Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 1059–1070
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Binding mode analysis and enrichment studies on homology models of the human histamine H4 receptor

Róbert Kiss [a], Béla Noszál [a], Ákos Rácz [a], András Falus [b], Dániel Erős [c], György M. Keserű [d,*]

[a] Department of Pharmaceutical Chemistry, Semmelweis University, Hőgyes Endre u. 9., H-1092, Budapest, Hungary
[b] Department of Genetics, Cell- and Immunobiology, Semmelweis University, Nagyvárad tér 4., H-1089, Budapest, Hungary
[c] Vichem Kft, Herman Ottó u. 15., H-1022, Budapest, Hungary
[d] Department of General and Analytical Chemistry, Budapest University of Technology and Economics, Szt. Gellért tér 4., H-1111, Budapest, Hungary

Received 16 May 2007; received in revised form 6 July 2007; accepted 17 July 2007
Available online 6 August 2007

Abstract

Ligand-supported homology models of the human histamine H4 receptor (hH4R) were developed based on the crystal structure of bovine rhodopsin and different known H4 ligands (histamine, OUP-16, JNJ7777120). Enrichment tests were performed to analyze whether our hH4R models can select known actives from random decoys. The impact of receptor conformation and the effect of different sets of random decoys, docking methods (FlexX, FlexX-Pharm) and scoring functions (FlexX-Score, D-Score, PMF-Score, G-Score, ChemScore) were investigated. We found that two agonists (histamine and OUP-16) form complementary interactions with Asp94 (3.32), Glu182 (5.46) and Thr323 (6.55), whereas JNJ7777120 interacts with Asp94 (3.32) and Glu182 (5.46) only. These results suggest a role of Thr323 (6.55) in ligand binding and presumably also in receptor activation. The models optimized in the presence of an agonist (histamine) and an antagonist (JNJ7777120) were compared in more detail. We conclude that the ligand used in the model building process can significantly influence the efficacy of virtual screening.

Keywords: Histamine; H4 receptor; Binding site; Enrichment; Docking; Scoring

1. Introduction

Histamine, the multifunctional messenger molecule, plays a crucial role in a number of biochemical processes. It serves as a mediator e.g. in hypersensitivity (allergic) responses, gastric acid secretion, neurotransmission, immuno-modulation, cell differentiation and embryonic development [1]. To date, four histamine receptor subtypes (H1R, H2R, H3R, H4R) have been identified [2]. All of these are members of the G-protein-coupled receptor (GPCR) superfamily. H1 and H2 antagonists have long been used in the treatment of inflammatorial and gastric hyperacidity diseases, respectively. H3R can stimulate the release of histamine and other neurotransmitters from neurons as a presynaptic autoreceptor [3,4], and H3 antagonists can be useful in cognitive and memory disorders [5] and obesity [6].

H4R is the novel member of the histamine receptor family. It was cloned by several groups independently in 2000 and 2001 [7–11]. H4R is closely related to H3R, but it shows as low homology for H1R and H2R as for other GPCRs [9]. Although H3R and H4R share high homology (with the sequence identity of 35%), their expression patterns are highly different. H3R is mainly expressed in the brain, while H4R has virtually no expression in the CNS [9,10]. H4R occurs in the bone marrow, spleen, eosinophils, mast cells, basophils, CD8 + T cells and dendritic cells [9,10,12–14]. It mediates

Abbreviations: GPCR, G-protein-coupled receptor; hH4R, human histamine H4 receptor; CNS, central nervous system; vHTS, virtual high throughput screening; IC, intracellular; EC, extracellular; HTS, high throughput screening; EF, enrichment factor; TM, transmembrane domain; SAR, structure–activity relationship.

* Corresponding author. Tel.: +36 1 4314605; fax: +36 1 4326002.
E-mail address: gy.keseru@richter.hu (G.M. Keserű).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.07.014

histamine-induced chemotaxis in mast cells and eosinophils, and this effect can be blocked by selective H4 antagonists [12,15,16]. H4R in common with H2R, has a role in the secretion of interleukin 16 (IL-16) from CD8 + T cells [17]. The expression pattern and the biological functions suggest a role of H4R in allergy and inflammation, being thus a potential target for drug design [18]. In spite of this, only a limited number of selective H4 ligands have been discovered so far.

H4R belongs to the class "A" of the GPCR superfamily. All of these receptors are transmembrane proteins and share similar structural features. They contain seven transmembrane helices and a helix which runs parallel to the cytosolic surface. Ligands as well as ligand binding sites of GPCRs are very heterogeneous [19]. The available mutational data indicate, however, that the binding site of most biogenic amines is located in the transmembrane region [19]. In fact, mutational data for H4R suggest that histamine has two major anchoring points: Asp94 (3.32 – Ballesteros–Weinstein numbering) and Glu182 (5.46) [20]. Mutating these residues to Ala caused a complete loss of affinity of histamine. These two residues are thought to interact with the protonated ethylamine side chain and the imidazole N(3)–H group of histamine. According to the proposed ligand–receptor model of Shin et al. [20] several residues (Thr178 (5.42), Ser179 (5.43), Ser320 (6.52), Asn147 (4.57)) have been mutated to identify the amino acid that interacts with N(1) in the imidazole ring of histamine. Nevertheless, the mutated residues have been found to have no significant effect on histamine binding. On the other hand, in this study Asn147 (4.57) and Ser320 (6.52) have been recognized to have a role in receptor activation.

G-protein-coupled receptors are the targets for 60–70% of drugs in development today [21], despite the fact that limited information is available on the 3D-structure of GPCRs. Until now, the only GPCR with an experimentally determined 3D-structure is bovine rhodopsin [22]. The high resolution crystal structure of bovine rhodopsin was a milestone in GPCR drug design; it allows building homology models of other GPCRs [23]. These models have been successfully used for the analysis of mutational data [24–38], mapping ligand binding sites and docking known ligands to these [24–33,38–41]. Some authors have recently reported GPCR homology models that had been suitable even for virtual screening [42–44]. Enrichment studies are usually used for testing whether a model can select known active compounds from a dataset containing actives and inactives. These tests show that homology models are usually less predictive than crystal structures [45]. They can be used, however, as cost effective alternatives of high throughput screening [46]. With this approach several novel active compounds have been already identified [29,43,44,47], showing the utility of these models for structure-based drug design.

It is generally accepted that the backbone conformations of GPCRs are very similar, especially in the transmembrane region, and can be modeled by using the bovine rhodopsin crystal structure. A much more difficult task is to determine the conformation of the amino acid side chains. Mutational data are useful to analyze the side chain conformations in the binding site of homology models. The orientation of the side chains can be either modified manually or by a more sophisticated approach using ligand information during the model building process [48]. Since several proteins have been cocrystallized with various ligands, it is well-known that the ligand has a great influence on the protein side chain conformations. The choice of the ligand can therefore have a large effect on screening efficacy, if the evident goal is to create a ligand-supported high quality homology model suitable for virtual high throughput screening (vHTS) [49].

We have developed homology models of the human histamine H4 receptor (hH4R) based on the crystal structure of bovine rhodopsin. The models with the available mutational data were used to understand the binding mode of several known H4 agonists and antagonists by means of docking calculations and subsequent optimizations of the complex. In this study we analyzed how model refinement and ligand information affect the quality and the predictivity of the homology models. An extensive enrichment study was carried out to determine whether any of the developed hH4R models is suitable for virtual screening.

2. Results and discussion

2.1. Sequence alignment

We have performed three automated sequence alignments by ClustalW [50] to make sure that the conserved GPCR elements are getting to the right positions. The amino acid sequence of bovine rhodopsin was aligned with (i) hH4R, (ii) all the available H4R sequences, and (iii) all the available H4R and H3R sequences. H3R was chosen because it shares the highest homology to H4R among GPCRs. In agreement with the proposed topology by Swiss-Prot (ID: Q9H3N8), the first and the third alignments indicated a large IC3 loop in hH4R having no equivalent in bovine rhodopsin. However, this loop region caused an error in the second alignment. We used the third alignment with the fewest gaps in the TM region as a starting point for the final sequence alignment. This alignment was manually modified to maintain the conserved (Cys-Cys) disulfide bridge, and the consistency of the transmembrane regions, that is believed to be more conserved than the EC or IC parts. The final alignment is depicted in Fig. 1.

