European Journal of Medicinal Chemistry 44 (2009) 2583–2596

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Homology modeling of MCH1 receptor and validation by docking/scoring and protein-aligned CoMFA

Areej Abu-Hammad [a], Waleed A. Zalloum [a], Hiba Zalloum [a], Ghassan Abu-Sheikha [b], Mutasem O. Taha [a,*]

[a] Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Queen Rania St, Amman, Jordan
[b] Department of Pharmaceutical Sciences, Faculty of Pharmacy, Al-Zaytoonah Private University of Jordan, Amman, Jordan

ARTICLE INFO

Article history:
Received 29 February 2008
Received in revised form
10 November 2008
Accepted 29 January 2009
Available online 5 February 2009

Keywords:
Homology modeling
GPCR
MCH1R inhibitors
Docking/scoring
CoMFA
Pharmacophore modeling

ABSTRACT

Homology modeling is becoming a valid method for obtaining three-dimensional coordinates for proteins. However, it is hard to judge the qualities of the resulting models warranting robust subsequent validations. In an attempt to evaluate the quality of Melanin-concentrating hormone 1 receptor (MCH1R) homology models, a number of homology structures were scanned for potential binding cavities. Subsequently, a group of 35 benzylpiperidines' MCH1R inhibitors were docked into each of the proposed binding sites via four different scoring functions. The docked structures were utilized to construct corresponding protein-aligned comparative molecular field analysis (CoMFA) models by employing probe-based (H^+, OH, CH_3) energy grids and genetic partial least squares (G/PLS) statistical analysis. The docking-based alignment succeeded in accessing self-consistent CoMFA models upon employing JAIN scoring function in one of the proposed binding pockets in a particular homology model. Furthermore, a ligand-based pharmacophore model was developed for the same set of inhibitors and was found to agree with the successful docking configuration. Therefore, we proved that the overall procedure of docking, scoring, and CoMFA evaluation can be a useful tool to validate homology models, which can be of value for structure-based design, *in-silico* screening, and in understanding the structural basis of ligand binding to MCH1R.

1. Introduction

There is clear discrepancy between the rate at which novel protein sequences are discovered and the rate at which detailed structural information on proteins can be obtained via X-ray diffraction or nuclear magnetic resonance spectroscopy (NMR) [1]. For this reason, there is a pressing need for theoretical methods to predict protein structure from sequence [1–4]. At present, it is not possible to reliably predict protein structure from sequence taking a truly *ab initio* approach. Instead, proteins are generally modeled based on the fact that sequence homology normally implies structural similarity [1]. Homology modeling involves aligning the protein of interest to one or more nonredundant homologous sequences of a template protein and evaluating the alignment using a scoring matrix. Differing side chains can be added from a side chain library according to known side chain preferences, followed by extensive energy minimization to eliminate steric clashes [1–4]. However, homology models are imprecise by definition, actually, considering the uncertainties involved in homology modeling, it is imperative that the initial three-dimensional (3D) model be carefully verified to assess its structural integrity and biological relevance before use in structure-based drug design projects [5].

In cases of low sequence homology, errors in the secondary structure definition and packing of secondary structure elements are common. Even the general fold may not be correct [6]. On the other hand, in cases of high sequence homology, the basic framework of the protein can normally be predicted with high accuracy. Nevertheless, errors still occur in variable loops, the relative orientations of secondary structure elements, and in the details of atomic packing. Unfortunately, even minor errors in critical regions are sufficient to prevent the use of models in sensitive applications such as rational drug design [7].

Present attempts to refine homology models are normally based either on energy minimization, limited conformational sampling using molecular dynamics (MD) in conjunction with a detailed force field, or more extensive sampling using simplified force fields. Still, these approaches are generally considered to be ineffective [8,9]. Failure of refinement schemes based on simplified representations and/or limited sampling is not surprising as proteins are densely packed rendering searching their conformational spaces rather difficult [7]. On the other hand, extended simulation times and the requirement of accurate representation of interactions within the

* Corresponding author. +962 777424750.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.01.031

protein and the surrounding environment impose significant limitations on MD-based refinement of homology models [7,10–12]. Therefore, there is a pressing need for more efficient approaches for homology modeling and subsequent structure validation.

The plethora of G-protein-coupled receptors' (GPCRs) druggable targets is in marked contrast to the availability of structural information for these receptors [13,14,62]. Thus, homology modeling of GPCRs based on bovine rhodopsin as a structural template continues to remain a desirable alternative for the development of reasonable 3D models of such proteins [15]. Furthermore, the presence of highly conserved residues in the transmembrane (TM) helices provides a guideline for the reasonable alignment of the GPCR sequences and to derive meaningful inferences from the developed homology models [17–19]. Bovine rhodopsin template structures are generally in their inactive conformation, therefore their homology models are more suited for probing binding of antagonists [15].

However, the generally poor sequence identity (usually less than 20%) among the primary amino acid sequences of target and template GPCR sequences combined with the previously mentioned problems inherent in homology modeling significantly complicates GPCR homology modeling process [16]. Moreover, it is rather difficult to transfer ligand binding site information from the rhodopsin template to the modeled GPCR [11,15].

The identification of potential binding sites for a particular GPCR is crucial not only to understand the intimate mechanism of ligand–receptor recognition, but also to speed up the discovery of new potent and selective GPCR-ligands [11].

The shortcomings of homology modeling combined with the general difficulty in identifying binding sites in GPCRs prompted us to propose an interesting ligand-guided computational approach to evaluate the validity of homology models and to explore the suitability of their putative surface cavities as potential binding sites. Melanin-concentrating hormone 1 receptor (MCH1R), which belongs to the rhodopsin superfamily A of GPCRs [20,21,63], was selected as a working example due to its critical role in food intake regulation and energy homeostasis [22–32].

We utilized the web-based homology modeling suit SWISS-MODEL [33] to generate a group of high-quality homology structures via several automatic and manually configured settings. Subsequently, the models were scanned to identify putative binding pockets at their surfaces. Thereafter, a group of MCH1R inhibitors (**1**–**35**, Fig. 1 and Table 1) were docked into the homology models' putative sites via 4 scoring functions. The resulting protein-aligned docked structures were then utilized to construct comparative molecular field analysis models (CoMFAs).

Success in accessing self-consistent and predictive CoMFA model(s) strongly suggests the validity of the corresponding docking/scoring configuration and the corresponding binding pocket and homology model, albeit at least for drug design and discovery purposes. This is not unreasonable as CoMFA models are highly sensitive to the way their training compounds are aligned in 3D space [34]. Actually, the success of docking/scoring/CoMFA combinations was recently reported [35,36] and showed high performance in predicting docked conformer/pose very close to that of the co-crystallized ligand [37].

