.3, water molecules located within the transmembrane region can significantly influence the outcome of a simulation. In July 2004, the crystal structure 1U19 with a resolution of 2.2 Å was released. [97] In this structure, coordinates were available for the entire rhodopsin structure including the complete intracellular loop region. Comparison of the resolved stretches 236–240 and 331–333 to the modelled residues revealed only minor deviations and again no significant changes occurred in the amino acid side chain placements. In 1U19 further water molecules could be resolved, the main number of which was however located in the extracellular loop region. It can be expected that this region is accessible to water molecules so that neglection of these molecules in the system setup will not influence an MD simulation. One difference that might be important in the simulation setup is an additional water molecule on the transmembrane level in proximity to residues W265/6.48 and S298/7.45. The possible significance of this solvent molecule will be discussed in section 5.3.

No *N*-terminal acetylation, glycosylations or a *C*-terminal palmitoyl moiety were considered in the models of bovine rhodopsin, as no comparable information is usually available when setting up a system of a GPCR homology model. Neglection of these residues did however not seem to introduce significant deviations during MD simulations.

5.2 Effects of an *N*-terminal Truncation

Construction of loop regions generally is a quite challenging task. [98] As residues of the extracellular loop region, especially the 2^{nd} extracellular loop, are known to participate in ligand binding [99, 100] or are responsible for receptor subtype specificity, [101, 102] an explicit consideration of these regions is crucial in the generation of GPCR models. However, sequence conservation — especially within the *N*-terminal region — is usually quite low so that the generation of reasonable conformations of this region would be quite time consuming and difficult. Numerous pharmaceutically interesting GPCR targets have a postulated binding pocket located within the transmembrane helix bundle distant from the *N*-terminal ending. Thus the effect of an *N*-terminal truncation upon the overall protein

integrity was tested in order to assess the possibility of a potential neglection of this region.

Truncation of the *N*-terminal region resulted however in pronounced deviations within the extracellular loop region including the 2^{nd} extracellular loop and of the position of helix 1, compared to the crystal structure. Structural changes within the extracellular loop region were due to the effect that residues originally buried were now brought to the protein surface facing the aqueous environment. The fact that also a structural rearrangement of helix 1 took place was unexpected, as helix 1 is anchored to helix 7 via two hydrogen bond interactions that were highly stable when the entire protein structure was simulated (see figures 4.7 and 4.8b). However, in the truncated models the interaction between Y43/1.38 and F293/7.40, located in proximity to the extracellular space, was broken up either by the intrusion of water molecules or by the increased flexibility of helix 1 lacking the *N*-terminal end.

The significantly larger structural damage following an *N*-terminal truncation observed in the CCl_4/water environment (see figure 4.8a), where deviations occurred due to the increased flexibility of the *N*-terminal end, was a consequence of the decreased viscosity of this membrane mimic compared to the DPPC/water (see section 5.4).

Rhodopsin-like GPCRs usually lack the proline kink in helix 1, [103] and no hydrogen bond linkage anchors the extracellular part of helix 1 to the helix bundle. Thus, helix 1 might consequently have an increased intrinsic flexibility compared to the one in bovine rhodopsin. Given the pronounced effects of a neglection of the *N*-terminal end that has been observed in the structure of bovine rhodopsin, an explicit consideration of the *N*-terminal end in GPCR homology models will be essential to preserve the topology of the helix bundle.

5.3 Effects of Internal Water Molecules

A proper incorporation of water molecules into the model of bovine rhodopsin turned out to be quite difficult and the outcome of the MD simulation does not imply a suitable tactic as to proceed in the case of GPCR homology models. When considered, water molecules located in protein regions rich in polar residues disrupted highly conserved hydrogen bond interactions and tended to catalyse the switch to alternatively possible interactions (as shown in figures 4.10 and 4.11). When neglected, amino acids that consequently lacked their interaction partner tended to form hydrogen bonds to residues located close-by (contacts marked with an asterisk in table 4.6).

In homology models, the situation will be even more complicated, as the exact position of (potential) water molecules is not known *a priori*. Additionally, there seems to exist

an interdependence between amino acid side chain conformations and the location of water molecules that will further hamper both a correct placement of side chains and water molecules. When applying the SCWRL algorithm [67] on the rhodopsin backbone, a large amount of residues is predicted correctly. Deviations were observed mostly for side chains facing the lipid or aqueous environment, for which however a high flexibility was observed during an MD simulation. Variations in the side chain conformations of residues facing the helix bundle interior were observed for Y306/7.53, E113/3.28, E181/4.70 and R135/3.50. Interestingly, the first three residues mentioned are in contact with internal water molecules. Thus, a potential function for these water molecules could be to restrain an otherwise suboptimal side chain conformation.

Once a reasonable homology model is obtained, the exact localisation of internal water molecules remains extremely difficult to evaluate and up until now no reliable methods exist to satisfactorily solve this task. Rarey *et al.* [104] have for example described a method for water molecule placement during ligand docking. However, the success of this strategy was limited; of the 200 PDB-protein-ligand complexes in which water molecules were involved in the active site, only 35% of the water locations could be predicted correctly. Alternatively, a crystal structure lacking internal water molecules could be scanned with a water probe, as for example implemented in the program GRID [105] that maps favourable interaction sites for H_2O molecules. However, here the results are often quite difficult to interpret due to the high number of indicated sites with small interaction fields and strongly depend on the contour level applied. Using this program, the strongest interaction fields, indicating favourable sites for water molecules, were generated for the sites corresponding to the location of water molecules 2020 (in contact with Y301/7.48 and N302/7.49) and 2015 (in contact with D83/2.50, G120/3.35 and N302/7.49) (see figure 5.1).

Even when the exact position of the H_2O-oxygen is known, the exact orientation of the water molecules is still not; and anyway seems to be highly flexible within regions of polar character, at least under MD simulation conditions. In order to avoid the described flexibility that can cause the disruption of conserved hydrogen bonds, a possible strategy would be the introduction of distance restraints fixing the position with respect to the interacting residues. Alternatively, the position of the H_2O-oxygen atom could be constrained. In the case of homology models this would be an applicable strategy for water molecules crosslinking highly conserved residues such as solvent molecule 2015 of the water cluster near D83/2.50.

In the structure 1U19, which became available only recently, the position of most water molecules at transmembrane level is almost identical to the placement in 1L9H. A main difference is however the presence of an additional water molecule in the water cluster near residue D83/2.50 (see figure 5.1, water molecule W 1U19). The consequences of

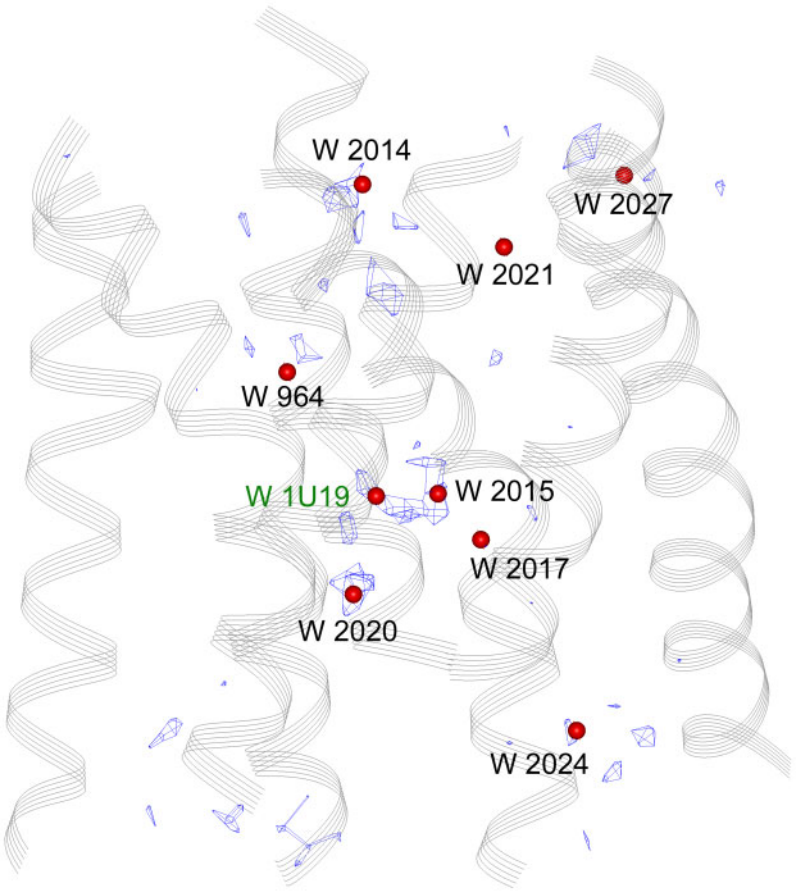


Figure 5.1: Positive interactions sites for H_2O molecules in bovine rhodopsin analysed with `GRID`. The sites where favourable interaction potentials were predicted for water molecules are depicted as volumes surrounded by blue lines. The oxygen atoms of water molecules as resolved in the structure 1U19 are represented in their van der Waals representation. Water molecule `W 1U19` was not predicted in the structure 1L9H. Apparently, prediction of the position of water molecules based only upon favourable interaction sites fails to reproduce all sites where solvent molecules are present.

an additional water molecule at this position on the output of the MD simulation is difficult to judge. In the paper of Crozier *et al.* the simulation was carried out with the structure 1F88, which had been the only available experimental structure at that time. Although it had been argued that the simulation should not depend on possible alternative starting points (referring to the structure 1L9H with more water molecules resolved) several factors analysed from the dynamics simulation, such as the interaction energy of E181/4.70 with retinal or the amount of transition observed for this residue, could be well influenced by a water molecule considered (1L9H) or neglected (1F88, W2014) interacting with E181/4.70. Although the intrusion and correct positioning of water molecules to the protein structure can occur during an MD simulation, the required time scales to ascertain correct placement of all water molecules may be impracticably high.

5.4 Influence of the Simulation Environment

Regarding the simulation environment, a CCl_4/water and a DPPC/water environment were compared. Theoretically, also simulations in vacuum can be carried out, however the GROMOS force field does not perform well on vacuum simulations and in any case, a complete neglection of the solvent environment is problematic. As described by Weinstein *et al.* [98] in vacuum there is a strong tendency that polar or charged residues on the protein surface do not exhibit extended conformations as can be expected for these residue types in contact with water and ions, but fold back on the protein surface producing artifacts in the hydrogen bonding pattern. In order to explicitly consider a more realistic environment of transmembrane proteins, several strategies have been published using either a CCl_4/water, [84] an octane/water, [106] or a phospholipid/water [106, 107] environment. Given that the proper function of a protein often requires a specific lipid composition, the use of CCl_4, octane or saturated lipid molecules as membrane mimic will apparently represent a simplification of the real circumstances. Nevertheless, a CCl_4/water and a saturated phospholipid/water system were chosen to simulate a model of bovine rhodopsin in order to analyse the influence of membrane mimics of different complexity and dissimilar physical properties.

As described in section 4.4, the environment used significantly influenced the protein flexibility, a consequence of the differences in viscosity of the two membrane mimics used. The outcome of the MD simulations in both solvent environments was similar when a mostly correct model of bovine rhodopsin was simulated (see figures 4.12 and 4.13). If however an incorrect model (i.e. for example a model with incorrectly adjusted titrable residues) was simulated, structural deviations became apparent much faster in a CCl_4/water environment due to the significantly lower structural hindrance imposed by this membrane mimic. It must be stressed that structural adaptations due to the introduced errors took place in both protein structures. However, in the DPPC/water environment adaptations only occurred on a local level, while in a CCl_4/water box entire helices would rearrange.

Thus, although the DPPC/water environment emulates the natural circumstances more closely, the increased viscosity of the medium will require much longer simulation periods in order to observe comparable deviations — if required. If the goal is to rapidly compare alternative GPCR homology models with respect to their structural stability, application of the more artificial CCl_4/water environment might be justified due to the significant lower hindrance of even large conformational changes.

5.5 Choice of Protonation States

As described in 2.1 the observed pK_a values within a protein can significantly differ from the pK_a in an aqueous environment. The `UHBD` program [69] can calculate these pK_a-shifts and outputs the calculated state of protonation at various pH values and pK_a-shifts for all titrable sites in the protein given.

For three acidic residues (D83/2.50, E122/3.37 and E1817/4.70) a considerable pK_a-shift was predicted thereby implying that these residues will be present in their neutral state. This result is in good agreement with experimental findings testing the protonation state of titrable sites in bovine rhodopsin. Fahmy *et al.* [108] have analysed rhodopsin and metarhodopsin II by means of Fourier-transform infrared difference spectra (FTIR) analysis which indicated a protonation of D83 and E122 while Raman vibrational spectra described in [109, 110] supported the view that also E181 is present in its protonated state.

D83/2.50 is a highly conserved residue throughout the GPCR family, and mutation studies revealed that this residue is involved in the activation process. As described in [81], D83/2.50 is stabilised by an interaction with water molecule 1a linking helices 2 (D83/2.50), 3 (G120/3.35) and 7 (S298/7.45). A protonation of D83/2.50 is also supported by the observation of the existence of a sodium binding site. [111] Presence of a sodium ion would result in a compensation of the potential negative charge of D83/2.50, thus the existence of the neutral form of D83/2.50 in absence of ions seems convincing.
If the protonated neutral forms of E122/3.37 and E181/4.70 were considered, E122 could form a hydrogen bond with the backbone carbonyl group of H211/5.46, and E181 could be stabilised by an additional hydrogen bond to Y268/6.51. The existence of E122/3.37 in its protonated neutral form was also supported by a multiple sequence alignment. In most sequences the corresponding position is occupied by a threonine, serine or asparagine, thus the hydrogen donor ability seems to be the key function of residues at this location rather than the introduction of a charge.
Finally, for residue E181/4.70 a potential implication as a counter ion switch for residue E113/3.28 has been discussed. [109, 110]

The experimental data available for bovine rhodopsin could be reproduced carrying out a combination of pK_a-shift prediction using the `UHBD` program, multiple sequence alignments and MD simulations. In the simulations carried out, the choice of an inappropriate state of protonation had a pronounced effect on the structural integrity of the protein structure and resulted in an overall distortion of the hydrogen bonding pattern including distortions in the backbone regions. This coincidence of calculated values with experimental findings stresses the usefulness of studying homology models of GPCRs by means of MD simulations and chemoinformatic tools. Still, the prediction of pK_a-shifts will be hampered

by the fact that the calculated pK_a-shift will strongly depend upon the exact side chain conformation that is not known in case of homology models. Besides, an interdependence might exist between the placement of water molecules and the calculated pK_a-shift.

5.6 Conformational Adaptations after Introduction of *all-trans*-Retinal

A lot of experimental and computational effort has been dedicated to elucidate the activation steps following the *cis-trans*-isomerisation of *11-cis*-retinal in bovine rhodopsin. It is known that after the absorption of a photon and the isomerisation of *11-cis*-retinal, structural changes occur in a stepwise manner resulting in the intermediates photorhodopsin, bathorhodopsin, lumirhodopsin, metarhodopsin I and metarhodopsin II (see photocycle in 1.4). Although attempted, [96] molecular dynamics simulations might not represent an appropriate tool to study the entire activation process, as already the initial *cis-trans*-isomerisation cannot be simulated using force field methods. Even if the simulation is started with a manually introduced *all-trans*-retinal, large conformational changes occurring during the transition from lumirhodopsin to metarhodopsin I and the subsequent de/protonation events are beyond a reasonable application of force field methods. An alternative approach has been described by Ishiguro *et al.* [112] who incorporated distance information known from electron spin resonance measurements. However, the simulations are carried out based on the initial assumption that TM3 must swing outwards. Structural rearrangements were manually introduced and the resulting model was minimised in order to account for the changes introduced.

In the work of Furutani *et al.* [7] it was shown that *all-trans*-retinylidene can be accommodated without large structural rearrangements. The relaxation of *all-trans*-retinylidene takes place during the transition from batho- to lumirhodopsin prior to large conformational changes and the de/protonation events mentioned. The resulting lumirhodopsin (containing a relaxed form of *all-trans*-retinylidene) could be photoconverted to rhodopsin (containing *11-cis*-retinylidene) at 77 K where molecular motions are largely impeded implying that no changes in the global structure of the protein have occurred during the flip-over of the ionone ring.

Thus, it should be possible to study at least the initial adaptations of amino acid side chain conformations starting from a *all-trans*-retinylidene chromophore using force field methods. An understanding of the initial amino acid rearrangements could potentially help in the design of antagonists or inverse agonists for other rhodopsin-like receptors that should impede these initial movements.

As described in section 4.6, the switch in amino acid side chain conformation of residue F212/5.47 would be the most pronounced, if large tethers were applied to both the manually introduced *all-trans*-chromophore and the backbone atoms. Apart from this adaptation, structural changes were strikingly small, which is conceivable as the backbone coordinates are a major determinant of the amino acid side chain conformation. This fact is used in programs such as SCWRL that incorporate backbone information in order to determine favourable side chain conformations. If the backbone atoms are frozen due to large tethers applied, also the switch of amino acid side chains will be significantly hampered. Phenylalanine 212/5.47 is a highly conserved residue that occurs in 70% of the sequences analysed by Baldwin *et al.* [80] The corresponding residue Y205/5.47 in the neurokinin-1 receptor has been proposed to play an essential role in the activation process. [113] In the rat 5-HT_{2A} receptor the F243/5.47A mutation reduced agonist stimulated phosphoinositide production [95] and in the human OPRX receptor, mutation of F224/5.47 to alanine resulted in a practically inactive mutant receptor. [114] These mutation studies (see GPCR-DB for a complete survey on mutated receptors [115]) together with the results from the MD simulation suggest that the conformational switch of F212/5.47 could represent one of the initial activation steps. This is in good agreement with numerous antagonist binding pockets suggested for rhodopsin-like GPCRs where the ligand extends from helix 3 towards helix 5 and extends into the gap between helix 5 and 6. This ligand placement could thus lock residue 5.47 in its original position and prevent a subsequent activation. In case of agonist binding, according to this model the gap between helix 5 and 6 should be closed by the alternative conformation of residue 5.47.

Apart from the flip-over of F212/5.47 also a structural adaptation of P171/4.60 was required in the MD simulation with backbone constraints. Deformation of the C167/4.56 — P171/4.60 region was as well reported by Ishiguro *et al.* [116] who simulated a similar setup in vacuum with tethers on the Cα atoms. Ishiguro *et al.* suggested a rearrangement of TM3 and TM4 in order to accommodate the cyclohexenyl ring. In analogy to the MD simulation carried out by Ishiguro [112], a rearrangement of helix 4 was observed in simulations carried out in a CCl_4/water environment without backbone tethers, whereas no outward swing of helix 3 could be observed (see figure 4.19). The most pronounced initial adaptation was again the change in side chain conformation of residue F212/5.47. Removal of backbone tethers now allowed the conformational switch of other amino acid side chains and a cascade of amino acid side chain adjustments resulted in the modification of hydrogen bond interactions between helices 4 and 5 as well as helices 3 and 5 (see figure 4.19b).

5.7 Deriving Interhelical Contacts as Potential Constraints in GPCR Simulations

Flohil *et al.* [117] have discussed that unconstrained molecular dynamics might not be an appropriate method for homology model refinement and they proposed to constrain those parts of the model that have a high likelihood to be modelled correctly. Yet, this strategy might be problematic when applied to GPCR homology models as there is evidence that some helices are devoid of kinks present in the template structure. Even if individual helices were modelled correctly, the helix bundle might still change its topology, which would be impeded by position restraints. Thus, an alternative might be to restrain only interhelical contacts that occur at high frequencies during a simulation of the template protein and are either highly conserved throughout the protein family or can be mapped easily in the target protein structure.

The goal was to identify interhelical contacts during a simulation of bovine rhodopsin in order to derive potential constraints for the simulation of GPCR homology models. For this purpose, MD simulations were only analysed up to a point where significant deviations from the crystal structure occurred. Apparently, the crystal structure does not represent a native structure due to packing effects and crystallisation adjuvants. However, packing effects due to protein-protein interactions will be restricted mainly to loop regions and residues located on the protein surface. This is supported by the low structural rearrangements that had been observed in the here presented simulation and in [94]. It can be therefore expected that the interhelical hydrogen bond pattern is not drastically influenced during a crystallisation procedure which is further supported by the low structural rearrangement within the transmembrane regions reported in all simulations of bovine rhodopsin published and the persistence of hydrogen bonds observed during simulations with low structural deviations from the crystal structure. Besides, structural rearrangements due to the removal of crystallisation constraints has been reported to occur rather fast during an MD simulation. [117] For the sake of completeness it must be mentioned that in some cases where unconstrained MD simulations were used for homology model refinement, a significant improvement could be achieved. [118] The required time-scales were however quite large (10-100 ns) and the focus was set to *ab-initio* protein models that tended to structurally rearrange during the course of the simulation.

As a measure for structural preservation, the positional root mean square deviation from the experimental structure was used. As stated in [118] the exclusive use of the RMSD value as an indicator of structural changes might be too vague. Thus, similarly to [118] the hydrogen bond pattern was analysed during the course of the simulation. In order to

keep the number of hydrogen bonds easy to survey, the contacts analysed were restricted to interhelical interactions that proved to be a good measure for structural preservation. In fact, changes in the hydrogen bonding pattern can often be directly related to changes in RMSD as becomes obvious when inspecting figures 4.9 or 4.12. For deriving interhelical constraints, the simulation of model RT2 was analysed up to 4.2 ns of unconstrained simulation. At this point a contact between residue Y223/5.58 and residues of helix 6 occurred which was not present in the reference crystal structure. Tyrosine 223 has been explicitly mentioned in the work of Huber *et al*. Here, Y223/5.58 exhibited however a movement towards the lipid/water interface. This supports the statement of Flohil *et al.* that sampling of the conformational space of a protein is too slow to allow sufficient sampling at the time scale of tens of nanoseconds. If however, the amount of configuration space sampled cannot be estimated, it will be difficult to judge the significance of the output of individual MD simulations. Additionally, it cannot be proven that configurations observed during a simulation are at all of biological relevance. Given that the bovine rhodopsin structure might be present as a dimer and that there are several indications for a G-protein precoupling for other GPCR proteins, the possibility exists that simulations of a simplified isolated rhodopsin structure will lead to artificial protein conformations. Given these uncertainties, it seems justified to use the similarity to the experimental crystal structure as a criteria for defining the time period appropriate for an analysis of interhelical contacts.

In figure 4.23a important interhelical contacts are graphically displayed. Residues N55/1.50, N78/2.45, D83/2.50, W160/4.50, N302/7.49 are highly conserved within family A of GPCRs. [80] It can thus be expected that the contacts in which these residues are involved will be as well.

Part II

Molecular Modelling Study of the Human Histamine H_3-Receptor

Chapter 6

Scope

The human histamine H_3 receptor (hH_3R) is predominantly expressed in the central nervous system where it is involved in the modulation of neurotransmitter release prompting interest in this target for several important therapeutic applications. H_3R agonists are believed to be potential therapeutics for insomnia or could be antinociceptives, while inverse agonists could be used to treat obesity, narcolepsy, attention-deficit hyperactivity disorder (ADHD), schizophrenia or Alzheimer's disease. Intense research is currently ongoing in order to identify and optimise suitable drugs. Knowledge on the binding site would be of tremendous advantage in designing new drugs, yet, G-protein coupled receptors, to which the hH_3R belongs, are difficult to crystallise and no crystal structure of the hH_3R is available. However, the recent crystallisation of the bovine rhodopsin structure has revealed a suitable basis for the approach of comparative modelling. Homology models of proteins can effectively assist a drug discovery process as they offer the possibility to understand receptor ligand interaction on an atomic level and can bring new impetus in lead-finding via database-screening methods or *de-novo* design strategies.

The goal of this work is to generate a homology model of the human histamine H_3-receptor, test its specificity for known hH_3R inverse agonists and apply this model for deriving new lead structures for targeting the hH_3R by means of virtual high throughput screening. Additionally, the binding site model shall help to understand currently known structure affinity relationships at an atomic scale. Ideally, the model will further explain species differences, propose general strategies for an inverse agonist design and identify amino acids involved in ligand binding. Due to the close relatedness to the hH_4R, which has been reported to be involved in inflammation, allergies, asthma and autoimmune diseases, the conclusions drawn for the hH_3R will be analysed also in the context of the hH_4R, for which the drug discovery process is still at an early stage.

Chapter 7

Results

7.1 Generation of a Homology Model of the Human Histamine H_3-Receptor

7.1.1 Sequence Analysis Tools

For the prediction of transmembrane regions, the programs TMHMM, TopPred, DAS, SPLIT4 and TMFINDER were used (see table 10.1 in appendix). Secondary structure was predicted using PHD, PROF, PSIPRED, SAM T99, SSPRO, JPRED and PREDATOR (see table 10.2 in appendix). The various programs were evaluated on their prediction results on the sequence of bovine rhodopsin. Regarding the TM prediction programs, DAS tended to underpredict helix length while TMFINDER failed to find helix 7. The results of TMHMM, TopPred, SPLIT4 and the SWISSPROT entry were comparable. Analysis of the secondary structure prediction results revealed that PHD and PROF were least suitable to reproduce the experimental data of the bovine rhodopsin structure. JPRED and PREDATOR were only applied to the sequence of the humane histamine H_3-receptor (hH_3R), however failed to assign helical secondary structure to all transmembrane segments. Thus, only PSIPRED, SAM T99 and SSPRO gave reasonable results. Figure 7.1 plots the sequence of the hH_3R and the expected location of transmembrane regions and lengths of helical segments.

In order to analyse the relatedness between human biogen aminergic receptors, a multiple sequence alignment of 34 sequences was generated with the program CLUSTALW using an identity scoring matrix and default parameters for the gap function. The representation of the resulting dendogram was carried out using the program drawgram (see figure 7.2). Apparently, the hH_3R shows the highest similarity to the hH_4R, while similarity to the hH_1R and hH_2R receptor is significantly smaller. Different from the muscarinic receptor family, the family of histamine receptors is much more divergent and it does not seem that all

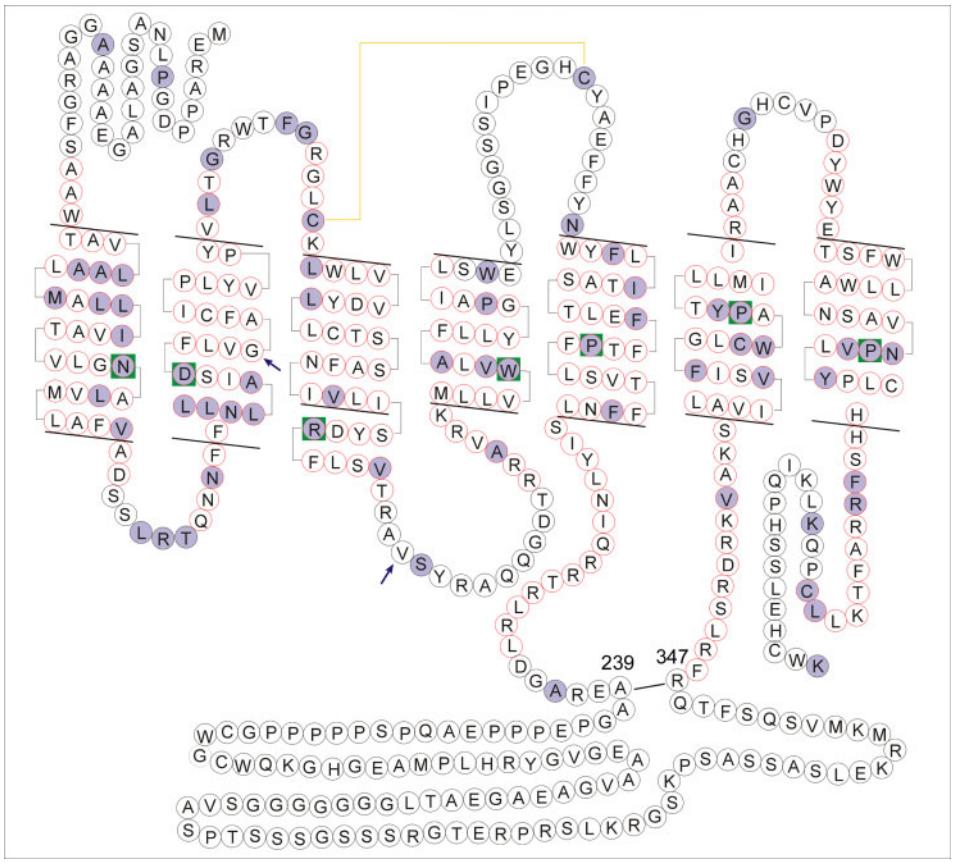


Figure 7.1: Snakeplot of the hH_3R (SWISSPROT-entry Q9Y5N1). Lengths of transmembrane segments are indicated by parallel lines, while regions predicted to have helical structure are written in red circles. The position of the conserved disulfide-linkage is indicated by a yellow line. The most conserved residue in each helix corresponding to position 50 in the Baldwin numbering scheme [80] is written in a green square. Splice points are indicated by arrows. Residues that match amino acids within the sequence of bovine rhodopsin are written in filled circles. Sequence identity within the transmembrane region amounts to 25%. The 3rd intracellular loop (stretch A240–Q346) was not included in the model.

members share a common ancestor. High sequence similarity exists to the muscarinic receptors and the $5HT_6$ receptor.

The genetic organisation of the hH_3R is quite similar to the hH_4R. The receptor is encoded by 3 exons, spanning the amino acid stretches M_1–G_{84}, A_{85}–V_{140} and S_{141}–K_{445} (splice sites are indicated by arrows in figure 7.1). In order to test the possibility that the hH_3R sequence might have evolved by rearrangement of different exons, BLAST searches with the individual amino acid stretches were carried out. High similarity of all exons to the hH_4R and several subspecies of the muscarinic receptor family was observed, however, there was no evidence of exon mixing. At the interface of exon 1 and 2, however, alternative splicing could be observed. In isoform 4 of the hH_3R sequence (SWISSPROT-entry: Q9Y5N1-4) the first 14 residues of exon 2 (AFCIPLYVPYVLTG) are missing that are

normally forming the extracellular part of helix 2. At the splice site between exon 2 and 3 no alternative isoforms are known to date.

BLAST searches were as well carried out with each helical segment taken individually. Apart from high similarities to the hH_4R, TM1-4 showed high similarity to corresponding regions in the M_1- and M_3-receptors and TM6-7 to the α_{1A}-, α_{2A}-, α_{2B}-, D_1- and D_5-receptors. Information on sequence homology to other biogen aminergic GPCRs is valuable, as it gives hints about which mutational data could be transferable as well as possible drug interactions that must be expected.

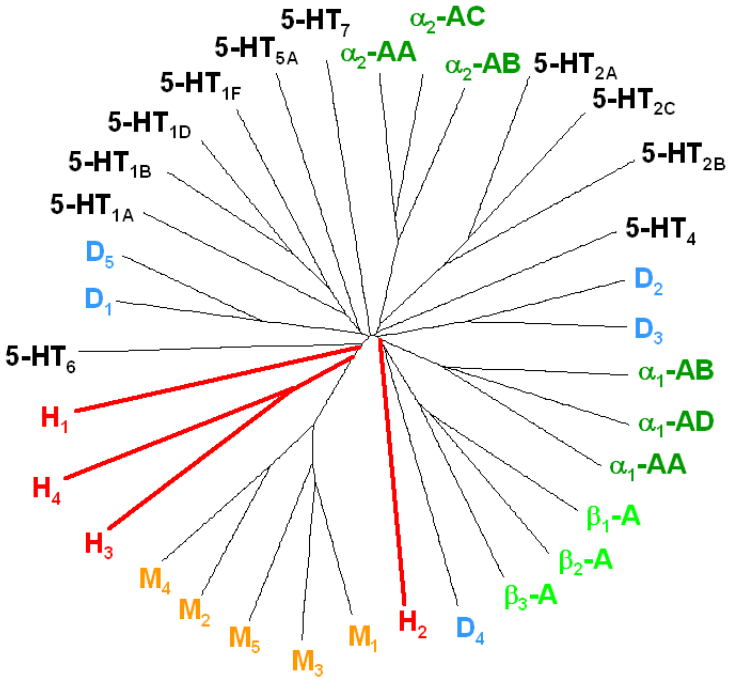


Figure 7.2: Phylogenetic tree of human biogen aminergic GPCRs. [119]

7.1.2 Sequence Structure Alignment

The sequence structure alignment between the hH_3R and bovine rhodopsin was mainly carried out based on the pinpoints identified by Baldwin *et al.*. Table 7.1 lists conserved residues in the transmembrane segments and their standard position number as given in reference [80].

Within helices 1–4 and 6–7 an undoubtable alignment was obtained already after assignment of the pinpoints. Within helix 5, that is least conserved within biogen aminergic GPCRs, the solely use of pinpoints did result in two alternatively possible alignments shown in table 7.2. In both models, three of the 5 pinpoints could be assigned. In model 1

Table 7.1: Conserved residues in the TM segments of biogen aminergic receptors as described by Baldwin *et al.* [80] The first number in each column enumerates the position, the second lists the amino acid in one-letter code, while the third column gives the percentage to which the named residue was observed within GPCR sequences analysed in reference [80].

I			II			III			IV		
21	V	76	9	N/S	88	29	I/V	94	11	W	96
18	N	99	10	L	96	26	Y	79	14	S/A	79
17	G	69	11	A	86	25	R	99	20	P	68
			13	A/S	85	24	D/E	94			
			14	D	94	22	S/A	87			
						21	I	60			
						18	L	74			
						14	S	77			

V			VI			VII			
25	I/V	72	0	K/R	75	29	R/K	79	
22	Y	91	12	F	82	28	F/Y	77	
18	I/M	79	15	C	71	21	Y	95	
14	P	85	16	W	85	18	P	98	
11	F	70	18	P	100	17	N/D	99	
						14	S/C	76	
						13	N/S	80	

the supposedly most conserved residue Y5.22 could be matched with a corresponding residue in the sequence of bovine rhodopsin. Model 2 was favoured due to the better overlap of predicted transmembrane regions. A MSA of biogen aminergic GPCRs further supported the alignment in model 2 (see table 7.3). Problems in the assignment of pinpoints within helix 5 arose due to the mutation Y5.22N in the sequences of hH_3R and hH_4R, that is — corresponding to reference [80] — the most conserved residue in TM5.

Table 7.2: Sequence structure alignment of the hH_3R sequence with the structure of bovine rhodopsin within TM5. Pinpoints that could be mapped in the sequence are coloured green, pinpoints that did not match are coloured red. Predicted transmembrane regions are indicated by plus marks. Two alternative alignments were obtained if merely pinpoints were considered for the alignment. In model 1 the most conserved residue Y5.22 could be assigned, model 2 coincided better with the predicted transmembrane segments.

Model 1	
rhodopsin (predicted TM)	++++++++++++++++++++++
rhodopsin (195–227)	HEETNNESFVIYMFVVHFIIPLIVIFFCYGQLV
hH_3R (194–226)	YNWYFLITASTLEFFTPFLSVTFFNLSIYLNIQ
hH_3R (predicted TM)	++++++++++++++++++++++++
Model 2	
rhodopsin (predicted TM)	++++++++++++++++++++++
rhodopsin (197–227)	NNESFVIYMFVVHFIIPLIVIFFCYGQLV
hH_3R (194–222)	YNWYFLITASTLEFFTPFLSVTFFNLSIY
hH_3R (predicted TM)	++++++++++++++++++++++++

The resulting alignment between the hH_3R (SWISSPROT-entry Q9Y5N1) and bovine rhodopsin is shown in table 7.4. In this alignment, the third intracellular loop had already been truncated as indicated in figure 7.1. Truncation of this loop to a comparable length as observed in the template structure was carried out as no structural information existed for this region. The I_3 loop is involved in G-protein coupling and lies distant from the binding pocket of hH_3R ligands on which this work focused. Additionally, isoforms with a significantly shorter 3^{rd} intracellular loop exist (e.g. isoform 2 described in [120] lacking 80 amino acids) that can still bind hH_3R inverse agonists — although with modified potency. In table 7.4 residues for which the backbone coordinates of bovine rhodopsin were directly adopted are held in capital letters, residues for which a loop search [83] had to be carried out are written in lower case characters.

7.1.3 Placing Amino Acid Side Chains

Amino acid side chains were added to the modelled hH_3R backbone using the program SCWRL3.0. [67] The suitability of this program to find reasonable conformations was again

Table 7.3: MSA of representatives of biogen aminergic GPCRs within helix 5. Matching pinpoints are coloured green, non-matching red. Conservation of pinpoints in the sequences of the hH_3R and hH_4R appears to be very low.