2.2. Model building and validation

Six initial hH4R models were built by MODELLER 6v2 [51]. In accordance with the sequence alignment, Cys87 (EC2) and Cys164 (EC3) have formed a disulfide bridge in the model. Hydrogens were added and multichannel surfaces were displayed for every model in Sybyl 7.0 [52]. According to the mutational data, we looked for a binding cavity near the crucial histamine binding residues Asp94 (3.32) and Glu182 (5.46). In four models, the position of Tyr95 (3.33) precluded the interactions with Asp94 (3.32) and Glu182 (5.46) simultaneously, which is in contrast with site-directed mutagenesis data [20]. One of the two other models possessed a much more favourable binding cavity, with reliable orientation of

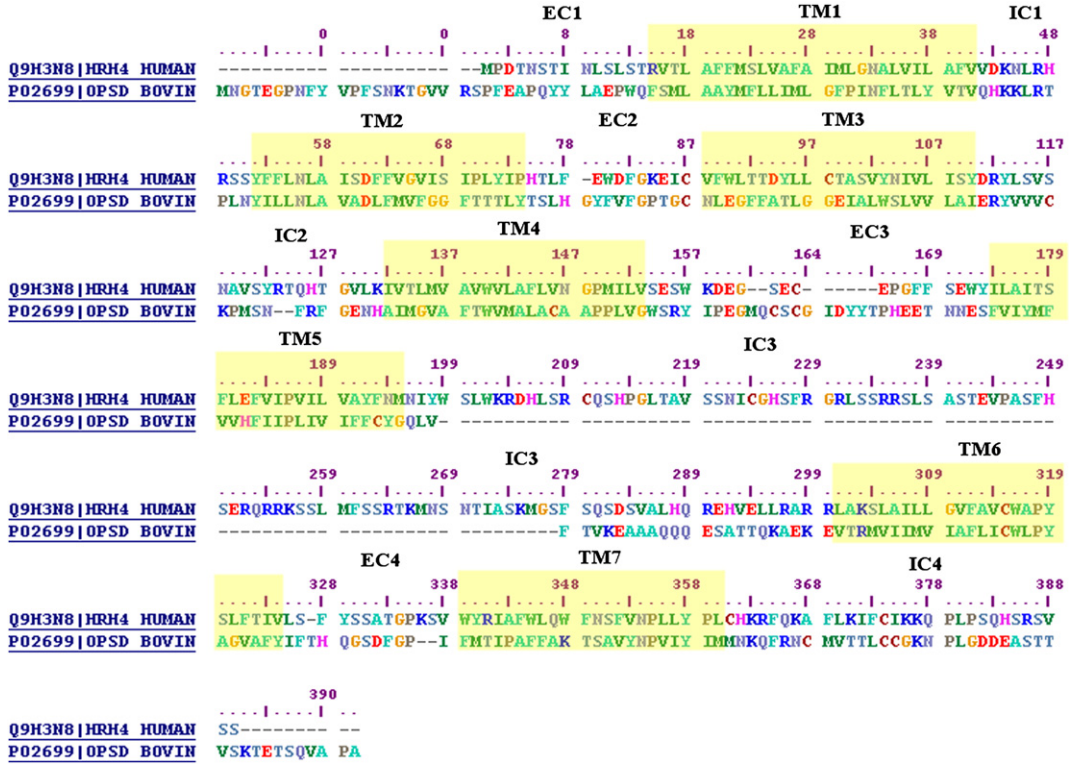

Fig. 1. Sequence alignment between the human histamine H4 receptor (HRH4_HUMAN) and bovine rhodopsin (OPSD_BOVIN).

Asp94 (3.32) and Glu182 (5.46) and sufficient space for ligand binding.

Several tests have been performed to check the quality of this hH4R model (Model A). The Ramachandran plots of the model and the template were generated by PROCHECK [53]. The Phi/Psi distribution shows that 97.5% of the residues in the model are in the most favoured or allowed regions, and only 5 and 2 residues are found in the generously allowed and disallowed regions, respectively. Similarly to the bovine rhodopsin crystal structure, most of the residues in the generously allowed and disallowed regions are located in the IC or the EC loops. Phe183 (5.47) in TM5 is the only exception, which faces the outer space, suggesting that this residue creates lipophilic interactions with the membrane, and therefore it stabilizes the protein. The unusual Phi/Psi values of Phe183 (5.47) are probable due to the lack of membrane environment in our model. Summarizing the results of the PROCHECK analysis, the stereochemical quality of the model is comparable to that of the template. After the docking and subsequent optimization the final models were analyzed again with PROCHECK. All the residues contributing directly to ligand binding were found in the most favoured or allowed regions of the Ramachandran plot.

Packing quality of the model was analyzed by the WHATIF Coarse Packing Quality Control test [54]. The overall quality of the model and the template is quite similar (−1.022 and −1.020, respectively). A residue with a score below −5.0 is probably problematic and needs further inspection. In the hH4R model and the bovine rhodospin template, 8 and 9 residues bear a score below −5.0, respectively. All of these poorly packed residues were found at the IC or EC loops. These results suggest a favourable packing environment in the hH4R model.

As a final test we used HARMONY [55] to evaluate whether the residues of the model adopt a structural environment that frequently occurs in protein structures. HARMONY indicates that the model and the template possess quite the same overall quality; however, some problematic sequences occur in both structures and most of these sequences overlap. A possible reason for this result is that high resolution experimental structures are available for only a few membrane proteins, hence these proteins are not well represented in the training set of HARMONY [56].

In summary, we can conclude that our hH4R model is in agreement with all the available mutational data and it possesses the common GPCR structural features. The results of the validation tests show that the model meets the criteria established for a reliable structure.

2.3. Binding site analysis and docking calculations

First, we docked histamine to the binding site by FlexX [57]. None of the resulting docking solutions were in accordance with the mutational data, as histamine formed only one interaction with Asp94 (3.32) or Glu182 (5.46). After inspection of the docking results we concluded that the carboxylate groups of Asp94 (3.32) and Glu182 (5.46) are too far to create interactions with donor groups of histamine, i.e. the protonated ethylamine side chain and the N(3)–H group. Moving the side chain of Glu182 (5.46) closer to Asp94 (3.32) was not

possible, on the other hand Asp94 (3.32) seemed to have more conformational freedom. Therefore we decided to rotate the carboxylate group of Asp94 (3.32) around the Cα–Cβ axis to accommodate histamine more favourably. Docking into this model (Model C) resulted in improved binding modes of histamine having simultaneous interactions with Asp94 (3.32) and Glu182 (5.46). Scores for these docking solutions were also improved. In 97% of the generated solutions the protonated ethylamine of histamine created an ionic interaction with Glu182 (5.46), and the N(3)–H formed an H-bond with Asp94 (3.32).

The H-bonding surfaces of the binding site and histamine show that Asp94 (3.32) and Glu182 (5.46) are hydrogen-accepting sites, and the protonated ethylamine and the N(3)–H moieties of histamine are two potential hydrogen donor groups (Fig. 2). One can therefore also hypothesize that two possible binding modes of histamine are possible at the binding site. Histamine protonated ethylamine side chain can interact with Glu182 (5.46) (binding mode I) or Asp94 (3.32) (binding mode II). The Glu182 (5.46) area of the binding site is more lipophobic than the surroundings of Asp94 (3.32). Thus, we proposed that binding mode I is favoured. Docking results revealed that in binding mode I the more lipophilic imidazole end of histamine prefers the more lipophilic environment of Asp94 (3.32). Our result, implicating a new binding mode of histamine in hH4R, underlines the importance of further experimental investigations of the hH4R–histamine complex.

We optimized the best scored histamine–hH4R complexes of both binding modes. Optimizations were carried out with rigid backbone atoms as well as without any constraints. However, in the former case clear planarity problems, caused by the close van der Waals contacts between aromatic amino acids (Tyr95 (3.33), Tyr319 (6.51), Trp348 (7.43)) and histamine occurred. Therefore, further studies were carried out only with structures optimized without backbone tethering.