Furthermore, we employed ligand-based pharmacophore modeling to validate our docking–CoMFA results based on the fact that ligand-based pharmacophores represent the chemical features of the ligand essential for interaction with the corresponding receptor and their 3D arrangement in space. The conformers/poses suggested by our homology modeling/docking/CoMFA studies were found to closely agree with their ligand-based pharmacophore counterparts; which further validates our work. Therefore, it can be concluded that coupling homology modeling/docking/CoMFA studies helps in discriminating appropriate from inappropriate homology models, predicting and understanding the binding mode of inhibitors in the absence of experimental structural information.

2. Result

2.1. Sequence alignments and homology modeling

Homology modeling was conducted employing SWISS-MODEL [33] *via* three separate approaches (see Section 5.2.2; Homology modeling under Methods). The resulting homology structures (**A**, **B**, **C**, **D**, **E** and **F**, Table 2) were evaluated by their WHATIF Quality Control values [61]. The WHATIF algorithm evaluates the consistency of the atomic model of each homologue against an updated protein structures library to identify possible outliers. Therefore, it can detect gross errors in protein structures (e.g., mistracing of chains) and local abnormalities of stereochemistry. Furthermore, WHATIF performs Ramachandran analysis to identify the number of residues with non-ideal torsion angles [38,39]. WHATIF generates a summary report of the overall quality of the structure as compared with current reliable structures presented in the form of RMS *Z*-scores [61]. Table 3 shows the summary output of WHATIF criteria for the generated homology structures (**A**, **B**, **C**, **D**, **E** and **F**).

To rank the resulting homology structures, we employed a consensus voting approach in which each individual WHATIF property casts a vote if the corresponding value of the considered structure falls within the highest 50% range obtained for that property across the different homology models. The consensus vote is the total number of votes received. Table 3 shows the number of votes given for each homology structure.

Despite that this voting approach might favor certain homology models based on trivial property differences; it seems to succeed in ranking the models according to their overall WHATIF qualities. Unsurprisingly, the automatically generated models **A** and **B** received the highest number of votes (9), while the ClustalW-based homology structure **D** scored second with only one vote difference (Tables 2 and 3). The remaining three models i.e., **C**, **E** and **F**, scored significantly lower (3 and 4 votes) suggesting their general inferior qualities. Therefore, it was decided to select homology models **A**, **B** and **D** for further analyses and subsequent docking/scoring and CoMFA modeling. The sequence identities and similarities between each of the three models and template structures are shown in Table 4 and Fig. 8.

2.2. Docking, scoring and CoMFA modeling

The molecular surfaces of the three MCH1R homology structures (**A**, **B**, and **D**) were scanned for potential binding sites employing the shape-based search approach implemented in LigandFit [44,45]. Several potential binding cavities were detected (Fig. 2). To validate the homology models and their potential binding pockets, we docked inhibitors **1**–**35** (Table 1 and Fig. 1) into each binding site of each model using LigandFit (see Section 5.3.2; Docking under Methods) [44,45]. The resulting conformers/poses were scored employing 4 scoring functions (PLP1, LigScore1, PMF and JAIN), and the highest-ranking conformers/poses, according to each scoring function, were aligned together for CoMFA modeling. Genetic partial least squares (G/PLS) was employed to search for optimal combinations of CoMFA descriptors capable of explaining variations in the bioactivities of the aligned training compounds. Several G/PLS settings were scanned for each molecular alignment, i.e., principal components and numbers of CoMFA descriptors. The qualities of each CoMFA model were assessed via five statistical criteria: (i) conventional regression coefficients against the training compounds (r^2_{28}), (ii) leave one-out regression coefficients (r^2_{LOO}),

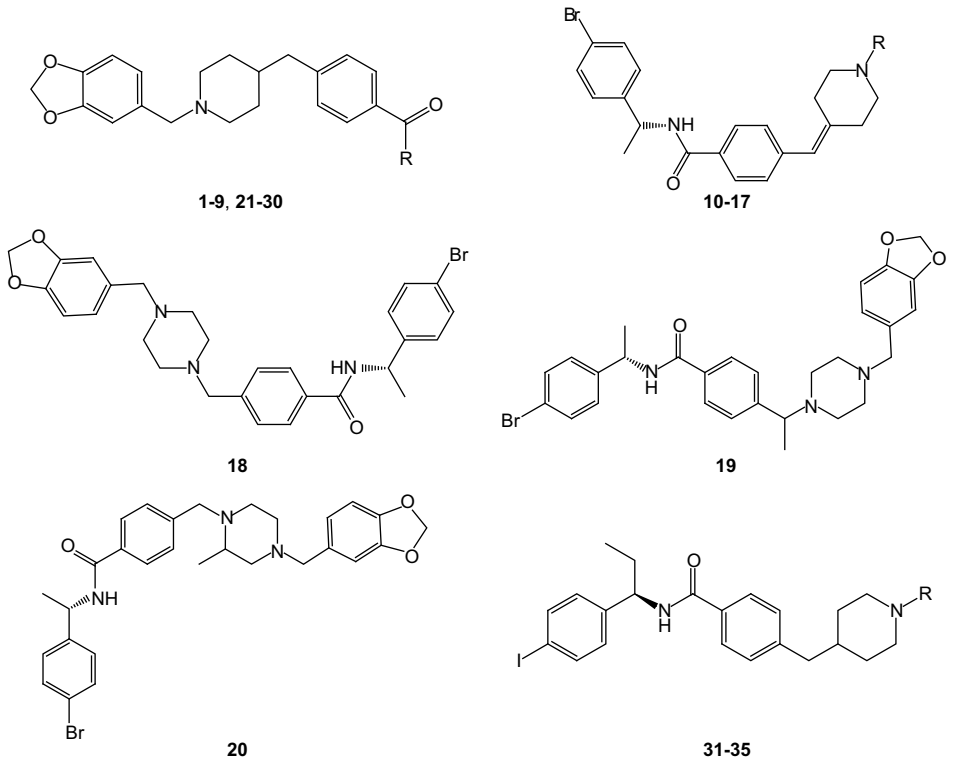


Fig. 1. The chemical scaffolds of different MCH1R inhibitors shown in Table 1.

(iii) bootstrapping regression coefficients (r^2_{BS}), (iv) predictive regression coefficients against the external sets of test compounds (r^2_{PRESS}), and (v) the sum of squared deviations between predicted and experimental bioactivities for molecules in the test set (PRESS). Despite these traditional cross-validation tests are very useful [64], they do not always pick up poor equations [65,66]. Accordingly, we decided to further validate superior models of r^2_{LOO} and $r^2_{PRESS} \geq 0.50$, by calculating their leave-20%-out ($r^2_{L-20\%-O}$) and randomization correlation coefficients (r^2_{random}). These tests ensure that the generated regression models were not produced by chance [67]. Therefore, a particular CoMFA model was deemed to be self-consistent and predictive if it passes the different correlation coefficients with values $\geq$0.5. Table 5 shows the statistical criteria of the best resulting CoMFA models.