M_4	TFGTAIAAFYLPVVIMTVLYIHI
α_1-AA	VLFSALGSFYLPLAIILVMYCRV
α_2-AA	VISSCIGSFFAPCLIMILVYVRI
β_1-A	AIASSVVSFYVPLCIMAFVYLRV
D_2	SS IVSFYVPFIVTLLVYIKI
D_4	VVYSSVCSFFLPCPLMLLLYWAT
5-HT_{1A}	TIYSTFGAFYIPLLLMLVLYGRI
5-HT_7	TIYSTAVAFYIPMSVMLFMYYQI
5-HT_{2C}	VLIGSFVAFFIPLTIMVITYCLT
5-HT_4	AITCSVVAFYIPFLLMVLAYYRI
5-HT_6	FVLVASGLTFFLPSGAICFTYCRI
H_1	FKVMTAIINFYLPTLLMLWFYAKI
H_2	LVDGLVTFYLPLLIMCITYYRI
H_3	FLITASTLEFFTPFLSVTFFNLSI
H_4	ILAITSFLEFVIPVILVAYFNMNI
rhodopsin	FVIYMFVVHFIIPLIVIFFCYGQL

validated on the structure of bovine rhodopsin. As already discussed in section 5.3, residues were predicted correctly to a large extent. While adding side chains to the hH_3R model, residues, that matched corresponding residues in bovine rhodopsin, were directly adopted in their conformation. The SCWRL algorithm was applied individually on each helical segment as well as the entire model. This way, suboptimal placements due to steric restraints were more easily detected. At two sites steric clashes could not be resolved by SCWRL:

- site 1: localised between helices 3 and 4 involving residues Y115/3.33 and Y167/4.57.
- site 2: located between helices 2, 3 and 7 involving residues Y91/2.61, W110/3.28, W292/7.40 and W295/7.43.

Regarding cluster 1, for residue Y167/4.57 a conformation was suggested, that produced a clash with backbone atoms of Y115/3.33. Only one alternative conformation taken from the Insight rotamer library [83] did not result in a clash with backbone atoms of helix 3 and was pointing outside of the binding pocket into the gap between helices 3 and 4. In order to decide which conformation was more compatible with a structural preservation of the model, MD simulations in a CCl_4/water environment were carried out, starting with alternative side chain conformations of Y167/4.57. Simulations were carried out with

Table 7.4: Alignment of hH_3R sequence (H3) with bovine rhodopsin (BR). The position of predicted transmembrane segments is indicated by plus marks. Pinpoints are depicted in green in case of a positive match, red if the residue did not coincide with the conserved pattern. Residues for which the backbone coordinates of bovine rhodopsin were directly adopted are held in capital letters, residues for which a loop search was carried out are written in lower case characters.

```
BR                                        ++++++++++++++++++++++++
BR  MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
H3    MERAPPDGPLNASGALAGEAAAAGGARGFSAAWTAVLAALMALLIVATVLGNALVML
H3                                        ++++++++++++++++++++

BR  +++          ++++++++++++++++++++++             ++++++++++
BR  VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
H3  AFVADSSLRTQNNFFLLNLAISDFLVGAFCIPLYVPYVLTGRWTFGRGLCKLWLVVDYLL
H3               ++++++++++++++++++++++              +++++++++

BR  ++++++++++++                      ++++++++++++++++++++
BR  GEIALWSLVVLAIERYVVVCKPMS  NFRFGENHAIMGVAFTWVMALACAAPPLVGWSRY
H3  CTSSAFNIVLISYDRFLSVTRavsyraqqgDTRRAVRKMLLVWVLAFLLYGPAILSWeyl
H3  ++++++++++                             ++++++++++++++++++

BR                          ++++++++++++++++++++++
BR  IPEG  MQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQ
H3  sggssipeGHCYaeff      ynWYFLITASTLEFFTPFLSVTFFNLSIYLniqrrtrl
H3                          ++++++++++++++++++++++

BR                      ++++++++++++++++++++             ++++++
BR  QQESA    TTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPA
H3  rldgarearFRLSRDRKVAKSLAVIVSIFGLCWAPYTLLMIIRAACHGHCVPDYWYETSF
H3                      ++++++++++++++++++++               +++

BR  ++++++++++++++++
BR  FFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
H3  WLLWANSAVNPVLYPLCHHSFRRAFTKLLCPQKLKIQPHSSLEHCWK
H3  ++++++++++++++++++
```

a trimethylammonium/hH_3R complex in order to neutralise the negative charge of the conserved aspartic acid D114/3.32. The simulations revealed that independent from the initial placement of Y167, TM4 moved outwards the helix bundle for about 2.5Å during the simulation. In case of Y167 pointing into the binding site (see figure 7.3(b)), hydrogen bonds were observed between TM3 and TM4 (T119–Y167, 62%), TM3 and TM5 (T119–E206, 47%) and TM4 and TM5 (Y167–E206, 38%) at the frequencies indicated in brackets.

When starting with the alternative conformation (Y167 pointing outside of the binding site), the carboxylate group of E206/5.46 was not involved in any interaction after 1 ns of simulation period. The conformation of E206 additionally varied noticeably from the conformation predicted by SCWRL and caused distortions of the helical geometry of TM5. In this simulation, residues from TM3 and TM4 were not involved in hydrogen bonds, between TM3 and TM5 the contact T119–L205 was observed frequently (86%) and between TM4 and TM5 the interaction L166–T201 took place in 78% of all frames analysed.

The placement of Y167/4.57 towards the inside of the binding site was also supported by mutational data of the M_1-, M_2- and H_4-receptors. [52, 121, 122] In the M_1- and M_2-receptors the corresponding tryptophane residues were involved in ligand binding, while in the H_4R the corresponding asparagine was reported to be involved in receptor activation.

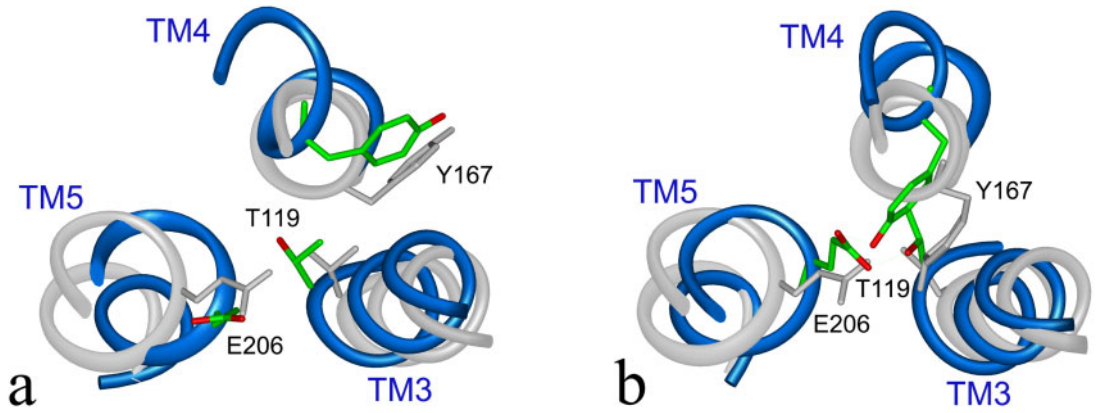


Figure 7.3: Possible orientations of residue Y167/4.57 in the hH_3R model. **(a)** Y167 is pointing outside of the binding site. An MD simulation with this start conformation (light grey) resulted in an outward shift of TM4 (blue). The resulting conformation for residue E206/5.46 differed significantly from the predicted conformation by SCWRL and the carboxyl moiety was not involved in any hydrogen bond interactions. **(b)** Y167 is pointing into the binding site. In the start conformation (light grey) the aromatic ring of Y167 clashes with the backbone of helix 3 resulting in pronounced distortions of the aromatic system during a minimisation procedure. After 1 ns of MD simulation, TM4 had shifted outwards and a hydrogen bond network had evolved between residues T119/3.37–Y167/4.57 and E206/5.46.

Regarding cluster 2, the side chain placement of residue W295/7.43 was problematic as this residue — in analogy to K296/7.43 in bovine rhodopsin that is involved in the Schiff-base linkage — was pointing towards D114/3.32 in helix 3. D114 is highly conserved within the family of biogen aminergic receptors and participates in ligand binding by forming a salt bridge interaction with protonated species. The suggested SCWRL placement blocked however the assumed binding pocket as can be seen in figure 7.4(a). Additionally, several clashes were observed for residues Y91/2.61, W110/3.28 and W292/7.40. In order to check whether structural adaptations would take place during an unconstrained MD simulation resulting in a ligand-compatible binding site, the original SCWRL side chain placement was used as a start conformation for a simulation in a CCl_4/water environment. As described above, a truncated ligand (trimethylammonium) was considered in order to neutralise the negative charge of D114/3.32. Figure 7.4(a) shows the start conformation (SCWRL output) where steric clashes are still present and cause deviations from ring planarity; figures 7.4(b) and (c) show two alternative end conformations resulting from the same setup (a). Within the backbone of helix 7, structural adaptations resulted in a more helix-like secondary structure compared to the backbone region adopted from bovine rhodopsin. However, in both setups W295/7.43 remained blocking the binding pocket for a subsequent ligand docking procedure.

Apparently, the two MD simulations resulted in different binding pocket conformations, although the same start conformation had been used. This is however comprehensible as velocities are randomly generated for all atoms at the start of a simulation. If the system has a low intrinsic stability (and thus a high corresponding potential energy) due to a suboptimal backbone conformation or wrong side chain placement, the initial velocity distribution can result in systems evolving in quite different orientations. Ideally, “all” conformations would be sampled if just the simulation period was long enough. In practice, it is more probable that a simulation will sample only the valley of potential energy that was reached already after a short MD simulation. Sampling from one “valley” to the next might be hampered due to large energetic barriers. Thus, in order to sufficiently sample the energetic landscape of a protein structure, several short lasting simulations (or a simulated annealing protocol) might result in more divergent conformations than one long lasting simulation.

Again, mutational data was analysed from other biogen aminergic GPCRs in order to infer which amino acids are likely participating in ligand binding and must therefore be oriented towards the binding site. Residues Y91/2.61 and W110/3.28 in the hH_3R most likely participate in an aromatic cage stabilising the positive charge of biogen amines as similarly observed for dopaminergic receptors. [123] Further residues of helix 7 (F291/7.39, W292/7.40 and W295/7.43) could contribute to this aromatic cage. Corresponding residues were mutated in several biogen aminergic GPCRs, such as the (rat M_3- (Y7.39,

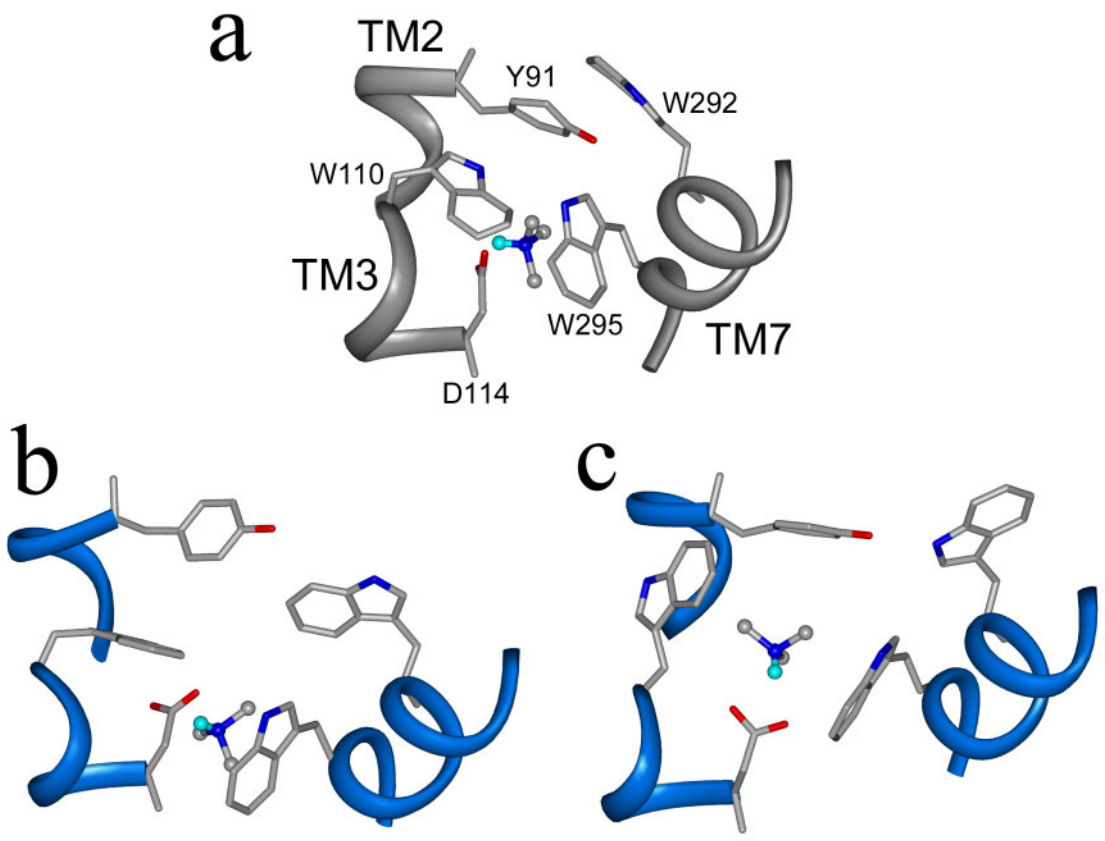


Figure 7.4: Structural adaptations around site 2 after an unconstrained MD simulation in a CCl_4/water environment. **(a)** Start conformation as obtained from `SCWRL`. **(b,c)** Conformations obtained during unconstrained MD simulations in a CCl_4/water environment starting from structure **(a)**.

W7.40, Y7.43), rat M_1- (Y7.39, W7.40, Y7.43), human $5HT_{1B}$- (7.39), rat α_1-AA- (F7.39), human D_3- (T7.39) and β_2-receptor (N7.39); see references [115, 124] for a complete list of mutational data. Interpretation of the mutational experiments was however difficult as the effect of the mutations depended on the point mutation carried out and the ant/agonists tested on the mutant receptor. Besides, point mutations may have a pronounced effect on ligand affinity although lying distant from the binding site simply by causing distortions of the binding site geometry.

In conclusion, addition of amino acid side chains with the program `SCWRL3.0`, subsequent MD simulation of alternative receptor models varying in the placement of selected amino acid side chains in a low-viscosity membrane mimic, and analysis of GPCR mutation data could favour a placement of residue Y167/4.57. In case of the aromatic cluster around residue W295/7.43 the same procedure did however not result in a clear preference of one amino acid side chain rotamer. Further constraints were thus required, e.g. in form of ligand information.

7.1.4 Simultaneous Side Chain and Ligand Placement

Although the MD simulations and analysis of mutational data described above could favour a placement of Y167, no definite placement of residues forming the aromatic cage could be derived. In order to find a reasonable conformation of the binding site compatible with the accommodation of a high affinity hH_3R ligand, a docking procedure of ligand FUB836 (compound 34 in [43], see figure 7.5) was carried out where several rotamers were considered for selected residues participating in the binding site. Distinct from a "normal" docking approach, where a binding pocket conformation is used as a filter in screening structure databases for compatible ligands, here, a ligand was used for retrieving the most suitable binding site. FUB836 is a high affinity ligand with a pK_i value of 10.04. Two conclusions should thus be allowed:

- The binding conformation of FUB836 must be quite close to the energetic minimum structure.
- The ligand fits well into the binding site, thus a large docking score is to be expected.

Using this approach of "inverse" docking, a binding site was searched for, which was capable of accommodating the high affinity ligand FUB836 in an energetically favourable conformation simultaneously producing a high docking score. The approximate position of FUB836 in the hH_3R binding pocket was known from mutational studies that had identified D114/3.32 and E206/5.46 as main interaction partners for histamine binding. [27] These residues have also been shown to be important in binding of inverse agonists. [28, 125] Thus, residues located in proximity to this putative binding pocket were selected. Table 7.5 lists residues for which distinct rotamers were considered and residues that were included in only one conformation in order to further border the approximate position of the binding pocket. Consideration of side chain flexibility is a combinatorial problem, thus it was not possible to consider different rotamers for all residues involved in the binding site as already the amount of flexibility considered resulted in 43740 alternative binding sites. Subsequently, residue FUB836 was docked into the binding pockets using the program `GOLD`. [126] As it is known that the protonated piperidyl moiety is in contact with D114/3.32, the distance between the piperidyl-nitrogen and the Cγ-atom of D114 was restrained to a value below five Angstroms. The resulting docking solutions were ranked corresponding to the `GOLD` scores obtained. As can be seen in figure 7.5, only a small fraction of docking solutions produced a high score. The corresponding binding pockets as well as the conformations of FUB836 were then further evaluated.

In the highest ranked complex, the piperidyl-nitrogen was in proximity to D114/3.32 (as required by the constraint) and the exocyclic nitrogen of the 4-aminoquinoline moiety was

Table 7.5: Residues included in the approach of “inverse” docking. Several conformations were considered for residues on the left, while residues on the right were only included to border the putative binding pocket. Conformations for W264/6.48 and Y267/6.51 were adopted in a conformation comparable to that observed in the structure of bovine rhodopsin.

flexible	no. (rotamers)	fixed
Y91/2.61	3	TM1: -
W110/3.28	3	TM2: V83/2.53, C87/2.57, I88/2.58
D114/3.32	3	TM3: L111/3.29, L117/3.35
Y115/3.33	3	TM4: -
C118/3.36	2	TM5: L199/5.39, A202/5.42, S203/5.43, F207/5.47
T119/3.37	3	TM6: W264/6.48, Y267/6.51, M271/6.55, I272/6.56
Y167/4.57	2	TM7: F291/7.39, W292/7.40, L294/7.42, S298/7.46
E206/5.46	3	E2: A190/5.30
T268/6.52	3	
W295/7.43	5	

in contact with E206/5.46. The sterically demanding quinoline system extended into the gap between helices 3, 4 and 5. For Y91/2.61, W110/3.28, Y167/4.57, F291/7.39 and W292/7.40 conformations pointing into the binding site were predicted. Also Y167/4.57 was predicted to point into the binding pocket consistent with the previously reported analysis. Most interestingly, W295/7.43 was predicted to point into the gap between helices 1 and 7. For all high scored ligands the spacer moiety showed an extended conformation implying that the passage between helix 3, 6 and 7 was sterically constrained. A critical residue in this respect is L294/7.42 in the sequence of the hH_3R which aligns to glycine in most other biogen aminergic GPCRs and a glutamine residue in the hH_4R.

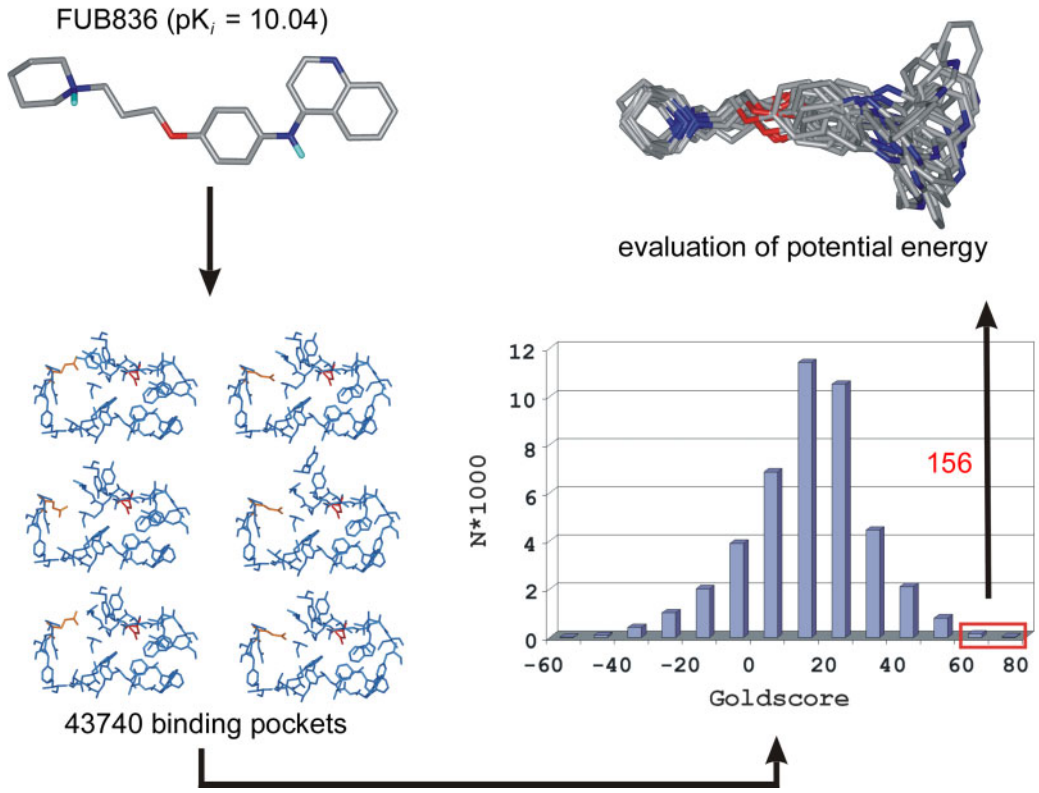


Figure 7.5: Approach of “inverse” docking as described in the text. Ligand FUB836 was used to screen alternative binding pockets. Complexes that were assigned a high docking score were further evaluated in respect to the potential energy of the ligand conformation. The superposition of top ranked solutions revealed that the position of the piperidyl moiety was quite stringent; additionally, an extended conformation for the spacer region seemed to be required for a high score. The position of the quinoline moiety, on the other hand, seemed to be more flexible.

7.2 MD Simulations of hH$_3$R Models

7.2.1 MD Simulations of Uncomplexed hH$_3$R Models

An hH$_3$R Model Based upon the Start Conformation Obtained by the Inverse Docking Procedure

In order to test the structural stability of a generated hH$_3$R model, reduce steric constraints and find energetically favourable conformations, the receptor model was studied by means of MD simulation. As unconstrained MD simulations for homology model refinement had been assessed critically in recent publications, [117] interhelical hydrogen bond contacts between well conserved residues were included in the simulation as constraints. For this purpose, a simulation of the template structure, bovine rhodopsin, had been previously analysed for these interhelical contacts (see section 4.7). Several well conserved residues

are present in the amino acid sequences of family A GPCRs and represent the basis for the generation of homology models of this class. These residues often play an important role in receptor activation via participating in the stabilisation of the off-state receptor conformation. The incorporation of constraints between well conserved residues seems thus reasonable. Figure 7.6 shows conserved contacts in the hH_3R model, table 7.6 additionally lists the contacts as included in the MD simulations. The force constant for the distance constraints was set to 2500 $kJmol^{-1}nm^{-2}$. Values defining the distance restraint potential were selected in a way, that an interference into the MD simulations would only occur in situations where a loss of conserved interactions was to expect, otherwise the simulation was uninfluenced (see table 7.6).

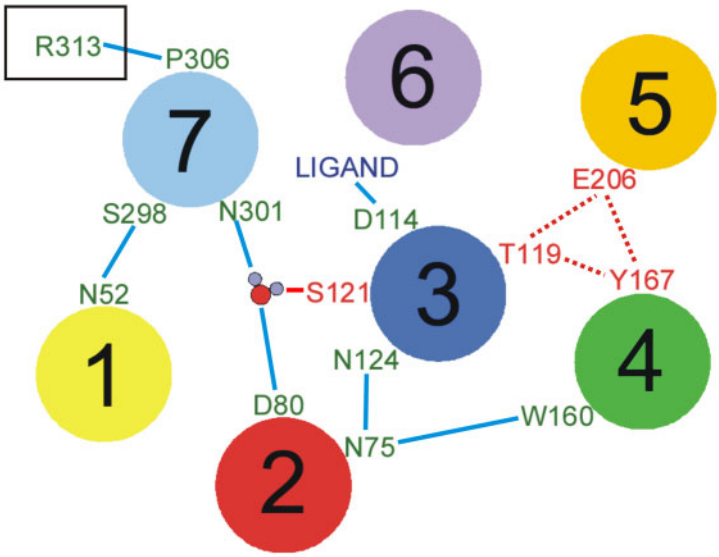


Figure 7.6: Conserved interhelical contacts in the hH_3R model. Contacts coloured in blue are comparable to those observed in the structure of BR (compare to figure 4.23). Contacts coloured in red are not observed between corresponding residues in BR, however would result in the same number of interhelical contacts.

Simulations of uncomplexed hH_3R models were carried out in a DPPC/water environment. The simulations were set up in a similar way as described for bovine rhodopsin. The receptor was merged with the DPPC/water box, into which a cylindrical hole had been inserted (see section 4.1.3). The initial equilibration phase of 1 ns was carried out using the `ffgmx` force field including the option pressure coupling in order to allow for the required adjustment of the box size. During this MD simulation, the entire hH_3R model was tethered with a force constant of 5000 $kJmol^{-1}nm^{-2}$ in order to avoid structural changes. Then, the system was transferred to the `ffG43a1` force field and further equilibrated. The highly conserved residue D80/2.50 was — in analogy to the simulation of bovine rhodopsin — considered in its protonated state. Similarly, one water molecule was included in the simulations, that cross linked helices 2, 3 and 7 (see figure 7.6) via interaction with the

Table 7.6: Interhelical contacts included as constraints in simulations of hH_3R models. Each constraint is defined by three values: a lower bound (`low`), an upper bound 1 (`up1`) and an upper bound 2 (`up2`) defining the distance restraint potential. If the distance between the pair of atoms, that shall be restrained, drops to a value below `low`, a harmonic potential will be set. Between `low` and `up1` no restraint is set, from `up1` to `up2` a quadratic potential is applied and beyond the largest bound `up2` the potential form is linear.

interactions			low [nm]	up1 [nm]	up2 [nm]
ASN52:ND2	—	SER298:O	0.250	0.280	0.290
ASN75:ND2	—	ASN124:OD1	0.243	0.273	0.283
ASN75:OD1	—	TRP:160:HE1	0.185	0.205	0.215
LEU307:O	—	ARG313:NH1	0.290	0.300	0.320
ASP83:OD2	—	H2O:OW	0.233	0.263	0.273
SER121:OG	—	H2O:HW2	0.106	0.136	0.146
ASN301:OD1	—	H2O:HW1	0.200	0.300	0.350
TYR167:HH	—	GLU206:OE2	0.240	0.300	0.500

conserved residues D2.50 and N7.49. In order to test if the protein model remained stable during an unconstrained MD simulation, the uncomplexed hH_3R model was simulated for 5 ns. The simulation protocol included — as already described for bovine rhodopsin — a stepwise reduction of tether forces in 100 ps time scales from 1000 to 500 to 200 to 100 $kJmol^{-1}nm^{-2}$. This proceeding had been reported to yield superior results, as initial bad contacts (e.g. between side chains or side chains and the model backbone) would not result in distortions of the backbone region but rather disappear during the initial equilibration period due to rearrangements of amino acid side chains. However, with respect to the placement of W295/7.43 this procedure was problematic. As previously described, W295/7.43 corresponded to the lysine residue in bovine rhodopsin, that was involved in the Schiff-base linkage to retinal. During the “inverse” docking procedure this residue was placed in a way that it pointed outside of the binding pocket into the gap between helices 1 and 7. Yet, after 100 ps equilibration time (with a backbone tether of 500 $kJmol^{-1}nm^{-2}$), the initial placement — pointing into the binding pocket — was recovered. This is, due to the interdependence between backbone coordinates and side chain placement, not surprising; however, *a priori* it is not known if the pronounced distortion of the helical region within helix 7 of bovine rhodopsin is due to the covalent linkage to the retinal ligand — and would thus be expected to disappear during an MD simulation — or represents a conserved structural feature required as well in other family A GPCRs. Imposing tethers on the backbone will in any case favour a placement of amino acids comparable to the template structure and required thus further analysis.

Another pronounced side chain rearrangement concerned F207/5.47. This residue — similarly to F5.47 in the simulation of the uncomplexed or *all-trans*-retinal complexed opsin

structure — moved into the binding site. In the hH_3R, mutation of the highly conserved residue F5.47 to alanine resulted in a 17-fold drop of potency, [27] which led to the speculation that this residue might be involved in upholding the receptor structure or reflect a role in receptor activation.

Figure 7.7 shows the RMSD within the helical backbone region (red curve) and the entire protein backbone (blue curve).

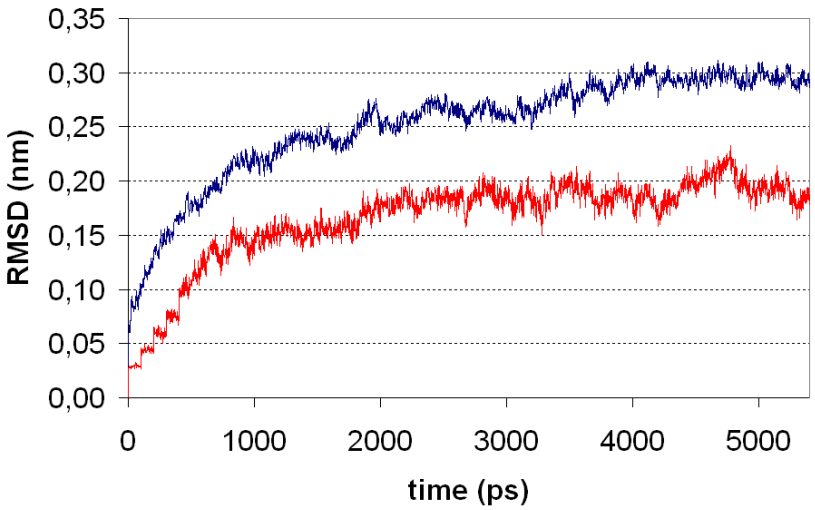


Figure 7.7: RMSD plot of a 5 ns simulation of an uncomplexed hH_3R model. The course of RMSD within the transmembrane helical region is coloured red, within the entire protein backbone blue.

Figure 7.8 shows the backbone of the modelled hH_3R overlayed on the backbone of bovine rhodopsin. Observed differences are:

- a shift of helix 1 starting at residue G1.49 resulting in an ideal helix structure. This adaptation, compared to the structure of bovine rhodopsin, was expected due to the mutation P1.49G.
- an outward shift of helix 4 as already observed in the simulation described in section 7.1.3.
- slight shifts of the intracellular parts of helices 5 and 6.
- a distortion of the extracellular part of helix 7 that seems to be a consequence of the recovered, yet questionable placement of W295/7.43.

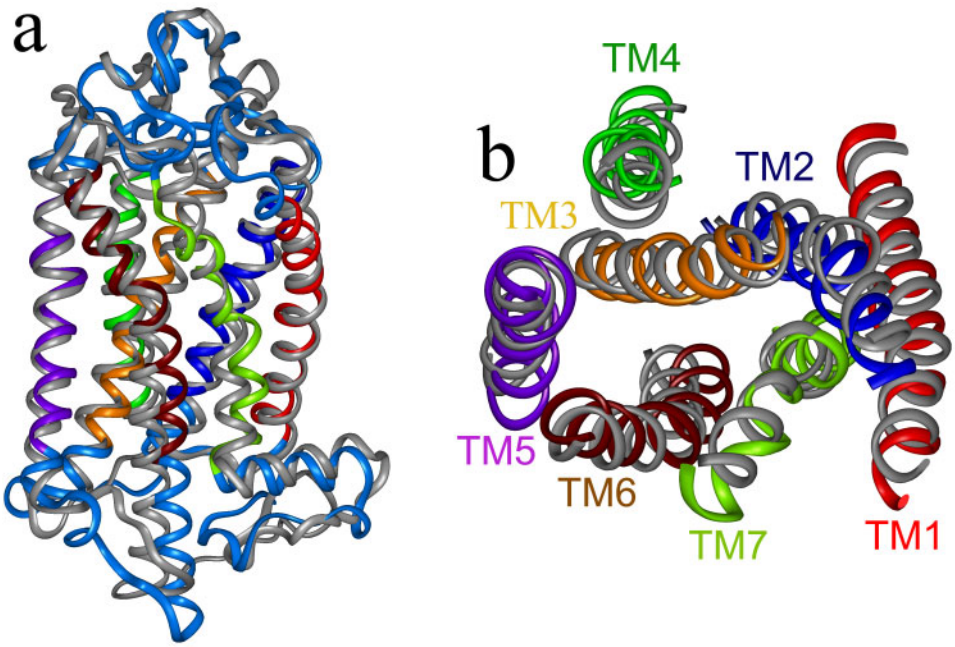


Figure 7.8: Comparison of the crystal structure of bovine rhodopsin (grey) and a model of an uncomplexed hH_3R after 5 ns unconstrained simulation in a DPPC/water environment. **(a)** side view **(b)** helix bundle viewed from the extracellular space (loops are neglected for reasons of clarity). Structural deviations are within a reasonable range except for helix 7 where pronounced distortions of the *N*-terminal region occurred.

A hH_3R Model with Modified Coordinates of Helix 7

As previously described, two simulation setups differing in the side chain rotamer of W295/7.43 were already analysed in respect of their influence on the structural preservation of the hH_3R model. In the CCl_4/water environment, W295/7.43 was placed as suggested by the program `SCWRL` pointing into the binding pocket. During the subsequent MD simulation, the kink around W295/7.43, which had been adopted from the structure of bovine rhodopsin, disappeared, and helix 7 adopted a more idealised helical conformation. As already mentioned, it is not clear, whether the kink around W295/7.43 represents a unique characteristics of the rhodopsin structure or a generalised structural feature of inactive GPCR structures. Yet, in this conformation W295/7.43 blocked the binding site for ligand docking. Within the approach of “inverse” docking a rotamer was suggested for W295/7.43 that pointed into the gap between helices 1 and 7. Yet, during the equilibration period, the original `SCWRL` placement was recuperated. After a 5 ns simulation in a DPPC/water environment, a good overall structural preservation was obtained, however the *N*-terminal part of helix 7 was distorted.

In order to analyse whether the idealised helix conformation obtained in the short (1 ns) simulation in a CCl_4/water environment would be stable during a longer simulation in

a DPPC/water environment, a helical geometry for helix 7 devoid of the characteristic kink was introduced and used as a start conformation. W295/7.43 was pointing into the binding site and the goal was to analyse if such a placement in combination with a helical geometry would result in a better structural stability. Although this setup had resulted in a blocked binding site in the CCl_4/water environment, inverse agonists could provoke a conformational change of this region. This is supported by known SAR of hH_3R agonists and inverse agonists: the binding site in proximity to D114/3.32 for agonist binding is sterically much more demanding than the binding pocket for inverse agonists. If inverse agonists triggered this adaptation, simulation of the uncomplexed receptor model would result in a blocked binding site. However, the goal of simulating the uncomplexed models was to find possible structurally stable conformations.

Thus, the trimethylammonium complexed hH_3R model was simulated for 5 ns in a DPPC/water environment as previously described. In figure 7.9a, the RMSD values obtained during this simulation (blue curve) and the simulation that resulted in deviations of helix 7 (red curve) are compared. As depicted in figure 7.9b, the hH_3-receptor model with modified start conformation of helix 7 remained stable during the entire simulation.

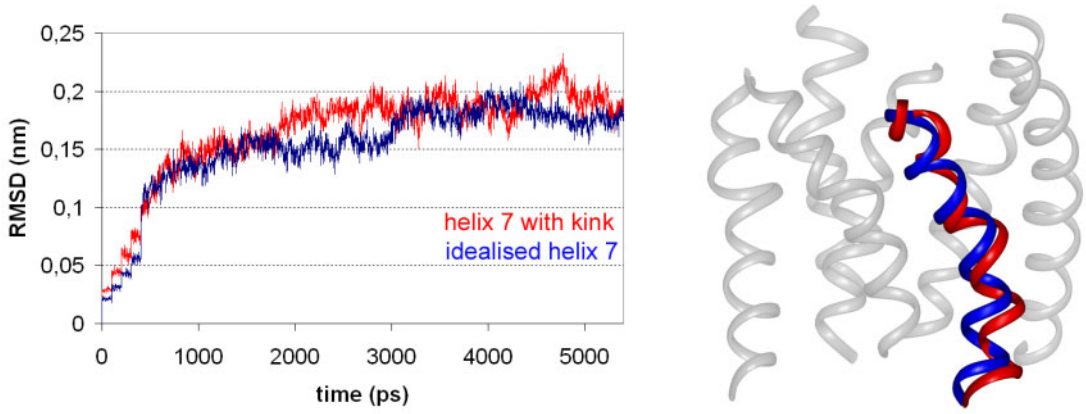


Figure 7.9: Comparison of RMSD and helix 7 from MD simulations of uncomplexed hH_3R models starting with different conformations of helix 7. **(a)** Course of RMSD within the transmembrane backbone region. The red curve corresponds to the calculation described under point 1. This setup resulted in distortions of secondary structure of helix 7 (see **(b)**, red helix). The blue curve describes the RMSD during an MD simulation starting from a conformation where helix 7 adopted backbone coordinates slightly varying from the template structure bovine rhodopsin around BR:K296. This setup resulted in a better preservation of the helical structure of TM7 as depicted in **(b)**, blue helix.

When comparing the obtained binding site geometries (see figure 7.10), it is apparent that the resulting placements of W295 are comparable towards the end of both simulations. Thus, if W295/7.43 pointed into the gap between helices 1 and 7, flipped however into the binding pocket during the equilibration period, distortions of the *N*-terminal region of helix 7 resulted. If a rotamer of W295 was chosen that pointed into the binding pocket,

this conformation was consistent with structural preservation if the kink present in the structure of bovine rhodopsin had been previously eliminated. Both start conformations resulted however in quite similar conformations for W295/7.43, with W295/7.43 pointing towards D114/3.32 thereby blocking the binding site for subsequent ligand docking. It can be speculated that upon binding of inverse agonists, W295/7.43 could be forced into a different side chain rotamer or modifications of the helical backbone region of helix 7 could result in an opening of the binding site. As mentioned before, this is supported by the observation that the agonist binding site is sterically much more demanding than the site for inverse agonists. Structural modifications close to helix 7 thus quite likely represent an essential mechanism in receptor activation.