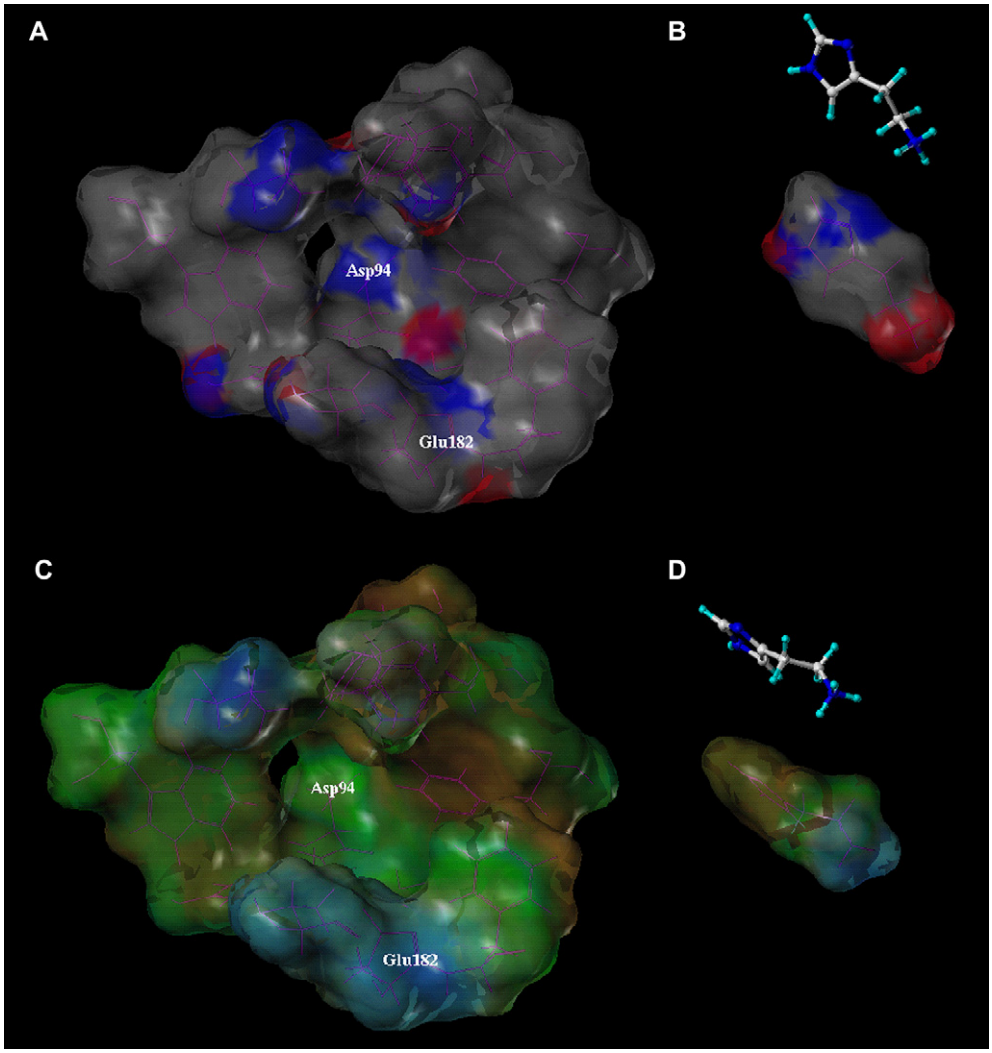

Fig. 2. The H-bonding surfaces of the hH4R binding site (A) and histamine (B). Potential H-bond acceptor and H-bond donor sites are coloured in blue and red, respectively. The lipophilic surfaces of the hH4R binding site (C) and histamine (D). Lipophilic, moderately lipophilic and lipophobic sites are coloured by brown, green and blue, respectively. The surfaces were generated by the multichannel and molecular surfaces modules in Sybyl7.0 (For interpretation of the references to color in these figure legends, the reader is referred to the web version of this article.).

The most important evaluating criterion of the resulting structures was whether they are in accordance with the site-directed mutagenesis data, i.e. histamine interacts with Asp94 (3.32) and Glu182 (5.46). Both resulting models preserved the specific features of GPCRs, such as the disulfide bond, and all the helices remained intact.

In binding mode II the H-bond between histamine and Asp94 (3.32) was still intact, however, the imidazole N(3)–H of histamine was far away from Glu182 (5.46). In binding mode I both interactions with Asp94 (3.32) and Glu182 (5.46) were intact, moreover, histamine N(1) formed an H-bond with Thr323 (6.55) (Fig. 3). Histamine created lipophilic interactions with Trp316 (6.48) and Tyr319 (6.51) after optimization of binding mode I (Model D).

Since no mutational data were reported on Thr323 (6.55), its role in histamine binding needs to be investigated in future. It has been previously proved that photoactivation of bovine rhodopsin can induce movements of TM3 and TM6, relative to the rest of the 7-helix bundle, a significant rigid body movement of TM6 in a counterclockwise direction when viewed from the extracellular side, and a movement of the cytoplasmic end of TM6 away from TM3 [19]. In their recent publication Salom et al. reported the crystal structure of the photoactivated bovine rhodopsin [58]. They stated that despite the low resolution of the obtained crystals, some conclusions can be drawn from the comparison of the ground-state and the photoactivated forms. In particular, in the photoactivated form the middle of TM3 is quite mobile, and the IC3 region (continuation of TM5 and TM6) is disordered having a significantly different architecture from that of the ground-state. The analog of Thr323 (6.55) in bovine rhodopsin is Ala272 (6.55). The backbone nitrogen of Ala272 (6.55) in bovine rhodopsin creates an H-bond with the backbone oxygen of Tyr268 (6.51), that is supposed to change its conformation significantly during activation [59]. These findings together with our results suggest a role of Thr323 (6.55) in ligand binding and also in receptor activation.

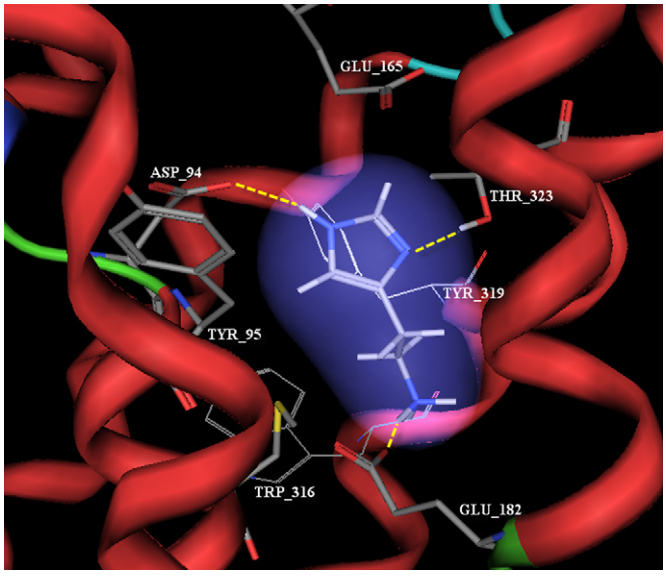

Fig. 3. Histamine at the hH4R binding site. Interactions with Asp94 (3.32), Glu182 (5.46) and Thr323 (6.55) are marked with dashed yellow lines.

In the light of these results, binding mode I of histamine, where the ethylamine part interacts with Glu182 (5.46), the imidazole N(3)–H interacts with Asp94 (3.32) and the imidazole N(1) interacts with Thr323 (6.55) seems to be more realistic. Shin et al. modeled histamine at the hH4R binding site in binding mode II, and proposed that the imidazole N(1) interacts with Thr178 (5.42) or Ser179 (5.43) in TM5 [20]. Experimental studies, however, did not validate the role of these residues in histamine binding [20]. This binding mode (binding mode II) in fact is similar to that in H1R and H2R. Site-directed mutagenesis studies revealed that the protonated ethylamine function of histamine forms H-bond to the conserved Asp (3.32) in both cases. In the case of H1R, the imidazole N(3)–H and N(1) interact with Asn198 (5.46) [60] and Lys191 (5.39) [61], respectively, while in H2R, the imidazole N(3)–H and N(1) form interactions with Asp186 (5.42) and Thr190 (5.46), respectively [62]. H4R has a similarly low level of homology for H1R and H2R as for any other GPCR [9]. This would suggest a convergent phylogenetic evolution of H4R, H1R and H2R, and therefore the proposed orientation of histamine at the binding site of H1R or H2R cannot be simply applied in the case of the H4R, which is further supported by the mutagenesis data [20].