Clearly from Table 5, LigandFit succeeded in docking inhibitors **1**–**35** into only site **A1** to access self-consistent and predictive CoMFA models upon alignment based on JAIN docking–scoring function (Figs. 3 and 4). The robustness of this alignment was further supported by the fact that it accessed several consistent and predictive CoMFA models despite variations in the settings of the G/PLS approach (i.e., number of latent variables and allowed field descriptors in the model), as shown in Table 6.

On the other hand, despite success in docking into site **A2**, all attempts to generate statistically valid CoMFA models from the docked structures proved futile, presumably due to their poor 3D alignments.

In contrast to model **A**, all binding sites of models **B** and **D** were invalidated either by failure in docking the inhibitors into the binding pockets or by yielding badly aligned docked structures, i.e., not appropriate for CoMFA modeling, as shown in Table 5.

2.3. Validation by ligand-based pharmacophore models

To gain confidence in the aforementioned docking–CoMFA results, we employed ligand-based pharmacophore modeling to explore the possibility of accessing binding models comparable to ligand–protein interactions suggested by the successful docking/scoring/CoMFA trial. For this purpose, we employed CATALYST-HYPOGEN to explore the pharmacophoric space of the inhibitors set **1**–**35**.

CATALYST-HYPOGEN models drug–receptor interaction using information derived only from the drug structure [57]. Molecules are described as collection of chemical functionalities arranged in 3D space. The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range. It identifies a set of chemical features common to active training molecules. This 3D array of chemical features provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen-bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups and aromatic planes (RingArom). Successful examples involving the use of CATALYST have been reported, wherein the CATALYST derived pharmacophore has been used efficiently as a query for database searching and in 3D-QSAR studies [68–70].

Pharmacophoric exploration of MCH1R inhibitors identified 10 high-quality binding hypotheses. Table 7 shows the pharmacophoric features and success criteria of the optimal pharmacophores (see Section 5.3.3.3; Pharmacophore Validation under Methods), while Table 8 and Fig. 5 show the features, corresponding weights, tolerances and 3D coordinates of the highest-ranking pharmacophore. Fig. 6 shows the experimental versus fitted bioactivities of the training compounds as calculated by mapping them against the highest-ranking pharmacophore (based on Eqs. (1) and (2) in the Methods section).

It is evident from Table 7 that all pharmacophore models exhibited plausible criteria, i.e., their total costs are reasonably separated from the null costs (see Section 5.3.3.3; Pharmacophore

Table 1
The structures of MCH1R compounds utilized in modeling (corresponding scaffolds are in Figure).

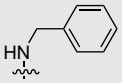
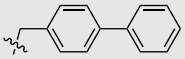
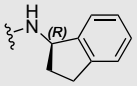
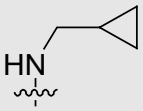
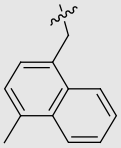
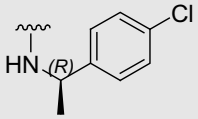
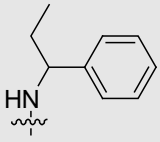
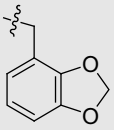
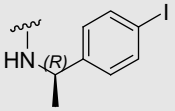
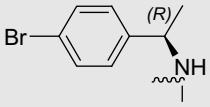
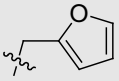
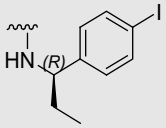
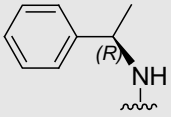
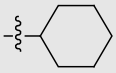
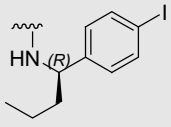
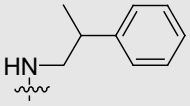
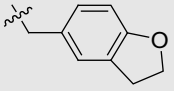
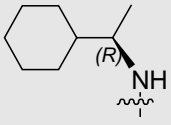
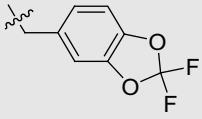
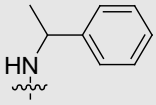
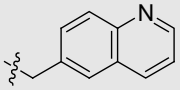
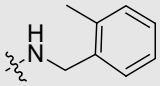
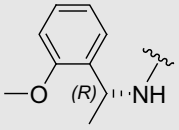
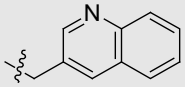
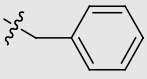
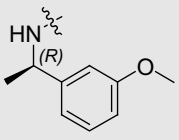
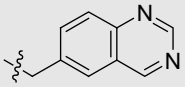
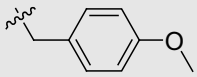
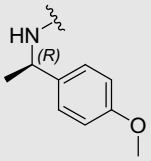
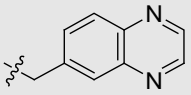
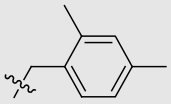
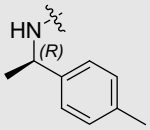
Compound	R	Ki (nM)	Compound	R	Ki (nM)	Compound	R	Ki (nM)
1		5642	**13**		5114	**25**		4678
2		13 890	**14**		1413	**26**		108
3		1831	**15**		17 380	**27**		51
4		101	**16**		15 606	**28**		20
5		3473	**17**		36 588	**29**		191
6		11 765	**18**	–	1951	**30**		133
7		3567	**19**	–	2334	**31**		795
8		4135	**20**	–	4560	**32**		5
9		9332	**21**		2056	**33**		312
10		2584	**22**		266	**34**		30

Table 1 (*continued*)

Compound	R	Ki (nM)	Compound	R	Ki (nM)	Compound	R	Ki (nM)
11		149	**23**		776	**35**		40
12		105	**24**		1078			

Validation under Methods for more details), and they have good correlation coefficients with reasonable confidence levels. However, the highest-ranking model is clearly superior to the remaining models (best overall criteria with 95% confidence and 0.91 correlation) prompting us to select this pharmacophore as representative of the binding process. Emergence of four hydrophobic features and a single hydrogen-bond acceptor in the highest-ranking pharmacophore, combined with the absence of any ionizable feature in the other models (except in the third pharmacophore, Table 7) despite allowing the possibility of positive ionizable features in the pharmacophore generation process, suggests that the inhibitors bind deep into a hydrophobic pocket within MCH1R. Hydrophobic microenvironments should discourage ionization. In fact, the pK_a values of various groups embedded within a receptor depend on the respective local microenvironment. For example, if an amine group is situated within a nonpolar local environment, its pK_a will be reduced because the cationic form is destabilized [72]. Fig. 7 shows how a potent inhibitor docks into the binding pocket according to the successful docking/scoring/CoMFA model, and how it fits the optimal pharmacophore.