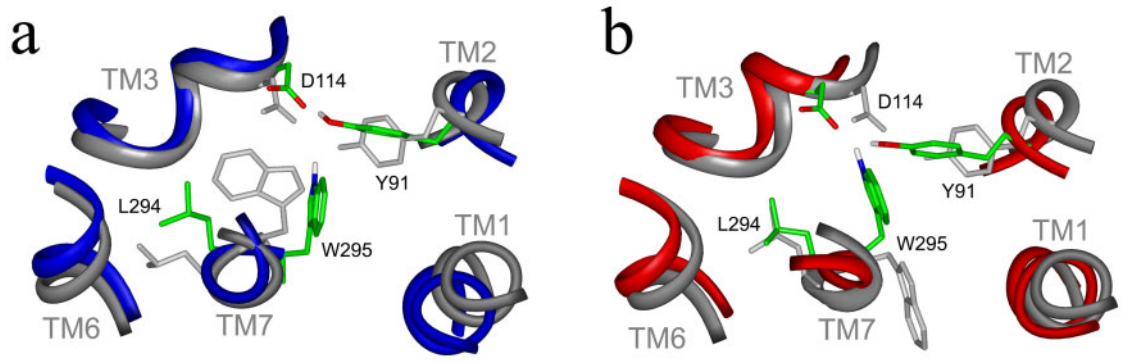


Figure 7.10: MD simulations of uncomplexed hH_3R models: influence of different start conformations of helix 7 on the resulting binding site conformation. **(a)** Slight modifications of the backbone of helix 7, resulted in a start conformation where W295/7.43 was located in the binding pocket bordering the aromatic cage towards the extracellular region (grey). In this case, the binding site is not blocked by W295/7.43, but a potential ligand would be located under the plane of the aromatic system. During a 5 ns simulation, this initial placement of W295/7.43 adopted however a different rotamer (colour coded). **(b)** Simulation starting from the binding site geometry as obtained by the “inverse” docking approach. During the equilibration period, W295/7.43 moved inside the binding site accompanied by distortions of the *N*-terminal region of helix 7. Interestingly, the two end conformations strongly resembled each other.

7.2.2 MD Simulation of Inverse Agonist/hH_3R Complexes

The observation that the available free volume in the binding site of agonists is significantly reduced compared to that of the inverse agonists urges to analyse inverse agonist/receptor complexes. The incorporation of ligands in the binding site imposes further constraints on the placement of amino acid side chains thereby reducing the degrees of freedom of amino acid side chains and guiding the simulation to ligand compatible conformations. For this purpose, the ligands ideally only contain a small number of rotable bonds and are sterically demanding in order to fill the assumed binding pocket. Several ligands (FUB833,

FUB836, thioperamide, GT2331, ciproxifan, clobenpropit, VUF5391, VUF5228, SCHpat, ABT-239, A-923, compound A4 from [127], compound 33 from [40], JNJ-5207852) were parameterised for the `ffG43a1` force field and were subsequently used in simulations of hH_3R complexes. It can be expected that the output of such simulations will strongly depend on the initial placement of both the amino acid side chains and the ligand as the free volume required for conformational changes is significantly reduced in ligand occupied binding pockets. It should however be possible to evaluate if the conformation of the complex is consistent with secondary structure preservation and the overall stability of the protein model. In the simulations of uncomplexed receptor models previously described, a good overall structural stability had been observed. Additionally, highly conserved residues such as W264/6.48 and Y267/6.51 adopted identical conformations as those in bovine rhodopsin indicating that the amino acid side chain placement was consistent with structural preservation and may therefore be used for a ligand placement.

In order to start MD simulations of complexed hH_3R models with a reasonable ligand conformation and placement, firstly, conformational analysis of hH_3R inverse agonists were carried out in order to find potential bioactive conformations and possible ligand placements were assessed.

Generation of Inverse Agonist/hH_3R Complexes

Agonists of biogen aminergic GPCRs are known to interact with a conserved aspartic acid in helix 3 (in the hH_3R: D114/3.32). Mutational studies of the hH_3R have shown that this conserved residue also participates in inverse agonist binding. [28, 125] Thus, a hydrogen bond reinforced salt bridge between the carboxylic group of D3.32 and the ligand could be assumed in all complexes generated. In the case of the hH_3R, E206/5.46 has been identified as another interaction partner in inverse/agonist binding. [28]
Some of the known hH_3R ligands are spatially quite demanding. For FUB836 *a priori*, two possible placements of the sterically demanding piperidyl-side chain are thinkable depicted in figure 7.11. The existence of branched hH_3R ligands such as SCHpat or VUF5228 (see figure 1.10) shows that a simultaneous occupancy of both subpockets is possible.

Conformational Analysis of hH_3R Inverse Agonists

Conformational analysis were carried out using the program `MOE` and the `MMFF94` force field including a solvation term. In order to carry out a systematic search, ligands with only few rotable bonds were selected. Conformational analysis is exemplarily described for a compound of Johnson&Johnson presented at the 33^{rd} Annual Meeting of the Histamine Research Society referred to as JNJ1 (pK_i = 8.8 on hH_3R; see figure 7.12) and GT2331

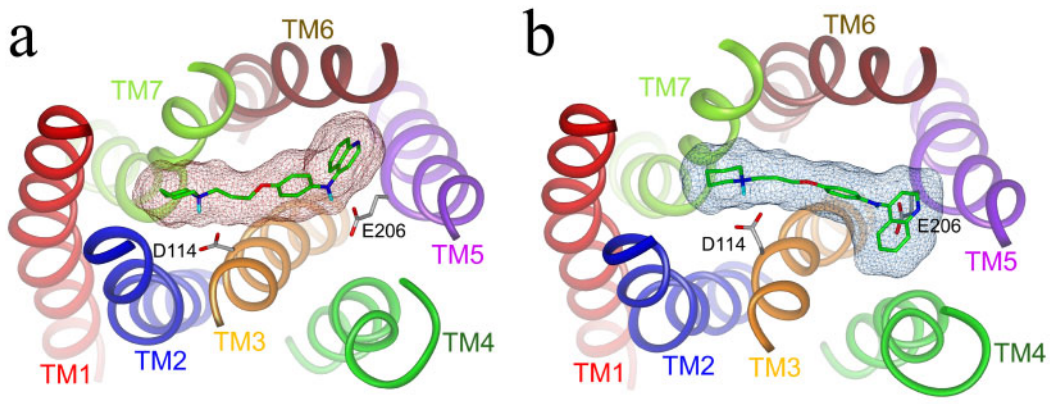


Figure 7.11: Two alternative placements of FUB836 in the hH_3R binding site. **(a)** The quinoline moiety lies in a pocket between helices 5 and 6. **(b)** The quinoline moiety lies in a pocket formed by helices 3, 4 and 5.

(pK_i = 9.6 for the (1S,2S) enantiomer on rat H_3R). [32] In the case of JNJ1, the $\alpha 1$ and $\alpha 2$ bond as well as the $\gamma 1$ and $\gamma 2$ bond were rotated in increments of 15° and the resulting conformations (within an energy cutoff of 10 $kJmol^{-1}$ from the global minimum) were minimised. Figure 7.12 plots the potential energy of the resulting structures versus the dihedral angle. Structures corresponding to low energy conformations are indicated by letters and the corresponding conformations are depicted. In case of the $\gamma 1$ and $\gamma 2$ bond, the lowest energy conformations b and c correspond to conformations where the protonated nitrogen is pointing towards the triple bond. These rotamers are however not consistent with the free volume in the binding pocket and other partially rigid hH_3R ligands such as GT2331 (see figure 7.13). Only conformation γd could be superimposed with conformation GT2331b (see figure 7.14). Selection of dihedral angles for $\alpha 1$ and $\alpha 2$ required further knowledge of the binding site and the position of E206/5.46.

FUB836 had been used in the approach of inverse docking and suitable complexes were selected based on the conformation of this ligand. Thus, also the conformational analysis of FUB836 shall be described briefly. Torsion angles were analysed in increments of 15° and the resulting conformations were minimised. For the $\alpha 123$ linker again the energetically most favourable conformation in solution was not consistent with an extended conformation described above, as the protonated nitrogen was pointing towards the aromatic ether atom. However, an extended rotamer (C1-N-C2-C3: -75/-174/-179 °) deviated by only 2.4 $kJmol^{-1}$ from the global minimum structure (C1-N-C2-C3: -159/68/-59). For the $\gamma 1$ torsion angle, values between [-30;30]° were energetically most favourable and for the $\gamma 2$ torsion angle four minima were observed at -150, -30, 30 and 150° (results not depicted). The energetic barriers between these minima were however quite small. Data from this conformational analysis was in good agreement with the CCD-entry VOTFIT (amodiaquine hydroxide dihydrochloride) where the corresponding $\gamma 1$ and

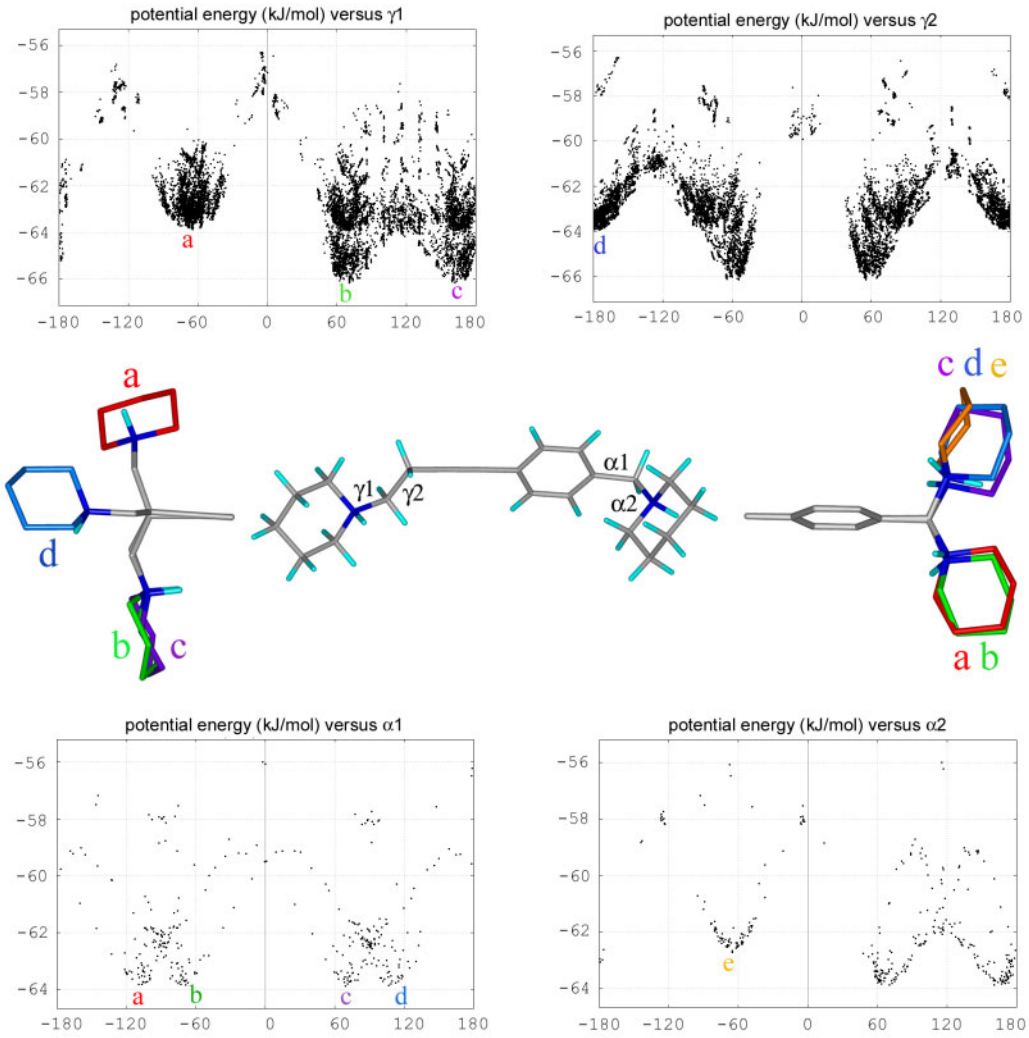


Figure 7.12: Systematic conformational analysis of compound JNJ1. Representatives of energetically favourable structures are indicated by coloured letters and correspond the equally colour coded structures.

γ2 angles were 28° and 165°, respectively.

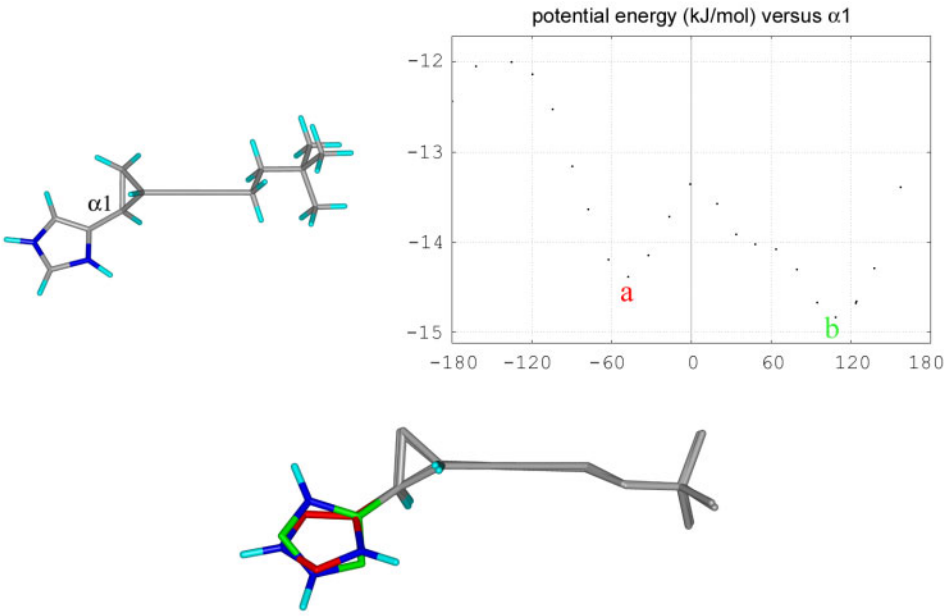


Figure 7.13: Dihedral energy plot for the α1 torsion angle of structure GT2331 analysed in increments of 15°. After energy minimisation of the resulting structures a global and a local minimum can be found indicated by coloured letters. The corresponding structures are depicted in corresponding colour codes below.

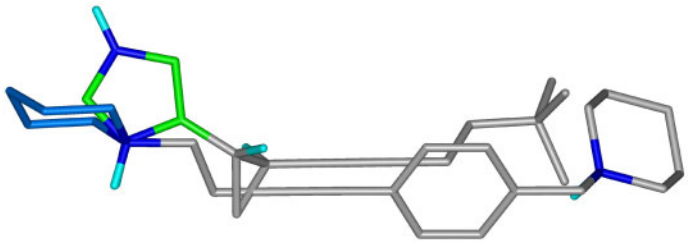


Figure 7.14: In order to generate a reasonable superposition of compounds JNJ1 and 1S,2S-GT2331, the global minimum structure of GT2331 and an extended conformation of JNJ1 (γ1/γ2 torsion angles corresponding to the local minimum d) had to be overlayed. As depicted, the Nπ-nitrogen atom of the imidazole moiety corresponds to the protonated nitrogen of the piperidine moiety.

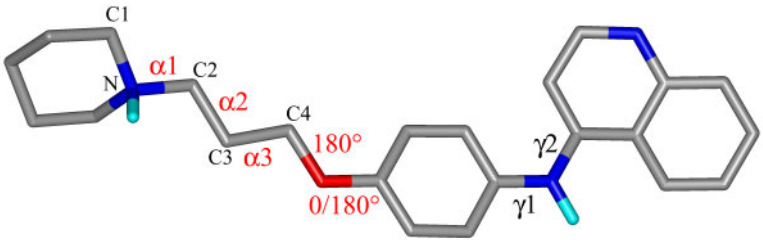


Figure 7.15: Systematic conformational analysis of compound FUB836. Dihedral angles analysed are indicated by greec letters.

MD Simulation of a FUB836/hH$_3$R Complex

In order to generate a start conformation, FUB836 was placed into the binding site by the "inverse" docking approach described above. The resulting complexes were then analysed regarding their potential energy and the docking score obtained. The — in this regard — optimal complex showed a ligand placement as in figure 7.11 with the quinoline moiety placed into the gap between helices 3, 4 and 5. The protonated nitrogen of the piperidyl moiety formed a salt bridge with D114/3.32 and was further stabilized by an aromatic cage formed by residues from helices 2 (Y91/2.61), 3 (W110/3.28) and 7 (F291/7.39, W292/7.40; W295/7.43 was pointing into the gap between helices 1 and 7). The flexible linker moiety lay in an extended conformation in the passage between helices 3 and 7 that was, due to residue L294/7.42 (that aligns to glycine residues in most other biogen aminergic GPCRs), sterically demanding. The phenyl moiety was stabilised by parallel displaced $\pi - \pi$ interactions with Y189 from the 2^{nd} extracellular loop and a T-shape interaction with Y267/6.51. The exocyclic nitrogen of the 4-aminoquinoline moiety interacted with E206/5.46 that was in turn stabilised by a hydrogen bond interaction with Y167/4.57.

In this regard, the type of interaction between E206/5.46 and the 4-aminoquinoline had to be further analysed. In [128] the pK_a value of a 4-aminoquinoline moiety was 7.53, indicating that this moiety would be present to almost equal parts in the protonated and neutral state at physiological pH. In order to evaluate the possibility of a pK_a-shift occurring in proximity to residue E206/5.46 that would favour binding of the ligand in its protonated state, calculations with the program `UHBD` were carried out. Prior to pK_a-shift calculations of ligand/protein complexes, the ligand had to be parameterised in the `UHBD` program. Suitable parameters were more easily obtained for ligands that resembled amino acid structures, thus, VUF5300 (K_i = 8.05 nM) [35] was used for calculating the pK_a-shift instead of FUB836. The pK_a of the imidazole moiety was set to 6.50, which is even less basic than the pK_a of the 4-aminoquinoline moiety. In case that the imidazole moiety should be predicted in its protonated state, this result would thus be transferable to the 4-aminoquinoline group.

In order to evaluate if the imidazole moiety sterically corresponded to the 4-aminoquinoline group, VUF5300 was superimposed on FUB836 using the program `FLEXS` and the piperidyl moiety as base fragment. In figure 7.16 the resulting superposition of FUB836 and VUF5300 is shown. Equal interaction points were predicted by `FLEXS` for the exocyclic nitrogen of the 4-aminoquinoline group and the imidazole moiety.

As an input structure for pK_a calculation, VUF5300 was docked into the binding pocket using the program `GOLD`. The resulting VUF5300/receptor complex was then used for calculating pK_a-shifts. As expected, on transmembrane level, only D80/2.50 significantly

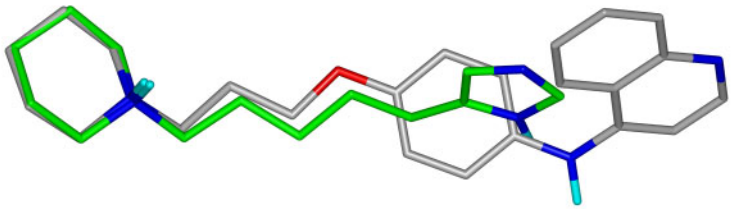


Figure 7.16: Superposition of hH_3R inverse agonists FUB836 and VUF5300 carried out using the program FLEXS. An identical interaction point is predicted for the exocyclic nitrogen function of the 4-aminoquinoline moiety and the imidazole moiety.

deviated from its pK_a in aqueous solution. For this residue, a pK_a-shift of more than 5 entities was predicted (4.0 – 9.1). Thus, this residue at a pH of 7.4 will be present in its neutral state with a probability of 93%. For the imidazole group of compound VUF5300 a pK_a-shift from 6.5 to 8.6 was predicted within proximity of E206/5.46 that was predicted to be charged with a probability of 86%.

Consequently, FUB836 was considered in its double protonated state. When protonating a 4-aminoquinoline, the positive charge will be delocalised between the endocyclic and exocyclic nitrogen as indicated by the mesomeric structures in figure 7.17. In case of an acidic residue in contact with the exocyclic nitrogen, the mesomeric structure depicted on the right could be favoured. Thus, for the parameterisation of the ligand, the charge was defined to be localised on the exocyclic nitrogen atom. Theoretically, the positive charge could be as well distributed over the system, yet, localising the charge at one site is required in order to guarantee that the interaction is recognised as a salt bridge during the force field simulation.

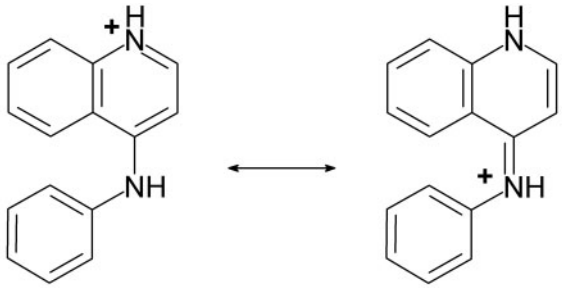


Figure 7.17: Mesomeric structures of 4-aminoquinoline.

In analogy to section 7.2.1, MD simulations were carried out in a DPPC/water environment. One water molecule was again included that was in contact with D80/2.50. Again, interhelical constraints were set between the residues given in table 7.6. In order to subsequently apply the resulting binding site conformation as a filter in docking procedures, the simulation period was restricted to 1 ns. In case of longer simulation periods, the binding site geometry tended to be over fitted to the ligand used in the

simulation resulting in lower docking scores for structurally divergent ligands.

The course of RMSD within the transmembrane backbone region (red curve) and the entire backbone (blue curve) is depicted in figure 7.18. Within the transmembrane region, a RMSD value of below 0.15 nm was reached which was only slightly below the value observed in the unconstrained simulation of the uncomplexed models after 1 ns (see figure 7.9). Thus, the placement of the ligand did not result in additional deviations from the start structure, indicating that the generated complex was in accordance with the binding pocket resulting from the uncomplexed models. Apparently, larger structural deviations occurred within the loop region. The rise of the value of RMSD up until the end of the simulation indicates, that this region did not fully equilibrate. However, flexibility within these regions was not expected to influence the binding site geometry.

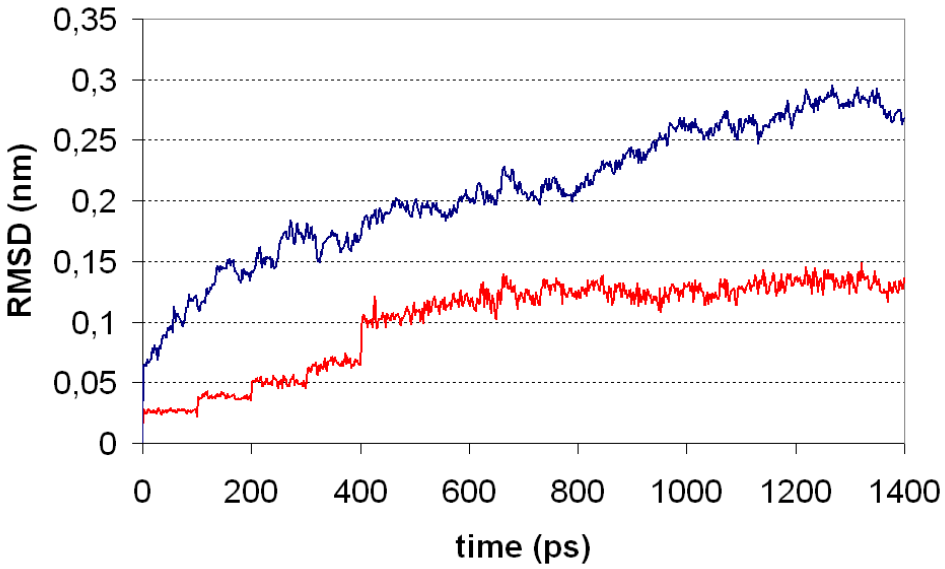


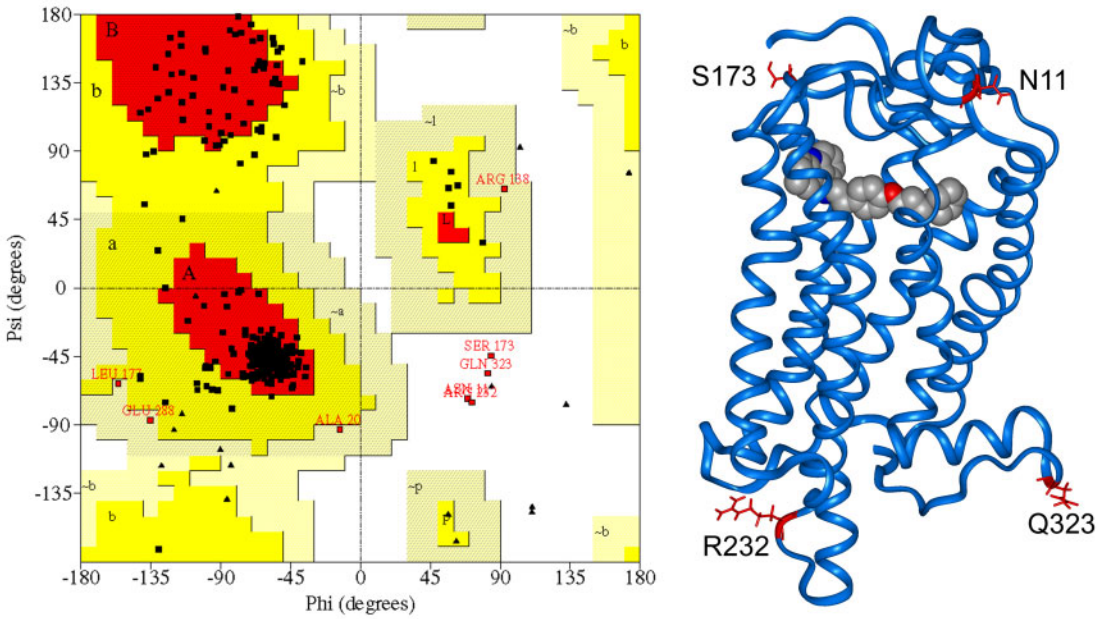
Figure 7.18: Course of RMSD during the simulation of a FUB836/hH_3R complex within the entire backbone (blue curve) and the backbone of the transmembrane region (red curve).

Visual inspection of 10 frames written out every 100 ps of the MD simulation showed that the flexibility of amino acids bordering the binding site was quite low. Main polar interactions between the ligand FUB836 and the hH_3R model were a hydrogen bond reinforced salt bridge between the piperidyl-nitrogen and residue D114/3.32 and a salt bridge between the exocyclic nitrogen of the 4-aminoquinoline moiety and E206/5.46. Tryptophane 295 pointed into the gap between helices 1 and 7 and was locked into that position by steric constraints imposed by the piperidyl moiety of the ligand. Under these conditions the secondary structure of helix 7 remained stable and differed only insignificantly from the one observed in bovine rhodopsin. Apart from the salt bridge, the protonated piperidyl moiety was stabilised by cation-π-interactions with W110/3.28, F291/7.39 and W292/7.40. The position of D114/3.32, in turn, was stabilised by

interactions with residues Y91/2.61 and Y189/5.29. Leucin 294/7.42 made van der Waals contacts to the flexible linker moiety and the adjacent aromatic ring was stabilised by $\pi - \pi$-interactions mainly with Y189/5.29 and Y267/6.51. The quinoline system could establish $\pi - \pi$ interactions with Y115/3.33. Finally, the endocyclic nitrogen interacted with Y194/5.37.

The placement of FUB836 did thus not vary significantly from the start conformation with exception of the flexible linker. Here, deviations from the extended start conformation were observed resulting in a conformation that deviated 3.42 kJmol^{-1} from the global minimum structure of FUB836 and 1 kJmol^{-1} from the extended conformation described in section 7.2.2. Given, that sterically more demanding linker groups can be accommodated in this part of the binding site, a significant conformational freedom of the flexible aliphatic linker had to be expected. Yet, the obtained conformation does — due to reasons explained in section 7.2.2 — not represent the expected bioactive conformation consistent with other ligand structures such as GT2331 or JNJ1.

Figure 7.19 shows the Ramachandran plot corresponding to the minimised end structure of the simulation of the FUB836/hH_3R complex. Four residues lay within disallowed regions of the Ramachandran plot, however distant from the binding site.



Residues in		
most favoured regions	256	84.5%
additional allowed regions	39	12.9%
generously allowed regions	4	1.3%
disallowed regions	4	1.3%

Figure 7.19: Ramachandran plot of a hH_3R model. The model comprised 338 residues, of which 19 were glycine residues (indicated by triangles within the plot) and 14 were prolines. The location of the remaining 303 residues within the Ramachandran plot was indicated by black squares. The localisation of the four disallowed residues within the hH_3R model is indicated on the right.

MD Simulation of a VUF5391/hH_3R Complex and a Complex of a Hybrid Compound Composed of FUB836 and VUF5391 and the hH_3R

Some inverse agonists such as VUF5391 (see figure 1.11) contain spatially demanding substituents on the piperidyl moiety that are expected to extend into the gap between helices 1 and 7. As previously mentioned, the initial amino acid side chain placement will significantly influence the outcome of an MD simulation as the conformational space is significantly reduced in a ligand occupied binding site. Again, the two thinkable side chain conformations of W295/7.43, either pointing into the binding pocket or into the gap between helices 1 and 7, were analysed.

A drawback of VUF5391 for an application as a ligand in MD simulations of ligand/protein complexes was the smaller size of its N-piperidine substituent. The ligand was thus unable to occupy the entire binding pocket and this lack of spatial constraints onto the binding site geometry together with the flexibility of the linker moiety increased the time required for an

adequate sampling. In contrast, compound FUB836, although containing the same p-(3-piperidine)-phenyl moiety could be positioned more easily within the binding pocket, as the exocyclic nitrogen could be assumed to interact with E206/5.46. Presence of two anchor points significantly decreased the conformational space available for the ligand. Apart from the benzothiazole moiety, VUF5391 and FUB836 share a high degree of similarity which facilitated the generation of a hybrid compound that combined structural features of both compounds (see figure 7.20).

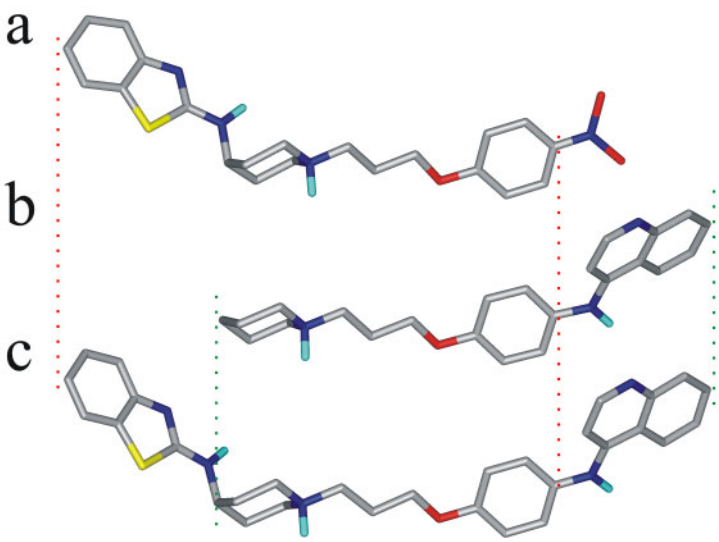


Figure 7.20: Generation of the hybrid compound FUB836-VUF5391 **(c)** from VUF5391 **(a)** and FUB836 **(b)**.

Simulation of the generated FUB836-VUF5391/hH_3R complex with W295/7.43 pointing into the binding pocket resulted in RMSD values within the backbone of the transmembrane region similar to those observed in the simulation of the FUB836/hH_3R complex (see figure 7.21). In this case, the kink in helix 7 was preserved during the simulation as can be seen in figure 7.22(b). If W295/7.43 pointed into the gap between helices 1 and 7 in the start conformation, the kink in helix 7 disappeared during the simulation (see figure 7.22(a)). The corresponding value of RMSD within the transmembrane region rose up to a value of 0.18 nm within the backbone of the transmembrane region.

Figure 7.23 shows the resulting binding sites of the two complexes. If W295/7.43 pointed into the binding pocket, this residue formed π-cation interactions with the protonated moiety. If W295/7.43 pointed outside the site, F291/7.40 assumed this function. Analysis of mutational data should theoretically favour one placement as in the case where residue 7.43 is pointing into the gap between helices 1 and 7 it will not be primarily involved in ligand binding and thus a point mutation should not have pronounced effects on binding affinity. Yet, for both ligands a participation in ligand binding was reported (see GPCR

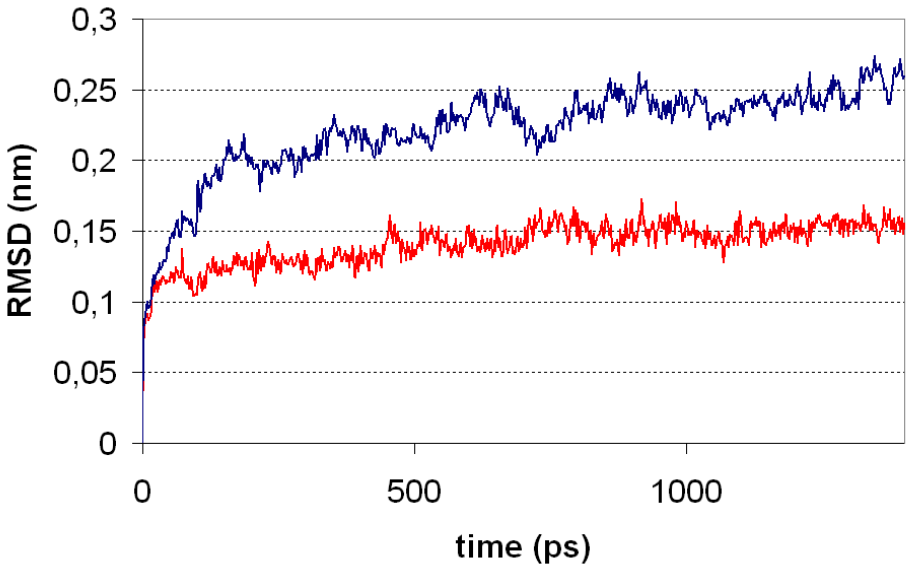


Figure 7.21: Course of RMSD during the simulation of a FUB836-VUF5391/hH_3R complex with W295/7.43 pointing into the binding site within the entire backbone (blue curve) and the backbone of the transmembrane region (red curve).

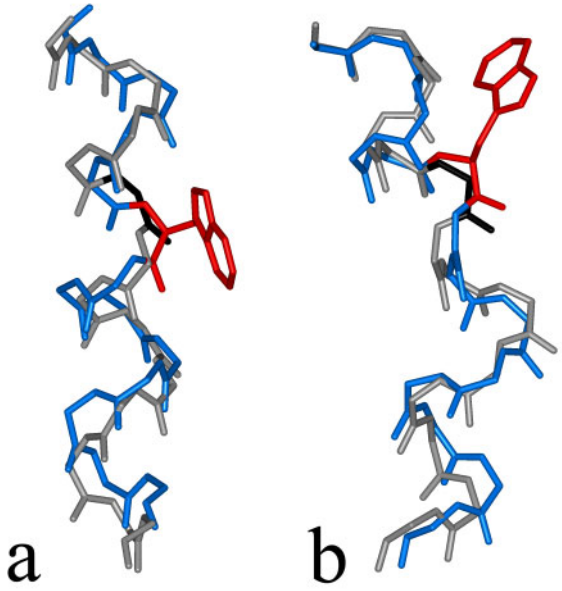


Figure 7.22: Comparison of resulting geometries for helix 7 (blue) depending on the conformation of residue W295/7.43 (red). **(a)** W295/7.43 was pointing outside the binding site; the kink disappears during the simulation. **(b)** W295/7.43 was pointing inside the binding site; the kink was preserved during the simulation. The backbone of helix 7 of bovine rhodopsin is overlayed in grey in both simulations; The backbone of residue K296/7.43 involved in the Schiff-base is coloured black.

database [124]). Additionally, interpretation of mutation studies was difficult as mutation of sterically demanding aromatic residues to alanine can always have indirect effects and a

reduction in binding affinity cannot be simply interpreted in a way that the corresponding ligand participates in ligand binding.

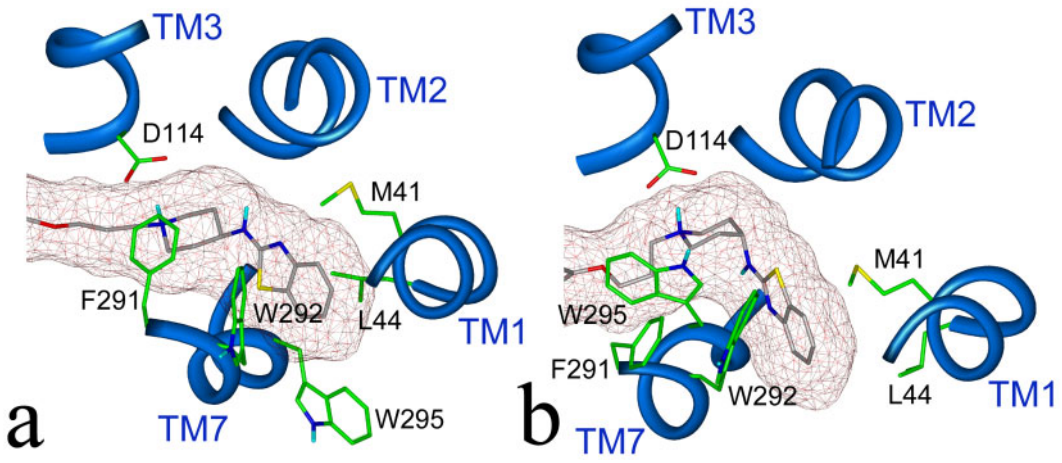


Figure 7.23: Comparison of resulting binding geometries resulting from different placements of W295/7.43.