We docked altogether fourteen known H4 agonists and antagonists (Fig. 4) into the hH4R model by FlexX. Docking to the receptor model, where the side chain conformation of Asp94 (3.32) was modified, resulted in significantly better solutions with every ligand, i.e. the scores were much lower, and most of the ligands have created interactions with both Asp94 (3.32) and Glu182 (5.46).

OUP-16 was the first reported H4 agonist with a considerable selectivity over H3R [63]. Here we analyzed whether it occupies a similar position at the binding site as histamine does. The best scored OUP-16–receptor complex was therefore optimized (Model E). Optimization left the interactions with Asp94 (3.32) and Glu182 (5.46) intact, and formed an additional interaction with Thr323 (6.55) (Fig. 5). Interestingly, it was not the N(1) of the imidazole ring but the CN-group of OUP-16 that created the interaction with Thr323 (6.55). Lipophilic interactions with Tyr95 (3.33), Trp157 (4.67), Trp316 (6.48), Tyr319 (6.51) and Phe344 (7.39) have also been formed.

JNJ7777120 was the first reported selective H4 antagonist [64]. SAR studies on JNJ7777120 and structural analogs have proven that both the piperazine nitrogen and the indole NH group are crucial for H4 affinity [64,65]. Since JNJ7777120 has two H-bond donors similar to histamine, we proposed that it creates two H-bonds or ionic interactions with Asp94 (3.32) and Glu182 (5.46). Docking JNJ7777120 only resulted in a single interaction between the piperazine part of the molecule and Glu182 (5.46) independently on the orientation of Asp94 (3.32). The indole part did not form any H-bond with the receptor. After optimization of the complex even the interaction with Glu182 (5.46) disappeared. Since the docking of JNJ7777120 by FlexX failed, we

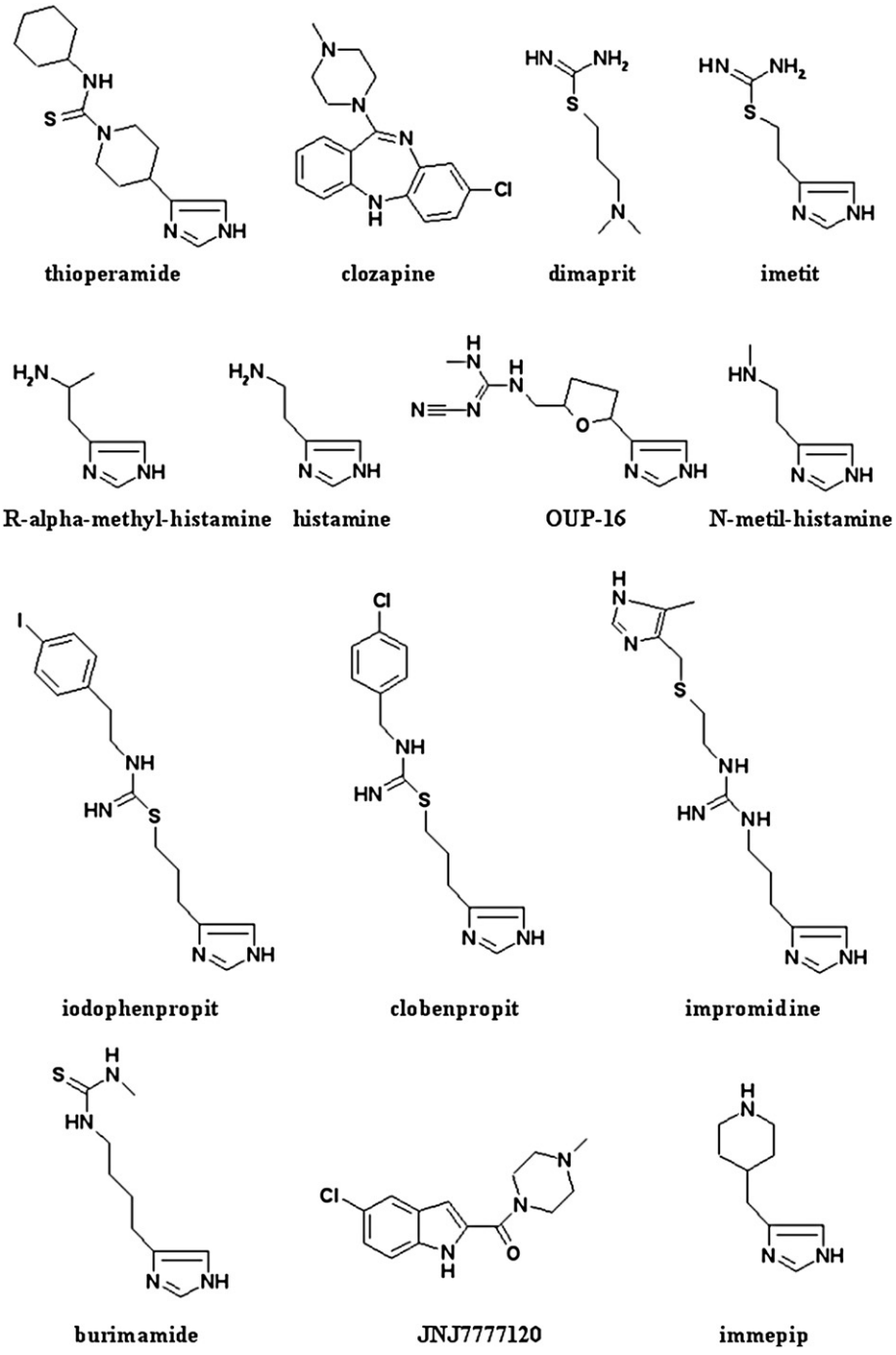

Fig. 4. H4 ligands used for the enrichment studies.

assumed that insufficient conformational flexibility of the protein side chains was allowed during the docking process. Therefore we docked JNJ7777120 to the binding site by FlexiDock, allowing all side chains within the binding site to rotate freely. As a result, the indole and the piperazine part of JNJ7777120 formed interactions with Asp94 (3.32) and Glu182 (5.46), respectively. After minimizing the FlexiDock output (Model F), both interactions remained intact, that finally yielded a JNJ7777120–receptor complex in agreement with the SAR data (Fig. 6). It is also remarkable that no interaction appeared between JNJ7777120 and Thr323 (6.55) in TM6; a distinctive aspect from the two agonists (histamine, OUP-16). JNJ7777120 has formed lipophilic interactions with Val64 (2.53), Phe312 (6.44), Trp316 (6.48), Tyr319 (6.51) and Trp348 (7.43).

2.4. Enrichment studies

To investigate the utility of the developed models for virtual screening, and the impact of ligand information on screening efficacy, we carried out several enrichment tests. In all of these studies we used FlexX that demonstrated an impressive track

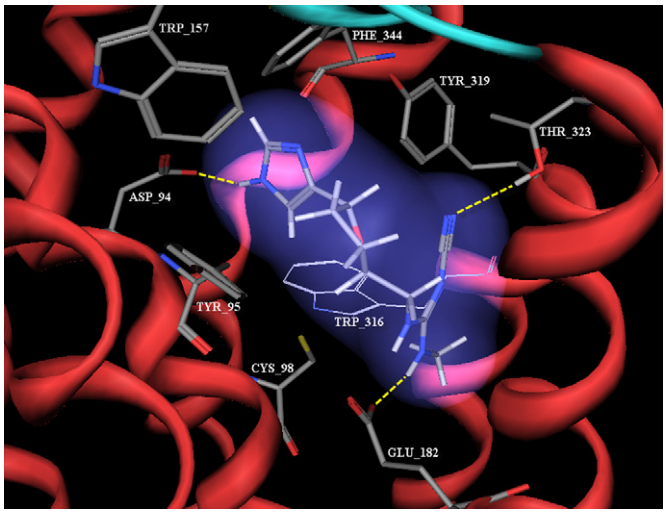

Fig. 5. OUP-16 at the hH4R binding site. Interactions with Asp94 (3.32), Glu182 (5.46) and Thr323 (6.55) are marked with dashed yellow lines.

record in structure-based virtual screening studies [66–71]. The high speed of FlexX enables testing large databases, even millions of compounds in a reasonable time frame. As one of our long-term objectives was to carry out a large scale virtual screening on the hH4R model, speed was also a very important aspect to consider. In summary, six different protein structures were evaluated for virtual screening by means of enrichment studies (Table 1). During the enrichment studies we also analyzed the influence of using different sets of inactive compounds. Furthermore, we investigated the effect of pharmacophore constraints (FlexX-Pharm) on the enrichment factors. As the pose extraction and ranking are separate steps of the ligand–receptor complex evaluation [72], a scoring function can handle these two steps with different efficiency. For that reason, all of the five scoring functions were applied for both extraction and ranking, consequently enrichment factors for 25 different scoring function combinations were calculated.