3. Discussion

The results of the aforementioned extensive docking–CoMFA exploration and pharmacophore analysis suggest that the inhibitors' activity can be explained via binding to site **A1** (Figs. 2A and 4). These results prompted us to hypothesize that homology structure **A** (Fig. 8) is the best representative of MCH1R for structure-based design efforts.

As apparent from Fig. 7, the iodobenzene moiety of **28** is placed by the docking model at close proximity to the hydrocarbon spacer arm of ARG352 and the phenyl moiety of PHE255 allowing their mutual interaction via hydrophobic and van der Waals' stacking, respectively. Both proposed interactions seem to correlate with hydrophobic (HY-2) and hydrophobic aromatic (HY-Ar) features mapping the iodobenzene moiety in the optimal pharmacophore (Fig. 7B).

The ethyl moiety in the docked pose of **28** is directed towards PRO267 hinting to certain hydrophobic interactions with the alicyclic backbone of this amino acid. This agrees with fitting the ethyl group against the hydrophobic feature HY-1 in the pharmacophore (Fig. 7B).

The central hydrophobic sphere in the pharmacophore model (Fig. 7B) coincides nicely with a middle hydrophobic pocket consisting of the lipophilic side chains of LEU348, LEU252, THR277, ILE 253 and LEU 273 in the homology structure (Fig. 7A). Agreeably, the piperidine center of **28** fits this middle hydrophobic region of both the pharmacophore model and the proposed binding pocket **A1** (Fig. 7). Finally, mapping one of the methylenedioxybenzene oxygens in **28** against a hydrogen-bond acceptor (HBA) feature (Fig. 7B) correlates with a proposed hydrogen-bonding interaction with the amidic HN of GLN344 (or GLN280 in other instances) in the docked complex (Fig. 7A).

Interestingly, loss of one or more of the ligand–receptor interactions in the docked complex seems to explain loss in activity

Table 2
Settings employed in the sequential alignment trials preformed for MCH1R.

Trial	Template(s)	Scoring matrix[c]	Gap penalties[d]			Homologue
			Opening	Extending	Separation	
1	1hzxA; 1l9hA; 1f88A; 1ln6A; 1u19A[a]		Automatically configured[e]			**A**
2	1l9h; 1f88; 1u19[b]		Automatically configured[e]			**B**
3	1u19[b]	Blosum	10	0.05	0.05	–[f]
4	1u19[b]	Blosum	100	0.10	0.10	**C**
5	1u19[b]	Gonnet	10	0.05	0.05	–[f]
6	1u19[b]	Gonnet	100	0.10	0.10	–[f]
7	1u19[b]	Identity	10	0.05	0.05	–[f]
8	1u19[b]	Identity	100	0.10	0.10	**D**
9	1u19[b]	Pam	10	0.05	0.05	**E**
10	1u19[b]	Pam	100	0.10	0.10	**F**

[a] Automatically selected templates from the SWISS-MODEL sequence library ExPDB.
[b] Manually selected templates from the Protein Databank.
[c] Scoring matrices used to assess sequence alignments (implemented in ClustalW).
[d] Manually configured ClustalW penalties imposed on gap placements in the alignment procedure. The "end gap penalty" parameter was fixed to the default value (i.e., 10).
[e] Alignment settings were automatically configured within SWISS-MODEL without ClustalW pretreatment.
[f] The resulting ClustalW-based alignments failed in producing corresponding homology models upon submission to SWISS-MODEL.

Table 3
Summary evaluation information produced by the WHAT-CHECK package for the generated homology structures.

Homologue[a]	Packing quality[b]		Ramachandran plot appearance[b]	chi-1/chi-2 rotamer normality[b]	Backbone conformation[b]	Bond[c]		Omega angle restraints[c]	Side chain planarity[c]	Improper dihedral distribution[c]	Inside/outside distribution[c]	Energy[d] (kJ/mol)	Consensus voting[e]
	1st generation	2nd generation				Lengths	Angles						
A	41.02	6.45	−0.676	−0.524	−3.576	0.874	1.131	1.104	1.558	4.373	1.045	−2708	9
B	25.00	3.62	−0.649	−0.482	−1.668	0.611	0.832	0.431	1.049	3.147	1.158	−1997	9
C	−1.77	−3.33	−3.704	−0.843	−2.877	0.901	1.524	0.718	2.517	2.108	1.257	2202	4
D	−1.74	−3.14	−2.917	−0.524	−2.74	0.833	1.508	0.748	1.729	1.723	1.261	2205	8
E	−2.33	−4.20	−3.672	−1.002	−3.926	1	1.574	0.852	1.943	1.805	1.33	8296	3
F	−1.94	−3.42	−3.153	−0.52	−3.467	0.875	1.536	0.758	1.783	1.752	1.268	4580	4

[a] Homologue labels as in Table 2.
[b] Positive values are better than average.
[c] Closer to 1.0 is better.
[d] The overall molecular mechanical energy of the model as calculated by SWISS-MODEL.
[e] A consensus voting approach in which each individual WHAT-CHECK property casts a vote if the value of the considered homology structure falls in the highest-ranking 50% of the range of values obtained for that property across the different homology models. The consensus vote is the total number of votes received.

among the inhibitors. For example, the combined absence of the hydrogen-bond forming group and shortening of the ethyl side chain rendered inhibitor **10** inactive (Ki = 2.6 μM). Fig. 9 shows the docked conformation/pose of this inhibitor according to the successful docking–scoring configuration.

From the above discussion we can envisage the binding pocket as a cleft within the helical domain of MCH1R consisting of three hydrophobic regions and a hydrogen-bonding polar region. Fig. 10 shows a representation of the binding pocket and the four essential regions that seem to interact with the ligands. The first hydrophobic region **R1** represented by the pharmacophoric feature HY-1 and composed of the side chain of PRO267. The second hydrophobic region **R2** includes both aliphatic and aromatic hydrophobic interactions (HY-2 and HY-Ar in the pharmacophore) and primarily consists of the gap formed between the hydrocarbon linker of ARG352 and the phenyl group of PHE255, while the third hydrophobic region **R3** occupies the center of the binding pocket and surrounded by the lipophilic side chains of LEU348, LEU252, THR277, ILE 253 and LEU 273. Finally, the hydrogen-bond forming region **R4** which consists of the amino side chain of GLN280 and GLN344.

The ability of this model in explaining bioactivity variations among the inhibitors, combined with its consistency with the optimal pharmacophore model reflects the validity of this homology structure and the binding pocket, thus, illustrating the reliability of the docking/scoring/CoMFA procedure as a validation check in selecting the most appropriate homology model for drug design and/or *in-silico* screening purposes.