7.3 Model Validation via Screening of a Focused Database

As described above, MD simulations with different inverse agonists could result in binding pockets varying in the placement of individual amino acid side chains. For a subsequent high throughput screening procedure, the consideration of alternative binding pocket geometries was however impracticable. Thus criteria had to be set in order to choose the most appropriate binding site geometry:

- RMSD, potential energy and stereochemical parameters of the entire protein model were considered although they are, if exclusively taken into account, too imprecise tools for a model evaluation. [118]
- Residues that matched analogue residues in the binding pocket of bovine rhodopsin and were additionally highly conserved within GPCR structures were required to be present in a comparable side chain rotamer as in the template structure (e.g. W264/6.48, Y267/6.51).
- The enrichment of ligands tested on the hH_3R model binding site against randomly selected drug-like ligands was evaluated in order to assess the suitability of the binding site conformation for a subsequent application as filter in a high throughput screening approach.

- A special focus was set to the resulting GOLD scores of docking sterically demanding ligands. This gave valuable hints upon if the binding site had realistic dimensions.

In order to test the enrichment of hH_3R ligands against randomly selected drug-like ligands taken from the WDI, several hH_3R binding sites were used. Exemplarily, the model validation is explained for the binding pockets resulting from the simulation of FUB836 and of the hybrid compound FUB836-VUF5391 with W295/7.43 located within the binding pocket.

The hH_3R ligand test data set comprised 418 ligands within a range of binding affinity from pK_i 5.29 to 10.04 and 8 inactive compounds (pK_i < 4). 473 compounds were randomly selected from the WDI. Ligands were docked into the binding pockets using the program GOLD with default settings for a “2-times speed up” genetic algorithm, and the corresponding GOLD scores were evaluated. All compounds were treated as being in the protonation state under physiological condition. All imidazole groups were considered in their protonated form. In order to account for the fact that D114/3.32 was the main interaction partner also for binding of inverse agonists, a protein hydrogen bond constraint was set which required establishment of a hydrogen bond with D114/3.32.

Figure 7.24 shows the population of GOLD score clusters by WDI and hH_3R ligands for both binding sites. 67 WDI ligands and 6 hH_3R ligands could not be docked into the binding sites. Most of the 67 WDI compounds that did not give a docking result, did not contain a protonated nitrogen capable of forming a hydrogen bond reinforced salt bridge with D114/3.32; others were too large or did not coincide with the common shape of hH_3R inverse agonists. Of the 6 hH_3R ligands that were not placed into the binding site, 4 were inactive (pK_i < 4) and two had binding affinities of 1000 and 1410 nM, respectively. All other hH_3R ligands could be placed into the binding site although no correlation could be observed between the affinity of the hH_3R inverse agonists and the corresponding GOLD score, which is, however, a well known problem of the GOLD scoring function that was designed rather to discriminate between different binding modes than towards the correlation of biological data and GOLD fitness scores.

When using the binding site geometry resulting from simulation of FUB836, hH_3R ligands were shifted to higher GOLD scores compared to the binding site of the FUB836-VUF5391/hH_3R complex. Further differences regarded the docking of the sterically demanding compound FUB833 (compound 42 in [43], see figure 1.11). Docking into the FUB836-VUF5391/hH_3R site resulted in a score of -22.85, while docking into the FUB836/hH_3R site in a score of +52.37 thus clearly favouring this site. The better fit into the binding site resulting from simulation of FUB836 was most likely due to the side chain conformation of L199/5.39 resulting in a larger free volume within this binding pocket. For subsequent screening procedures, binding pocket FUB836/hH_3R was thus preferred.

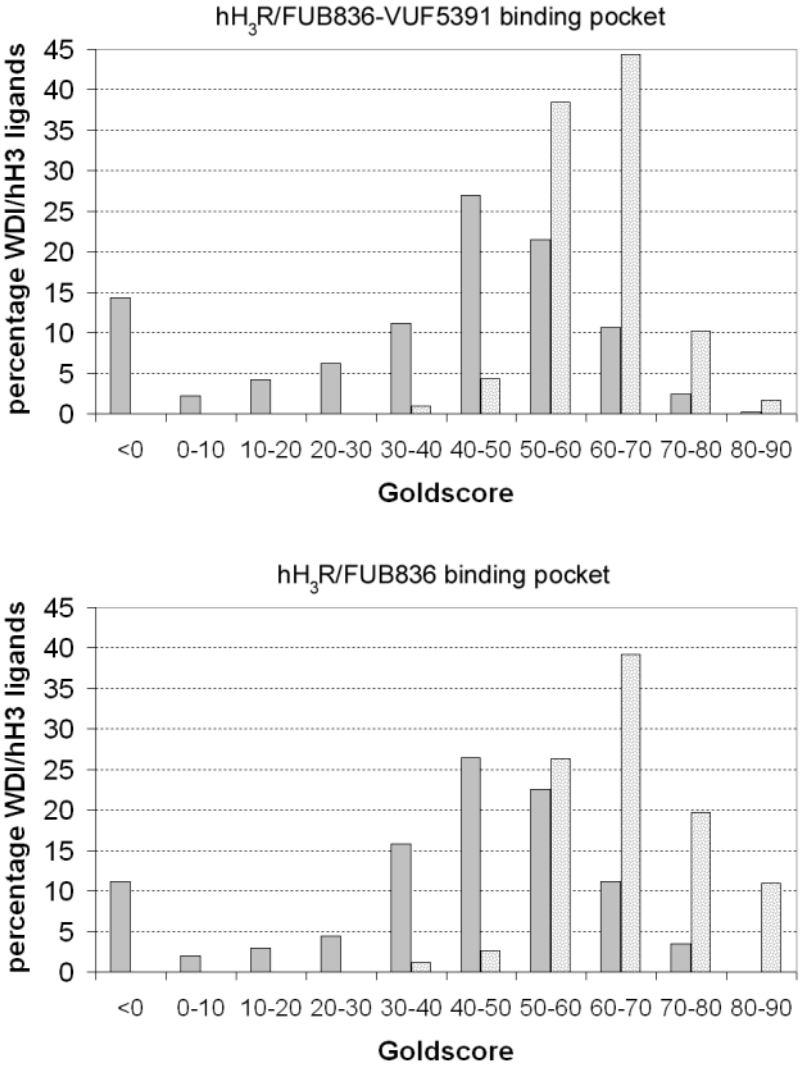


Figure 7.24: Enrichment of hH_3R inverse agonist tested against randomly selected WDI compounds evaluated by GOLD docking. The light grey columns represent hH_3R ligands, the dark grey columns randomly selected WDI compounds. Apparently, the histogram corresponding to the hH_3R compounds was shifted to higher GOLD scores. Thereby, the binding pocket resulting from simulation of the complex FUB836/hH_3R performed better than the site from the FUB836-VUF5391/hH_3R complex, as the GOLD score clusters [70;80] and [80;90] were higher populated.

The binding site conformations of complex FUB836/hH_3R is shown in figure 7.25a. Even more informative than the exact placement of amino acid side chains is an analysis of interaction fields. For this purpose the binding pocket was scanned with the phenolic OH-probe and the DRY probe (see figure 7.25b). Interaction fields produced by the DRY-probe, that maps hydrophobic sites, are coloured red, interaction fields that resulted from scanning the binding site with the phenolic OH-probe, that maps polar sites, are coloured in green. In the binding pocket of the FUB836/hH_3R complex, three hydrophobic regions exist

(D1-D3). Site D1 resulted from residues L199/5.39 and M271/6.55, D2 from W264/6.48, Y189/5.29 and Y267/6.51 while the hydrophobic character of site D3 was due to residues Y91/2.61 and W292/7.40.

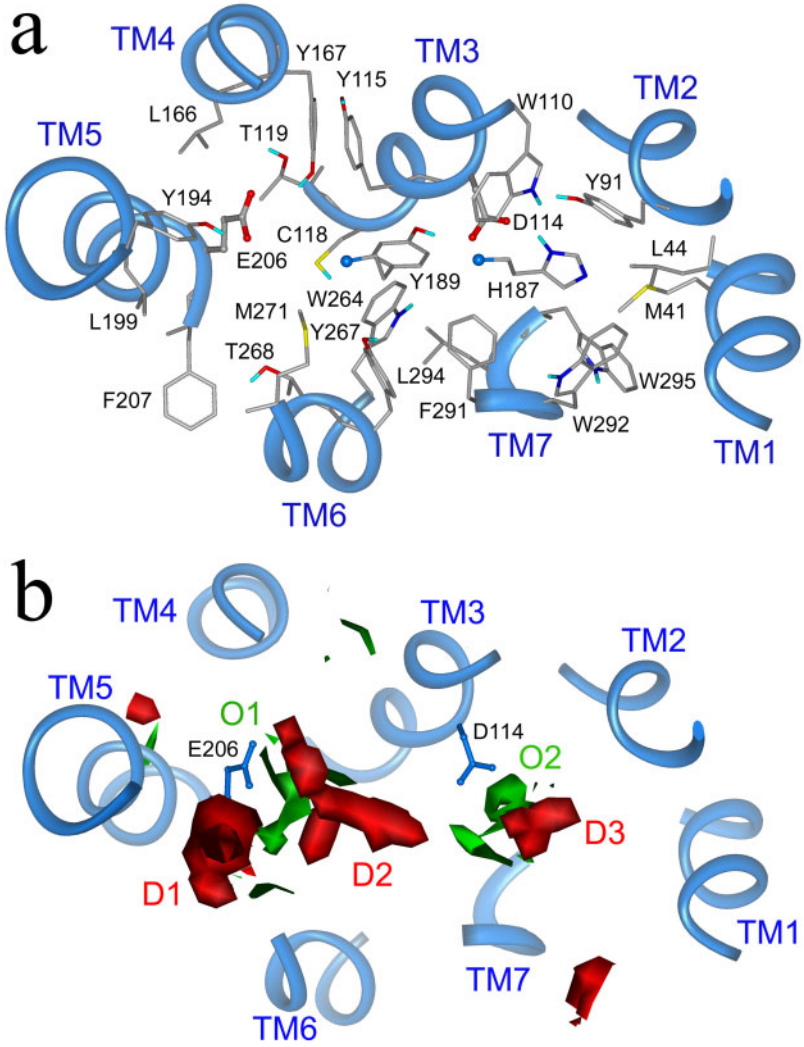


Figure 7.25: hH_3R binding pocket resulting from the simulation of the complex FUB836/hH_3R. **(a)** Amino acids participating in the ligand binding site. **(b)** Interaction fields for the DRY- and phenolic OH-GRID-probes enumerated D1-D3 and O1-O2, respectively.

In order to reduce the bias of docking programs towards higher molecular weight compounds, the resulting GOLD scores were multiplied by the correction term $1/\sqrt{N}$, where N is the number of non-hydrogen atoms. Application of this correction term significantly improved the enrichment of hH_3R ligands as can be seen from figure 7.26. The blue curve corresponds to the original GOLD scores obtained when docking to the FUB836/hH_3R binding site, while the red curve depicts the enrichment after applying the correction term. Thus, 11.4% WDI ligands must be tolerated, if the goal is to retrieve 80% of validated hH_3R ligands.

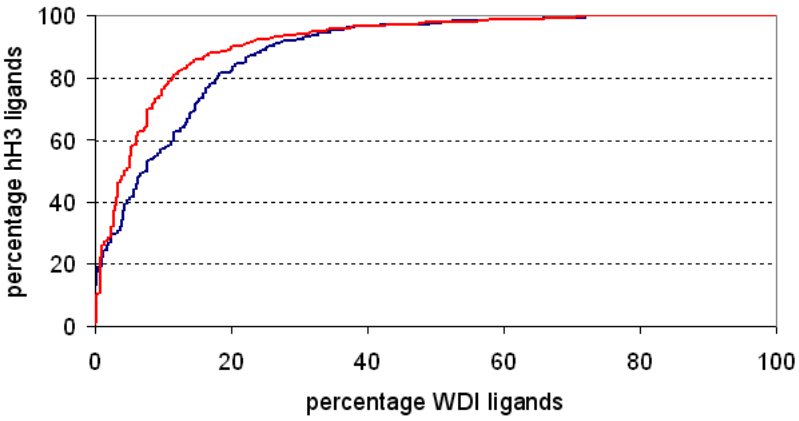


Figure 7.26: Enrichment of hH_3R inverse agonists using the binding site geometry resulting from the simulation of FUB836 with (red curve) and without (blue curve) application of a correction term $1/\sqrt{N}$. After application of the term $1/\sqrt{N}$, 11.4% WDI ligands had to be tolerated, if the goal was to retrieve 80% of validated hH_3R ligands.

Several of the compounds taken from the WDI did not contain a protonated moiety or were too large to fit into the hH_3R binding pocket. In other terms, the chemical space spanned by the hH_3R ligands and WDI compounds was quite distinct. It was thus not surprising that a good discrimination was obtained. In order to evaluate if the generated binding site could distinguish between hH_3R compounds and WDI compounds with similar 1D properties, a focused library of WDI compounds was generated similarly to [129]. First, for each active, i, the distance to the nearest other actives $D(i)$ was calculated using equation 7.1 with N_D being the number of hydrogen bond donors, N_A the number of hydrogen bond acceptors and N_{NP} the number of non hydrogen atoms. $D(i)$ was then averaged over all actives to give the average distance to the nearest other active D_{min}. In case of the hH_3R ligands, 138 compounds with a binding affinity K_i < 10 nM were considered to evaluate the distance $D(i)$. Values for N_{NP}, N_A and N_D were calculated for all ligands using the program `MOE`. The average distance between all hH_3R compounds amounted to 5.85 while the averaged minimal distance $D(i)$ to 0.55 (this small value arises due to the fact that several compounds had distance zero to the nearest other active).

$$D(i,j) = \sqrt{[N_D(i) - N_D(j)]^2 + [N_A(i) - N_A(j)]^2 + [N_{NP}(i) - N_{NP}(j)]^2} \qquad (7.1)$$

Assessment of the 470 randomly selected WDI compounds used in the screening procedure above showed, that only 52 of the 470 compounds had the required distance $D(i)$ < 0.55 to at least one active hH_3R compound. The discrepancy between the chemical spaces populated by randomly selected WDI compounds and hH_3R compounds is depicted in figure 7.27.

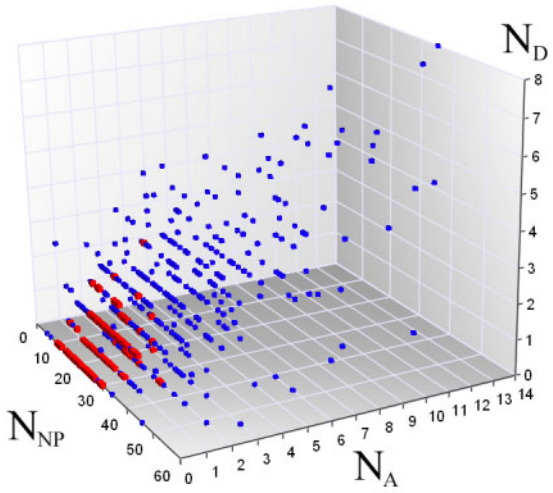


Figure 7.27: Comparison of chemical spaces populated by randomly selected WDI compounds and hH_3R ligands. WDI compounds correspond to blue, hH_3R ligands to red cubes.

In order to better match the chemical spaces and to get a realistic idea if the generated hH_3R pocket could discriminate between compounds having similar 1D properties, a focused library was generated. Additionally to similar parameters N_D, N_A and N_{NP}, the compounds were required to contain a secondary or tertiary amine moiety or an imidazole group. For this purpose, firstly only compounds containing one of the moieties named were preselected from the WDI resulting in 10721 compounds. For these structures, the descriptors N_D, N_A and N_{NP} were evaluated and the distance of each WDI compound to all 137 hH_3R ligands of the test data set (K_i < 10 nM) calculated. Finally, WDI compounds lying within the distance $D(i)$ < 0.55 from at least one active hH_3R compound were chosen, resulting in a selection of 3923 WDI compounds (of which 625 ligands only differed in the counter ion of the protonated compound). 470 compounds were again selected from this WDI subset and were — analogous to the unfocused approach — docked into the binding pocket of the FUB836/hH_3R complex. In figure 7.28 enrichment curves using a non/focused library are compared. Apparently, a higher enrichment of hH_3R ligands was obtained if the screening had been carried out against randomly selected compounds. Here, in order to retrieve 80% of the hH_3R compounds, 11% WDI compounds had to be accepted. If compared to screening the focused library, the enrichment curves run congruent up to 60% hH_3R. In order to retrieve 80% hH_3R compounds, now 23% WDI ligands had to be accepted. Due to the comparable physicochemical properties, these ligands represented also interesting hits for further visual inspection. In figure 7.29 some high ranked compounds are depicted.

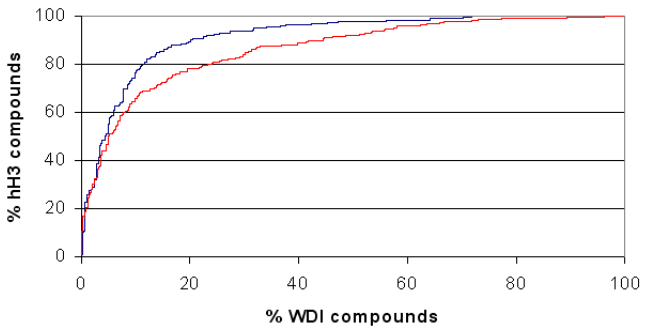


Figure 7.28: Comparison of enrichment curves using a non/focused library for validation purposes.

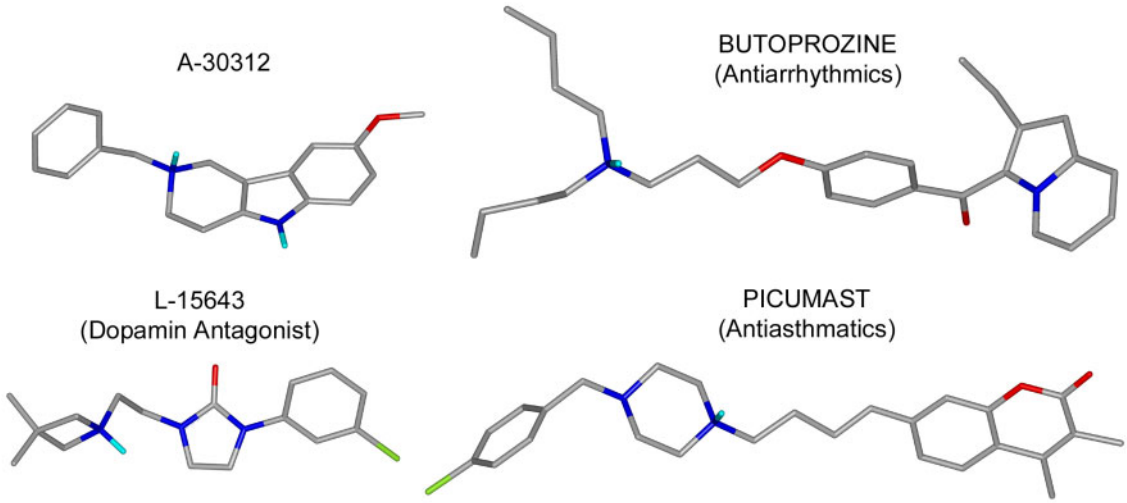


Figure 7.29: WDI compounds of the focused library producing high scores when docked into the hH_3R binding pocket. Due to the similarity to known H_3R compounds, an evaluation of their affinity towards the hH_3 receptor might be interesting.

7.4 Application of the Generated Binding Pocket Conformation as Filter in HTS

As described in section 7.3, the binding pocket resulting from the simulation of ligand FUB836 could serve as a filter for high throughput screening. Active hH_3R compounds are known to contain a protonated moiety; thus, such compounds were preselected resulting in 7101 compounds from the Maybridge Database (MDB), and 16048 WDI-compounds. Primary amines, guanidines, pyrimidinium salts and amidines were excluded from the database as they are quite polar and cannot be expected to penetrate the blood-brain barrier. Additionally, a molecular weight filter with a cutoff of 600 Da was applied, which further reduced the number of compounds to 13524.

These ligands were then docked into the hH_3R binding pocket using the program GOLD

with default parameters for a “library screening” genetic algorithm. This algorithm was computationally significantly less demanding than the “2 times speed-up” algorithm applied during the model validation. The drawback of applying this method is depicted in figure 7.30. When screening the hH_3R compound library, the average score obtained with the “2 times speed-up” algorithm was 66, when applying the “library screening” method only 45. As can be seen from the distribution of data points in figure 7.30(a), the distribution is not only shifted by a value of approximately 20, but the “library screening” algorithm tended to underpredict some compounds. More important, also the enrichment of the hH_3R ligand data set against the focused library described in section 7.3 depended on the algorithm applied. Thereby, the “library screening” algorithm resulted in a worse separation of hH_3R compounds from WDI/MDB ligands. While in case of the “2 times” accelerated algorithm 23.0% WDI compounds had to be accepted in order to retrieve 80% of the hH_3R ligands, application of the library screening algorithm increased the number of high scored WDI compounds to 32.8%. The correlation between reliability and speed has been described in the manual to the program GOLD. For flexible ligands a “2 times speed-up” or “3 times speed-up” algorithm had been recommended that is however computationally too demanding for virtual high throughput screening.

The resulting docking solutions were then ranked according to the GOLD scores obtained. In order to get an idea, up to which score ligands had a high probability of being hH_3R compounds, the distribution of scores for hH_3R ligands was overlayed to the histogram of WDI and MDB compounds. Apparently, both histograms depict a Gaussian distribution with a mean value of [20;30] in case of WDI/MDB compounds and [40;50] in case of hH_3 compounds. In figure 7.31 the number of hH_3R compounds is multiplied by 10 for reasons of clarity.

The cutoff value chosen for visual inspection of high scored WDI/MDB compounds was a compromise between feasibility of the subsequent visual inspection and the increasing rate of hH_3R compounds that would be withheld by this filter. If a cutoff of 40 was chosen, 66.5% of the validated hH_3R ligands would pass the filter, meanwhile 87% of the WDI and MDB compounds could be filtered out. Still, 1839 ligands had to be visually inspected with this cutoff setting.

In biogen aminergic receptors, a considerable degree of conservation is observed for residues participating in the binding site. As an example, in muscarinic as well as in dopamine receptor subtypes an aromatic cage has been observed formed by helices 2, 3 and 7. Additionally, the shape of the binding pocket is influenced to a certain amount by the position of transmembrane helices. Screening of the focused library in section 7.3 resulted in high scores for compounds that bind to biogen aminergic GPCRs. Thus, in order to evaluate, if the generated binding site is able to discriminate also between hH_3R

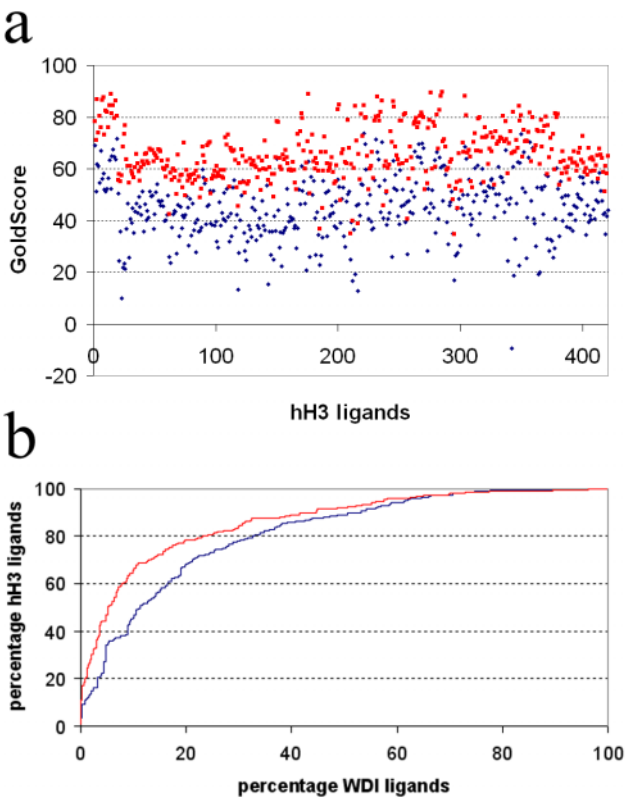


Figure 7.30: Comparison of GOLD scores and enrichment curves depending on the screening algorithm used within GOLD. **(a)** GOLD scores obtained by docking identical ligands into one binding pocket applying either a "2 times speed-up" (red squares) or the algorithm adapted for "library screening" (blue squares). **(b)** Enrichment curves obtained when repeating the screening procedure described in section 7.3 (red curve) with the "library screening" settings (blue curve).

ligands and compounds targeting biogen aminergic GPCRs, the resulting histograms were generated for this WDI subset. For this purpose, 1105 WDI compounds that were described to bind to biogen aminergic GPCRs were selected and their GOLD scores compared to those of the hH_3R data set (see figure 7.32).

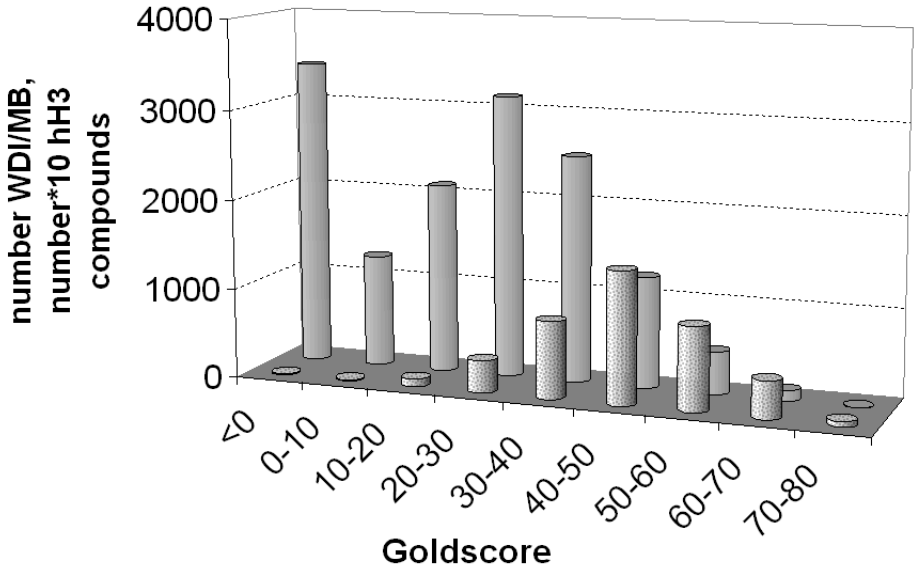


Figure 7.31: Comparison of GOLD scores obtained when docking WDI/MDB (grey column) and hH_3 (grey/white column) compounds. Note that the number of hH_3R compounds is multiplied by 10 for reasons of clarity. The distribution of hH_3R compounds is shifted by a value of 20 to higher GOLD scores indicating that the discrimination between validated binders and compounds from WDI/MDB is satisfactorily.

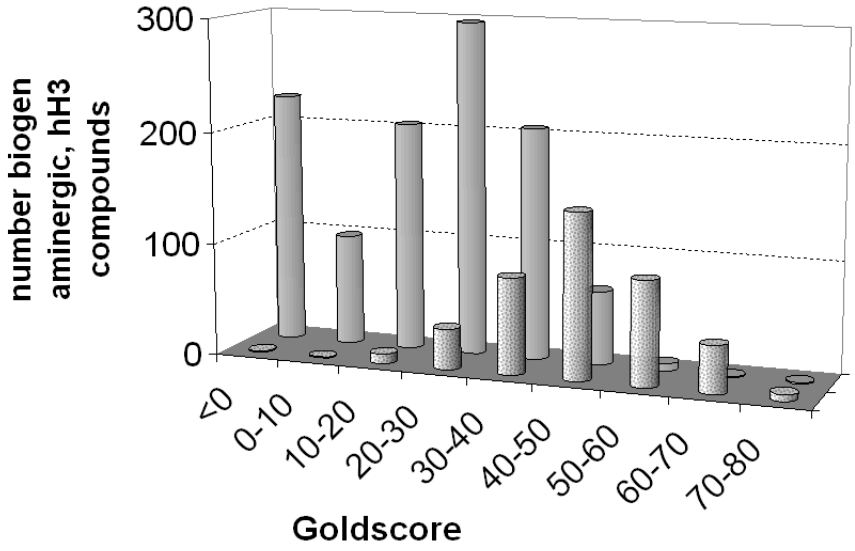


Figure 7.32: Comparison of GOLD scores obtained when docking WDI compounds known to address biogen aminergic GPCRs (grey columns) and hH_3R compounds (grey/white columns). The mean value of the distribution of WDI compounds still lies within the cluster [20;30] indicating that the hH_3R binding site does not just enrich compounds targeting biogen aminergic GPCRs but is specific to hH_3R compounds.

7.5 Pharmacophore Based Screening

Pharmacophore based screening was carried out using the program `Catalyst`. A database of 426 ligands tested on the hH_3R was generated and conformational models generated for all compounds. An energy cutoff of 20 $kcalmol^{-1}$ from each energetic minimum structure was set in order to avoid high energetic structures. Three pharmacophore models were then defined based on the template molecules FUB836, FUB833 and FUB209 in their supposedly bioactive conformation that was derived from conformational analysis and MD simulation or docking of these compounds into the hH_3R binding site. Defining a pharmacophore model upon a ligand has the advantage that this way the individual features are already correctly aligned in space. In order to account for the great structural variability of hH_3R inverse agonists, the pharmacophoric filters were defined as loose as possible in order to still retrieve most of the validated hH_3R ligands as hits. Once a filter capable of retrieving known hH_3R inverse agonists had been defined, it could be used in a subsequent screening procedure of commercial structural libraries (WDI, MDB).

Figure 7.33 shows the position of pharmacophoric features defined on molecule FUB836. Each sphere comprised a variety of chemical moieties as listed below:

- red sphere (xyz = (-12.596/2.347/2.042); r=1.5 Å)
  - any positively charged element (predefined in `Catalyst`)
  - imidazole moieties
- orange sphere (xyz = (-10.609/3.009/0.708); r = 1.5 Å)
  - ethers, thioethers, disulfides
  - aliphatic un/saturated hydrocarbon chains
  - cyclopropyl groups
  - aromatic ring systems (predefined in `Catalyst`)
  - hydrophobic groups (predefined in `Catalyst`)
- magenta sphere (xyz = (-6.183/2.726/-1.495); r = 1.5 Å)
  - aromatic ring systems (predefined in `Catalyst`)
  - π-electron rich systems: carbamate-, ester-, urea-, thiourea-groups
  - t-butyl groups

The choice of chemical moieties was based on chemical functionalities observed in validated hH_3R inverse agonists and inspection of the binding pocket. As previously described, the linker moiety and the adjacent hydrophobic/π-electron rich system lay in a cleft between helices 3, 6 and 7 of the hH_3R. In this region, residues Y189/5.29, Y267/6.51, F291/7.39 and L294/7.42 border the binding site. Analysis of a binding pocket (obtained after MD simulation of FUB836) with GRID probes showed that this region was rather hydrophobic. The seemingly contradicting observation that also polar groups could be accommodated in this cleft could be explained if one assumed that potential hydrogen bond donor functions were present in this region, however involved in intramolecular hydrogen bond interactions. Thus, in order to interact with a ligand hydrogen bond acceptor, first, an intramolecular interaction had to be broken, resulting in a negligible netto gain of enthalpic binding energy due to the introduced hydrogen bond acceptor. In case of polar groups such as carbamate-, ester-, urea- or thiourea-moieties, the correspondence to aromatic/hydrophobic systems could lie in the presence of π-electrons available for π-π-interaction with Y189 and T-shape interaction with Y267. No pharmacophoric features were defined upon the 4-aminoquinoline moiety as a high degree of chemical diversity is observed in active hH_3R ligands within this region. Any restriction of chemical features was thus avoided.

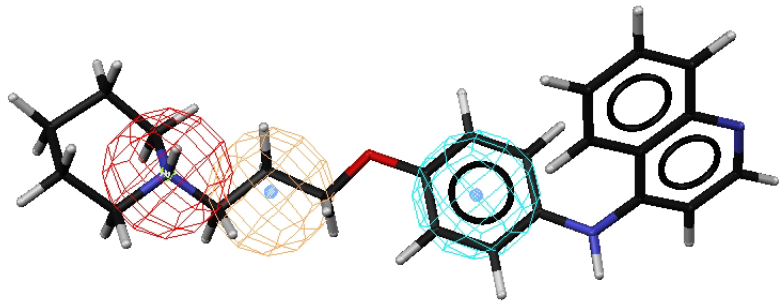


Figure 7.33: Pharmacophoric features defined based upon compound FUB836.

Apparently, the derived pharmacophore model is too loose-fitting for screening a database. Thus, the van der Waals volume of ligand FUB836 was included as an additional constraint into the pharmacophore model. Default parameters were used for the definition of the shape query, except for the value of similarity tolerance, which was adjusted to a minimum value of 0.45 instead of 0.5. Lowering this value, 316 instead of 257 hH_3R ligands could be retrieved as hits when applying the pharmacophore model described on the hH_3R ligand database, as a greater variation of the ligands from the template shape was now allowed.

Finally, also forbidden volumes (black spheres) were defined in order to account for the fact that ligands (such as FUB697 and FUB741) extending into these areas were inactive although resembling other active compounds (see figure 7.34). The forbidden volume defined upon FUB697 was centered at the site xyz = (-2.199/2.157/1.732) and

spanned a sphere with a radius of 0.9 Å. Volume 2 defined upon FUB741 was centered at (-6.744/-0.107/-1.514) and had a radius of 1.5 Å. Finally, an additional forbidden volume (r = 0.25Å) was defined in proximity to the space occupied by the piperidyl nitrogen (xyz = (-11.786/0.616/2.510)) in order to avoid larger substituents at this site that would — if the pharmacophore model was seen in its context with the binding site — produce a clash with aspartate 114/3.32.

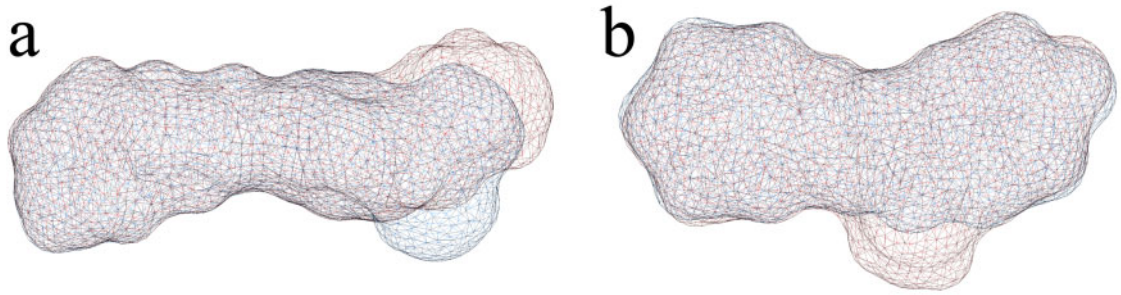


Figure 7.34: Superposition of **(a)** the active compound FUB395 with nanomolar binding affinity (blue grid) and the inactive compound FUB697 (red grid) in their lowest energy conformations. FUB697 extends the volume occupied by FUB395. This additional space is however not available for ligand binding. **(b)** Superposition of the active FUB742 (blue grid) and the inactive FUB743 (red grid). Again, FUB743 extends the volume occupied by FUB742.

Figure 7.35 shows the ligand FUB836 fitted into the complete pharmacophore model. Using this model, 316/428 ligands from the hH_3R database were found as hits. The hH_3R ligand database contained compounds spanning a pK_i-range of 5 orders of magnitude. Even 17 inactive compounds (with a pK_i <6) were included. Ideally, the pharmacophore model should be stringent enough to filter out these inactive compounds. The capability of the pharmacophore model to enrich active compounds is assessed in figure 7.36a. 93% of the ligands with highest activity were retrieved by the pharmacophore model; less satisfactorily, also 54% of the inactive compounds could pass the pharmacophoric filter. However, the filter was still quite loose so that a subsequent definition of further pharmacophoric features could result in a better separation of in/actives.

Application of the pharmacophore filter in screening the Maybridge Database and the World Drug Index resulted in 249 and 929 hits, respectively. Thus, 70% of the active and moderate active hH_3R ligands (with a pK_i > 7) were retrieved by the FUB836-filter, meanwhile from the pool of MDB and WDI ligands (MDB: 59194 compounds, WDI: 48405 compounds) 98.9% could already be excluded.