The results show, that the effect of different sets of inactives was limited on either the highest enrichment factor achieved

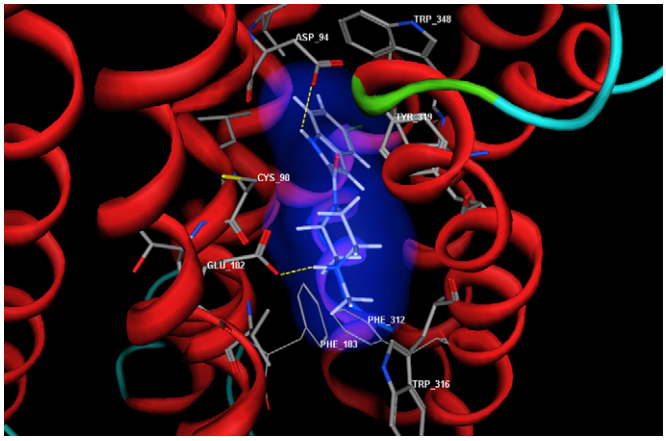

Fig. 6. JNJ7777120 at the hH4R binding site. Interactions with Asp94 (3.32) and Glu182 (5.46) are marked with dashed yellow lines.

Table 1
Designation of the hH4R models used in the enrichment studies

Model A	MODELLER output
Model B	MODELLER output optimized
Model C	MODELLER output with a modified Asp94 side chain conformation
Model D	Histamine docked by FlexX and optimized
Model E	OUP-16 docked by FlexX and optimized
Model F	JNJ7777120 docked by FlexiDock and optimized

or the minimal % of compounds that had to be tested to recover all the actives (Fig. 7). This means that – at least in our study – very similar conclusions could be drawn from the tests with an active ratio of 1.38%, 0.94% or 0.46%. These results also strengthen that the calculated enrichment factors are determined by the quality of the given hH4R model, and are not products of serendipity. The MODELLER output (Model A) and its optimized version (Model B) performed quite well in terms of the enrichment factors, however, with these models 60–65% of the compounds needed to be tested to recover all the actives. Modification of the side chain conformation of Asp94 (3.32) (Model C) had a significant effect on the highest enrichment factors achievable. Furthermore, all actives could be recovered by testing only 25–30% of the whole compound set. The OUP-16-supported model (Model E) did not show any improvement compared to the original MODELLER output structure (Model A). The JNJ7777120-supported model (Model F) performed moderately relative to the other ones. On the other hand, the optimization of the protein in the presence of histamine (Model D) achieved the highest enrichment factors.

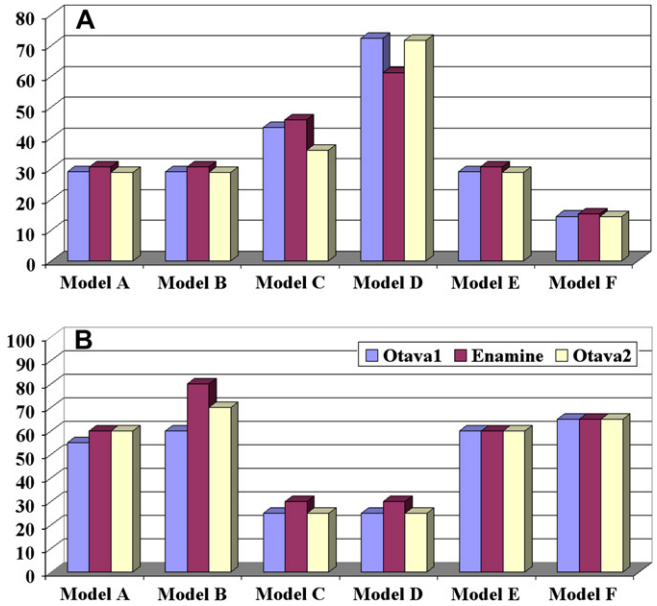

Fig. 7. Highest enrichment factors could be achieved by the six different models (A). The minimum % of the compound database to be tested to recover all actives (B). Enrichment tests were performed with three different sets of inactives (Otava1, Enamine, Otava2). The highest enrichment factors could be achieved when the top 0.5% of the database was analyzed.

Table 2
Enrichment factor improvement using pharmacophoric constraints (FlexX-Pharm)

	Otava1 inactive set		Enamine inactive set	
	FlexX	FlexX-Pharm	FlexX	FlexX-Pharm
Model A	28.9	31.2	30.6	32.9
Model C	43.4	50.7	45.8	49.9
Model D	72.4	92.1	61.1	77.8
Model E	28.9	50.7	30.6	35.6
Model F	14.5	16.9	15.3	17.8

Enrichments studies were also performed with pharmacophore constraints introduced by FlexX-Pharm. In the case of ''Model B'' numerous actives could not be recovered. This is probably due to the absence of a ligand during the optimization process, which resulted in the shrinkage of the binding crevice. Consequently, only the three smallest active ligands could form an interaction with either Asp94 (3.32) or Glu182 (5.46). Interestingly, no ligands of the inactive set Enamine and only four ligands of the inactive set Otava1 could be docked to ''Model B'' with pharmacophore constraints. For that reason, enrichment factors for this model were not calculated. In the case of the other models, some improvement in the enrichment factors could be obtained with FlexX-Pharm, especially with ''Model D'' (Table 2). On the other hand, it also has to be mentioned that recovering all the actives was only possible with ''Model C'', whereas ''Model A'', ''Model D'', ''Model E'' and ''Model F'' could only find 13, 11, 12 and 12 actives, respectively. This means that virtual screening on these models with FlexX-Pharm could loose a couple of actives. However, even in the case of the best model (Model D), without using the Pharm-module, 25% of the database has to be tested to recover all the actives. Since usually 1–2% of a virtually screened database is tested in vitro, it is clear that some of the actives will not be found in either way. Therefore using the protocol achieving the highest enrichment factors (Model D by FlexX-Pharm) at the top 0.5–2% is probably the best choice (Fig. 8).

In agreement with the results of Schulz-Gasch et al. [72] the FlexX-FlexX, FlexX-ChemScore and FlexX-G-Score combinations were found to provide the best results. These authors found that an incremental docking method like FlexX, in combination with a ''hard'' scoring function like the own score of FlexX or the closely related, robust scoring function Chemscore can be used with excellent results.

Comparing the enrichment factors achieved by our models to others reported on GPCR homology models, the performance is significant. In a comparative vHTS study, Evers et al. reported the homology models of four GPCRs [73]. Their enrichment factors calculated at the top 1%, 5% and 10% vary from 4 to 12. In another paper, Evers and his coworkers report a homology model of the Alpha1A receptor [44]. The highest enrichment factor achieved was 13.1 calculated at the top 1% of their database. It is difficult to compare enrichment factors calculated by different groups using different protocols, however, we can conclude that our models can select known actives from random decoys effectively.