4. Conclusions

This project included extensive exploration of several high-quality homology models to define potential binding sites in MCH1R. The exploratory process included docking a group of inhibitors into candidate binding sites within the best homology structures. The docked structures in each site were aligned together for CoMFA evaluation. One of the binding sites in one of the homology structures succeeded in accessing valid CoMFA models. In this case, homology structure **A** performed best. The subsequent docking/scoring/CoMFA analyses identified site **A1** as consistent binding pocket for **1**–**35**. This conclusion is further supported by ligand-based pharmacophore modeling.

5. Methods

5.1. Hardware and software

Homology modeling was performed through the web-based port of the Automated Comparative Protein Modeling Server (SWISS-MODEL, http://swissmodel.expasy.org//SWISS-MODEL.html) [33]. The resulting homology structure was assessed using the protein structure verification WHAT-CHECK [38,39] module of the WHATIF

Table 4
Percentage of similar and identical anchors employed in the alignment of MCH1R against bovine rhodopsin templates in homology models **A**, **B** and **D**.

Model	Percent alignment		
	Number of MCH1R residues used in alignment[a]	Identical residues	Similar residues[b]
A	296	12.84%	22.64%
B	296	22.64%	28.04%
D	348	9.48%	23.85%

[a] Total number of residues in MCH1R raw sequence is 422.
[b] As defined by SWISS-MODEL.

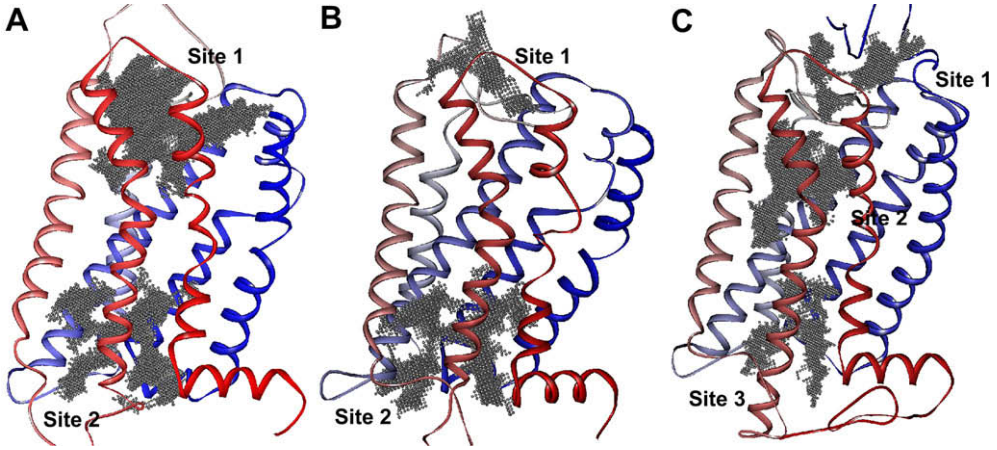


Fig. 2. Perspective views of the homology models: (A) structure **A**, (B) structure **B**, and (C) structure **D**. The figure also shows the corresponding binding pockets (gray dots) as suggested by the shape search approach of LigandFit. Binding sites' labels (1, 2, and 3) are as in Tables 5 and 6.

on-line server (http://swift.cmbi.kun.nl/WIWWWI/modcheck.html). The model was visualized on a personal computer using the Deep-View pdbviewer shareware (version 3.7) provided on-line via SWISS-MODEL web-site (http://ca.expasy.org/spdbv/).

Docking, scoring, and molecular field analysis studies were performed using CERIUS2 suite of programs (version 4.10) from Accelrys Inc. (San Diego, CA, www.accelrys.com). Pharmacophore modeling studies were performed using HYPOGEN module of CATALYST software (version 4.11) from Accelrys Inc. Both CERIUS2 and CATALYST were installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system.

Table 5
Statistical criteria of the best CoMFA models developed for inhibitors **1**–**35** via docking-based alignments into promising cavities of homology structures **A**, **B** and **D**.[a]

Homologue	Potential binding site	Scoring function	CoMFA parameters									
			LV[b]	Number of terms[c]	r^2_{28}[d]	r^2_{BS}[e]	r^2_{LOO}[f]	r^2_{PRESS}[g]	PRESS[h]	$r^2_{L\text{-}20\%\text{-}o}$[i]	r^2_{Random}[j]	Success rate[k]
A	1	PLP1	1[l]	7	0.804	0.802	0.663	0.379	4.047	–	–	0/24
		LigScore1	5	7	0.890	0.707	0.585	0.391	3.924	–	–	0/24
		PMF	2	6	0.781	0.776	0.521	0.314	4.469	–	–	0/24
		JAIN	**3**	**6**	**0.900**	**0.859**	**0.725**	**0.812**	**1.227**	**0.637**	**0.176**	**4/24**
	2	PLP1	The docked poses/conformers are extremely misaligned									
		LigScore1										
		PMF	3	7	0.938	0.931	0.880	0.322	4.419	–	–	0/24
		JAIN	The docked poses/conformers are extremely misaligned									
B	1	PLP1	1	5	0.680	0.679	0.546	0.24	4.498	–	–	0/24
		LigScore1	4	8	0.923	0.910	0.739	0.001	6.508	–	–	0/24
		PMF	3	5	0.842	0.746	0.657	0.434	3.680	–	–	0/24
		JAIN	2	6	0.791	0.788	0.535	0.380	4.039	–	–	0/24
	2	PLP1	1	7	0.888	0.887	0.764	0.318	4.441	–	–	0/24
		LigScore1	The docked poses/conformers are extremely misaligned									
		PMF										
		JAIN										
D	1	PLP1	1	5	0.676	0.668	0.379	0.269	4.861	–	–	0/24
		LigScore1	The docked poses/conformers are extremely misaligned									
		PMF										
		JAIN										
	2	PLP1			0.879	0.871	0.695	0.289	4.629	–	–	0/24
		LigScore1	1	6	0.754	0.753	0.589	0.071	6.058	–	–	0/24
		PMF	3	8	0.961	0.949	0.864	0.185	5.315	–	–	0/24
		JAIN	The docked poses/conformers are extremely misaligned									
	3	PLP1	1	6	0.934	0.931	0.885	0.672	2.138	−0.961	0.184	0/24
		LigScore1	1	9	0.881	0.875	0.726	−0.774	11.363	–	–	0/24
		PMF	3	8	0.936	0.933	0.730	0.587	2.694	0.200	0.245	0/24
		JAIN	2	9	0.911	0.906	0.713	0.431	3.705	–	–	0/24

[a] The best CoMFA model are bolded.
[b] Number of latent variables (principle components) in the optimal CoMFA model.
[c] Number of terms in the optimal CoMFA model.
[d] Non-cross-validated correlation coefficient determined for 28 training compounds.
[e] Bootstrapping correlation coefficient.
[f] Leave one-out cross-validation correlation coefficients.
[g] Predictive r^2 determined for 7 test compounds.
[h] The sum of squared deviations between predicted and actual activity values for every molecule in the test set.
[i] Cross-validation correlation coefficients determined by the leave-20%-out technique.
[j] The average randomization correlation coefficients.
[k] Number of successful CoMFA models (of r^2_{LOO} and $r^2_{PRESS} \geq 0.5$) per number of evaluated models as in Eq. (1) and related text.