In order to further increase the percentage of active hH_3R ligands found during the screening procedure, further pharmacophore models were defined in a similar way based upon compounds FUB833 (K_i = 0.33 nM) and FUB209 (K_i = 69 nM [[130], compound 43]). For the definition of the FUB833-pharmacophore, again the three features described above

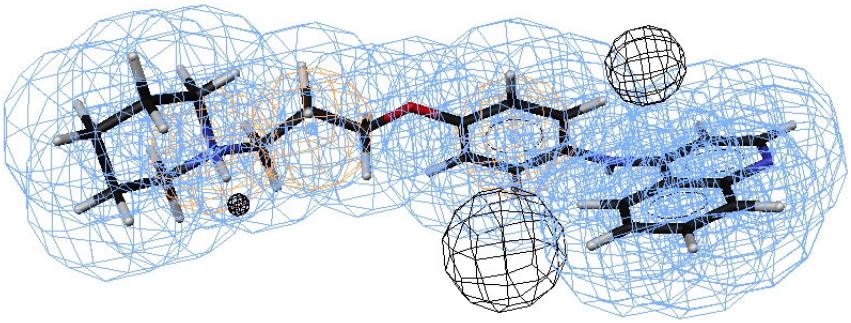


Figure 7.35: Pharmacophore model based upon FUB836.

were used in combination with a shape query and forbidden volumes. Figure 7.36b shows compound FUB833 and the enrichment resulting when applying the generated pharmacophore in screening the hH_3R ligand database. Significantly less compounds (100/426) were obtained as hits when applying this model. This might seem surprising at first sight, as the same features had been used and the molecule had a larger volume. Yet, in order to be assessed as a hit, a molecule had to match a certain percentage of the available volume set to values between 70 and 130%. Thus, small ligands — although fitting the features and into the template molecule's shape — were too small to be considered as a hit. If the enrichment in each pK_i cluster was evaluated, the FUB833-model performed quite well as it was capable of retrieving mainly high affinity binders. Screening of commercial databases resulted in 9 hits in the MDB and 115 hits in the WDI. Three hit structures from the MDB screening are depicted in figure 7.37.

Yet another pharmacophore model was defined based on FUB209 (see figure 7.36c) capable of retrieving 68% of ligands deposited in the hH_3R database. Applied to MDB and WDI, 524 and 1294 hits were found, respectively.

By combining the three pharmacophore models, 369 of 398 (93%) hH_3 ligands with a pK_i > 7 could be obtained, while only 2668 (2.5%) compounds were obtained as hits when screening the MDB and WDI with together 107599 structures deposited. The population of different pK_i clusters is shown in figure 7.36d. The small percentage of structures from commercial databases matching the pharmacophores showed that the generated models were stringent enough for a reasonable screening.

Alternatively, pharmacophore searches were carried out, where the stringency of the pharmacophore filter was achieved by including more, and more stringent features. Again, FUB836 was used as a template for the definition of these pharmacophores. With a model defined by

- a protonated moiety or imidazole group,

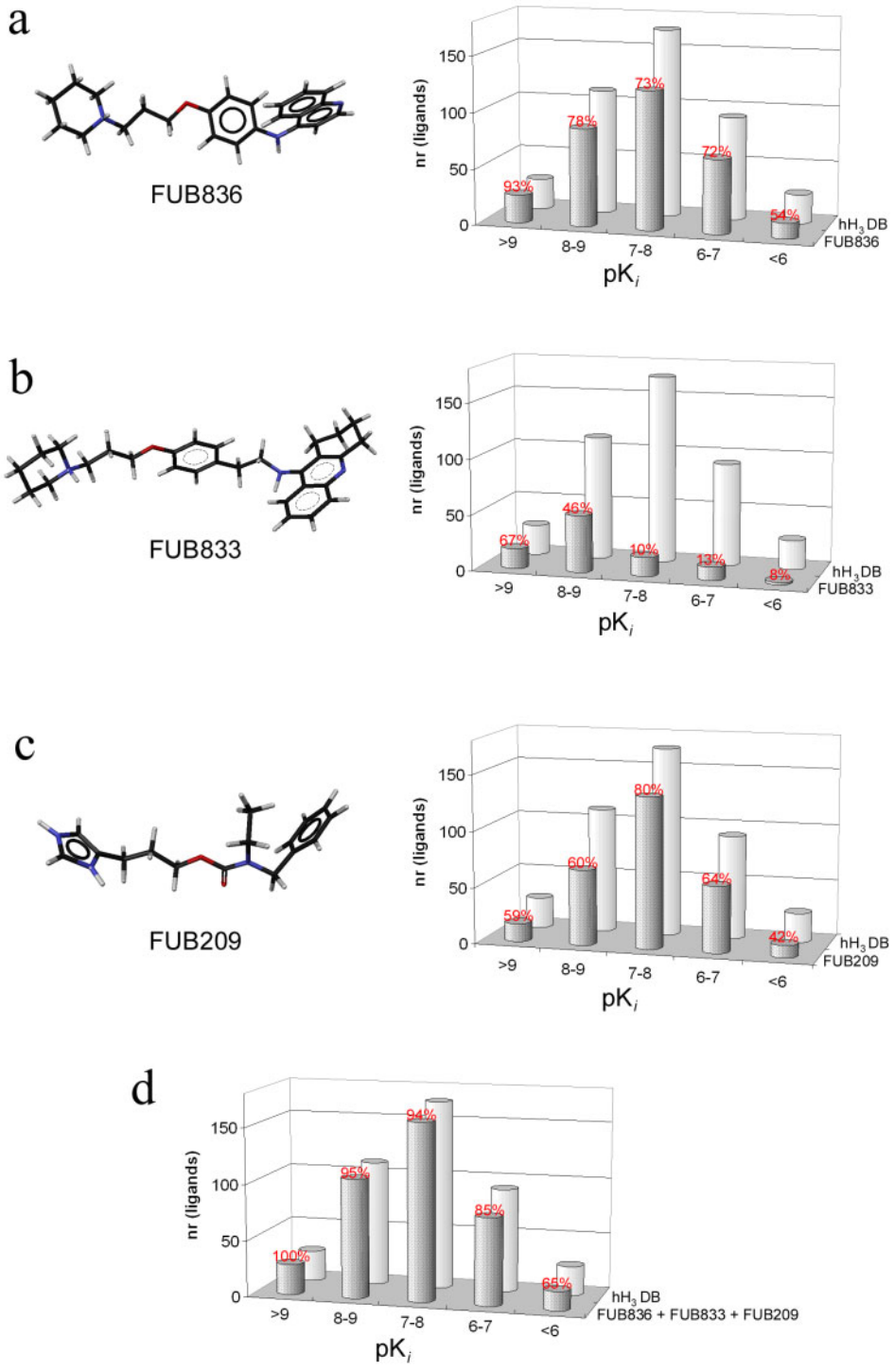


Figure 7.36: Enrichment of hH_3R ligands by pharmacophore model based on FUB836, FUB833, and FUB209. In figure **(d)** the percentage of hH_3R ligands retrieved by a combination of all three pharmacophore models within each pK_i-cluster is depicted. The percentage of ligands found in each cluster (dark columns) is written in red numbers and compared to the population of pK_i clusters of all hH_3R compounds in the data set (light grey columns).

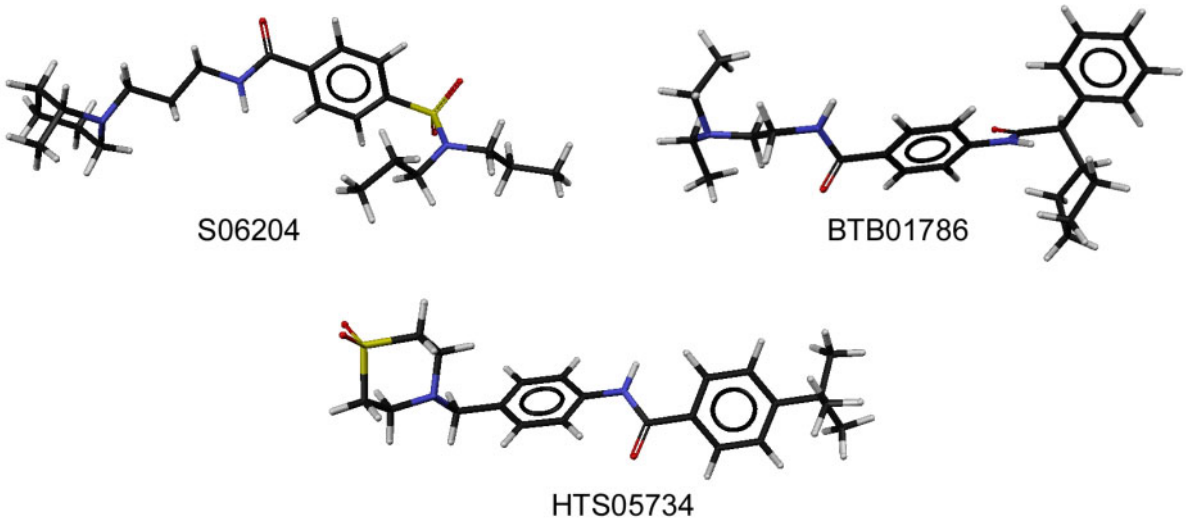


Figure 7.37: Hits when screening the MDB with a pharmacophore model based upon FUB833.

- an aromatic system adjacent to the spacer group,
- a hydrogen bond donor or protonated nitrogen (corresponding to the exocyclic nitrogen),
- a hydrogen bond donor or acceptor (corresponding to the endocyclic nitrogen),

10 hits were found in the MDB and 462 in the WDI. Most of the WDI compounds were however high molecular weight compounds (56% had a MW > 600 Da) and the lack of shape requirements resulted in proposed structures that did not resemble the shape of hH_3R ligands.

More stringent filters were however useful to further screen the subset of 2668 compounds obtained by screening with the pharmacophore models based upon FUB836, FUB833 and FUB209. Based on FUB836, a leave-one-out pharmacophore model was defined. For this purpose, firstly, all features were defined upon FUB836 as depicted in figure 7.38. Many compounds — although active hH_3R ligands — do not possess all features depicted. Hence, the stringency of the pharmacophore model was reduced by defining a set of leave-one-out models where one feature (except for the essential protonated moiety and a spacer group) was neglected at a time. Although this filter could only retrieve 106 compounds from the hH_3R ligand database, it was capable of filtering out 96% of the inactive compounds (pK_i < 6) and 84% of the moderately active compound cluster (6 < pK_i < 7). A more complex pharmacophore model also has the advantage that this way, ligands with more features are selected that should correspond to higher selective compounds, due to the

better match with the hH_3R binding site. When screening the preselected 2668 hits, 320 compounds matched the leave-one-out filter.

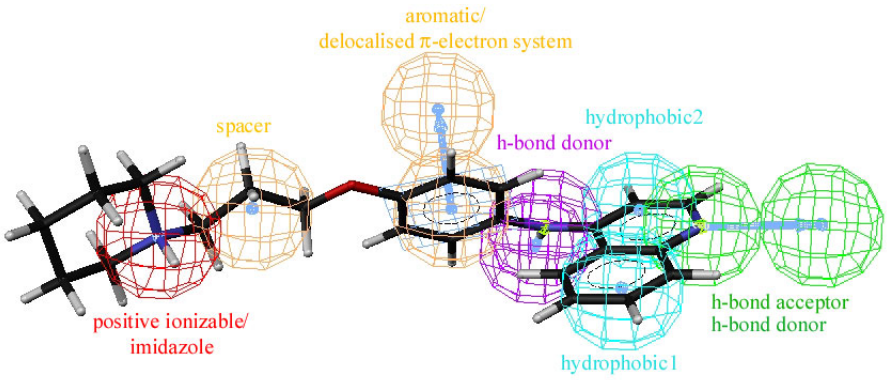


Figure 7.38: Features in FUB836 used in a combinatorial way for the definition of a leave-one-out pharmacophore model. The protonated moiety/imidazole group and the spacer were required in all models.

In figure 7.39a-d some hH_3R compounds mapped onto the leave-one-out pharmacophore model are depicted. In the same figure also some compounds from commercial databases that match the leave-one-out model are shown. Apparently, the compounds depicted share similarity with known hH_3R binders. Thus, they represent interesting hits for further evaluation.

The generated pharmacophore models allowed molecules to be still regarded as hits even if they exceeded the shape query to a certain amount (130%). Additionally, functionalities of the ligands may be misplaced if seen in the context of the hH_3R binding pocket. Thus, the selected 2668 hits from WDI and MDB were docked into the hH_3R binding site using the program `GOLD`. For this docking procedure, again the “library screening” genetic algorithm was preferred to the “2 times speed-up” algorithm due to its computationally higher performance. After docking, the ligands were scored and clustered accordingly to the `GOLD` score obtained. Figure 7.40 shows the population of each cluster with hH_3R ligands and WDI/MDB compounds. The distribution of hH_3R ligands had its maximum in the cluster [40;50] while the histogram representing the WDI and MDB compounds had the maximum at [30;40]. Compared to histogram 7.31 where the maximum of the distribution lied within the cluster [20;30], a prescreening with `Catalyst` reduced the number of potential hits significantly and shifted the distribution of higher scores. If the cutoff for visual inspection was again set to a value of 40, 66% of the hH_3R compounds passed the filter and were thus found applying this combined procedure of pharmacophore search and subsequent docking. The absolute number of WDI and MDB compounds which passed the filter now amounted to 714 which is significantly less than the 1839 ligands that were suggested for visual inspection when only applying the docking procedure as described in

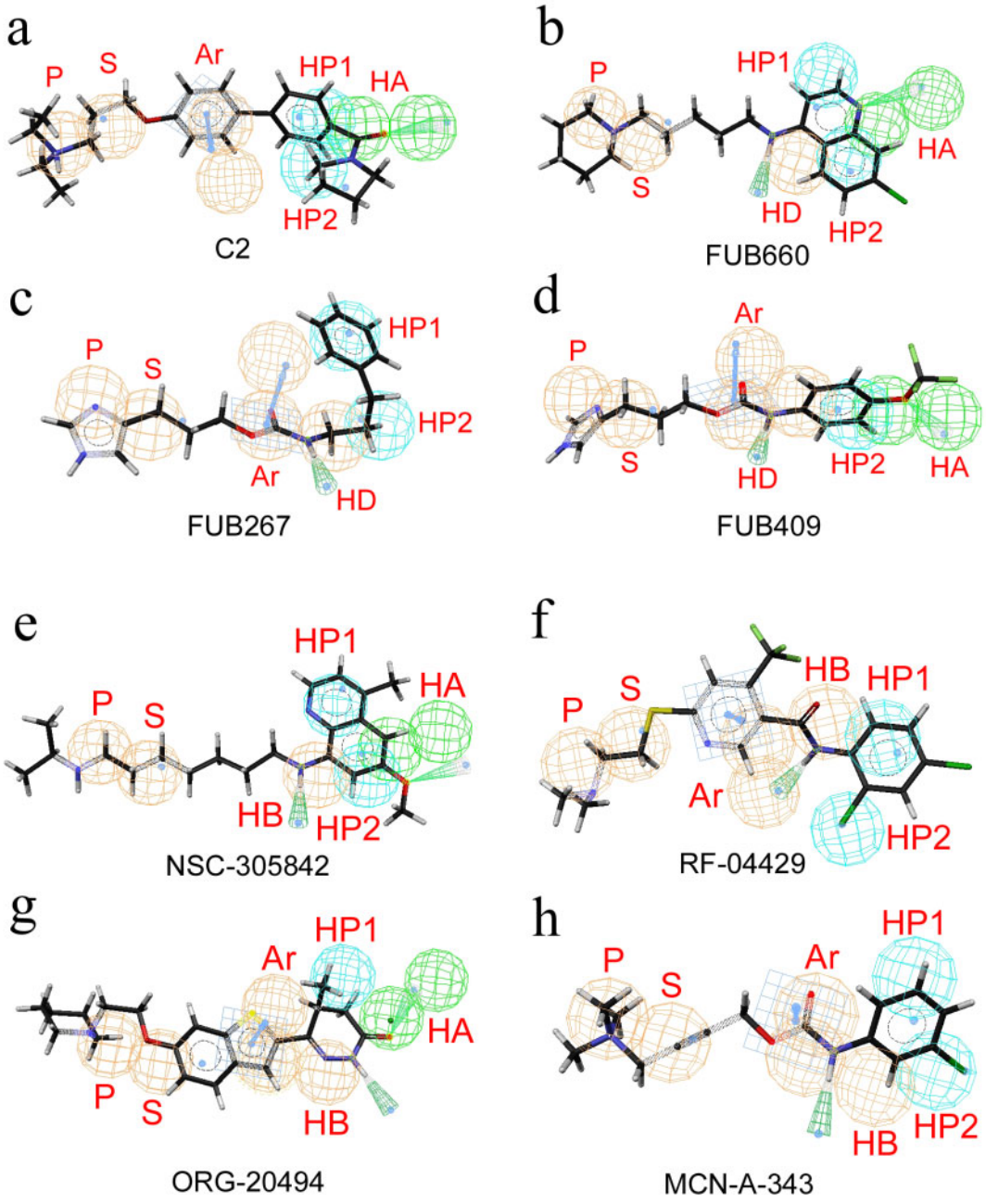


Figure 7.39: **(a)**-**(d)** hH_3R compounds mapping the leave-one-out pharmacophore model based upon compound FUB836. **(a)** C2 [42]; **(b)** FUB660 [43]: compound 20; **(c)** FUB267 [130]: compound 40; **(d)** FUB409 [130]: compound 11. **(e)**-**(h)** Compounds from WDI and MDB matching the leave-one-out pharmacophore model based upon compound FUB836.

section 7.4.

An interesting question is, if the molecules selected will form part of the focused WDI and MDB library, or in other terms, if they will have similar 1D physicochemical parameters to known hH_3R ligands. For this purpose, again descriptors N_D, N_A and N_{NP} were evaluated

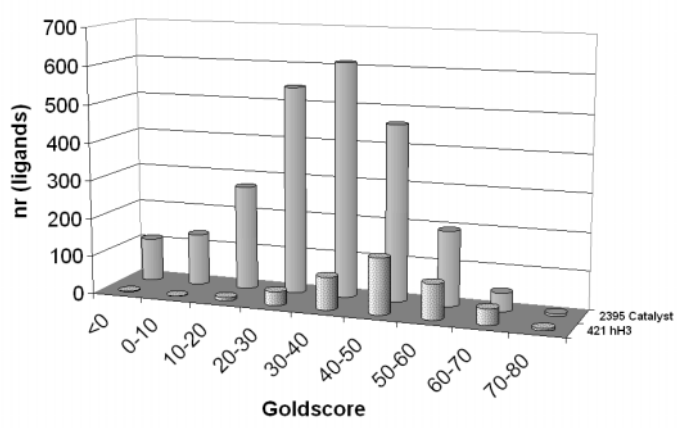


Figure 7.40: GOLD scores obtained for 2668 compounds preselected by pharmacophore search with Catalyst in comparison with known hH_3R binders.

and the amount of WDI and MDB ligands lying within the distance $D(i)$ from active hH_3R compounds was evaluated. Of the 2668 Catalyst compounds 804 lay within the required distance and formed thus part of the focused library described in section 7.3. A limitation to molecules with similar 1D physical properties would hence result in the neglection of a considerable part of hits found.

In respect of a subsequent testing, hits from the Maybridge Database were more easily available. Due to the higher performance of the “2 times speed-up” algorithm, a high scored MDB subset of 216 hits were again docked and further evaluated. Figure 7.41 shows the distribution of GOLD docking scores for hits found via the pharmacophore search in comparison with the distribution of hH_3R compounds. Apparently, the two distributions are congruent. This becomes visually even more clear, when — instead of the absolute number of compounds (see figure 7.41, left) — the percentage of Maybridge and hH_3R compounds is depicted (see figure 7.41, right) in order to account for the smaller number of MDB hits. Thus, a pharmacophore based search was capable of selecting ligands from the MDB that produced scores comparable to validated hH_3R binders which is a good indication that at least some of these molecules can be expected to show affinity towards the hH_3R.

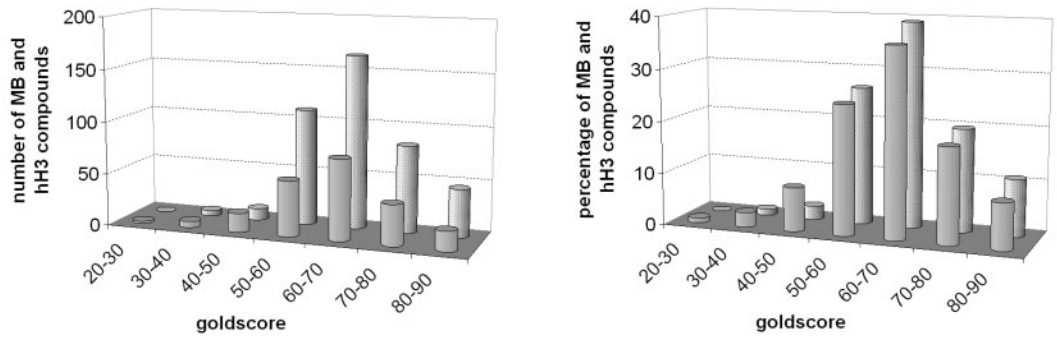


Figure 7.41: Comparison of the distribution of GOLD scores for hH_3R ligands (grey/white columns) and selection of 216 MDB compounds (grey columns) output as hits by Catalyst.

7.6 The hH_3R Binding Site, Suggested Structures and Implications for the hH_4R

7.6.1 The hH_3R Binding Site

The resulting binding pocket of the hH_3R used for high throughput screening is depicted in figure 7.25. Important structural features and ligand interactions are listed below:

- Inverse agonists such as FUB836 spanned from helix 3 to helix 5. The sterically demanding 4-aminoquinoline moiety lay between helices 3, 4 and 5. This orientation was consistent with available mutation data that will be discussed in detail in section 8.4.1.
- Protonated moieties, such as piperidine, piperazine, and pyrrolidine groups as well as imidazole moieties, which were considered in their protonated state were interacting with D114/3.32. A hydrogen reinforced salt bridge was formed with piperidine, piperazine and pyrrolidine moieties while the protonated imidazole moiety was able to form an additional interaction. Depending on the inverse agonist under study, a hydrogen bond interaction has been observed with either the backbone carbonyl moiety of F291/7.39 or the NH moiety was pointing towards the aromatic plane of W292/7.40 (see figures 7.42 and 7.43, respectively).
- The position of D114/3.32 was restrained by interactions with Y91/2.61 and Y189/5.29. Y91/2.61 corresponded to lysine residues in the D_1R and D_5R. A spatial proximity of these residues seemed thus convincing.
- An aromatic cluster consisting of W110/3.28, F291/7.39 and W292/7.40 further stabilised protonated moieties.

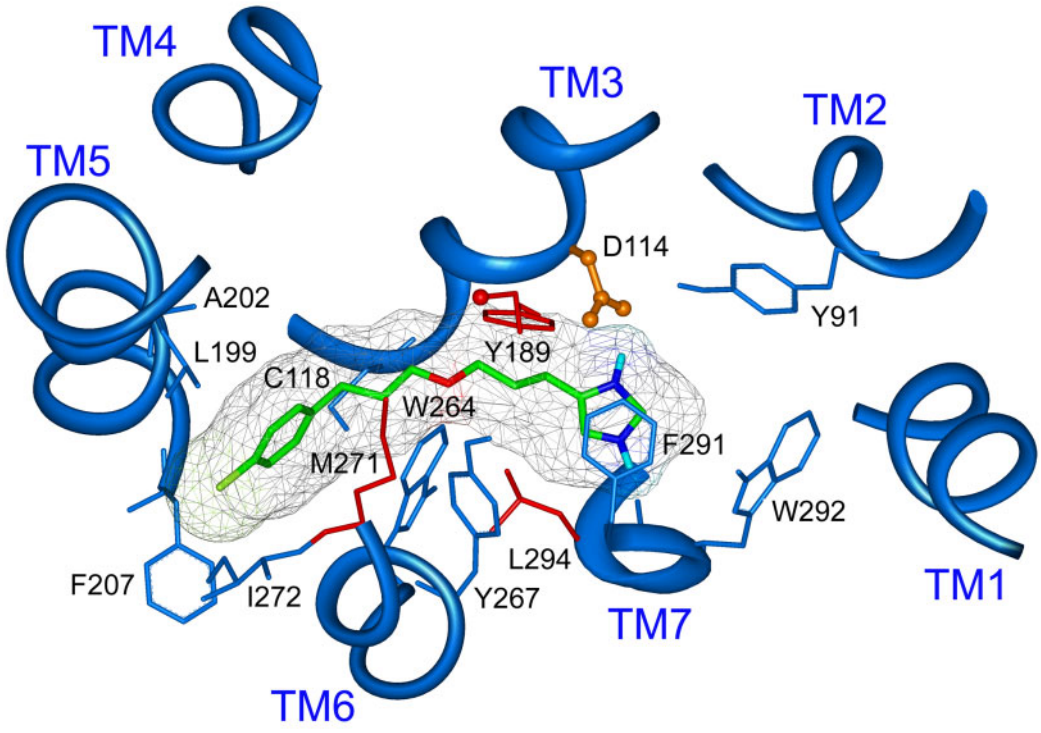


Figure 7.42: FUB181 [33] in the hH_3R binding site. Amino acids that supposedly account for the subtype specificity compared to the hH_4R are held in red.

- The flexible linker moiety lay in a cleft between helices 3 and 7. In this respect, L294/7.42 was important that aligned to glycine residues in most other human biogen aminergic GPRC sequences, except for the muscarinic receptor family where a cysteine was found at position 7.42, and the human histamine H_4R where a glutamine residue resides at site 7.42. It can thus be expected that the breadth of the cleft is significantly reduced due to this mutation. Aromatic ring systems in the linker moiety (e.g. in SCH79687 (see figure 1.10)) are expected to be oriented orthogonal to the membrane plane.

- Many hH_3R inverse agonists contain a p-(3-piperidinopropyloxy)-phenyl moiety which simply arises from the fact that in compounds of this class the imidazole moiety could be replaced without affinity loss. In the herein described model, the aromatic system lies parallel displaced to Y189/5.29 from the 2^{nd} extracellular loop. A parallel displaced orientation has been reported to be energetically more favourable in ligand/protein complexes than a π-π-stacking interaction that was according to the authors almost never observed. [72] At the same time, a T-shaped interaction could be formed with Y267/6.51.

  Polar groups such as carbamate-, ester-, urea-, thiourea-, isothiourea-, amide-moieties or unsaturated hydrocarbon chains could similarly form π-π-interactions (see figure 7.43). Y189/5.29 and Y267/6.51 could also represent potential hydrogen

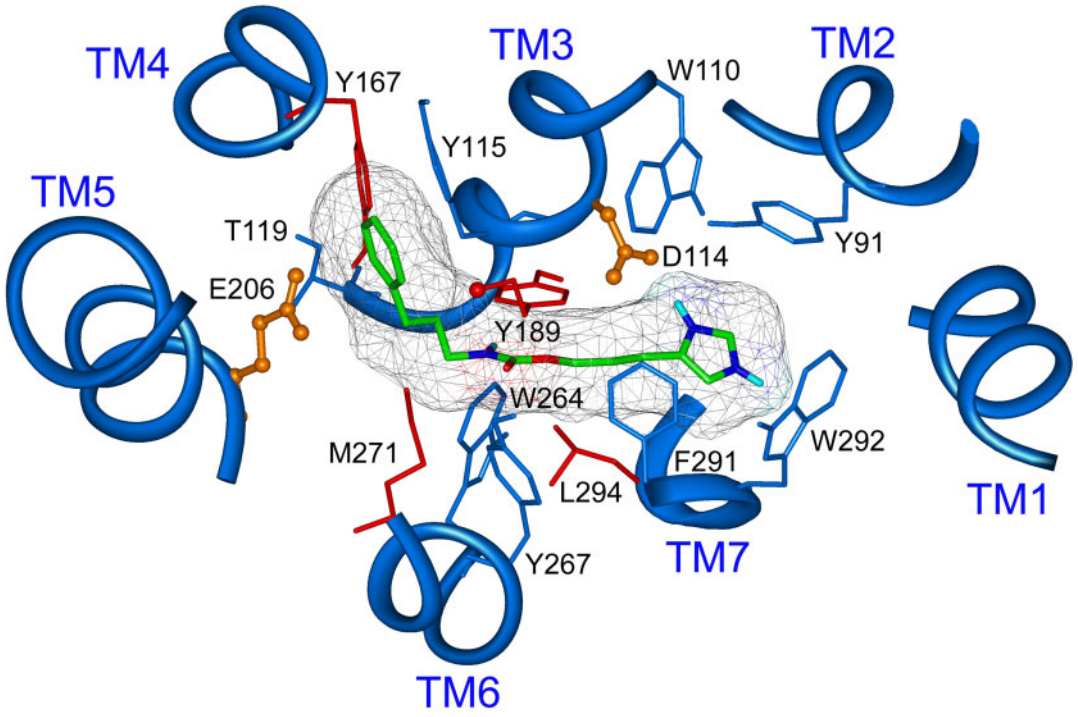


Figure 7.43: FUB267 in the hH_3R binding site. Amino acids that supposedly account for the subtype specificity compared to the hH_4R are held in red.

bond donors (e.g. for an interaction with ether moieties). In absence of a polar moiety, Y267/6.51 was involved in a hydrogen bond to Y287/7.35 in the E_3 loop and Y189 interacted with D114/3.32. Thus, in order to interact with polar moieties, these intramolecular hydrogen bond interactions had to be disrupted, thus giving a reasonable explanation, why either polar or apolar moieties could be accommodated at identical sites in the binding pocket.

- A hydrogen bond cluster was formed by residues of helix 3 (T119/3.37), 4 (Y167/4.57) and 5 (E206/5.46). T119/3.37 was hydrogen bonding to Y167/4.57 which in turn interacted with E206/5.46. Y167/4.57 was thus highly restrained in its position and ligand interactions were hampered. This could give a reasonable explanation for the decreased affinity of e.g. ciproxifan on the human H_3R compared to the rat H_3R. In the rat H_3R, T119/3.37 is mutated to an alanine that would not be capable of interacting with Y167. Thus, Y167 could more easily get involved in hydrogen bond interactions with hydrogen acceptor moieties present in ligands known to be strongly affected by species heterogeneity (e.g. ciproxifan, A-304121, thioperamide). See section 8.4.2 for a more detailed discussion of species differences and figure 7.44 for an example of a ligand that is expected to be affected by species differences.

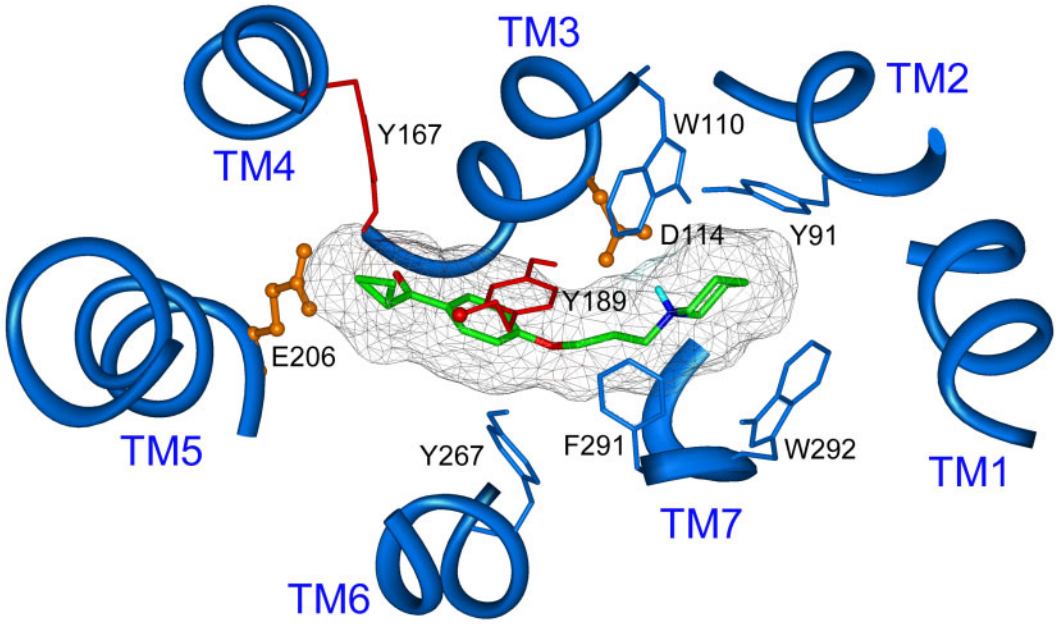


Figure 7.44: Ligand UCL2190 [33] in the hH_3R binding site. The carbonyl moiety is located in hydrogen bond distance to Y167/4.57. This residue in turn is interacting with T119/3.37 and E206/5.46. In the rat H_3R, a greater conformational freedom for Y167/4.57 would result in a facilitated interaction with ligand-hydrogen bond acceptor moieties. Amino acids that supposedly account for the subtype specificity compared to the hH_4R are held in red.

- Two lipophilic pockets were present for inverse agonist binding located between helices 3, 4 and 5 and helices 5 and 6 as suggested by De Esch and coworkers. [36] In figure 7.42 the chlorine-substituted aromatic ring system points into the gap between helices 5 and 6. Obviously, this placement will lock F207/5.47 — that had been shown by mutational analysis to be involved in the activation process — in its position. Although M271/6.55 was in this binding pocket not in an optimal position to form a charge-transfer interaction with the aromatic system, it could easily adopt a distinct side chain rotamer. The often observed lack of preference for a certain spacer length between the protonated moiety and an aromatic ring system [131] could be a consequence of the consecutive placement of Y189/5.29 and M271/6.55 and the ability of M271 to perform an induced fit.

- Compounds with two protonated moieties could simultaneously interact with D114/3.32 and E206/5.46 as exemplarily depicted for compound JNJ1 (see figure 7.45). The orientation of JNJ1 in the binding site might depend on the exact placement of M271/6.55.

- Y194/5.34 from the 2^{nd} extracellular loop region could be an interaction partner for

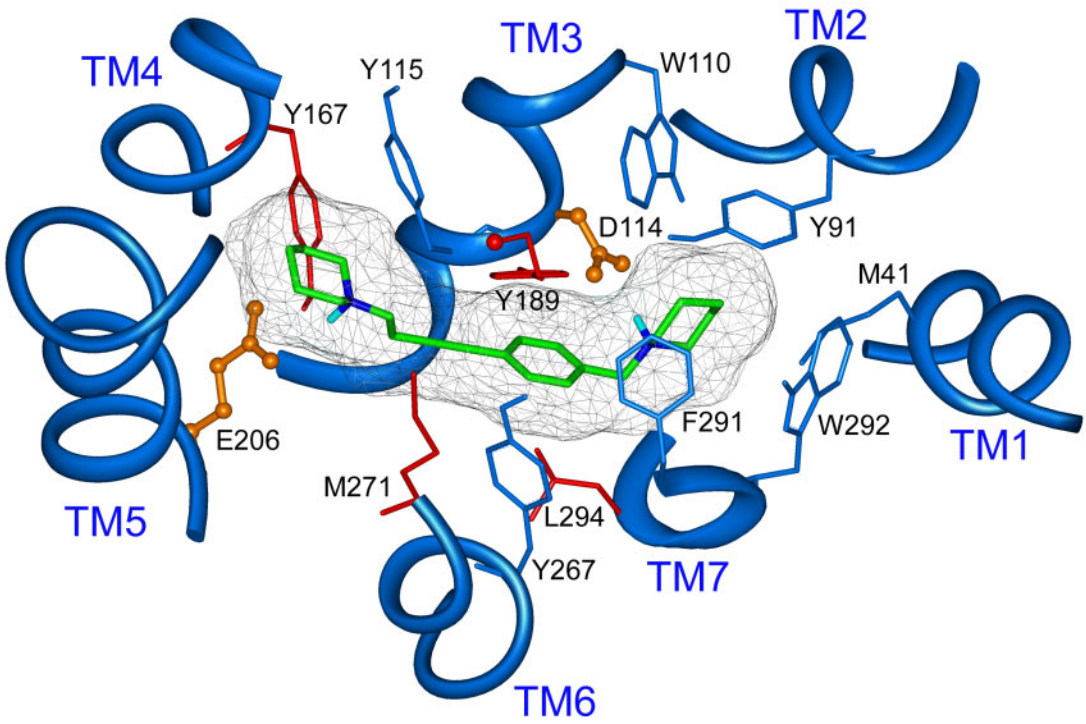


Figure 7.45: Ligand JNJ1 in the hH_3R binding site. Amino acids that supposedly account for the subtype specificity compared to the hH_4R are held in red.

the endocyclic nitrogen of the quinoline moiety in FUB836, the cyano moiety of the aminoalkoxybiphenylnitrile series published by Faghih and coworkers [42] as well as amide groups in the aminoalkoxy-biaryl-4-carboxamide series (see figure 1.11, compound A-349821 for a representative of this series).

7.6.2 Suggested Structures for Experimental Testing

Some interesting compounds for experimental evaluation have been already depicted in figures 7.29 and 7.37, more hits are depicted in figure 7.46. HTS docking of compounds from the WDI and MDB or the molecules preselected by pharmacophore based searches suggested several modifications to result in high docking scores:

- In many high scored docking solutions, the protonated nitrogen atom had shifted to the 4-position and a bulky substituent was attached to the nitrogen atom (e.g. a phenethyl moiety; see figure 7.46: HTS-09319, HTS-07217, TB-00019, HTS-02216, AW-00833, HTS-03517). This way, the flexibility of the linker chain would be reduced and the linker at the same time enlarged, which could improve interactions with L294/7.42. An attached aromatic system could increase affinity as a good stabilisation within the aromatic cluster can be expected.