''Model D'' and ''Model F'' were created by the optimization of the protein in the presence of two different H4 ligands. The enrichment results showed the largest differences in these two models, therefore we compared the scores of the active and inactive compounds. In the histamine supported model (Model D) 12 of the 14 active compounds were scored better

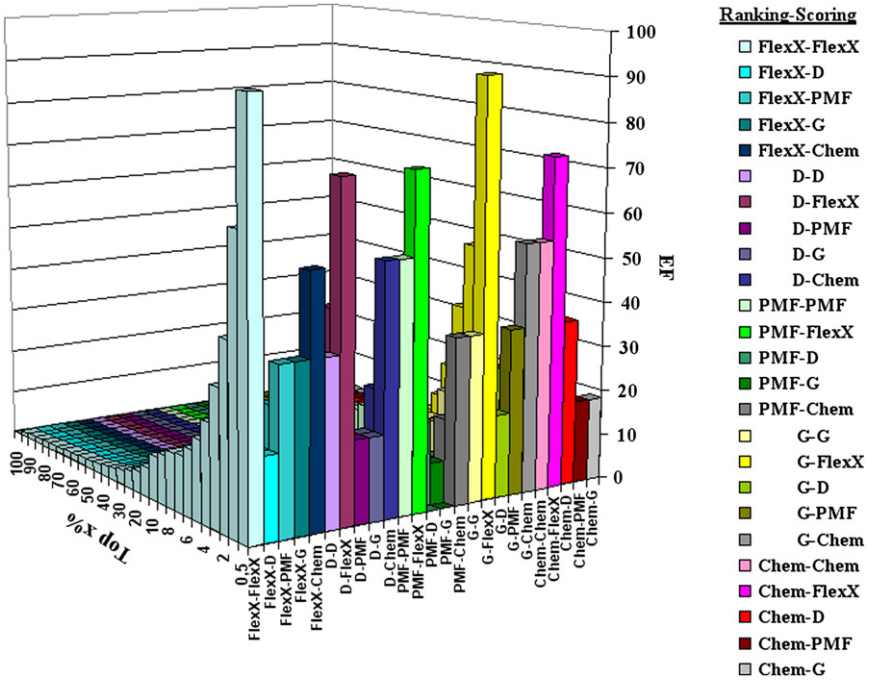

Fig. 8. Enrichment factors calculated by 25 different scoring function combinations for ''Model D'' using FlexX-Pharm.

by the majority of the scoring function combinations. Only clozapine and JNJ7777120 were found to fit better into the JNJ7777120-supported model (Model F). On the other hand, 67.1% (Otava1), 63.3% (Enamine) and 68.0% (Otava2) of the successfully docked inactive compounds were scored better by the majority of the scoring function combinations in "Model F". Comparing the number of inactive compounds that were impossible to dock into the models, 2.46 (Otava1), 2.14 (Enamine) and 2.39 (Otava2) times more inactives could not be docked by FlexX into "Model D" than into "Model F".

Investigating the structural background of the different performances, we found three significant conformational differences between the two protein structures (Fig. 9). One of these is the side chain orientation of Asp94 (3.32). In the model optimized with histamine (Model D), Asp94 (3.32) faces the interior of the binding crevice, and therefore it decreases its volume relative to "Model F". Another important difference is that Trp316 (6.48) in "Model F" occupies a totally different position than that in "Model D". This means a considerable increase in the volume of the binding crevice. The third important difference is that in "Model D", Cys98 (3.36) forms an H-bond with one of the carboxylate oxygens of Glu182 (5.46), therefore ligands are restricted to interact with the other oxygen of Glu182 (5.46), which also reduces the number of possible binding modes in "Model D". We speculate that these three differences contribute to the higher selectivity of "Model D" over "Model F", since less random decoys can occupy the significantly smaller binding site of "Model D". The reduced number of accessible interacting sites of Glu182 (5.46) in "Model D" also makes the binding of inactives more difficult. On the other hand, we propose that the better binding of actives in "Model D" is mainly related to the side chain conformation of Asp94 (3.32) and to some extent also to Trp316 (6.48), since it creates a lipophilic interaction with most of the actives in this model.

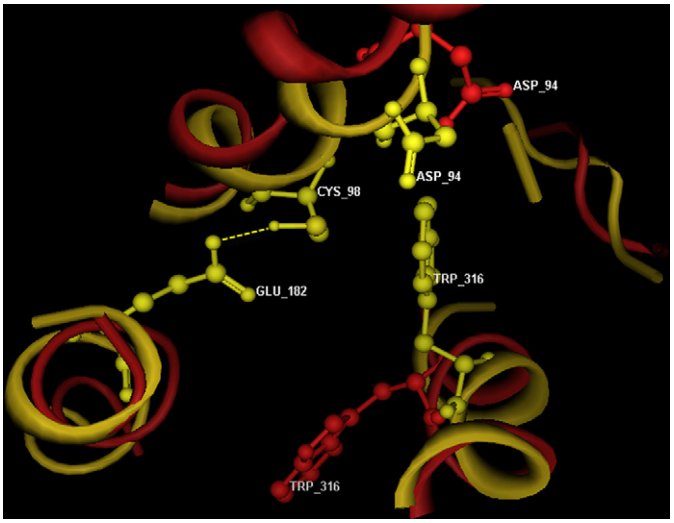

Fig. 9. Main differences between the binding sites obtained by the optimization with histamine ("Model D" in yellow) and JNJ7777120 ("Model F" in red).

3. Conclusions

In conclusion, we have created homology models of the human histamine H4 receptor, also in its complexes with known active ligands. Two H4 agonists, histamine and OUP-16 were found to form interactions with Asp94 (3.32), Glu182 (5.46) and Thr323 (6.55). On the other hand JNJ7777120, an H4 antagonist formed interactions with Asp94 (3.32) and Glu182 (5.46) only. These models were in agreement with all literature data. Our results suggest, that histamine has a different orientation at the hH4R binding site than it was previously proposed. However, the verification of this binding mode and also the role of Thr323 (6.55) needs further support from both theoretical (membrane-based MD simulations) and experimental (site-directed mutagenesis) studies. Membrane-based MD simulation of the hH4R–histamine complex is in progress at our laboratory. Enrichment tests were performed on six different models, using five different scoring functions, and three different inactive sets with docking program FlexX. We found that ligand information can significantly affect the performance of the homology models in virtual screening. Enrichment factors obtained were reasonably high to conclude that some of the presented models (especially "Model D") can be useful for virtual screening to find new H4 ligands.

4. Experimental protocols

4.1. Sequence alignment

The amino acid sequences of bovine rhodopsin and the histamine H4 (human, mouse, rat, cavpo, pig) and H3 receptors (rat, mouse, cavpo, human) were downloaded from the Swiss-Prot database. All automated sequence alignments were carried out by ClustalW1.83 on the freely available automated web server of EBI (European Bioinformatics Institute; http://www.ebi.ac.uk). The alignments were analyzed and modified in BioEdit version 7.0.5.3, considering the position of conserved GPCR specific amino acids. The long IC3 loop of hH4R was not modeled, since it has no equivalent in bovine rhodopsin and it does not contribute to ligand binding.

4.2. Model building

Using the final alignment six initial models were built by MODELLER 6v2. MODELLER was running with a slightly modified python file to achieve more thoroughly refined models, than the default ones (LIBRARY_SCHEDULE = 1; MAX_VAR_ITERATIONS = 500; MD_LEVEL = 'REFINE3'; RSTRS_REFINED = 2; REPEAT_OPTIMIZATION = 5). With this protocol, geometry optimization and molecular dynamics simulations were performed at a higher level during the MODELLER model building process. It is also important to mention that as the molecular dynamics were performed by MODELLER, the constraints derived from the template structure prevent the homology model from unfolding in vacuum. Hydrogens were added and the models were examined in Sybyl 7.0.

The model, which seemed to possess the most favourable ligand binding site, was chosen for further investigations. This model was subjected to several tests to make sure that it was of reasonable quality for analyzing ligand binding, and to check its ability for virtual screening.

4.3. Model validation

The hH4R homology model was validated by three independent tests. In all three tests the quality of the model was compared to that of the template crystal structure to make sure that we did not commit an error during the sequence alignment or the model building process.

The PROCHECK software was applied to analyze the stereochemical quality of the models. The Phi/Psi Ramachandran plot indicated the amino acids with an unusual backbone conformation, comparing them to the preferred values of proteins in the Protein Data Bank.

The packing quality of the models was inspected by another validation test, using the WHATIF Coarse Packing Quality Control value. The packing environment of every single residue is compared to the average distribution observed in the Protein Data Bank.

The third test assessed the compatibility of a particular amino acid sequence in the model with a proposed 3D-structure. In this test (accomplished by HARMONY, http://caps.ncbs.res.in/harmony) the tendency of an amino acid to adopt a putative structural environment is quantified by its occurrence in a large number of known structures. The reverse sequence of the tested protein and its scores are used as a control to identify local errors in the proposed protein model.

4.4. Docking calculations

Docking calculations were performed by FlexX version 1.13.5 and FlexiDock in Sybyl7.0. FlexX was running with formal charges assignment and every other parameter was set as default. Based on the visualization of the multichannel surfaces in our hH4R model we have defined the active site as 5 Å radius circles around residues Asp94 (3.32), Glu182 (5.46) and Asn350 (7.45). When pharmacophoric constraints were applied, docking results with at least one interaction with the carboxylate group of Asp94 (3.32) or Glu182 (5.46) were allowed only.