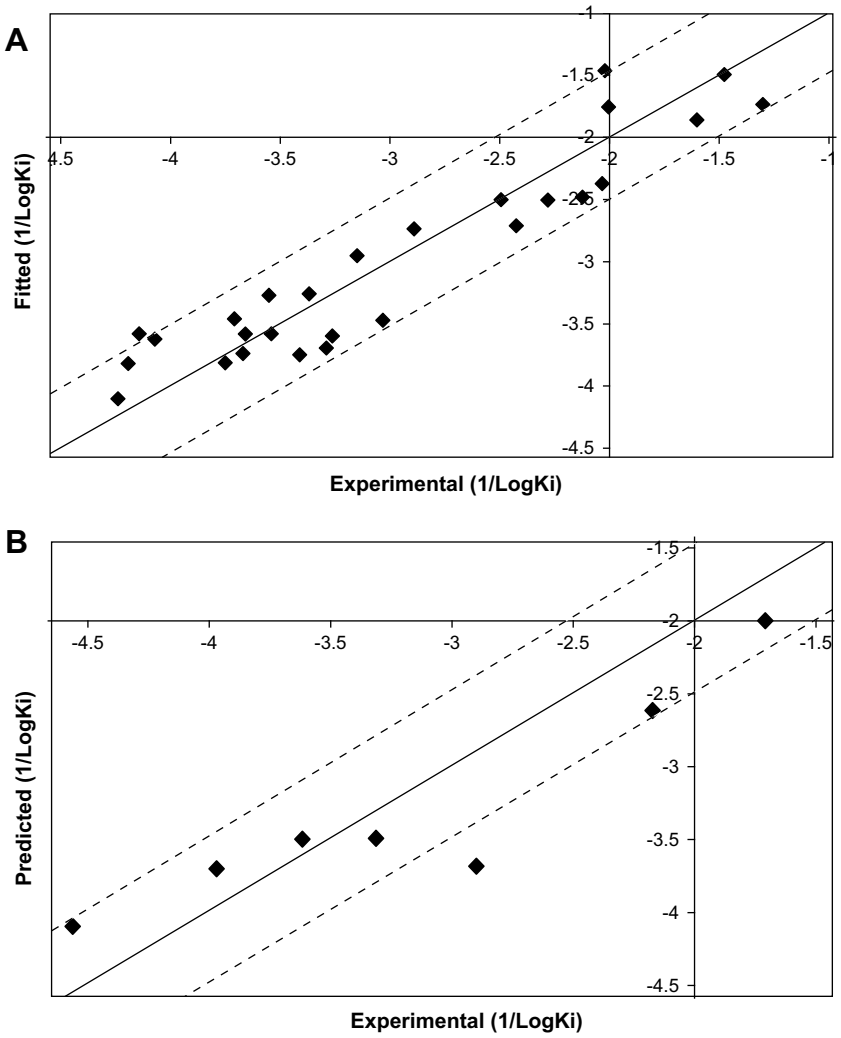


Fig. 3. Experimental versus fitted (**A**, 28 compounds, $r^2_{LOO} = 0.725$) and predicted (**B**, 7 compounds, $r^2_{PRESS} = 0.812$) bioactivities calculated from the best CoMFA model obtained after 30 000 iterations of G/PLS performed on compounds (**1**–**35**) aligned as docked by LigandFit into site **A1** and scored by JAIN docking–scoring function (Table 5). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the ±0.5 log point error margins.

5.2. Protein modeling

5.2.1. MCH1R sequence

The amino acid sequence of MCH1R was collected on-line from the Molecular Class-Specific Information System (MCSIS) project (http://www.gpcr.org/7tm/htmls/entries.html) [40].

5.2.2. Homology modeling

SWISS-MODEL applies a rigid fragment assembly approach for homology modeling [33,41]. This method involves the following steps: (i) Template selection: suitable template(s) can be automatically [42] or manually selected for a given target protein from sequences within the templates library of SWISS-MODEL (ExPDB) or from the Protein Databank. (ii) Alignment: a local pair-wise alignment of the target sequence to the main template structures is calculated [42]. The placement of insertions and deletions is performed taking into consideration the template structure context. (iii) Model building: the templates are weighted by their sequence similarity to the target sequence, while significantly deviating atom positions are excluded [33] (iv) Side chain modeling: the reconstruction of the model side chains is based on the weighted positions of corresponding residues in the template structures. Possible side chain conformations are selected from a backbone dependent rotamer library [33]. (v) Energy minimization: deviations in the protein structure geometry due to joining rigid fragments are cleaned by steepest descent energy minimization while molecular dynamics methods are used only to regularize the structure [33].

We employed the SWISS-MODEL server to generate different MCH1R homology models via the following approaches:

(1) The fully automatic procedure: using the amino acid sequence of MCH1R as input data the server automatically selected the following bovine rhodopsin crystal structures (from ExPDB) as suitable templates (sequence identities with MCH1R in brackets): 1hzxA (27.06%), 1l9hA (27.06%), 1f88A (27.00%), 1ln6A (28.22%) and 1u19A (26.6%). Subsequently, SWISS-MODEL automatically generated the first homology model (**A**, Table 2).

(2) A semi-automatic procedure: using selected bovine rhodopsin crystal structures (downloaded from the Protein Databank): 1l9h, 1f88, and 1u19 as templates into the SWISS-MODEL server, which automatically generated the second homology model for MCH1R (**B**, Table 2). The three template structures were selected as they exhibit the best resolution among different bovine rhodopsin crystallographic structures in the Protein Databank.

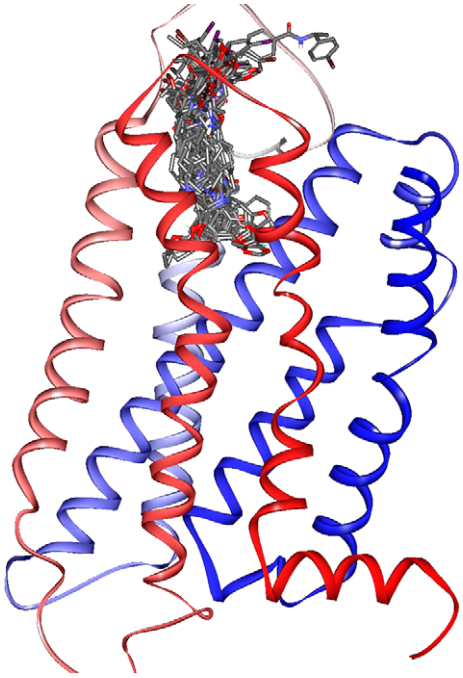


Fig. 4. The alignment of the training set inhibitors into binding pocket **A1** as suggested by the successful docking–scoring configuration.