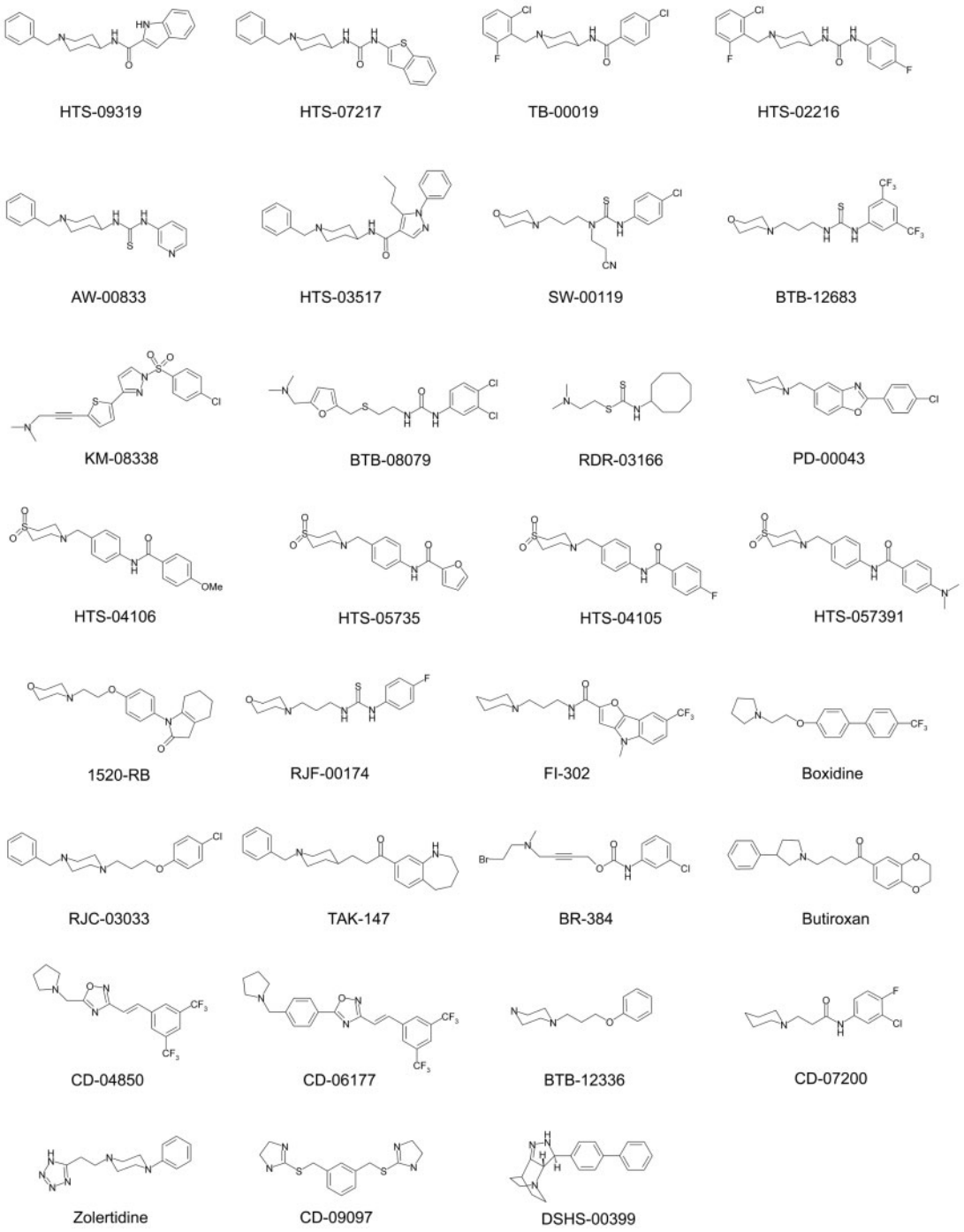


Figure 7.46: Hits found by docking or pharmacophore based search of the MDB and WDI.

Similar piperazine containing compounds have been recently published by Novo Nordisk (see figure 7.47). [132] In the resulting piperazine derivatives, a bulky substituent at the 4-position was required for affinity (e.g. 3-pentyl or cyclopropyl) while heteroatoms at this position led to inactive compounds. This SAR is in well agreement with the proposed model. The bulky substituent will be accommodated within the aromatic cage bordered by F291/7.39, W292/7.40, W110/3.28 and

Y91/2.61. Additionally, a bulky substituent allows the positive charge be distributed on the hydrocarbon atoms, resulting in a decreased polarity. On the quinoline group, substituents at position 3, 4, 5, 7 and 8 were not tolerated. Only at position 6 substituents were well accepted. Here, lipophilic substituents (Cl, CF_3) or sterically demanding moieties (cyclohexyl, 4-fluorobenzoyl) lowered affinity. Small polar groups (OMe, CN, pyrazoyl, cyclopropanoyl) increased affinity. If this SAR is seen in the context of the binding site, the substituted aromate would be in contact with M271/6.55. Small polar groups could decrease electron density and facilitate a charge transfer complex. At the same time, the space at this site is limited due to L199/5.39.

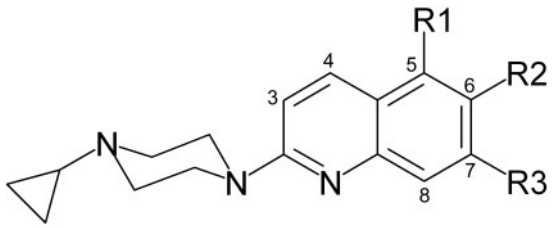


Figure 7.47: A series of Novo Nordisk compounds recently published. [132]

- As will be extensively discussed in section 8.4.1, imidazole-containing compounds that lack a protonated moiety in the side chain likely interact with D114/3.32. From superimposing FUB836 and 1S,2S-GT2331 (see figure 7.14) it can be expected that the imidazole moiety will form a hydrogen bond reinforced salt bridge with its Nπ-atom. Yet, a second hydrogen bond acceptor site can be expected for the interaction with the Nτ-atom. In the herein described complexes, the Nτ-atom of the imidazole-moieties was observed to interact with either the backbone carbonyl moiety of F291/7.39 or pointing towards the aromatic plane of W292/7.40.
  Some ligands with sterically demanding headgroups capable of simultaneously interacting with D114/3.32 and F291/7.39 were found during the HTS, which could clarify if an interaction with F291/7.39:O is likely to happen (see figure 7.46: DSHS-00399).

- Also cyclic amidines could represent an interesting headgroup. This moiety strongly resembles imidazole due to the presence of two hydrogen bond donor moieties and steric similarities. An amidine moiety has been observed also in antiallergic drugs targeting the hH_1R (see CCD-entry: TUDSEQ). Amidines are however rather polar. In order to not limit the use of such compounds to targeting the peripheral hH_3R, oxamidines could lower the pK_a value (pK_a (amidine) = 11-12; pK_a (oxamidine) = 10.46). Using such moieties could also be interesting in agonists as it could potentially replace the imidazole moiety.

- Several hits had sterically demanding lipophilic moieties attached to the linker group. As will be discussed in section 8.4.3, blocking the side chain conformational switch of F207/5.47 could represent a strategy for inverse agonist design. Ligand extending into the gap between helices 5 and 6 could accomplish this task.

The proposed model of the hH_3R binding site will certainly not be able to explain all structure affinity relationships observed for hH_3R ligands. Yet, models can help to better understand experimental data and imply further experiments to prove the model either right or wrong. A possible explanation for SAR observed for a series of piperidine-containing carbamates published by [131] (see figure 7.48) shall be given: No significant influence of the spacer length *m* and *n* could be observed, and the affinity of the ether containing lead compound (pK_i = 7.77 on rat cortex) could neither be reached by the spacer length combination (*m* = 3, *n* = 3) nor (*m* = 1, *n* = 3) that should result in the same separation of the protonated nitrogen atom and the aromatic moiety. Lazewska *et al.* stated that also the lack of the 4-Cl-substituent — present in the ether analogue — could not be fully responsible for this observation as the hydrogen analogue of the lead compound still displayed only 60% of the affinity of the lead.

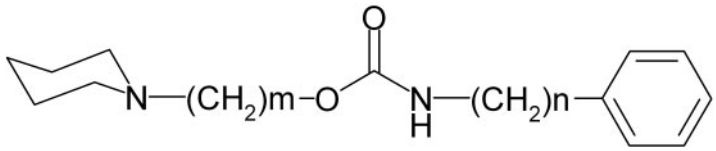


Figure 7.48: For *m* and *n* the combinations 3/1 (5,88), 3/2(5.76), 3/3(6.28), 3/4(6.39), 3/5(6.47), 4/4(6.28), 5/4(6.48), 6/4(6.41), 7/4(7.02) and 8/4(<6.9) were tested. pA_2 values determined on rat H_3R are given in brackets.

In context with the suggested hH_3R model proposed, the lack of preference for a specific spacer length might indicate that the carbamate moiety is not exclusively involved in hydrogen bonding interaction but can form π-π contacts with Y189/5.29 and/or Y267/6.51. The low affinity of compound (*m* = 1,*n* = 3) can be explained as in this case the polar carbamate moiety were located in the hydrophobic cleft bordered by L294/7.42. The lower affinity of the ligand with (*m* = 3, *n* = 3) could be explained by the different position of the carbamate moiety compared to the ether moiety. Most likely, the polar ether group places the aromatic ring in a reasonable way for a charge-transfer complex with M271/6.55. Presence of a carbamate moiety might result in a parallel-displaced arrangement with Y189/5.29 resulting in a lower position of the entire molecule. While in the ether lead compound the oxygen-atom might be involved in a hydrogen bond with Y267/6.51, in the carbamate-containing compound, the carbonyl-oxygen might be involved in hydrogen bonds as the largest negative partial charge is located on this atom. Why however imidazole-containing carbamate ligands are higher affine at the H_3R

than piperidine-containing carbamate ligands is difficult to answer and might depend on the more favourable steric constraints an imidazole moiety imposes on the side chain compared to the piperidyl moiety.

Chapter 8

Discussion

8.1 Generation of a Homology Model of the Human Histamine H_3-Receptor and Comparison with other hH_3R Models

The human histamine H_3-receptor forms part of the family of biogen aminergic G-protein coupled receptors. In this protein family, the structure of bovine rhodopsin is the only available X-ray structure. Although sequence conservation within the GPCR family is quite low, the arrangement of the seven helices as well as several motifs (e.g. D(E)RY in helix 3, NPXXY in helix 7) and numerous amino acids at the transmembrane level [80] seem to be conserved among family A receptors, and there is evidence that not only the overall topology but also the mechanism of activation is conserved. The fact that bovine rhodopsin was crystallised in its inactive state has important implications for the applicability of models based upon this structure as it will limit their use to study inverse agonists that are thought to stabilise the inactive receptor conformation. Several activation models of bovine rhodopsin exist and indicate that upon activation a cleft opens at the cytoplasmatic end of the helix bundle and exposes various regions for interaction with the G-protein. [133] Translational movements, rotation of helices and kinking of helices accompany the opening of this cleft. Thus, significant changes have to be expected within the binding site which is located in the upper half of the helix bundle. Studies described herein thus focused solely on inverse agonist/receptor complexes although it was sought to find a more general explanation for agonism versus inverse agonism. For simulating ligand/receptor complexes, a model of the human H_3R receptor was built based on the crystal structure of bovine rhodopsin. The model lacked part of the 3^{rd} intracellular loop region, which is known to couple to G-proteins. Neglection of this loop was however not expected to influence the binding pocket that lies distant from the intracellular loop region

within the helix bundle. One intrinsic water molecule was considered in the model, which is located in proximity to D83/2.50 (see figure 7.6) and links helices 2, 3 and 7. Regarding potential pK_a-shifts, analysis of the VUF5300-complexed hH_3R model with the program UHBD suggested — in analogy to D80/2.50 in bovine rhodopsin — a deviant pK_a value for D83/2.50, which was predicted to be present in its protonated state. D114/3.32 and E206/5.46 were predicted to be negatively charged and produced a pK_a-shift of close by basic moieties such as imidazole-groups favouring their presence in the protonated state.

The sequence of the hH_3R was analysed by applying several transmembrane prediction and secondary structure prediction algorithms. Secondary structure prediction algorithms are however an imprecise tool, with a prediction accuracy of only 76%. [64] Still, combined application of several algorithms that had been evaluated on the sequence of bovine rhodopsin allowed to deduce that also in the hH_3R a helix 8, parallel to the membrane plane, would be present. Prediction of putative transmembrane helices has been reported to have a prediction accuracy of 90-95% [64] and was valuable in order to decide between two alternative alignments depicted in figure 7.2. The low conservation of helix 5 of the hH_3R was also mentioned by Uveges and coworkers, [27] who built a model of the hH_3R as well based on the structure of bovine rhodopsin. After a sequence-structure alignment, the model of Uveges *et al.* was optimised using energy minimisation algorithms and subsequently a histamine ligand was manually docked into the model. Manual adjustments of helix 5 were however required in order to force an interaction of E206/5.46 with the imidazole moiety of histamine. All loop regions were then removed and the complex further optimised. The fact that manual adjustments of helix 5 were required in order to achieve an interaction with E206/5.46 quite probably arose from the fact that a model based on an inactive receptor conformation had been used to simulate an agonist/receptor complex. Given the complex activation process, it can however not be anticipated whether this procedure will be successful. In contrast to the receptor model of Uveges, in the model herein presented all extracellular loop regions were explicitly considered, as several mutation studies showed that residues from the second extracellular loop (especially residues adjacent to the cysteine residue involved in the disulfide linkage) could be involved in ligand binding [99, 100] or were responsible for receptor subtype specificity. [101, 102] In the hH_3R and hH_4R sequences, residues adjacent to the conserved cysteine residue in the E_2 loop differ: in the hH_3R at position 5.27 a histidine residue and at 5.29 a tyrosine residue is found while in the hH_4R two glutamic acids reside at corresponding positions. In the models herein proposed these residues were pointing into the binding pocket and participated in ligand binding. Yet, mutational analysis of these residues will be required in order to prove their participation in ligand binding.

Further models of the hH_3R based upon the structure of bovine rhodopsin were published

in 2001 by Sippl [134] and in 2003 by Yao *et al.*. [46] Both models were built in order to explain species differences observed between the rat and human H_3R that differ only in 5 amino acids on the transmembrane level. Both models were relaxed by means of energy minimisation and MD simulation, showed however striking differences in ligand placement. In the model of Yao and coworkers, the inverse agonist A-304121 extended from D114/3.32, which formed a salt bridge with the piperidyl moiety orthogonal to the membrane plane down to residue D80/2.50. While ligands in this model were located between helices 2, 3 and 7, ligands in the model of Stark and coworkers extended from D114/3.32 towards E206/5.46, accounting better for mutational data.

In none of the published models, problems in generating the ligand/receptor complexes were mentioned. Yet, after a sequence structure alignment had been carried out and amino acid side chain were added using `SCWRL3.0`, clashes were observed at two sites as described in section 7.1.3 and compound FUB836 could not be automatically docked into the binding pocket. In order to resolve these unfavourable contacts, unconstrained MD simulations were carried out in a CCl_4/water environment. This membrane mimic had been previously analysed in MD simulations of bovine rhodopsin (see section 4.4) and a decreased viscosity compared to a DPPC/water environment had been observed. As the scope of these simulations was to identify incompatible amino acid placements upon distortions of the receptor model, a CCl_4/water environment seemed more suitable for these simulations.

Regarding the placement of Y167/4.57 a side chain rotamer pointing into the binding site was favoured. Starting with this conformation, a hydrogen bond cluster emerged, involving residues T119/3.37, Y167/4.57 and E206/5.46; additionally, this placement was consistent with mutational data available for other biogen aminergic GPCRs. As will be described in detail in section 8.4.2, due to this hydrogen bond cluster, Y167/4.57 was highly constrained in the human H_3R-model and was not fully available for ligand interaction. In contrast, in the rat H_3R the mutation T3.37A would break up this hydrogen bond network and Y167/4.57 is assumed to better interact with compounds containing a hydrogen bond acceptor moiety at this site.

Clashes between amino acid side chain were also observed within an aromatic cluster formed by residues Y91/2.61, W110/3.28, F291/7.39, W292/7.40 and W295/7.43. Clashes arose, as in bovine rhodopsin the corresponding residues were often sterically less demanding (Y2.61T, W3.28E, F291A, W292F, W295K). Special attention was given to W295/7.43, which corresponded to the lysine residue K296/7.43 in bovine rhodopsin that was involved in the Schiff-base linkage. In the crystal structure of bovine rhodopsin a kink was observed around K296/7.43 in helix 7. Adaptation of the backbone of bovine rhodopsin during the sequence structure alignment resulted in a predicted side chain conformation for W295/7.43 that in analogy to K296/7.43 in bovine rhodopsin pointed into

the binding pocket. This however blocked the binding pocket and inverse agonists could not be accommodated in an orientation where a hydrogen bond reinforced salt bridge interaction with D114/3.32 could be established. Resolving the clashes within this aromatic cluster and deriving a ligand compatible binding site mainly depended on the side chain rotamer for W295/7.43. Choosing a placement for this residue was however difficult as several uncertainties had to be faced:

- In the structure of bovine rhodopsin a kink was observed at position 7.43, which corresponds to W295/7.43 in the hH_3R. K296/7.43 in bovine rhodopsin is involved in the Schiff-base linkage to retinal. A covalent linkage to a ligand is unique for the rhodopsin structures. It is difficult to judge if the observed kink is a consequence of this covalent linkage — and thus a characteristic feature of the structure of bovine rhodopsin — or represents a generalised structural feature in off-state receptors.

- Two alternative placements for W295/7.43 were thinkable, independent of whether the kink was present or not. W295/7.43 could either point into the binding pocket or into the gap between helices 1 and 7. Mutational analysis of these residues could potentially favour a placement, yet, such data is not available for the hH_3R. Point mutations were however carried out in other biogen aminergic GPCRs containing a similar aromatic cluster; yet, could not give definite results.

- Given the different SAR of agonists and inverse agonists described in section 1.4.1 it can be expected that the binding site in proximity to D114/3.32 differs in the on- and off-state of the receptor. It is however not known whether inverse agonists are capable of triggering a suitable conformation of the aromatic cluster, or, in other terms, whether an induced fit might be expected in this region. Under such conditions, simulation of uncomplexed receptor models will not result in reasonable binding site conformations.

- Whichever start conformation was chosen for W295/7.43, the initial placement was energetically unfavourable, and during a minimisation procedure not all constraints could be resolved. Starting however with a structure high in potential energy can result in different pathways on the energy landscape that might not be connected due to high energy barriers. Thus the outcome of one simulation might not be statistically representative. This effect has been described in section 7.1.3, where different binding site geometries were obtained starting from the same start geometry (see figure 7.4). This effect is however not contradictory to the deterministic behaviour of MD simulations as velocities were randomly generated for all atoms at the start of a simulation. A reproducible simulation would only be obtained if the simulations started from an identical coordinate- and velocity-file. However, the fact that the

MD simulations resulted in significantly different binding pockets is an indication that the initial placement had a low intrinsic stability and high corresponding potential energy. The simulation setups could thus evolve in quite different orientations resulting finally in different binding sites. Little is usually known on the shape of the energy landscape of a protein structure. Depending on the number, width and depth of potential energy valleys it might be impossible to sample the entire conformational space, especially at low temperatures, where the kinetic energy is too small to overcome higher barriers in potential energy. Potential workarounds for this “sampling problem” would be a significantly increased system temperature, the application of stochastic sampling methods (e.g. Monte Carlo Methods) or the application of simulated annealing protocols. Several short lasting simulations might in any case result in better sampling of the potential energy hyperface than one long lasting simulation.

Several setups were analysed in regard to the structural preservation of the model (especially of helix 7) and ligand compatibility:

- A trimethylammonium complexed hH_3R model with W295/7.43 in the placement suggested by SCWRL, pointing into the binding site, was analysed for 1 ns in a CCl_4/water environment. The start conformation as well as the end conformation blocked the binding pocket for docking of inverse agonists. Helix 7 adopted an idealised helical conformation devoid of the kink around W295/7.43.

- In order to obtain a binding pocket compatible with the accommodation of high affinity inverse agonists, further information was required in order to decide which amino acid side chain conformations should be used in the start conformation. For this purpose, ligand information was incorporated into the placement of amino acid side chains in an “inverse” docking approach (see section 7.1.4). Application of this protocol clearly suggested that W295/7.43 had to point into the cleft between helices 1 and 7 in order to accommodate ligand FUB836 in an energetically favourable manner. This approach proved to be a valuable tool in objectively generating ligand/receptor complexes, yet it also had a number of drawbacks. Possible rotamers will always depend on the backbone geometry. Thus, if adaptations in the backbone region occur, the set of compatible amino acid side chain conformations will change. Another point that has to be considered in this approach is that complexes were selected based upon their docking score. It is however a well known problem of scoring functions that ligand affinity is not well correlated to the obtained docking scores. Yet, in order to find reasonable complexes, this strategy is promising as it compares scores for one ligand involved in different complexes and not between different structures.

- When starting with the amino acid placement suggested by the “inverse” docking approach (W295/7.43 was pointing into the gap between helices 1 and 7), during the equilibration period, the rotamer suggested by SCWRL reestablished. This was quite likely due to the backbone tethers applied during equilibration and the interdependence of backbone geometry and amino acid side chain placement. After a 5 ns MD simulation, structural deviations within the *N*-terminal region of helix 7 were observed (see figure 7.8).

- An idealised helical conformation of helix 7 was introduced in the start conformation, as it had been observed after simulation in the CCl_4/water environment under point 1. In the start conformation, W295/7.43 was pointing into the binding pocket. However, a free volume for accommodation of inverse agonists was available below the aromatic plane. After a 5 ns MD simulation, the ideal helix conformation of TM7 had been preserved, however W295/7.43 had moved, thus blocking the binding pocket.

- Studying ligand/receptor complexes appeared to be advantageous as the ligand would impose further conformational constraints on the binding site. This held especially true for the hybrid compound FUB836-VUF5391, which was expected to extend into the gap between helices 1 and 7. Interestingly, distinct to the simulations described above, if W295/7.43 was pointing into the binding site, the kink was preserved in this simulation; if pointing outside, the kink disappeared.

In conclusion, it seems that helix 7 can adopt either an idealised helical conformation devoid of the kink around W295/7.43 or the kink can be preserved. The switch from one conformation to the other seems to be quite delicate and not well reproducible. It might be speculated, due to known SAR of agonists and inverse agonists, that the switch between helical geometries and amino acid side chain placement could be involved in the activation process. While in the agonist binding site the available volume seems to be significantly reduced, resulting in a significant loss in activity for histamine derived agonists such as Nα-ethyl- and -propyl-histamine (see figure 1.9), the inverse agonist binding pocket can accommodate bulky piperidine, pyrrolidine or piperazine moieties. The MD simulations carried out failed however to clearly favour one binding site. Analysis of mutational data should theoretically favour one placement, as in the case where residue W7.43 is pointing into the gap between helices 1 and 7, it will not be primarily involved in ligand binding and thus a point mutation should not have pronounced effects on binding affinity. As no mutational data exists for this residue in the hH_3R, mutational data of other biogen aminergic GPCRs was analysed. Lu and coworkers [121] have carried out point mutations of residues Y7.39 and Y7.43 in the M_1 muscarinic acetylcholine receptor. Mutations to phenylalanine resulted in a 6 to 9-fold decrease of NMS (N-methylscopolamine) affinity,

while affinity of QNB (quinuclidinyl benzilate) was not affected by the mutations to alanine. For the α1-adrenergic receptor, F7.39 was reported to be involved in inverse agonist binding. [135] Thus for both F291/7.39 and W295/7.43 a participation in ligand binding was reported. Interpretation of mutation studies is additionally complicated, as mutation of a sterically demanding aromatic residue to alanine can always have indirect effects; and a reduction in binding affinity cannot be simply interpreted in a way that the corresponding residue participates in ligand binding. Additionally, the highly conserved residue W292/7.40 seems to be involved in receptor activation, [136] which further complicates interpretation of the data.

Thus, MD simulations with different inverse agonists and different start conformations for the binding pockets geometries resulted in alternative binding sites. For a subsequent high throughput screening procedure, the consideration of alternative binding pocket geometries was however impracticable. Thus criteria had to be set in order to choose the most appropriate binding site geometry. Besides a good overall structural preservation of the model, side chain that were conserved with bovine rhodopsin were required to adopt a comparable conformation. Additionally, the enrichment of hH_3R inverse agonists, with special focus to sterically demanding ligands against randomly selected WDI compounds applying the hH_3R binding site was analysed. This finally resulted in the selection of a binding pocket that had resulted from simulation of the complex FUB836/hH_3R. The binding pocket is depicted in figure 7.25. Several ligand/receptor complexes are depicted in figures 7.42, 7.43, 7.44 and 7.45 and important structural features are described in section 7.6.

Docking scoring functions are usually biased towards higher molecular weight compounds, as such compounds have more possibilities to interact with a given binding site. In order to reduce this bias, the resulting `GOLD` scores were multiplied by the correction term $1/\sqrt{N}$, where N is the number of non-hydrogen atoms. [137] Application of this correction term significantly improved the enrichment of hH_3R ligands, as now smaller ligands with highly specific interactions were favoured (see figure 7.26). Screening against an unfocused library has been reported to strongly depend on the choice of the random compound library. [129]. Significant enrichments can be easily obtained if the test library and the random compound library have different 1D ligand properties. In fact, the chemical space spanned by the randomly selected WDI compounds and the hH_3R inverse agonists varied significantly (see figure 7.27). Thus, application of a focused library was recommended. While in case of the unfocused approach, 11% WDI compounds were ranked among the top scored 80% hH_3R compounds, now, the value increased to 23% (see figure 7.28). Yet, due to the similarity in 1D properties, high scored structures within this WDI subset also represented interesting structures with potential affinity towards the hH_3R. In figure 7.29 some hits are depicted. Butoprozine partly resembles ciproxifan (see figure 1.11) and

piperazine-containing ligands with a bulky 3-substituent such as A-303112 have been recently suggested by NovoNordisk (e.g. NNcmp1 in figure 1.11).

In conclusion, based upon the screening against a non-focused and focused WDI subset, a satisfactory enrichment of validated hH_3R inverse agonists could be obtained for a binding site, which had resulted from simulation of the complex FUB836/hH_3R. This binding site was subsequently used in a HTS of WDI and MDB compounds.

8.2 High Throughput Screening by Docking

In order to accelerate the search for new drug molecules, the strategy of high throughput screening (HTS) can be applied. In a solely experimental drug discovery project, HTS is carried out subsequent to the target protein identification and the development of an appropriate screening assay. Large numbers of chemical compounds are then tested in a "try and error" approach for potential binding to the protein target. Although significantly improved in recent years, experimental HTS remains to be expensive. In the virtual HTS counterpart, ligand molecules can be flexibly placed into a modelled binding site and the quality of the fit is evaluated by mathematical scoring functions. Commonly used structure databases for virtual HTS in medicinal chemistry are the Maybridge Database (MDB, currently containing about 60000 ligands) and the World Drug Index (WDI, containing 58000 ligands). Both databases have the advantage that they contain mostly drug-like compounds. The terminus *drug-like* thereby embraces a combination of various molecular properties (such as hydrophobicity, hydrogen bonding characteristics, molecular size, ...), which determines whether a particular molecule will be a drug or a non-drug. Some ranges experimentally observed for drug molecules are for example a molecular weight below 500 Da, a lipophilicity range from $logP = -2$ to 5 and a small number of hydrogen bond donors and acceptors.[1]

As described in section 7.4, the validated binding pocket was used as a filter in screening the WDI and the MDB. Ligands with piperidine, piperazine, pyrrolidine and imidazole moieties were preselected and docked into the binding site using the program `GOLD`. For this HTS approach, the "library screening" genetic algorithm had to be applied due to its higher performance. Compared to the "2 times speed-up" algorithm, application of this algorithm resulted however in a worse separation of validated binders from compounds of the focused library (see figure 7.30). In this regard, hH_3R inverse agonists are especially problematic, as they usually contain a high number of rotable bonds. For such

[1] Lipinski's rule of five, that is often mentioned in this context, only describes the passive diffusion across membranes and predicts a low bioavailability if 2 of the following rules are violated: $MW > 500$, $logP > 5$, number of hydrogen bond donors > 5, number of hydrogen bond acceptors > 10.

compounds application of the “2 times speed-up” or “3 times speed-up” algorithm had been recommended by the authors of GOLD, which is however computationally too demanding for virtual high throughput screening. A potential workaround would lie in prescreening of WDI and MDB with the faster algorithm and screening a high scored subset with the more accurate, but slower algorithm as carried out in section 7.5 for a WDI subset.

The docked solutions were ranked according to the GOLD scores obtained. Also the XSCORE scoring function was tested, resulted however in a worse separation of hH_3R inverse agonists and WDI and MDB compounds. The corresponding histograms are depicted in figure 7.31. The distribution of validated hH_3R inverse agonists is shifted by a value of 20 towards higher docking scores. This indicates a satisfactory separation; at the same time, however, depending on the choice of the cutoff score, a significant percentage of hH_3R ligands will be withheld by this filter. This is especially problematic as no correlation exists between the docking score and the ligand affinity. A GOLD score of 40 seemed however to be a good compromise as it resulted in retrieval of 66.5% of the validated hH_3R ligands and limitation of WDI and MDB compounds to 1839 structures (1.7% of the complete number of ligands deposited in WDI and MDB).

Additionally, the enrichment over a subset of biogen aminergic receptor ligands was tested in order to avoid artificial enrichment of ligands targeting biogen aminergic GPCR. The enrichment proved however to be similarly significant as the enrichment over the entire WDI and MDB (see figure 7.32).

8.3 Pharmacophore Based Screening

The hH_3R inverse agonist data set is quite challenging for an application in pharmacophore based screening. The high flexibility of the ligands together with the huge structural diversity makes an application of classical pharmacophore approaches difficult: e.g. in many pharmacophore projects, initially, common structural features essential for high affinity have to be identified either manually or automatically. Comparison of the hH_3R ligands FUB335 [34] and compound 7n [138] (see figure 8.1) indicates that the identification of common features is not a trivial problem.

As shown in equation 2.4, the affinity of a drug (K_i) is correlated to the free energy of binding (ΔG). The free energy of binding, in turn, comprises an enthalpic and an entropic part:

$$\Delta G = \Delta H - T\Delta S \tag{8.1}$$

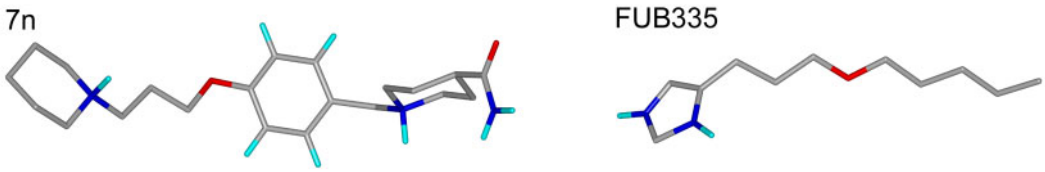


Figure 8.1: Compound 7n [138] and FUB335 [34] differ significantly in the number of pharmacophoric features.

While the enthalpic part (ΔH) describes the match in steric and physico-chemical properties between the ligand and the binding site, the entropic part accounts for desolvation of the receptor binding site and the ligand that must take place prior to binding. Additionally, $T\Delta S$ accounts for the loss of conformational freedom of both ligand and receptor once the bioactive conformation is "frozen" in the binding site and reduced to only vibrational motions.
For tight binding, a ligand molecule must form specific interactions with the binding site and fit the shape of the binding pocket which will also importantly influence the selectivity of a ligand for a given binding pocket. Additionally, the desolvation costs of the compound should be small, which can be achieved by an overall apolar character of the molecule, and the molecule should be rather rigid in order to have small losses of conformational freedom upon binding.

The concept of enthalpic and entropic binding is stressed by molecules FUB335 and 7n: binding of FUB335 will be to a significant extent entropic-driven. Desolvation of the long aliphatic hydrocarbon chain will lead to a loss in entropy as surrounding water molecules are freed from these unfavourable interactions. Additionally, the loss of conformational freedom in the binding site can be assumed to be small due to the fact that much larger ligands can be accommodated in the same binding pocket so that no tight binding is to be expected. The enthalpic contribution to binding affinity will be restricted to the interactions of the imidazole moiety, the ether moiety and van der Waals contacts to the hydrocarbon chain in regions where a sufficiently tight interaction with the binding pocket is achieved. Compound 7n on the other hand is rich in functional moieties thus representing an enthalpic binder while entropic terms such as desolvation will most likely even reduce binding affinity.

In order to optimise affinity both the enthalpy contribution (via a good steric fit and specific interactions) as well as the entropic part (via a low desolvation and a high rigidity) must be optimised. The netto effect of an additional hydrogen bond acceptor or donor moiety is due to the simultaneous influence of enthalpy and desolvation difficult to judge *a priori*. However, in terms of ligand specificity, a high contribution of enthalpy to the free energy is required. These concepts are however not taken into account when identifying common

structural features in programs such as Catalyst. An additional basic moiety in ligand molecule 7n would quite likely not be regarded as an essential feature in an automated hypothesis generation, as no significant improvement of binding affinity could be observed compared to compound FUB335.

An alternative strategy was thus chosen, consisting in the definition of few and loose pharmacophoric features (in order to account for the great structural variability of hH_3R inverse agonists) and the incorporation of the molecules' shape as an additional feature in order to generated a sufficiently restrictive pharmacophore model. The conformation of the molecules on which the pharmacophoric features were defined, was derived by conformational analysis of the ligand molecules and docking into the hH_3R binding site. Combination of three pharmacophore models and further incorporation of forbidden volumes resulted in retrieval of 93% of the hH_3R inverse agonists simultaneously reducing the number of WDI and MDB compounds to 2668 (2.5%). The percentage of hH_3R ligands within each pK_i-cluster is depicted in figure 7.36d. While the cluster of high affinity compounds could be retrieved almost complete, in the cluster of inactive compounds (pK_i < 6), 35% of the ligands could be excluded by the filter.
While screening with pharmacophore models consisting of more, and more stringent features in absence of a shape query did not result in interesting hits, further screening of the subset of 2668 compounds of the WDI and MDB with a leave-one-out filter defined on FUB836 could further improve results. Although only 106 compounds passed the leave-one-out filter, application of this filter resulted in a better enrichment of high affinity compounds and could withheld 96% of inactive hH_3R inverse agonists (pK_i < 6) present in the hH_3R library. When screening the preselected 2668 hits, 320 compounds matched the leave-one-out filter. The higher number of features included in this pharmacophore model is also advantageous in terms of higher selective compounds as previously explained.
Hits found by pharmacophore based searches were subsequently docked into the modelled binding site using the program GOLD. This way, molecules that exceeded the volume of the binding site or contained misplaced functional moieties could be excluded.

Compared to the HTS by docking, application of the pharmacophore based search resulted thus in significant better results. While in the docking approach 66.6% of the hH_3R ligands were retrieved, while limiting the number of WDI and MDB compounds to 1.7%, application of a pharmacophore based search allowed retrieval of 93% of active compounds, while reducing the number of WDI and MDB structures to 2.5%. Yet, both methods in combination might give the best result as HTS by docking imposes no constraints regarding ligand moieties or size while in the pharmacophore based approach herein described, the definition of a reasonable shape required knowledge of a sterically demanding ligand.

8.4 Analysis of the hH_3R Binding Pocket

The hH_3R binding site was analysed in order to answer a number of questions, which have been discussed controversially in the available literature, i.e. the orientation of hH_3R ligands in the binding site and potential reasons for observed species differences. Additionally, strategies for an inverse agonist design shall be discussed.

8.4.1 Orientation of H_3R Ligands in the Binding Pocket

For biogen aminergic GPCRs it is commonly assumed that the protonated moieties of biogen amines — the natural agonists — are in contact with a conserved aspartic acid in helix 3. A corresponding residue, D114/3.32, is also present in the hH_3R and it is expected that this residue is in contact with the Nα-amino group of histamine. The point mutation D3.32E resulted in a drop of affinity for histamine from pK_i = 7.74 to pK_i = 6.65 [27, 125], while the mutation D3.32N resulted in an inactive receptor [27, 28] supporting the idea that a salt bridge interaction is formed between the basic moiety in the histamine side chain and D114/3.32.

Another important residue in histamine binding is glutamate E206/5.54, in helix 5. This residue is thought to interact with the imidazole moiety of histamine (see figure 8.2a). Mutation to alanine decreased histamine affinity for more than one order of magnitude. [27, 125] Interestingly, the point mutation E206Q resulted in a constitutively active receptor [28] with a complex pharmacological profile: no activity was observed for the agonists R-α-methylhistamine and iodoproxyfan on the E206Q mutant. However, when inspecting affinity, a drop of three orders of magnitude was observed for R-α-methylhistamine (pK_i(E5.46) = 9.0; pK_i(Q5.46) = 5.9), while the affinity of iodoproxyfan was not changed drastically (pK_i(E5.46) = 10.3; pK_i(Q5.46) = 9.1). Jacobsen and coworkers thus speculated that some agonists such as R-α-methylhistamine interacted with both D114/3.32 and E206/5.46 simultaneously, while others such as iodoproxyfan did not. These results are consistent with the work of Uveges *et al.*, where the mutation E206A did not significantly affect binding of [^{125}I]iodoproxyfan but the mutation of D114N or D114E abolished binding.[2] [27] It might be thus expected that agonists containing only an imidazole moiety and no further basic moieties in the side chain interact with D114/3.32 rather than E206/5.46 (see figure 8.2c).