FlexiDock was run with a modified parameter file, to examine a more thorough sampling of the conformational space available for the ligand at the binding site; number of islands was set from 1 to 50, maximum number of members saved per island was set from 20 to 50, the population size for all islands was set from 500 to 1500 and the full mutation chance per gene was increased from 0.1 to 0.3.

4.5. Energy minimization

Since MODELLER uses energy terms, atom and residue types based on the CHARMM22 force field, all the optimization processes were carried out with the CHARMM22 force field and charges in MOE [74]. Parameters were set as default, except the convergence gradient was set to 0.01 kcal/mol instead of 0.05 kcal/mol. All the calculations were terminated when this convergence criterion had been reached. In calculations when only the side chain atoms were allowed to move freely, tethering was set to 100 000 on every backbone atom of the system.

4.6. Enrichment tests

In the enrichment studies three different sets of inactive compounds were used to test whether the active/inactive ratio of the compounds influenced the results (14 active compounds vs. 999, 1483 and 3002 inactive ones). As one of our further goals was to perform virtual screening of huge compound databases, the inactive molecules were collected from the drug-like and diverse sets of two major chemical companies (Otava: www.otavachemicals.com; Enamine: www.enamine.net). All of these compounds were ''washed'' by MOE, to generate the most probable protonation state of them. 3D coordinates were generated by Concord 4.0 in Sybyl7.0. The active compounds were collected from the available literature data [11,18,64]. Since the H4 receptor was cloned quite recently, there are only a few H4 agonists and antagonists with diverse structures. Considering this, and that the hit ratio in high throughput screening (HTS) is usually 0.5–1%, we have chosen 14 active compounds for the enrichment studies (Fig. 4). pK_a values of all the actives were calculated by CompuDrug's Pallas software [75]. Histamine was considered as a monocation (protonated on the ethylamine side chain) and all imidazole derivatives were considered in the N(τ) form (protonated on N(3) and deprotonated on N(1); see depiction on Fig. 4), supported by the Pallas calculations as well as other available theoretical [76,77] and experimental literature data [78,79] on histamine and its derivatives.

Enrichment tests were performed on six different H4 models with all three inactive sets by the docking program FlexX. Pharmacophoric constraints were also applied in case of every model with two inactive sets (Otava1, Enamine).

For evaluating the docked poses of the compounds, five different scoring functions were used. The CScore package contains four scoring functions (D-score, G-score, PMF-score, Chemscore), and the own score of FlexX was also used.

The enrichment factors were calculated as follows: EF (%) = $(N_{\text{active}(\%)} \times N_{\text{all}})/(N_{(\%)} \times N_{\text{active}})$ where $N_{\text{active}(\%)}$ is the number of actives found in the top x%, N_{all} is the number of the compounds used in the test, $N_{(\%)}$ is the x% of the compounds used in the calculation of EF (%), N_{active} is the number of all actives used in the test. As one scoring function can be used to rank the docked poses for a specific compound, and another scoring function can be applied to compare the best poses of the different compounds, we have calculated the enrichment factors for every 25 scoring function combinations.

Acknowledgement

The financial support provided by the Grants OTKA (T 042933 and T 43579) of the Hungarian National Research

Fund is gratefully acknowledged. We are grateful for the outstanding support of BioSolveIT GmbH the sole distributor of FlexX.

Appendix. Supplementary data

Phi/Psi Ramachandran plot, WHATIF score distribution and static images generated by HARMONY for both hH4R (Model A) and bovine rhodopsin. Physicochemical properties such as molecular weight and log *P* calculated by MOE for all the actives and inactives. All active and inactive compound sets are available from the authors upon request. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2007.07.014.

References

[1] M.J. Nissinen, K. Karlstedt, E. Castren, P. Panula, J. Histochem. Cytochem. 43 (1995) 1241–1252.
[2] L.B. Hough, Mol. Pharmacol. 59 (2001) 415–419.
[3] J.M. Arrang, M. Garbarg, J.C. Schwartz, Nature 302 (1983) 832–837.
[4] E. Schlicker, B. Malinowska, M. Kathmann, M. Gothert, Fundam. Clin. Pharmacol. 8 (1994) 128–137.
[5] R. Leurs, P. Blandina, C. Tedford, H. Timmerman, Trends Pharmacol. Sci. 19 (1998) 177–183.
[6] K. Takahashi, H. Suwa, T. Ishikawa, H. Kotani, J. Clin. Invest. 110 (2002) 1791–1799.
[7] T. Nguyen, D.A. Shapiro, S.R. George, V. Setola, D.K. Lee, R. Cheng, L. Rauser, S.P. Lee, K.R. Lynch, B.L. Roth, B.F. O'Dowd, Mol. Pharmacol. 59 (2001) 427–433.
[8] T. Oda, N. Morikawa, Y. Saito, Y. Masuho, S. Matsumoto, J. Biol. Chem. 275 (2000) 36781–36786.
[9] C. Liu, X. Ma, X. Jiang, S.J. Wilson, C.L. Hofstra, J. Blevitt, J. Pyati, X. Li, W. Chai, N. Carruthers, T.W. Lovenberg, Mol. Pharmacol. 59 (2001) 420–426.
[10] K.L. Morse, J. Behan, T.M. Laz, R.E. West Jr., S.A. Greenfeder, J.C. Anthes, S. Umland, Y. Wan, R.W. Hipkin, W. Gonsiorek, N. Shin, E.L. Gustafson, X. Qiao, S. Wang, J.A. Hedrick, J. Greene, M. Bayne, F.J. Monsma Jr., J. Pharmacol. Exp. Ther. 296 (2001) 1058–1066.
[11] Y. Zhu, D. Michalovich, H. Wu, K.B. Tan, G.M. Dytko, I.J. Mannan, R. Boyce, J. Alston, L.A. Tierney, X. Li, N.C. Herrity, L. Vawter, H.M. Sarau, R.S. Ames, C.M. Davenport, J.P. Hieble, S. Wilson, D.J. Bergsma, L.R. Fitzgerald, Mol. Pharmacol. 59 (2001) 434–441.
[12] C.L. Hofstra, P.J. Desai, R.L. Thurmond, W.P. Fung-Leung, J. Pharmacol. Exp. Ther. 305 (2003) 1212–1221.
[13] U. Lippert, M. Artuc, A. Grutzkau, M. Babina, S. Guhl, I. Haase, V. Blaschke, K. Zachmann, M. Knosalla, P. Middel, S. Kruger-Krasagakis, B.M. Henz, J. Invest. Dermatol. 123 (2004) 116–123.
[14] D. Voehringer, K. Shinkai, R.M. Locksley, Immunity 20 (2004) 267–277.
[15] K.F. Buckland, T.J. Williams, D.M. Conroy, Br. J. Pharmacol. 140 (2003) 1117–1127.
[16] P. Ling, K. Ngo, S. Nguyen, R.L. Thurmond, J.P. Edwards, L. Karlsson, W.P. Fung-Leung, Br. J. Pharmacol. 142 (2004) 161–171.
[17] F. Gantner, K. Sakai, M.W. Tusche, W.W. Cruikshank, D.M. Center, K.B. Bacon, J. Pharmacol. Exp. Ther. 303 (2002) 300–307.
[18] I.J. de Esch, R.L. Thurmond, A. Jongejan, R. Leurs, Trends Pharmacol. Sci. 26 (2005) 462–469.
[19] K. Kristiansen, Pharmacol. Ther. 103 (2004) 21–80.
[20] N. Shin, E. Coates, N.J. Murgolo, K.L. Morse, M. Bayne, C.D. Strader, F.J. Monsma Jr., Mol. Pharmacol. 62 (2002) 38–47.
[21] K. Lundstrom, Trends Biotechnol. 23 (2005) 103–108.
[22] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Science 289 (2000) 739–745.
[23] J. Ballesteros, K. Palczewski, Curr. Opin. Drug Discov. Devel. 4 (2001) 561–574.
[24] A. Pedretti, M. Elena Silva, L. Villa, G. Vistoli, Biochem. Biophys. Res. Commun. 319 (2004) 493–500.
[25] K. Johren, H.D. Holtje, J. Comput. Aided Mol. Des. 16 (2002) 795–801.
[26] K.E. Furse, T.P. Lybrand, J. Med. Chem. 46 (2003) 4450–4462.
[27] B.M. Broer, M. Gurrath, H.D. Holtje, J. Comput. Aided Mol. Des. 17 (2003) 739–754.
[28] S. Costanzi, L. Mamedova, Z.G. Gao, K.A. Jacobson, J. Med. Chem. 47 (2004) 5393–5404.
[29] D.E. Clark, C. Higgs, S.P. Wren, H.J. Dyke, M. Wong, D. Norman, P.M. Lockey, A.G. Roach, J. Med. Chem. 47 (2004) 3962–3971.
[30] O.M. Salo, M. Lahtela-Kakkonen, J. Gynther, T. Jarvinen, A. Poso, J. Med. Chem. 47 (2004) 3048–3057.
[31] C. Montero, N.E. Campillo, P. Goya, J.A. Paez, Eur. J. Med. Chem. 40 (2005) 75–83.
[32] S.K. Kim, Z.G. Gao, L.S. Jeong, K.A. Jacobson, J. Mol. Graph. Model. 25 (2006) 562–577.
[33] M.R. Braden, J.C. Parrish, J.C. Naylor, D.E. Nichols, Mol. Pharmacol. 70 (2006) 1956–1964.
[34] P.L. Freddolino, M.Y. Kalani, N. Vaidehi, W.B. Floriano, S.E. Hall, R.J. Trabanino, V.W. Kam, W.A. Goddard, Proc. Natl. Acad. Sci. USA 101 (2004) 2736–2741.
[35] J.J. Chambers, D.E. Nichols, J. Comput. Aided Mol. Des. 16 (2002) 511–520.
[36] X.Q. Xie, J.Z. Chen, E.M. Billings, Proteins 53 (2003) 307–319.
[37] H. Gutierrez-de-Teran, N.B. Centeno, M. Pastor, F. Sanz, Proteins 54 (2004) 705–715.
[38] J. Onuffer, M.A. McCarrick, L. Dunning, M. Liang, M. Rosser, G.P. Wei, H. Ng, R. Horuk, J. Immunol. 170 (2003) 1910–1916.
[39] A.M. ter Laak, H. Timmerman, R. Leurs, P.H. Nederkoorn, M.J. Smit, G.M. Donne-Op den Kelder, J. Comput. Aided Mol. Des. 9 (1995) 319–330.
[40] B. Mouillac, B. Chini, M.N. Balestre, J. Elands, S. Trumpp-Kallmeyer, J. Hoflack, M. Hibert, S. Jard, C. Barberis, J. Biol. Chem. 270 (1995) 25771–25777.
[41] M.F. Hibert, S. Trumpp-Kallmeyer, A. Bruinvels, J. Hoflack, Mol. Pharmacol. 40 (1991) 8–15.
[42] A. Evers, G. Klebe, Angew. Chem., Int. Ed Engl. 43 (2004) 248–251.
[43] A. Evers, G. Klebe, J. Med. Chem. 47 (2004) 5381–5392.
[44] A. Evers, T. Klabunde, J. Med. Chem. 48 (2005) 1088–1097.
[45] S.L. McGovern, B.K. Shoichet, J. Med. Chem. 46 (2003) 2895–2907.
[46] J.L. Jenkins, R.Y. Kao, R. Shapiro, Proteins 50 (2003) 81–93.
[47] J. Varady, X. Wu, X. Fang, J. Min, Z. Hu, B. Levant, S. Wang, J. Med. Chem. 46 (2003) 4377–4392.
[48] A. Evers, H. Gohlke, G. Klebe, J. Mol. Biol. 334 (2003) 327–345.
[49] A. Urzhumtsev, F. Tete-Favier, A. Mitschler, J. Barbanton, P. Barth, L. Urzhumtseva, J.F. Biellmann, A. Podjarny, D. Moras, Structure 5 (1997) 601–612.
[50] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[51] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[52] Sybyl, version 7.0; Tripos Associates: St. Louis, MO.
[53] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thorton, J. Appl. Crystallogr. 26 (1993) 283–291.
[54] G. Vriend, C. Sander, J. Appl. Crystallogr. 26 (1993) 47–60.
[55] G. Pugalenthi, K. Shameer, N. Srinivasan, R. Sowdhamini, Nucleic Acids Res. 34 (2006) W231–W234.
[56] A. Patny, P.V. Desai, M.A. Avery, Curr. Med. Chem. 13 (2006) 1667–1691.
[57] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[58] D. Salom, D.T. Lodowski, R.E. Stenkamp, I. Le Trong, M. Golczak, B. Jastrzebska, T. Harris, J.A. Ballesteros, K. Palczewski, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 16123–16128.