(3) Modeling requests to the SWISS-MODEL server based on several ClustalW [41] pair-wise alignments: the MCH1R was aligned against the crystallographic structure 1u19 within DeepView. This structure was selected as template as it exhibits the highest resolution (2.2 Å) among other bovine rhodopsin structures in the Protein Databank. Several sequential alignments of 1u19 and MCH1R were performed by means of ClustalW via four scoring matrices (Blosum, Pam, Gonet and Identity) and different gap penalty combinations, as in Table 2 [41]. The resulting alignments, including deletion and insertions, were submitted to SWISS-MODEL as DeepView project files. However, only 4 out of 8 ClustalW-based alignments succeeded in yielding corresponding homology models (**C**, **D**, **E** and **F** in Table 2).

Table 6
Effects of variable G/PLS settings on statistical criteria of CoMFA models generated from successful docking-based alignment.[a]

Successful homologue/site[a]	Scoring function[b]	CoMFA parameters						
		LV[c]	Number of terms[d]	r^2_{28}[e]	r^2_{BS}[f]	r^2_{LOO}[g]	r^2_{PRESS}[h]	PRESS[i]
A1	JAIN	1	8	0.916	0.914	0.850	0.533	3.041
		3	8	0.965	0.964	0.932	0.576	2.762
		1[j]	**6**	**0.900**	**0.859**	**0.725**	**0.812**	**1.227**
		3	6	0.900	0.859	0.812	0.637	0.420

[b]Docked inhibitors as in Table 1 and Fig. 1.
[a] Homology models and docking conditions as in Tables 2 and 5.
[b] Docking-scoring function.
[c] Number of latent variables (principle components) in the CoMFA model.
[d] Number of terms in the CoMFA model.
[e] Non-cross-validated correlation coefficient.
[f] Bootstrapping correlation coefficient.
[g] Leave one-out cross-validation correlation coefficients.
[h] Predictive r^2 as determined for test compounds.
[i] The sum of squared deviations between predicted and actual activity values for every molecule in the test set.
[j] The best CoMFA models are shown in bold.

Table 7
Success criteria of the pharmacophoric hypotheses generated for MCH1R inhibitors (Table 1 and Fig. 1).

Rank of hypothesis[a]	Features[b]	Total cost	Null cost	Residual cost	Correlation coefficient (r)[c]	Cat.Scramble confidence[d]
1	HBA, 3 × HY, HY-Ar	147	187	40	0.91	95%
2	HBA, 3 × HY, HY-Ar	154	187	33	0.86	95%
3	HBA, 2 × HY, HY-Ar, Pos-Ion	154	187	33	0.85	95%
4	HBA, 2 × HY, 2 × HY-Ar	155	187	32	0.85	95%
5	3 × HY, 2 × HY-Ar	156	187	31	0.84	95%
6	HBA, 2 × HY, 2 × HY-Ar	157	187	30	0.83	95%
7	2 × HBA, 2 × HY, HY-Ar	158	187	29	0.83	90%
8	HBA, HBD, 2 × HY	157	187	30	0.82	90%
9	HBA, 3 × HY, HY-Ar	158	187	29	0.83	90%
10	HBA, 3 × HY, HY-Ar	159	187	28	0.82	90%

[a] By default, CATALYST-HYPOGEN yields 10 optimal pharmacophoric hypotheses for each automatic run. This rank represents the position of the particular hypothesis within the list of generated models as ranked by their total costs.
[b] HBA: hydrogen-bond acceptor; HBD: hydrogen-bond donor; HY: aliphatic hydrophobic; HY-Ar: aromatic hydrophobic; Pos-Ion: positive ionizable.
[c] The correlation coefficients between bioactivity estimates, as determined from each pharmacophore (using Eqs. (1) and (2)), and the bioactivities of the corresponding training compounds.
[d] Confidence level of each pharmacophore as determined from 19 Cat.Scramble iterations.

5.2.3. Analysis of the homology models

The resulting homology structures were evaluated employing the WHAT-CHECK module of the WHATIF on-line server. To rank the resulting homology structures, we employed a consensus voting approach in which each individual WHAT-CHECK property casts a vote if the corresponding value of the considered homology structure falls within the highest-ranking 50% of the range of values obtained for that property across the different homology models. The consensus vote is the total number of votes received.

Table 3 shows the summary outputs of WHAT-CHECK calculated for the generated homology structures (**A**, **B**, **C**, **D**, **E** and **F**) and their consensus votes. And Table 4 shows the percentage of similar and identical anchors in these models.

5.3. Ligand-based evaluation of the homology models

5.3.1. Dataset

A set of 35 MCH1R inhibitors belonging to benzylpiperidines [43] (compounds **1**–**35**, Fig. 1, Table 1) were used for both docking-based modeling and ligand-based pharmacophore validation. Racemic inhibitors (i.e., **3**, **6**, **8**, **19** and **20**, Table 1 and Fig. 1) were assigned arbitrary absolute chiral configurations for subsequent modeling.

Table 8
Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of the highest-ranking pharmacophore.

Definitions		Chemical features					
		HBA		HY-1	HY-2	HY-3	HY-Ar
Weights			1.69	1.69	1.69	1.69	1.69
Tolerance		1.6	2.2	1.6	1.6	1.6	1.6
Coordinates	*X*	2.92	34.02	−6.72	−3.98	3.18	−5.14
	Y	−3.52	−3.19	0.66	−7	1.78	−3.76
	Z	−5.36	−8.13	4.62	1.54	0.89	1.34