This observation is especially interesting as it has been assumed for a long time that a specific imidazole binding site involving E206/5.46 might exist in proximity to helix 5 for both agonists and inverse agonists supporting orientation 8.2b. [36] This speculation

[2] In the work of Uveges *et al.* iodoproxyfan is still referred to as an antagonist. However, this ligand has been recently reassessed as an agonist.

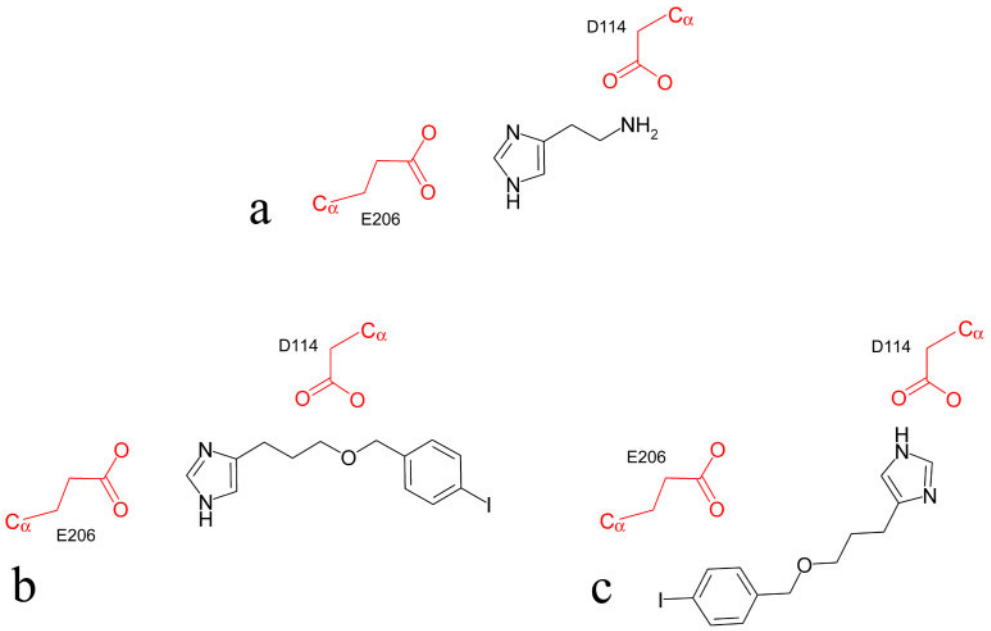


Figure 8.2: Possible orientations of H_3R agonists in the binding site. **(a)** Histamine is expected to interact with the basic Nα-amino group with D114/3.32 while the imidazole moiety is in contact with E206/5.46. **(b,c)** For agonists such as iodoproxyfan that do not contain a basic moiety in the imidazole side chain, two possible orientations are thinkable.

has originated from the observation that a 4(5)-substituted imidazole moiety was essential in agonist and inverse agonist binding. While this does no longer hold true for inverse agonists, all known agonists do still contain an imidazole moiety. Still, the mutation data strongly suggest that the imidazole moiety may as well be in contact with D114/3.32 casting doubt on the existence of a imidazole specific binding site. An orientation as in figure 8.2c is further supported by known SAR of H_3R agonists. While one or two methyl groups on the histamine Nα moiety increased affinity, larger moieties such as a ethyl- or propyl-substituent resulted in a significant loss in activity probably indicating that the space in the agonist binding site is very limited. [24] Accommodation of an aromatic ring as in iodoproxyfan or a *t*-butyl-group of FUB 475 (see figure 1.9) would be inconsistent with this SAR.

Also some inverse agonists were tested on mutant hH_3-receptors reported by [28]. The point mutation E206Q significantly affected the binding of iodophenpropit, clobenpropit and NNC-0038-1035 but not of ciproxifan, thioperamide, GT2016 and GT2394. A common feature for all affected inverse agonists is that they contain structural groups (isothiourea and guanidine) which can interact with E206/5.46. Thus, also inverse agonists are capable of interacting with both acidic residues in the binding site.

No mutational data is available of how the mutation D114N and E206Q/A affects binding of imidazole-containing inverse agonists without a protonated moiety in the imidazole side

chain (e.g. ciproxifan). Thus, the orientation of these compounds in the ligand binding site is *a priori* not clear. However, given the fact that the imidazole group has been shown to interact with D114/3.32 in structurally related agonists (e.g. iodoproxyfan), this might be also expected for imidazole-containing inverse agonists. Additionally, imidazole-containing inverse agonists can reach a considerable size (see figure SCHpat in figure 1.10) and more space is available in proximity to E206/5.46 than to D114/3.32 for accommodation of sterically demanding lipophilic substituents.

In this work, imidazole-containing compounds lacking a basic moiety in the side chain were considered to interact with D114/3.32. Upon ligand docking, for some ligands both orientations were however observed. Yet, the similarity of imidazole-containing compounds (e.g. ciproxifan) with non-imidazole ligands that also contain a propyloxy-benzoic moiety (e.g. FUB833, FUB836) is striking and favours the orientation assumed.

8.4.2 Species Differences for the H_3R

Species differences, although not affecting agonist binding, [29] can influence inverse agonist binding. Significant differences in binding affinities for some ligands (e.g. ciproxifan) have been observed especially comparing rat and human H_3R test systems, while for other compounds (e.g. proxyfan, clobenpropit) comparable binding affinities on the hH_3R and rH_3R were observed. [29, 30] The human and the rat hH_3R only differ in five amino acids on the transmembrane level. [29] Carrying out the point mutations A119T and V122A in the rat H_3R was sufficient to restore the binding profile of the hH_3R. This was shown by [41] for compound A-304121 (see figure 1.11), which is one of the most sensitive ligands in this respect: while a pK_i of 8.6 was observed on the rat H_3R, affinity was 300-fold reduced on the human H_3R.

Two models were so far published that try to explain species differences. In the model published by Sippl, [134] alanine 119 and valine 122 in the rat receptor were located in proximity to the binding pocket, indirectly influencing the geometry of the binding site. In the model suggested by Yao *et al.* agonists and inverse agonists were accommodated in a cleft between helices 2, 3, 6 and 7. The most basic moiety (including imidazole groups in case of compounds lacking a protonated moiety in the side chain) was interacting with D114/3.32. Agonists could additionally interact with D80/2.50 in helix 2. In this model, the ligands were directly interacting with D114/3.32 and V122/3.40 while A119/3.37 was located in proximity to the binding site. Yet, the proposed orientation for histamine — spanning from D2.50 to D3.32 — does not account well for mutational data as E206/5.46, which was reported to significantly influence histamine binding, does not interact with histamine at all.

The herein described model combines features of both models described above. Basic moieties and the imidazole group of compounds, such as ciproxifan (see figure 1.10), that lack a basic moiety in the imidazole side chain are proposed to interact with D114/3.32 via a hydrogen bond reinforced salt bridge interaction. Although the imidazole group is only slightly basic, in proximity to the acidic residue D114/3.32 a significant pK_a-shift can occur that favours binding of the protonated species. An orientation of the imidazole moiety towards D114/3.32 is essential in order to explain why species differences do not affect agonist binding. If the imidazole moiety interacted in a similar way with E206/5.46 in both agonist and inverse agonist complexes, all ligands would be similarly affected by species differences. In the proposed human H_3R model, a hydrogen bond cluster is formed by residues T119/3.37, Y167/4.57 and E206/5.46 (see section 7.1.3). While Y167/4.57 is anchored to T119/3.37 in the human H_3R, the mutation A119/3.37 would disrupt this interaction resulting in an increased conformational freedom for Y167/4.57, which could thus more easily interact with functionalities such as carbonyl moieties present in ciproxifan or A-304121 (see figure 7.44)

According to this model, species differences should be most striking for compounds capable of interacting with Y167/4.57. Compounds containing only a lipophilic moiety or a protonated group in the side chain should interact in a comparable way with both the rat and human receptors as these interactions are less directional. This is in good agreement with several H_3R inverse agonists. Yet, the presence of a carbonyl moiety or other hydrogen bond acceptor groups does not necessarily result in pronounced differences in binding affinity. Depending on the lipophilic group attached to the carbonyl group or the nature of the linker moiety, the entire ligand can reorient within the binding site resulting in different SAR.

8.4.3 Agonism versus Inverse Agonism

In 2000, De Esch and coworkers proposed a model in which the conformation of the aspartic acid D114/3.32 differed according to whether it was binding to an agonist or inverse agonist. [36] In this study, only inverse agonists were included that contained a basic moiety in the imidazole side chain. Compounds lacking this protonated moiety were excluded as it was not clear if such ligands would bind to the same receptor site. In this model, imidazole groups of all agonists and inverse agonists interacted with the same receptor site. The proposed model suggested a molecular switch mechanism based on the observation that the linker groups connecting the basic moieties in inverse agonists are significantly longer than in agonists. Since this model had been proposed, further knowledge was acquired on H_3R ligands, receptor activation, constitutive receptor activity and mutational experiments. As previously described, the existence of an imidazole

specific binding site must be doubted due to an increasing number of H_3R ligands lacking basic moieties in the side chain and point mutation experiments that suggest an interaction with D114/3.32 for these compounds. In order to account for constitutive activity, based upon this model, a spontaneous conformational change of the aspartic residue in absence of any interaction partner would be required. Additionally, there is increasing evidence that upon receptor activation large conformational changes occur that would decrease the distance between helices 3 and 5 thereby making additional conformational changes of D114/3.32 unnecessary.

More recently, Yao *et al.* proposed a model where the possibility to interact with D80/2.50 would decide if a ligand is an agonist or an inverse agonist. Histamine would thus interact with its basic moiety with D114/3.32 and the imidazole group with D2.50. Yet, also this model is not consistent with all information available on biogen aminergic GPCRs and H_3R ligands. In analogy to bovine rhodopsin, it might be thus expected that a water molecule is located in proximity to D2.50, which is present in its protonated state. The agonistic behaviour of ligands lacking a protonated moiety is also difficult to explain based on this model, as such ligands would interact with their protonated imidazole moiety with D114/3.32 but the non-basic side chain would extend towards D80/2.50. Additionally, sterically demanding ligands such as FUB833 might be difficult to accommodate in the proposed binding site if the extracellular loop region is explicitly considered in the model.

The complexity of the classification of ligands into agonists, neutral antagonists and inverse agonists has been increased by homologue series such as those depicted in figures 1.12 and 1.13. Minor structural variations pronouncedly influenced the pharmacological profile. In preliminary studies using the structure of bovine rhodopsin, protein adaptations after introducing an *all-trans*-retinal have been studied by means of MD simulation. Although a realistic simulation of the entire activation process is certainly beyond the scope of MD simulations, Furutani and coworkers generated the basis for a reasonable application of MD simulations in studying initial steps in the activation process by reconverting lumirhodopsin to rhodopsin at 77K. [7] At 77K, structural motions are to a large extent impeded, indicating that *all-trans*-retinylidene can be accommodated without large structural rearrangements. The initial side chain adaptations following the isomerisation but prior to larger conformational changes and de/protonation events were thus studied in a model of bovine rhodopsin. Observed adaptations consisted in a switch in side chain conformation of residue F212/5.47 and a cascade of amino acid side chain adjustments, resulting in the modification of the hydrogen bond pattern between helices 4 and 5 as well as helices 3 and 5. F212/5.47 is highly conserved within the family of biogen aminergic receptors and also present in the sequence of the hH_3R. The point mutation F207/5.47A in the hH_3R was shown to significantly reduce potency [27] making an implication in receptor activation probable. If a similar activation mechanism is assumed

for the hH_3R model, an essential parameter determining over an agonistic or inverse agonistic behaviour of ligands would be the volume of a lipophilic group attached to the linker moiety. This could reasonably explain the differences in pharmacological profiles of FUB373, FUB335, FUB407 and FUB397 (see figure 1.12). Although varying only in one or two methyl groups, the increased length of the lipophilic tails in FUB335 and FUB397 could block the side chain switch of F212/5.47. Interpretation of the profile within the VUF-series (see figure 1.13) is more difficult, as the orientation of the ligands within the binding site is not clear. Due to the increased basicity of the secondary or tertiary amines, this moiety should interact with D114/3.32 while the imidazole group could be in contact with E206/5.46. Accommodation of a bulky piperidine substituent as in VUF5300 contradicts however the SAR observed for histamine derived agonists where a n-propyl-substituent on the Nα-amino group abolished affinity. Regarding the available volume of the binding site in proximity to D114/3.32, significant differences may be expected between the in/activated state. As mentioned, Nα-propyl-histamine and compound 40δ (see figure 1.9) cannot trigger H_3R activation due to their bulky substituents. On the other hand, inverse agonists frequently bear a piperidine or piperazine moieties that can be additionally substituted. Thus, the free space available for ligand binding in the off-state model must be significantly larger than in the activated model.

It seems thus likely that several strategies for an inverse agonist design exist. Ligands orientated in a comparable placement as *11-cis*-retinal in bovine rhodopsin could impede the rotamer switch of F207/5.47 and block the activation at an initial stage. Compounds with bulky substituents on the protonable moiety in contact with D114/3.32 might avoid activation by impeding movements of helices 1 and 7. Inverse agonists with an increased distance between the basic moieties observed in agonists could finally avoid an approximation of helices 3 and 5.

In simulations with an *all-trans*-retinal, the rotamer switch of F212 provoked a modification of the off-state hydrogen bond pattern between helices 3, 4 and 5. This might also occur in the H_3R. Given that species differences affect only inverse agonist binding, and a possible model to explain species differences can be based on a modified hydrogen bond cluster between helices 3, 4 and 5, it is likely that E206/5.46 will not be involved in a hydrogen bond cluster linking helices 3, 4 and 5 after agonist binding. Disruption of the hydrogen bond cluster between helices 3, 4 and 5 could thus represent a further strategy for an inverse agonist design.

Chapter 9

Summary

The human histamine H_3-receptor is a G-protein coupled receptor and modulates the liberation of various neurotransmitters in the central and peripheral nervous system. The hH_3R is therefore a potential target in the therapy of numerous diseases. Although ligands addressing this receptor are already known, the discovery of alternative lead structures represents an important goal in drug design and an understanding of how drugs interact with their receptor at an atomic scale helps to find strategies for ligand optimisation and to propose mutational studies. The goal of this work was to study the human histamine H_3-receptor and its inverse agonists by means of molecular modelling tools.

In the first part of this work, extensive molecular dynamics studies were carried out with bovine rhodopsin, which was used as a template structure for the generation of the hH_3-receptor model. In order to find suitable parameters for a subsequent MD simulation of the hH_3R homology model, the effect of an *N*-terminal truncation, the influence of internal water molecules and the impact of different membrane mimics (CCl_4/water, DPPC/water) upon the structural stability of the bovine rhodopsin structure were analysed. The possibility of correctly predicting protonation states of buried residues was studied using a combination of MD simulations, multiple sequence alignments and application of chemoinformatic software (UHBD-program). Additionally, interhelical constraints upholding the inactive receptor conformation in bovine rhodopsin were analysed in order to derive constraints for the simulation of the hH_3R homology model. Finally, amino acid sidechain adaptations after the introduction of an *all-trans*-retinal were studied in order to rationalise an inverse agonist design in other biogen aminergic GPCRs.

In the second part of this work, a homology model of the hH_3R was generated and simulated in a membrane mimic incorporating the knowledge gained in part 1. The resulting binding pocket of the receptor was validated by evaluating the enrichment of tested hH_3R ligands when screening against a focused library of drug-like compounds. Subsequently, this binding pocket was applied as a filter in a virtual high throughput

screening by ligand docking.

Based upon the bioactive conformation of three hH_3R inverse agonists pharmacophore models were built for a ligand-based high throughput screening. Several interesting compounds could be suggested that shall now be experimentally tested.

Finally, visual inspection of the hH_3R binding site resulted in an identification of amino acids potentially involved in ligand binding and a possible explanation of observed species differences at an atomic scale.

Chapter 10

Appendix

10.1 Force Field Terms

10.1.1 Bonded Interactions

Bond stretching

Bond stretching is most adequately described by the Morse function (see equation 10.1). This function is however usually not applied in common force fields as three parameters D_l, α and r_0 would be required for each bond type and exponential terms are computationally quite demanding.

$$\mathrm{E(r)} = \mathrm{D_l}\left[1 - e^{-\alpha(r-r_0)}\right]^2$$

$$\begin{aligned}
&E(r) \quad \ldots \quad \text{potential energy} \\
&D_l \quad \ldots \quad \text{depth of potential energy minimum} \\
&\alpha = \omega\sqrt{\mu/2D_l} \quad \text{with} \quad \mu \quad \ldots \quad \text{reduced mass} \\
&\qquad\qquad\qquad\qquad\qquad\quad \omega \quad \ldots \quad \text{frequency of bond vibration} \\
&r \quad \ldots \quad \text{actual bond length} \\
&r_0 \quad \ldots \quad \text{reference bond length}
\end{aligned} \tag{10.1}$$

The relation between the Morse function and Hook's law:

$$E(r) = \frac{k}{2}(r - r_0)^2$$

$$\begin{array}{lll} E(r) & \ldots & \text{potential energy} \\ k & \ldots & \text{stretching constant of the bond} \\ r & \ldots & \text{actual bond length} \\ r_0 & \ldots & \text{reference bond length} \end{array} \tag{10.2}$$

that is commonly used to describe bond stretching becomes clear when the Taylor expansion of the Morse function is developed:

$$E(r) = D_l \alpha^2 (r - r_0)^2 \left[1 - \alpha(r - r_0) + \frac{7}{12}\alpha^2 (r - r_0)^2 - \ldots\right] \tag{10.3}$$

$$E(r) = \frac{k}{2}(r - r_0)^2 \left[1 - k'(r - r_0) + k''(r - r_0)^2 + \ldots\right] \tag{10.4}$$

Hook's law represents the first term of the Taylor expansion. Further consideration of a cubic or quartic term results in a more accurate approximation of the original Morse function and thus in more realistic models for bond stretching. However, close to the equilibrium distance of the bond, a harmonic potential reproduces the Morse function quite well. In the GROMOS-96 force field, Hook's law is used as the default method for calculating bond stretching. In case of highly distorted bond lengths one can however switch to the computationally more demanding Morse potential.

Angle Bending

Also the bond angle vibration between a triplet of atoms $i - j - k$ is commonly presented as a harmonic potential:

$$E(\theta) = \frac{k}{2}(\theta - \theta_0)^2$$

$$\begin{array}{lll} E(r) & \ldots & \text{potential energy} \\ k & \ldots & \text{bending constant of theangle} \\ \theta & \ldots & \text{actual angle} \\ \theta_0 & \ldots & \text{reference angle} \end{array} \tag{10.5}$$

Torsion Angle Rotation

The potential energy associated with the rotation around a bond is usually modelled as a cosine function:

$$E(\tau) = a\,[1 + \cos(n\tau - \Phi)]$$

$E(\tau)$	...	potential energy	
a	...	parameter describing rotational barriers	(10.6)
n	...	multiplicity	
τ	...	torsion angle	
Φ	...	phase factor	

Out-of-plane Bending Motions

In order to keep hydrogen atoms on aromatic rings, ester- or carbamate groups planar — or more general: in all cases where three atoms ABC span a plane and a fourth atom D linked to atom B is required to lay within this plane — an out-of-plane term must be included. Out-of-plane bending can be treated as an improper torsion angle, resulting in a cosine function similar to 10.6. Alternatively, the distance that atom D lies above the plane spanned by atoms ABC can be described by an harmonic potential similar to 10.2.

10.1.2 Non-Bonded Interactions

Van der Waals Interactions

If two non-bonded atoms approach, first an attractive interaction will be observed that has a distance dependency of r^{-6}. If the distance between the atom gets increasingly small, an overlap of the atoms is avoided through a strong repulsive term with a distance dependency of r^{-12}. Van der Waals interactions are in most cases described by a 12-6 Lennard Jones Potential:

$$E(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$E(r)$	...	LJ energy	
ε	...	well depth	(10.7)
σ	...	collision diameter	
r	...	separation between two atoms	

Electrostatic Interaction

The electrostatic interaction between two molecules or within different parts of the same molecule can be calculated by Coulomb's law:

$$E(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

$$\begin{array}{lll} E(r) & \ldots & \text{Coulomb energy} \\ q_i, q_j & \ldots & \text{charges} \\ \varepsilon_0 & \ldots & \text{dielectric constant} \\ r_{ij} & \ldots & \text{separation between charges} \end{array} \tag{10.8}$$

10.2 Example Parameter Input File for an MD Simulation in GROMACS

```
; VARIOUS PREPROCESSING OPTIONS
```

```
title = H836/hH3 complex in a DPPC/water environment
```

Names the title of the MD simulation.

```
cpp = /lib/cpp
```

```
include = -I/home/schlegel/INCLUDEFILES
```

Names directories that shall be included into the simulation, e.g. directories where user-defined parameters are deposited.

```
define = -DPOSRES
```

Gives the command to incorporate position restraints into the simulation, e.g. during the equilibration period.

```
; RUN CONTROL PARAMETERS
```

```
integrator = md
```

A leap-frog algorithm will be used for integrating Newton's equation of motion.

```
tinit = 0
```

Start time of the simulation. This value will be normally set to 0 ps. Only if a previous simulation is continued, a different start time will be set.

```
dt = 0.002
```

Time-step used during the simulation; here set to 2 fs.

The maximum time-step in MD simulations is limited by the smallest oscillation period that can be found in the simulated system. In biological systems, the highest frequencies of motion are usually observed for hydrogen bond stretching and hydrogen angle bending that limit the time-step to 2 or 2.5 fs, respectively. [139] As these motions are not expected to influence the macroscopic behaviour and are additionally not modelled properly by a classical treatment, [139] several strategies exist to eliminate these motions from the system. The most common is the application of constraints such as SHAKE or LINCS (see

“OPTIONS FOR BONDS” parameter block)

```
nsteps = 50000
```

Number of steps calculated. 50000 steps with a time-step of 2 fs will result in a simulation period of 50000 $\times$ 2fs = 100 ps.

```
comm-mode = Linear
```

Names mode for center of mass motion removal. In case of a linear mode, the center of mass translation is removed periodically. This will avoid that the entire system translates out of the box boundaries during an MD simulation.

```
nstcomm = 1
```

Frequency for center of mass removal; here set to every step.

```
comm-grps = Protein H2O H836 NA+ CL- SOL DPPC
```

The center of mass is removed from all groups in the system. If no value is given, the center of mass motion is removed for the entire system.

```
; OUTPUT CONTROL OPTIONS
nstxout = 5000
nstvout = 5000
nstfout = 5000
```

Frequency to write system coordinates (x), velocities (v) and forces (f) into the trajectory file ($*$.trr). $*$.trr files are required to continue an MD simulation, are however significantly larger files than the $*$.xtc files that do not contain velocity and force information. The output rate is thus set to larger intervals.

```
nstlog = 1000
```

Frequency to write system energy into $*$.log file.

```
nstenergy = 1000
```

Frequency to write system energy into $*$.edr file.

```
nstxtcout = 1000
```

Frequency to write system coordinates into $*$.xtc trajectory. The $*$.xtc file does not contain information on atoms’ velocities and can thus not be used to continue an MD simulation.

```
xtc_precision = 1000
```

Precision to write to the $*$.xtc trajectory.

```
xtc_grps = Protein H2O H836 NA+ CL- SOL DPPC
```

Groups to write to the xtc trajectory.

```
energygrps = Protein H2O H836 NA+ CL- SOL DPPC
```

Groups to write to the $*$.edr energy file.

```
; NEIGHBOURSEARCHING PARAMETERS
```

The most time consuming part of an MD simulation is the calculation of non-bonded ener-

gies and/or forces. While the number of bonded interactions is proportional to the number of atoms N in the system, the number of non-bonded terms increases with N^2. [70] In principle, non-bonded interactions need to be calculated between every pair of atoms in the system. Due to the fast decline of the Lennard Jones potential (see equation 10.7) this is however not justified and it is sufficient to take into account interactions within a cutoff distance. In order to avoid that in every integration step the distances between all atoms have to be calculated, a non-bonded neighbour list is generated that stores all atoms that lie within the cutoff distance of each atom. In each integration step, only the atoms in the neighbour list are included in the calculation of Lennard Jones potential. The neighbour list is then updated at regular intervals throughout the simulation.

`nstlist = 5`

Frequency to update the neighbour list.

`ns_type = grid`

If a new neighbour list is constructed, the simulation box will be subdivided into grid cells. Only atoms in neighbouring grid cells will be checked for being potential neighbours for an atom in the center cell.

`pbc = xyz`

Periodic boundary conditions (PBC) will be used in all directions.

The introduction of PBC is essential in order to avoid so called “edge effects”: the behaviour of molecules located at the boundaries of the simulation box will be influenced by the “hard” box walls. If PBC are introduced, the simulation box is surrounded by 9 mirror images. The box boundaries are now permeable as every molecule leaving the simulation box at one side will enter into the box from a mirror image so that the total number of atoms in the system remains unchanged. Mirror images of the box are also included in the generation of the neighbour list. The *minimum image convention* thereby assures that each atom “sees” only one image of every other atom in the system.

`rlist = 0.9`

Enumerates the cutoff distance [nm] for the generation of the short-range neighbour list for calculating the van der Waals interactions. This cutoff distance must be smaller than half of the smallest box vectors in order to avoid that an atom “sees” its own image.

`; OPTIONS FOR ELECTROSTATICS AND VDW`

As regards non-bonded interactions, due to the fast decline of the Lennard Jones potential van der Waals interactions may be truncated beyond a cutoff distance. Electrostatic interactions are however of significantly longer range thus a cutoff based approach is usually avoided as it would result in the introduction of truncation effects. Here, the Paricle Mesh Ewald summation method was applied as this method was described to giver superior results than standard cutoff models. [140] (see [70] for a detailed description of the PME

method.)

```
coulombtype = PME
```

The Particle Mesh Ewald method was used to calculate long-range electrostatic interactions.

```
fourierspacing = 0.12
```

Grid spacing in nm for the FFT grid as required for the PME calculations.

```
pme_order = 4
```

```
ewald_rtol = 1e-05
```

```
ewald_geometry = 3d
```

```
optimize_fft = yes
```

Default parameters for calculating the FFT grid for PME calculations.

```
vdw_type = Cut-off
```

Van der Waals interactions are treated by a cutoff method.

```
rvdw = 0.9
```

Distance for the Lennard Jones cutoff in [nm].

```
; OPTIONS FOR WEAK COUPLING ALGORITHMS
```

Common macroscopic ensembles for MD simulations are the NVT and the NPT ensemble. In both ensembles the number of atoms N and the system temperature T are kept constant. In the NVT (canonical) ensemble the volume of the simulation box is kept fixed. This ensemble was used for simulations of protein models in a CCl_4/water environment. When using the NPT (isobaric-isothermal) ensemble, the value of pressure is adjusted to a constant value allowing the simulation box to adapt in its size. The NPT ensemble was used for simulations of protein models in a DPPC/water environment. The compliance with a set value of temperature is thereby achieved by coupling the system to a temperature bath with the reference temperature ref_t [K] applying the Berendsen algorithm.

```
tcoupl = Berendsen
```

```
tc_grps = Protein H2O H836 NA+ CL- SOL DPPC
```

Groups within the simulation system coupling individually to the temperature bath.

```
tau_t = 0.1 0.1 0.1 0.1 0.1 0.1 0.1
```

Time constant for temperature coupling [ps].

```
ref_t = 310 310 310 310 310 310 310
```

Reference temperature for individual groups.

```
pcoupl = Berendsen
```

Method for obtaining the set value of pressure.

```
pcoupltype = isotropic
```

Type of pressure coupling applied.

```
tau_p = 0.5
```

Time constant for pressure coupling.

`compressibility = 4.5e−5`

Compressibility of the system in [bar^{-1}]. The value corresponds to water at 1 atm and 300 K. In biphasic systems several parameters would be required for the individual phases. This is however not implemented into the system.

`ref_p = 1.0`

Reference pressure [bar].

`; SIMULATED ANNEALING`

No simulated annealing was carried out as indicated for every group.

`annealing = no no no no no no no`

`annealing_npoints =`

`annealing_time =`

`annealing_temp =`

`; GENERATE VELOCITIES FOR STARTUP RUN`

At the start of a simulation velocities are randomly generated for all atoms in the system according to the system temperature [K].

`gen_vel = yes`

Velocities shall be generated. The option “no” is only set in case an MD simulation is continued and velocities shall be taken from the $*$.trr file.

`gen_temp = 310`

System temperature [K].

`gen_seed = 173529`

Integer used to initialize the random generator for random velocities.

`; OPTIONS FOR BONDS`

Depending on the time-step applied, setting a constraint to bonds may be required. As described above, the time-step depends on the highest frequency of motion within the system. If a time-step superior to 1 fs is chosen, bonds involving hydrogen will not be sampled sufficiently, thus they should be constrained. Two algorithms, `Lincs` (for `Lin`ear `C`onstraint `S`olver) and `Shake` exist to accomplish this task of which the `Lincs` algorithm is faster and more stable. In the herein described simulations all-bonds were constrained with the Lincs algorithm. The default settings for this case are given below.

`constraints = all-bonds`

`constraint_algorithm = Lincs`

`unconstrained_start = yes`

```
lincs_order = 4
lincs_warnangle = 30
```

```
; NMR refinement stuff
```

If distance restraints between pairs of atoms are used during MD simulations, the following parameters must be set. If the parameter disre_weighting is set to “equal”, the restraint force is divided equally over the atoms pair. The force constant is enumerated via the parameter disre_fc [kJmol^{-1}nm^{-2}].

```
disre = Simple
disre_weighting = Equal
disre_mixed = no
disre_fc = 500
disre_tau = 0
nstdisreout = 1000
```

10.3 One Letter Code for Amino Acids

A alanine

C cysteine

D aspartic acid

E glutamic acid

F phenylalanine

G glycine

H histidine

I isoleucine

K lysine

L leucine

M methionine

N asparagine

P proline

Q glutamine

R arginine

S serine

T threonine

V valine

W tryptophane

Y tyrosine

10.4 List of Abbreviations

BR bovine rhodopsin

CA constitutive activity

cAMP cyclic 3',5'-adenosine monophosphate

CCD Cambridge Crystallographic Database

CNS central nervous system

DAO diamine oxidase

DPPC dipalmitoylphosphatidylcholine

ER endoplasmatic reticulum

GABA γ-aminobutyric acid

GDP guanine nucleotide diphosphate

G-protein guanine nucleotides binding protein

GOLD Genetic Optimisation for Ligand Docking

GPCR G-protein coupled receptor

GROMACS Groningen Machine for Chemical Simulations

GTP guanine nucleotide triphosphate

H_1R histamine H_1 receptor

H_2R histamine H_2 receptor

H_3R histamine H_3 receptor

H_4R histamine H_4 receptor

hH_3R human histamine H_3 receptor

hH_4R humane histamine H_4 receptor

HMM Hidden Markov Model

HNMT histamine N-methyltransferase

IL-16 interleukin-16

MAO monoamine oxidase

MAP mitogen-activated protein

MD molecular dynamics

MDB Maybridge Database

MSA multiple sequence alignment

NT neurotransmitter

PLC phospholipase C

PME Particle Mesh Ewald

POPC palmitoyloleoylphosphatidylcholine

RMSD root mean square deviation [nm]

SAR structure affinity relationship

TM transmembrane

TNFα tumor necrosis factor α

TopPred topology Prediction

UHBD University of Houston Brownian Dynamics

WDI World Drug Index

Table 10.1: Prediction of transmembrane regions in the sequence of the hH_3R.

TMHMM	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
TopPred	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
DAS	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
Split4	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
TMFinder	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
SWISSPROT	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
TMHMM	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
TopPred	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
DAS	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
Split4	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
TMFinder	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
SWISSPROT	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
TMHMM	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
TopPred	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
DAS	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
Split4	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
TMFinder	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
SWISSPROT	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
TMHMM	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
TopPred	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
DAS	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
Split4	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
TMFinder	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
SWISSPROT	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
TMHMM	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
TopPred	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
DAS	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
Split4	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
TMFinder	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
SWISSPROT	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
TMHMM	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
TopPred	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
DAS	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
Split4	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
TMFinder	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
SWISSPROT	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
TMHMM	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
TopPred	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
DAS	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
Split4	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
TMFinder	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
SWISSPROT	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK

Table 10.2: Prediction of secondary structure in the sequence of the hH_3R.

PHD	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
PROF	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
PSIPRED	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
SAM T99	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
SSPRO	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
JPRED	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
PREDAT	MERAPPDGPL NASGALAGEA AAAGGARGFS AAWTAVLAAL MALLIVATVL
PHD	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
PROF	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
PSIPRED	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
SAM T99	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
SSPRO	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
JPRED	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
PREDAT	GNALVMLAFV ADSSLRTQNN FFLLNLAISD FLVGAFCIPL YVPYVLTGRW
PHD	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
PROF	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
PSIPRED	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
SAM T99	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
SSPRO	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
JPRED	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
PREDAT	TFGRGLCKLW LVVDYLLCTS SAFNIVLISY DRFLSVTRAV SYRAQQGDTR
PHD	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
PROF	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
PSIPRED	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
SAM T99	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
SSPRO	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
JPRED	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
PREDAT	RAVRKMLLVW VLAFLLYGPA ILSWEYLSGG SSIPEGHCYA EFFYNWYFLI
PHD	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
PROF	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
PSIPRED	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
SAM T99	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
SSPRO	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
JPRED	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
PREDAT	TASTLEFFTP FLSVTFFNLS IYLNIQRRTR LRLDGAREAR FRLSRDRKVA
PHD	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
PROF	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
PSIPRED	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
SAM T99	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
SSPRO	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
JPRED	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
PREDAT	KSLAVIVSIF GLCWAPYTLL MIIRAACHGH CVPDYWYETS FWLLWANSAV
PHD	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
PROF	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
PSIPRED	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
SAM T99	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
SSPRO	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
JPRED	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK
PREDAT	NPVLYPLCHH SFRRAFTKLL CPQKLKIQPH SSLEHCWK

Table 10.3: Prediction of transmembrane regions in the sequence of bovine rhodopsin.

TMHMM	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
TopPred	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
DAS	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
SPLIT4	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
TMFinder	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
SWISSPROT	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
TMHMM	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
TopPred	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
DAS	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
SPLIT4	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
TMFinder	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
SWISSPROT	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
TMHMM	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
TopPred	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
DAS	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SPLIT4	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
TMFinder	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SWISSPROT	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
TMHMM	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
TopPred	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
DAS	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
SPLIT4	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
TMFinder	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
SWISSPROT	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
TMHMM	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
TopPred	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
DAS	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
SPLIT4	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
TMFinder	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
SWISSPROT	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
TMHMM	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
TopPred	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
DAS	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
SPLIT4	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
TMFinder	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
SWISSPROT	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
TMHMM	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
TopPred	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
DAS	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
SPLIT4	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
TMFinder	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
SWISSPROT	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA

Table 10.4: Prediction of secondary structure in the sequence of bovine rhodopsin.

PHD	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
PROF	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
PSIPRED	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
SAM T99	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
SSPRO	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
SWISSPROT	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
1HZX	MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
PHD	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
PROF	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
PSIPRED	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
SAM T99	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
SSPRO	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
SWISSPROT	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
1HZX	GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
PHD	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
PROF	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
PSIPRED	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SAM T99	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SSPRO	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SWISSPROT	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
1HZX	GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
PHD	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
PROF	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
PSIPRED	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
SAM T99	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
SSPRO	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
SWISSPROT	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
1HZX	NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
PHD	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
PROF	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
PSIPRED	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
SAM T99	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
SSPRO	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
SWISSPROT	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
1HZX	ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
PHD	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
PROF	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
PSIPRED	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
SAM T99	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
SSPRO	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
SWISSPROT	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
1HZX	TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
PHD	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
PROF	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
PSIPRED	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
SAM T99	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
SSPRO	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
SWISSPROT	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
1HZX	YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA

Bibliography

[1] H. Rang, M. Dale, J. Ritter, and P. Moore. *Pharmacology*. Churchill Livingstone, 2003.

[2] E. Archer, B. Maigret, C. Escrieut, L. Pradayrol, and D. Fourmy. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol Sci*, 24:36–40, 2003.

[3] MJ. Marinissen and SJ. Gutkind. G-protein-coupled receptors and signaling networks: Emerging paradigms. *Trends Pharmacol Sci*, 22(7):368–376, 2001.

[4] K. Palczewski, T. Kumasaka, T. Hori, CA. Behnke, H. Motoshima, BA. Fox, I. Le Trong, DC. Teller, T. Okada, RE. Stenkamp, M. Yamamoto, and M. Miyano. Structure of rhodopsin: A G-protein-coupled receptor. *Science*, 289:739–745, 2000.

[5] M. Caffrey. Membrane protein crystallization. *J Struct Biol*, 142(1):108–132, 2003.

[6] OP. Ernst and FJ. Bartl. Active states of rhodopsin. *ChemBioChem*, 3:968–974, 2002.

[7] Y. Furutani, H. Kandori, and Y. Shichida. Structural changes in lumirhodopsin and metarhodopsin I studied by their photoreactions at 77 K. *Biochemistry*, 42(28):8494–8500, 2003.