[59] P.J. Spooner, J.M. Sharples, S.C. Goodall, P.H. Bovee-Geurts, M.A. Verhoeven, J. Lugtenburg, A.M. Pistorius, W.J. Degrip, A. Watts, J. Mol. Biol. 343 (2004) 719–730.
[60] K. Ohta, H. Hayashi, H. Mizuguchi, H. Kagamiyama, K. Fujimoto, H. Fukui, Biochem. Biophys. Res. Commun. 203 (1994) 1096–1101.
[61] R. Leurs, M.J. Smit, R. Meeder, A.M. Ter Laak, H. Timmerman, Biochem. Biophys. Res. Commun. 214 (1995) 110–117.
[62] I. Gantz, J. DelValle, L.D. Wang, T. Tashiro, G. Munzert, Y.J. Guo, Y. Konda, T. Yamada, J. Biol. Chem. 267 (1992) 20840–20843.
[63] S. Harusawa, L. Araki, H. Terashima, M. Kawamura, S. Takashima, Y. Sakamoto, T. Hashimoto, Y. Yamamoto, A. Yamatodani, T. Kurihara, Chem. Pharm. Bull. (Tokyo) 51 (2003) 832–837.
[64] J.A. Jablonowski, C.A. Grice, W. Chai, C.A. Dvorak, J.D. Venable, A.K. Kwok, K.S. Ly, J. Wei, S.M. Baker, P.J. Desai, W. Jiang, S.J. Wilson, R.L. Thurmond, L. Karlsson, J.P. Edwards, T.W. Lovenberg, N.I. Carruthers, J. Med. Chem. 46 (2003) 3957–3960.
[65] N. Terzioglu, R.M. van Rijn, R.A. Bakker, I.J. De Esch, R. Leurs, Bioorg. Med. Chem. Lett. 14 (2004) 5251–5256.
[66] T. Polgar, G.M. Keseru, J. Med. Chem. 48 (2005) 3749–3755.
[67] D.S. Patel, P.V. Bharatam, J. Comput. Aided Mol. Des. 20 (2006) 55–66.
[68] P.D. Lyne, P.W. Kenny, D.A. Cosgrove, C. Deng, S. Zabludoff, J.J. Wendoloski, S. Ashwell, J. Med. Chem. 47 (2004) 1962–1968.
[69] J.F. Guichou, J. Viaud, C. Mettling, G. Subra, Y.L. Lin, A. Chavanieu, J. Med. Chem. 49 (2006) 900–910.
[70] M. Forino, D. Jung, J.B. Easton, P.J. Houghton, M. Pellecchia, J. Med. Chem. 48 (2005) 2278–2281.
[71] M. Stahl, M. Rarey, J. Med. Chem. 44 (2001) 1035–1042.
[72] T. Schulz-Gasch, M. Stahl (Online), J. Mol. Model. 9 (2003) 47–57.
[73] A. Evers, G. Hessler, H. Matter, T. Klabunde, J. Med. Chem. 48 (2005) 5448–5465.
[74] Molecular Operating Environment (MOE), Version 2005.06, Chemical Computing Group, Inc., Montreal, Canada; 2005.
[75] Pallas, Version 3.1; CompuDrug International, Inc., 115 Morgan Drive, Sedona, AZ 86351, USA; 2003.
[76] F.J. Ramirez, I. Tunon, J.A. Collado, E. Silla, J. Am. Chem. Soc. 125 (2003) 2328–2340.
[77] J.A. Collado, I. Tunon, E. Silla, F.J. Ramírez, J. Phys. Chem. 104 (2000) 2120–2131.
[78] B. Noszal, D.L. Rabenstein, J. Phys. Chem. 95 (1991) 4761–4765.
[79] C.R. Ganellin, J. Pharm. Pharmacol. 25 (1973) 787–792.