[8] H. Auterhoff, J. Knabe, and H.-D. Höltje. *Lehrbuch der Pharmazeutischen Chemie*. Wissenschaftliche Verlagsgesellschaft mbH Stuttgart, 1999.

[9] JR. Horton, K. Sawada, M. Nishibori, X. Zhang, and X. Cheng. Two polymorphic forms of human histamine methyltransferase: Structural, thermal, and kinetic comparison. *Structure*, 9:837–849, 2001.

[10] R. Leurs, RA. Bakker, H. Timmerman, and IJ. De Esch. The histamine H_3 receptor: from gene cloning to H_3 receptor drugs. *Nature*, 4:107–120, 2005.

[11] I. Gantz, G. Munzert, T. Tashiro, M. Schaffer, L. Wang, J. Del Valle, and T. Yamada. Molecular cloning of the human histamine H_2 receptor. *Biochem Biophys Res Commun*, 178:1386–1392, 1991.

[12] H. Fukui, K. Fujimoto, H. Mizuguchi, K. Sakamoto, Y. Horio, S. Takai, K. Yamada, and S. Ito. Molecular cloning of the human histamine H_1 receptor gene. *Biochem Biophys Res Commun*, 2001:894–901, 1994.

[13] TW. Lovenberg, BL. Roland, SJ. Wilson, X. Jiang, J. Pyati, A. Huva, MR. Jackson, and MG. Erlander. Cloning and functional expression of the human histamine H_3 receptor. *Mol Pharmacol*, 55:11101–11107, 1999.

[14] T. Nakamura, H. Itadani, Y. Hidaka, M. Ohta, and K. Tanaka. Molecular cloning and characterization of a new human histamine receptor, HH_4R. *Biochem Biophys Res Commun*, 279:615–620, 2000.

[15] T. Oda, N. Morikawa, Y. Saito, Y. Masuho, and SI. Matusomto. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J Biol Chem*, 47:36781–36786, 2000.

[16] C. Liu, XJ. Ma, X. Jiang, SJ. Wilson, CL. Hafstra, J. Blevitt, J. Pyati, X. Li, W. Chai, N. Carruthers, and TW. Lovenberg. Cloning and pharmacological characterization of a fourth histamine (H_4) receptor expressed in bone marrow. *Mol Pharmacol*, 59:420–426, 2001.

[17] T. Nguyen, DA. Shapiro, SR. George, VM. Setola, DK. Lee, R. Cheng, L. Rauser, SP. Lee, KR. Lynch, BL. Roth, and BF. O'Dowd. Discovery of a novel member of the histamine receptor family. *Mol Pharmacol*, 59:427–433, 2001.

[18] Y. Zhu, D. Michalowich, HL. Wu, KB. Tan, GM. Dytko, IJ. Mannam, R. Boyce, J. Alstron, LA. Tierney, and X. Li. Cloning, expression, and pharmacological characterization of a novel human histamine receptor. *Mol Pharmacol*, 59:434–441, 2001.

[19] JM. Arrang, M. Garbarg, and JC. Schwartz. Auto-inhibition of brain histamine release mediated by a novel class (H_3) of histamine receptor. *Nature*, 302(5911):832–837, 1983.

[20] RA. Bakker. Histamine H_3-receptor isoforms. *Inflamm Res*, 53(10):509–516, 2004.

[21] AA. Hancock, TA. Esbenshade, KM. Krueger, and BB. Yao. Genetic and pharmacological aspects of histamine H_3 receptor heterogeneity. *Life Science*, 79:3043–3072, 2003.

[22] M. Cowart, R. Altenbach, L. Black, R. Faghih, C. Zhao, and AA. Hancock. Medicinal chemistry and biological properties of non-imidazole histamine H_3 antagonists. *Mini Rev Med Chem*, 4(9):979–992, 2004.

[23] H. Stark, M. Kathmann, E. Schlicker, W. Schunack, B. Schlegel, and W. Sippl. Medicinal chemistry and pharmacological aspects of imidazole-containing histamine H_3 receptor antagonists. *Mini Rev Med Chem*, 4(9):965–977, 2004.

[24] IJ. De Esch and KJ. Belzar. Histamine H_3 receptor agonists. *Mini Rev Med Chem*, 4(9):955–963, 2004.

[25] ER. Kandel, JH. Schwartz, and TM. Jessell. *Principles of Neural Science.* McGraw-Hill, 2000.

[26] F. Gbahou, A. Rouleau, S. Morisset, R. Parmentier, S. Crochet, JS. Lin, X. Ligneau, J. Tardivel-Lacombe, H. Stark, W. Schunack, CR. Ganellin, JC. Schwartz, and JM. Arrang. Protean agonism at histamine H_3 receptors *in vitro* and *in vivo*. *PNAS*, 100:11086–11091, 2003.

[27] AJ. Uveges, D. Kowal, Y. Zhang, TB. Spangler, J. Dunlop, S. Semus, and PG. Jones. The role of transmembrane helix 5 in agonist binding to the human H_3 receptor. *J Pharmacol Exp Ther*, 301(2):451–458, 2002.

[28] ML. Jacobsen, K. Rimvall, S. Hastrup, IV. Pettersson, and BS. Wulff. Construction of a fully constitutively active human histamine H_3 receptor. In *Poster presentation: XXXII Annual Meeting of the European Histamine Research Society; 7-11 May, Noordwijkerhout, the Netherlands*, 2003.

[29] TW. Lovenberg, J. Pyati, H. Chang, SJ. Wilson, and MG. Erlander. Cloning of rat histamine H_3 receptor reveals distinct species pharmacological profiles. *J Pharmacol Exp Ther*, 293(3):771–778, 2000.

[30] H. Stark, K. Purand, X. Ligneau, A. Rouleau, JM. Arrang, M. Garbarg, JC. Schwartz, and W. Schunack. Novel carbamates as potent histamine H_3 receptor antagonists with high *in vitro* and oral *in vivo* activity. *J Med Chem*, 39:1157–1163, 1996.

[31] CE. Tedford, M. Hoffmann, N. Seyedi, R. Maruyama, R. Levi, SL. Yates, SM. Ali, and JG. Phillips. High antagonist potency of GT-2227 and GT-2331, new histamine H_3 receptor antagonists, in two functional models. *Eur J Pharmacol*, 351(3):307–311, 1998.

[32] H. Liu, FA. Kerdesky, LA. Black, M. Fitzgerald, R. Henry, TA. Esbenshade, AA. Hancock, and Bennani YL. An efficient multigram synthesis of the potent histamine

H_3 antagonist GT-2331 and the reassessment of the absolute configuration. *J Org Chem*, 69:192–194, 2004.

[33] G. Meier, J. Apelt, U. Reichert, S. Graßmann, X. Ligneau, S. Elz, F. Leurquin, CR. Ganellin, JC. Schwartz, W. Schunack, and H. Stark. Influence of imidazole replacement in different structural classes of histamine H_3 antagonists. *Eur J Pharm Sci*, 13:249–259, 2001.

[34] T. Nickel, U. Bauer, E. Schlicker, M. Kathmann, M. Göthert, A. Sasse, H. Stark, and W. Schunack. Novel histamine H_3-receptor antagonists and partial agonists with a non-aminergic structure. *Br J Pharmacol*, 132(8):1665–1672, 2001.

[35] K. Wieland, G. Bongers, Y. Yamamoto, T. Hashimoto, A. Yamatodani, WM. Menge, H. Timmerman, TW. Lovenberg, and R. Leurs. Constitutive activity of histamine H_3 receptors stably expressed in SK-N-MC cells: display of agonism and inverse agonism by H_3 antagonists. *J Pharm Exp Ther*, 299(3):908–914, 2001.

[36] IJ. De Esch, H. Timmerman, WM. Menge, and PH. Nederkoorn. A qualitative model for the histamine H_3 receptor explaining agonistic and antagonistic activity simultenously. *Arch Pharm Phram Med Chem*, 333:254–260, 2000.

[37] Schering Corporation. Imidazole compounds substituted with a six or seven membered heterocyclic ring containing two nitrogen atoms. *International patent application*, WO 00/53596, 2000.

[38] IJ. De Esch, JE. Mills, TD. Perkins, G. Romeo, M. Hoffmann, K. Wieland, R. Leurs, WM. Menge, PH. Nederkoorn, P. Dean, and H. Timmerman. Development of a pharmacophore model for histamine H_3 receptor antagonists, using the newly developed molecular modeling program SLATE. *J Med Chem*, 44:1666–1674, 2001.

[39] A. Vasudevan, SE. Conner, RG. Gentles, R. Faghih, H. Liu, W. Dwight, L. Ireland, CH. Kang, TA. Esbenshade, YL. Bennani, and AA. Hancock. Synthesis and evaluation of potent pyrrolidine H_3 antagonists. *Bioorg Med Chem Lett*, 12:3055–3058, 2002.

[40] R. Faghih, W. Dwight, A. Vasudevan, J. Dinges, SE. Conner, TA. Esbenshade, YL. Bennani, and AA. Hancock. Aminoalkoxybiphenylnitriles as histamine-3 receptor ligands. *Bioorg Med Chem Lett*, 12:3077–3079, 2002.

[41] TA. Esbenshade, KM. Krueger, TR. Miller, CH. Kang, LI. Denny, DG. Witte, BB. Yao, GB. Fox, R. Faghih, YL. Bennani, M. Williams, and AA. Hancock. Two novel and selective nonimidazole histamine H_3 receptor antagonists A-304121 and A-317920: I. In vitro pharmacological effects. *J Pharmacol Exp Ther*, 305:887–896, 2003.

[42] R. Faghih, K. Phelan, TA. Esbenshade, TR. Miller, CH. Kang, KM. Krueger, BB. Yao, GB. Fox, YL. Bennani, and AA. Hancock. D-alanine piperazine-amides: novel non-imidazole antagonists of the histamine H_3 receptor. *Inflamm. Res. Supplement 1*, 52:47–48, 2003.

[43] J. Apelt, X. Ligneau, HH. Pertz, JM. Arrang, CR. Ganellin, JC. Schwartz, W. Schunack, and H. Stark. Development of a new class of nonimidazole histamine H_3 receptor ligands with combined inhibitory histamine N-methyltransferase activity. *J Med Chem*, 45(5):1128–1141, 2002.

[44] M. Cowart, R. Faghih, MP. Curtis, GA. Gfesser, YL. Bennani, LA. Black, L. Pan, KC. Marsh, JP. Sullivan, TA. Esbenshade, GB. Fox, and AA. Hancock. 4-(2-[2-(2(*R*)-methylpyrrolidin-1-yl)ethyl]benzofuran-5-yl)benzonitrile and related 2-aminoethylbenzofuran H_3 receptor antagonists potently enhance cognition and attention. *J Med Chem*, 48:38–55, 2004.

[45] ID. Linney, IM. Buck, EA. Harper, SB. Kalindjian, MJ. Pether, NP. Shankley, GF. Watt, and PT. Wright. Design, synthesis, and structure-activity relationship of novel non-imidazole histamine H_3 receptor antagonists. *J Med Chem*, 43:2362–2370, 2000.

[46] BB. Yao, CW. Hutchins, TL. Carr, S. Cassar, JN. Masters, YL. Bennani, TA. Esbenshade, and AA. Hancock. Molecular modeling and pharmacological analysis of species-related histamine H_3 receptor heterogeneity. *Neuropharmacology*, 44:773–786, 2003.

[47] H. Stark, K. Purand, A. Hüls, X. Ligneau, M. Garbarg, JC. Schwartz, and W. Schunack. [^{125}I]Iodoproxyfan and related compounds: a reversible radioligand and novel classes of antagonists with high affinity and selectivity for the histamine H_3 receptor. *J Med Chem*, 39:1220–1226, 1996.

[48] RA. Clark, JI. Gallin, and AP. Kaplan. The selective eosinophil chemotactic activity of histamine. *J Exp Med*, 142:1462–1476, 1975.

[49] JA. Jablonowski, NI. Carruthers, and RL. Thurmond. The histamine H_4 receptor and potential therapeutic uses for H_4 ligands. *Mini-Reviews in Med Chem*, 4:993–1000, 2004.

[50] JA. Jablonowski, CA. Grice, W. Chai, CA. Dvorak, JD. Venable, AK. Kwok, KS. Ly, J. Wei, SM. Baker, PJ. Desai, W. Jiang, SJ. Wilson, RL. Thurmond, L. Karlsson, JP. Edwards, TW. Lovenberg, and NI. Carruthers. The first potent and selective non-imidazole human histamine H_4 receptor antagonist. *J Med Chem*, 46:3957–3960, 2003.

[51] T. Hashimoto, S. Harusawa, L. Araki, OP. Zuiderveld, MJ. Smit, T. Imazu, S. Takashima, Y. Yamamoto, Y. Sakamoto, T. Kurihara, R. Leurs, RA. Bakker, and A. Yamatodani. A selective human H_4-receptor agonist: (-)-2-cyano-1-methyl-3-(*2R,5R*)-5-[1*H*-imidazol-4(5)-yl]tetrahydrofuran-2-ylmethylguanidine. *J Med Chem*, 46:3162–3165, 2003.

[52] N. Shin, E. Coates, NJ. Murgolo, KL. Morse, M. Bayne, CD. Strader, and FJ. Monsma. Molecular modeling and site-specific mutagenesis of the histamine-binding site of the histamine H_4 receptor. *Mol Pharmacol*, 62(1):38–47, 2002.

[53] FC. Bernstein, TF. Koetzle, GJ. Williams, EF. Meyer, MC. Brice, JR. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112(3):535–542, 1977.

[54] C. Bissantz, P. Bernard, M. Hibert, and D. Rognan. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins: Structure, Function, and Genetics*, 50(1):5–25, 2003.

[55] T. Klabunde and G. Hessler. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem*, 3:928–944, 2002.

[56] BM. Bröer, M. Gurrath, and H.-D. Höltje. Molecular modelling studies on the ORL1 receptor and ORL1-agonists. *J Comput Aided Mol Des*, 17(11):739–754, 2003.

[57] S. Möller, MD. Croning, and R. Apweiler. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 17(7):646–653, 2001.

[58] J. Kyte and RF. Doolittle. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157(1):105–132, 1982.

[59] C. Zhou, Y. Zheng, and Y. Zhou. Structure prediction of Membrane Proteins. *Geno Prot Bioinfo*, 2(1):1–5, 2004.

[60] M. Cserzo, JM. Bernassau, I. Simon, and B. Maigret. New alignment strategy for transmembrane proteins. *J Mol Biol*, 243(3):388–96, 1994.

[61] CM. Deber, C. Wang, LP. Liu, AS. Prior, S. Agrawal, BL. Muskat, and AJ. Cuticchia. TM Finder: A prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Science*, 10:212–219, 2001.

[62] EL. Sonnhammer, G. von Heijne, and A. Krogh. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*, 1998(6):175–182, 1998.

[63] D. Juretic, L. Zoranic, and Zucic D. Basic charge clusters and predictions of membrane protein topology. *J Chem Inf Comput Sci*, 42:620–632, 2002.

[64] B. Rost. Review: Protein secondary structure prediction continues to rise. *J Struc Biol*, 134(2-3):204–218, 2001.

[65] R. Durbin, SR. Eddy, A. Krogh, and G. Mitchison. *Biological sequence analysis.* Cambridge University Press, 1998.

[66] JD. Thompson, F. Plewniak, and O. Poch. A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res*, 27(13):2682–2690, 1999.

[67] AA. Canutescu, AA. Shelenkov, and RL. Dunbrack. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Science*, 12(9):2001–2014, 2003.

[68] CN. Schutz and A. Warshel. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins: Structure, Function, and Genetics*, 44(4):400–417, 2001.

[69] JM. Briggs, JD. Madura, ME. Davis, MK. Gilson, J. Antosiewicz, BA. Luty, RC. Wade, B. Bagheri, A. Ilin, RC. Tan, and JA. McCammon. *UHBD (University of Houston Brownian Dynamics)*. University of Houston, 5.1 edition.

[70] A. Leach. *Molecular Modelling. Principles and Application.* Prentice Hall, 2001.

[71] H.-D. Höltje, W. Sippl, D. Rognan, and G. Folkers. *Molecular Modeling: Basic principles and applications.* Wiley-VCH, 2003.

[72] C. Hunter, J. Singh, and JM. Thornton. $\pi - \pi$ Interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J Mol Biol*, 218:837–846, 1991.

[73] PJ. Goodford. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J Med Chem*, 28:849–857, 1985.

[74] RA. Laskowski, MW. McArthur, DS. Moss, and JM. Thornton. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst*, (26):283–291, 1993.

[75] CG. Wermuth. *The practice of medicinal chemistry.* Elsevier, 2003.

[76] C. Lemmen, T. Lengauer, and G. Klebe. FLEXS: A method for fast flexible ligand superposition. *J Med Chem*, 41:4502–4520, 1998.

[77] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. Predicting receptor-ligand interactions by an incremental construction algorithm. *J Mol Biol*, 261:470–489, 1996.

[78] DS. Goodsell and AJ. Olson. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Structure, Function, and Genetics*, 8:195–202, 1990.

[79] G. Jones, P. Willett, and RC. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol*, 245:43–53, 1995.

[80] JM. Baldwin, GF. Schertler, and VM. Unger. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol*, 272(1):144–164, 1997.

[81] T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, EM. Landau, and Y. Shichida. Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. *PNAS*, 99(9):5982–5987, 2002.

[82] DC. Teller, T. Okada, CA. Behnke, K. Palczewski, and RE. Stenkamp. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry*, 40(26):7761–7772, 2001.

[83] *INSIGHT II*. Biosym/MSI, San Diego, USA, 2000. HOMOLOGY/FDISCOVER module.

[84] T. Wymore and TC. Wong. Molecular dynamics study of substance P peptides in a biphasic membrane mimic. *Biophys J*, 76(3):1199–1212, 1999.

[85] U. Essmann, L. Perera, ML. Berkowitz, T. Darden, H. Lee, and LG. Pedersen. A smooth particle mesh Ewald potential. *J Chem Phys*, 103:8577–8592, 1995.

[86] GT. Ibragimova and RC. Wade. Importance of explicit salt ions for protein stability in molecular dynamics simulation. *Biophys J*, 74(6):2906–2911, 1998.

[87] DP. Tieleman and HJ. Berendsen. Molecular dynamics simulations of a fully hydrated dipalmitoylphasphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *J Chem Phys*, 105(11):4871–4880, 1996.

[88] LD. Schuler, X. Daura, and WF. van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comp Chem*, 22(11):1205–1218, 2001.

[89] I. Chandrasekhar, M. Kastenholz, RD. Lins, C. Oostenbrink, LD. Schuler, DP. Tieleman, and WF. van Gunsteren. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. *Eur Biophys J*, 32(1):67–77, 2003.

[90] http://www.moose.bio.ucalgary.ca.

[91] JD. Faraldo-Gómez, GR. Smith, and MS. Sansom. Setting up and optimization of membrane protein simulations. *Eur Biophys J*, 31(3):217–227, 2002.

[92] LF. Braganza and DL. Worcester. Hydrostatic pressure induces hydrocarbon chain interdigitation in single-component phospholipid bilayers. *Biochemistry*, 25(9):2591–2596, 1986.

[93] E. Egberts, SJ. Marrink, and HJ. Berendsen. Molecular dynamics simulation of a phospholipid membrane. *Eur Biophys J*, 22(6):423–436, 1994.

[94] T. Huber, AV. Botelho, K. Beyer, and MF. Brown. Membrane model for the G-protein-coupled receptor rhodopsin: Hydrophobic interface and dynamical structure. *Biophys Journal*, 86(4):2078–2100, 2004.

[95] DA. Shapiro, K. Kristiansen, WK. Kroeze, and BL. Roth. Differential modes of agonist binding to 5-hydroxytryptamine(2A) serotonin receptors revealed by mutation and molecular modeling of conserved residues in transmembrane region 5. *Mol Pharmacol*, 58(5):877–886, 2000.

[96] PS. Crozier, MJ. Stevens, LR. Forrest, and TB. Woolf. Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: Coupling between local retinal and larger scale conformational change. *J Mol Biol*, 333(3):493–514, 2003.

[97] T. Okada, M. Sugihara, AN. Bondar, M. Elstner, P. Entel, and V. Buss. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol*, 342(2):571–583, 2004.

[98] EL. Mehler, X. Periole, SA. Hassan, and H. Weinstein. Key issues in the computational simulation of GPCR function: representation of loop domains. *J Comput Aided Mol Des*, 16(11):841–853, 2002.

[99] U. Voigtländer, K. Jöhren, M. Mohr, A. Raasch, C. Tränkle, S. Buller, J. Ellis, H.-D. Höltje, and K. Mohr. Allosteric site on muscarinic acetylcholine receptors: identification of two amino acids in the muscarinic M_2 receptor that account entirely

for the M_2/M_5 subtype selectivities of some structurally diverse allosteric ligands in *N*-methylscopolamine-occupied receptors. *Mol Pharmacol*, 64:21–31, 2003.

[100] MT. Piascik and DM. Perez. Alpha1-adrenergic receptors: new insights and directions. *J Pharmacol Exp Ther*, 298(2):403–410, 2001.

[101] MM. Zhao, J. Hwa, and DM. Perez. Identification of critical extracellular loop residues involved in alpha 1-adrenergic receptor subtype-selective antagonist binding. *Mol Pharmacol*, 50(5):1118–1126, 1996.

[102] T. Wurch and PJ. Pauwels. Coupling of canine serotonine 5-HT(1B) and 5-HT(1D) receptor subtypes to the formation of inositol phosphates by dual interactions with endogenous G(i/o) and recombinant G(alpha15) proteins. *J. Neurochem.*, 75(3):1180–1189, 2000.

[103] L. Shi, MM. Simpson, JA. Ballesteros, and JA. Javitch. The first transmembrane segment of the dopamine D2 receptor: Accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry*, 40(41):12339–12348, 2001.

[104] M. Rarey, B. Kramer, and T. Lengauer. The particle concept: Placing discrete water molecules during protein-ligand docking predictions. *Proteins: Structure, Function, and Genetics*, 34(1):17–28, 1999.

[105] M. von Itzstein, WY. Wu, GB. Kok, MS. Pegg, JC. Dyason, B. Jin, T. Van Phan, ML. Smythe, HF. White, and SW. Oliver. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, 363(6428):418–423, 1993.

[106] DP. Tieleman, HJ. Berendsen, and MS. Sansom. Voltage-dependent insertion of alamethicin at phospholipid/water and octane/water interfaces. *Biophys J*, 80(1):331–346, 2001.

[107] JO. Trent, ZX. Wang, JL. Murray, W. Shao, H. Tamamura, N. Fujii, and SC. Peiper. Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists. *J Biol Chem*, 278(47):47136–47144, 2003.

[108] K. Fahmy, F. Jäger, M. Beck, TA. Zvyaga, TP. Sakmar, and F. Siebert. Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: A Fourier-Transform infrared spectroscopy study of site-directed mutants. *PNAS*, 90(21):10206–10210, 1993.

[109] EC. Yan, MA. Kazmi, Z. Ganim, JM. Hou, D. Pan, BS. Chang, TP. Sakmar, and RA. Mathies. Retinal counterion switch in the photoactivation of the G-protein-coupled receptor rhodopsin. *PNAS*, 100(16):9262–9267, 2003.

[110] M. Ishiguro. Conformational analysis of Glu181 and Ser186 in the metarhodopsin I state. *Chembiochem*, 5(9):1204–1209, 2004.

[111] KA. Neve, MG. Cumbay, KR. Thompson, R. Yang, DC. Buck, VJ. Watts, CJ. Durand, and MM. Teeter. Modeling and mutational analysis of a putative sodium-binding pocket on the dopamine D2 receptor. *Mol Pharmacol*, 60(2):373–381, 2001.

[112] M. Ishiguro, O. Yoshiaki, and T. Hirano. Structural models of the photointermediates in the rhodopsin photocascade, lumirhodopsin, metarhodopsin I, and, metarhodopsin II. *ChemBioChem*, 5(3):298–310, 2004.

[113] RR. Huang, H. Yu, CD. Strader, and TM. Fong. Interaction of substance P with the second and seventh transmembrane domains of the neurokinin-1 receptor. *Biochemistry*, 33(10):3007–3013, 1994.

[114] L. Mouledous, CM. Topham, C. Moisand, C. Mollerau, and JC. Meunier. Functional inactivation of the nociceptin receptor by alanine substitution of glutamine 286 at the C terminus of transmembrane segment VI: evidence from a site-directed mutagenesis study of the ORL1 receptor transmembrane-binding domain. *Mol Pharmacol*, 57(3):495–502, 2000.

[115] F. Horn, J. Weare, MW. Beukers, S. Horsch, A. Bairoch, W. Chen, O. Edvardsen, F. Campagne, and G. Vriend. GPCRDB: an information system for G-protein-coupled receptors. *Nucleic Acids Res.*, 26(1):275–279, 1998.

[116] M. Ishiguro, T. Hirano, and Y. Oyama. Modelling of photointermediates suggest a mechanism of the flip of the β-ionone moiety of the retinylidene chromophore in the rhodopsin photocascade. *Chembiochem*, 2-3:228–231, 2003.

[117] JA. Flohil, G. Vriend, and HJ. Berendsen. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins: Structure, Function, and Genetics*, 48(4):593–604, 2002.

[118] H. Fan and AE. Mark. Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci*, 13:992–999, 2004.

[119] H. Stark, B. Schlegel, and W. Sippl. News of the old histamine. I. A widening of the receptor family. *Pharm Unserer Zeit*, 32(2):92–93, 2003.

[120] P. Wellendorph, MW. Goodman, ES. Burstein, NR. Nash, MR. Brann, and DM. Weiner. Molecular cloning and pharmacology of functionally distinct isoforms of the human histamine H_3 receptor. *Neuropharmacology*, 42(7):929–940, 2002.

[121] ZL. Lu, JW. Saldanha, and EC. Hulme. Transmembrane domains 4 and 7 of the M(1) muscarinic acetylcholine receptor are critical for ligand binding and the receptor activation switch. *J Biol Chem*, 276(36):34098–34104, 2001.

[122] F. Heitz, JA. Holzwarth, JP. Gies, RM. Pruss, S. Trumpp-Kallmeyer, MF. Hibert, and C. Guenet. Site-directed mutagenesis of the putative human muscarinic M2 receptor binding site. *Eur J Pharmacol*, 380(2-3):183–195, 1999.

[123] MM. Simpson, JA. Ballesteros, V. Chiappa, J. Chen, M. Suehiro, DS. Hartman, T. Godel, LA. Snyder, TP. Sakmar, and JA. Javitch. Dopamine D4/D2 receptor selectivity is determined by a divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. *Mol Pharmacol*, 56(6):1116–1126, 1999.

[124] F. Horn, E. Bettler, L. Oliveira, F. Campagne, FE. Cohen, and G. Vriend. GPCRDB information system for G-protein-coupled receptors. *Nucleic Acids Res.*, 31(1):294–297, 2003.

[125] G. Bongers, A. Uveges, P. Jones, RA. Bakker, and R. Leurs. Preliminary investigation of the histamine H_3-receptor ligand binding pocket. In *Poster presentation: XXXII Annual Meeting of the European Histamine Research Society; 7-11 May, Noordwijkerhout, the Netherlands*, 2003.

[126] G. Jones, P. Willett, RC. Glen, AR. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible ligand docking. *J Mol Biol*, 267:727–748, 1997.

[127] R. Faghih, W. Dwight, JB. Pan, GB. Fox, KM. Krueger, TA. Esbenshade, JM. McVey, K. Marsh, YL. Bennani, and AA. Hancock. Synthesis and SAR of aminoalkoxy-biaryl-4-carboxamides: novel and selective histamine H_3 receptor antagonists. *Bioorg Med Chem Lett*, 13:1325–1328, 2003.

[128] DC. Warhurst, JC. Steele, IS. Adagu, JC. Craig, and C. Cullander. Hydroxychloroquine is much less active than chloroquine against chloroquine-resistant *Plasmodium falciparum*, in agreement with its physiochemical properties. *J of Antimicrobial Chemotherapy*, 52:188–192, 2003.

[129] ML. Verdonk, V. Berdini, MJ. Hartshorn, WT. Mooij, CW. Murray, RD. Taylor, and P. Watson. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci*, 44(3):793–806, 2004.

[130] H. Stark, K. Purand, X. Ligneau, A. Rouleau, JM. Arrang, M. Garbarg, JC. Schwartz, and W. Schunack. Novel carbamates as potent histamine H_3 receptor antagonists with high *in vitro* and *in vivo* activity. *J Med Chem*, 39:1157–1163, 1996.

[131] D. Lazewska, K. Kiec-Kononowicz, HH. Pertz, H. Stark, W. Schunack, and S. Elz. Piperidine-containing histamine H_3-receptor antagonists of the carbamate series: variation of the spacer length. *Pharmazie*, 56:927–932, 2001.

[132] F. Zaragoza, H. Stephensen, B. Peschke, and K. Rimvall. 2-(4-alkylpiperazin-1-yl)quinolines as a new class of imidazole-free histamine H_3 receptor-antagonists. *J Med Chem*, 48:306–311, 2005.

[133] EC. Meng and HR. Bourne. Receptor activation. What does the rhodopsin structure tell us? *Trends Pharm Sci*, 22(11):587–593, 2001.

[134] H. Stark, W. Sippl, X. Ligneau, JM. Arrang, CR. Ganellin, JC. Schwartz, and W. Schunack. Different antagonist binding properties of human and rat histamine H_3 receptors. *Bioorg Med Chem Lett*, 11:951–954, 2001.

[135] DJ. Waugh, RJ. Gaivin, MJ. Zuscik, P. Gonzalez-Cabrera, SA. Ross, J. Yun, and DM. Perez. Phe-308 and Phe-312 in transmembrane domain 7 are major sites of alpha 1-adrenergic receptor antagonist binding. Imidazoline agonists bind like antagonists. *J Biol Chem*, 276(27):25366–25371, 2001.

[136] BL. Roth, DL. Willins, K. Kristiansen, and WK. Kroeze. 5-Hydroxytryptamine2-family receptors (5-hydroxytryptamine2A, 5-hydroxytryptamine2B, 5-hydroxytryptamine2C): where structure meets function. *Pharmacol Ther*, 79(3):231–257, 1998.

[137] Y. Pan, N. Huang, S. Cho, and AD. MacKerell. Consideration of molecular weight during compound selection in virtual target-based database screening. *J Chem Inf Comput Sci*, 43:267–272, 2003.

[138] R. Apodaca, CA. Dvorak, W. Xiao, AJ. Barbier, JD. Boggs, SJ. Wilson, TW. Lovenberg, and NI. Carruthers. A new class of diamine-based human histamine H_3 receptor antagonists: 4-(aminoalkoxy)benzylamines. *J Med Chem*, 46:3938–3944, 2003.

[139] KA. Feenstra, B. Hess, and HJ. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich system. *J Comput Chem*, 20(8):786–798, 1999.

[140] RL. Norberto de Souza, O. Ornstein. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *J Biomol Struct Dyn*, 16(6):1205–1218, 1999.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mir während der Jahre meiner Promotion hilfreich zur Seite standen, allen voraus meinen Kolleginnen und Kollegen, die für ein angenehmes Arbeitsklima gesorgt haben, in dem nicht nur das wissenschaftliche Arbeiten Spass gemacht hat.

Britta Bröer und Gerd Menche möchte ich für die jahrelange Gesellschaft beim Mittagessen und beim Kaffee danach danken, die einen die zweifelhaften kulinarischen Genüsse der Cafete vergessen ließen. Bei Britta möchte ich mich für ihre ansteckende Fröhlichkeit, ihren Enthusiasmus und ihre Anteilnahme bei allen größeren und kleineren Problemen bedanken. Vielen Dank auch für diverse Pizzaabende und das Korrekturlesen der Arbeit. Gerd danke ich für seine scharfen und zielgenauen Analysen der Welt und allem was dazu gehört, denen ich auch viele wertvolle Anregungen für meine Arbeit verdanke. In guter Erinnerung sind mir auch etliche Semester gemeinsamer Studentenbetreuung unter deiner Leitung, die viel Spaß gemacht haben.

Bei Bernd Rupp bedanke ich mich für die Einweisung in die hohe Kunst von Gromacs und des Entlüftens von HPLC-Säulen. Unsere Feierabend-Bierchen in Modigliani und Tigges gehen mir schon ab, aber nach Berlin schaffen wir’s auch noch.

Besonders möchte ich mich auch bei allen Mitgliedern des Hallenser Arbeitskreises bedanken, allen voran bei Sonja Schlimme, mit der ich die Erinnerungen an leider viel zu wenige Monate als Zimmerkollegin in Düsseldorf, Chili-Kuchen und gemeinsame Sushi-Abende teile. Bei Rene Meier und Björn Windshügel möchte ich mich für die super Systemadministration des Hallenser Clusters bedanken, und schöne Erinnerungen ans Abendprogramm der Sommerschule in Wien.

Sampo, I will never forget the combination of exploring Düsseldorf´s Saturday night life and chilling out in Sunday´s cinema sessions and the great time we had. It was a such a pity you didn´t stay for longer.

Bei meinen Kollegen des 4. Semesters, Dirk Betscheider und Riccardo Bollig, möchte ich mich für die freundschaftliche und schöne Zusammenarbeit bedanken.

Bei Birte Brandt, Gisela Jessen und Peter Sippel bedanke ich mich für die sehr gute

technische Unterstützung während all den Jahren.

Tambien quiero decir gracias a mi novio Bernardo Reino por el gran interes que ponia en mi trabajo y por la ayuda sobre todo en asuntos informaticos. Gracias tambien por haber corregido mi tesis.

Personal Data

Name	Birgit Schlegel
Date of birth	30.06.1976
Place of birth	Linz, Austria
Nationality	Austrian

Education

11/2001 - 05/2005	PhD-Student at the Institute for Pharmaceutical Chemistry, Heinrich-Heine-University of Düsseldorf, Germany in the group of Prof. H.-D. Höltje under the supervision of Prof. W. Sippl. Topic of research: The human histamine H3-receptor: A molecular modelling study of a G-protein coupled receptor.
10/1999 - 08/2000	Graduation Project at the Centre for Ultrastructural Research and Nanotechnology, Vienna, Austria in the group of Prof. U. B. Sleytr under the supervision of Prof. M. Sará. Topic of research: Recombinant S-layer-fusion proteins: cloning, expression and characterization of S(urface)-layer-streptavidin fusion-proteins for development of functionalized crystalline surfaces.
10/1994 - 09/2001	Study of Biotechnology and Food technology, University of Agricultural Science (Universität für "Bodenkultur"), Vienna, Austria.
01/2001 - 06/2001	Erasmus term at the University of Barcelona, Faculty of Biology, Spain.
06/1998 - 06/1998	Award of university scholarship for exceptional grades.
09/1997 - 07/1998	Quality management and techniques course, technical grammar school, Vienna; final examination passed with distinction.
09/1986 - 06/1994	High school of Kreuzschwestern, Linz, Austria. Final examination passed with distinction.

Professional Experience

06/2005 - present	Patent Examiner at the European Patent Office, The Hague, The Netherlands.
11/2001 - 05/2005	Scientific assistant at the Heinrich-Heine-University of Düsseldorf, Germany.
02/2004 - 04/2004	Trainee at BASF, Ludwigshafen, Germany. Department of structure-based drug design.
09/1999 - 09/2001	Several jobs on fairs in Germany and Italy for "Agrarmarkt Austria" (Austrian Ministry of Agriculture.)

09/1999 - 09/1999	Trainee at Lactoprot, Pasching, Austria. Laboratory for microbiology.
07/1999 - 08/1999	Trainee at Nestlé, Athens, Greece. Laboratory for quality management.
06/1998 - 07/1998	Trainee at Ybbstaler Fruchtsaft Ges.m.b.H., Kröllendorf, Austria. Laboratory for research and quality management.
08/1997 - 08/1997	Trainee at Nestlé, Linz, Austria. Laboratory for quality management.
09/1996 - 09/1996	Trainee in the hospital of Elisabethinen, Linz, Austria. Department of microbiology.

Publication List

- Molecular Dynamics Simulations of Bovine Rhodopsin: Influence of Protonation States and Different Membrane Mimicking Environments. B. Schlegel, W. Sippl, H.-D. Höltje. J Mol Model (2005) [accepted]
- Setting up a simulation system for GPCR homology modelling. B. Schlegel, H. Stark, H.-D. Höltje, W. Sippl. Proceedings of the EURO-QSAR 2004 (Istanbul, Turkey).
- Model of a specific human histamine H3 receptor (hH3R) binding pocket suitable for virtual drug design. B. Schlegel, H. Stark, W. Sippl, H.-D. Höltje. Inflamm. Res. 54, Supplement 1 (2005): 50-51.
- Medicinal Chemical and Pharmacological Aspects of Imidazole-Containing Histamine H3 Receptor Antagonists. H. Stark, M. Kathmann, E. Schlicker, W. Schunack, B. Schlegel, W. Sippl. Mini-Reviews in Medicinal Chemistry, 2004, Vol 4: 965-977.
- Neues zum alten Histamin I. Ein weiteres Mitglied in der Rezeptorfamilie. H. Stark, W. Sippl, B. Schlegel. Pharm. Unserer Zeit. 32. Jahrgang 2003: Nr.2: 92-93
- S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. D. Moll, C. Huber, B. Schlegel, D. Pum, U. B. Sleytr, M. Sará. PNAS, 2002, Vol. 99, 23: 14646-14651.