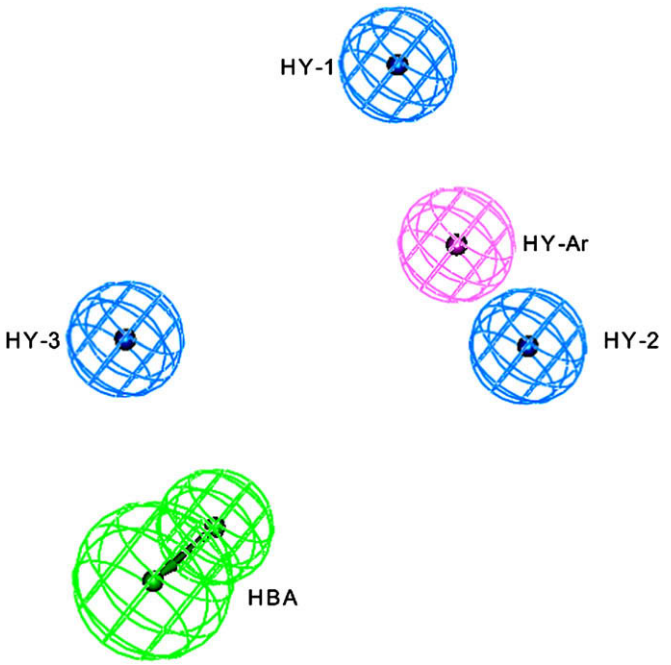


Fig. 5. The best hypothesis model Hypo1 produced by the HYPOGEN module in CATALYST4.10 software package. Pharmacophore features are color-coded with light-blue for hydrophobic groups, pink for hydrophobic aromatic group, and green for hydrogen-bond acceptor. HY-1, hydrophobic group 1; HY-2, hydrophobic group 2; HY3, hydrophobic group 3; HBA, hydrogen-bond acceptor; HY-Ar, hydrophobic aromatic group. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This assumption is based on the fact that the enantiomeric pairs of these ligands behaved similarly under the implemented docking–scoring and pharmacophore modeling configurations, i.e., they yielded visually similar docked conformers/poses. Despite that each inhibitor (**1**–**35**, Table 1 and Fig. 1) contains one ionizable group, i.e., a tertiary amino substituent, our pharmacophore modeling efforts (see Section 5.3.3.2. Pharmacophore exploration below) revealed a predominant majority of optimal binding models lacking any positive ionizable features (see Section 2.3. Validation by ligand-based pharmacophore models under Results) suggesting that ligand–MCH1R binding is not mediated by electrostatic attraction, and therefore it was decided to dock the ligands in their unionized forms.

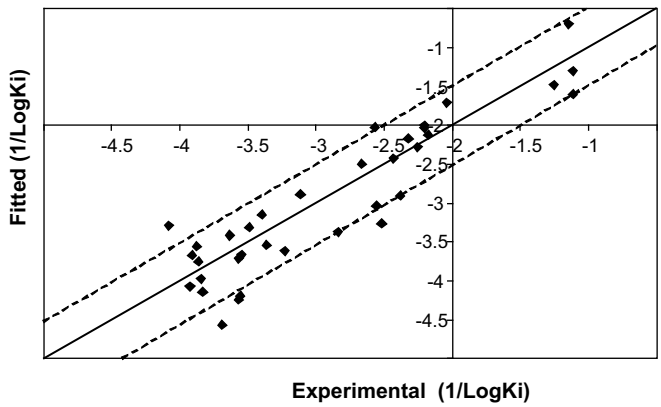


Fig. 6. Experimental versus fitted bioactivities of inhibitors **1**–**35** (Table 1 and Fig. 1) calculated by fitting their chemical structures against the highest-ranking pharmacophore model according to Eqs. (1) and (2). The solid line is the regression line for the fitted bioactivities of the training compounds, whereas the dotted lines indicate the ±0.5 log point error margins.

The *in vitro* bioactivities of MCH1R inhibitors are reported in Table 1 as Ki values spanning from 5.0 nM to 36.6 μM. The logarithm transformations of measured Ki (nM) values were used in 3D-QSAR modeling, thus correlating the data linearly to the free energy change. Since it is essential to assess the predictive power of the resulting CoMFA models on external sets of inhibitors, seven inhibitors (*ca.* 20% of the modeled compounds) were employed as external testing subset for validating the corresponding CoMFA models: **8**, **9**, **11**, **17**, **21**, **27** and **31** (Fig. 1 and Table 1). The same set of MCH1R inhibitors (**1**–**35**) was utilized to generate HYPOGEN hypotheses.

5.3.2. Docking, scoring and CoMFA analyses

5.3.2.1. Binding site search and docking simulations. The LigandFit module [44,45] in CERIUS2 (version 4.10) was used for the docking study and binding site search. LigandFit gives the best poses at the binding site by a stochastic conformational search and evaluation of the energy of the ligand–protein complex. It uses a grid method when evaluating interactions between the protein and the ligand. In our case the binding site search was performed by the shape-based mode (flood filling method) for the successful homology models [44,45]. The grid resolution was set to 0.5 Å and only sites of more than 800 grid points were considered as potential binding pockets for subsequent docking studies.

For initial exploration of binding site, various sites were searched and analyzed by docking of the 35 MCH1R inhibitors into the different binding sites employing the following docking configuration [44,45]: (i) Monte Carlo search parameters: number of trials was set to variable-table [44] and search step for torsions with polar hydrogens = 30.0°. (ii) The RMS threshold for ligand-to-binding site shape match was set to 2.0 employing a maximum of 2.0 binding site partitions and 1.0 site partition seed. (iii) The interaction energies were assessed employing CFF force field (version 95 implemented in CERIUS2 4.10) with a nonbonded cutoff distance of 10.0 Å and distance dependent dielectric. An energy grid extending 3.0 Å from the binding site was implemented. (iv) Rigid-body ligand minimization by 10 iterations of Steepest Descend (SD) minimization followed by 20 BFGS iterations applied to every orientation of the docked ligand. (v) A maximum of 10 diverse docked conformations/poses of optimal interaction energies were saved. The similarity threshold was set to a DockScore of 20 kcal/mole and an RMS value of 1.5 Å. (vi) The saved conformers/poses were further energy-minimized within the binding site for a maximum of 100 rigid-body iterations. (vii) Eventually, the final conformers/poses were clustered into 5 groups using the Leader Algorithm implemented in CERIUS2 at similarity RMS threshold of 1.5 Å.

5.3.2.2. Scoring of docked conformers/poses. Structures of the optimal docked conformers/poses predicted by LigandFit were ranked with built-in scoring functions in CERIUS2.10. As each scoring function may rank binding poses differently, we were prompted to use four different scoring functions in our evaluation LigScore1 [46], PLP1 [47], PMF [48], and JAIN [49]. Considering each scoring function in turn, the highest scoring docked conformer/pose was selected for each inhibitor for subsequent CoMFA modeling. This resulted in four sets of 35 docked molecules with scores corresponding to each scoring function. LigScore1 scores were calculated employing CFF force field (version 95) [50] and using grid-based energies with a grid extension of 5.0 Å across the binding site. PMF scores were calculated employing cutoff distances for carbon–carbon interactions and other interactions of 12.0 Å.

5.3.2.3. Molecular field analysis. The molecular field analysis (MFA) and G/PLS modules within CERIUS2 were used to perform 3D-QSAR analyses [51]. The alignments of different inhibitors came directly

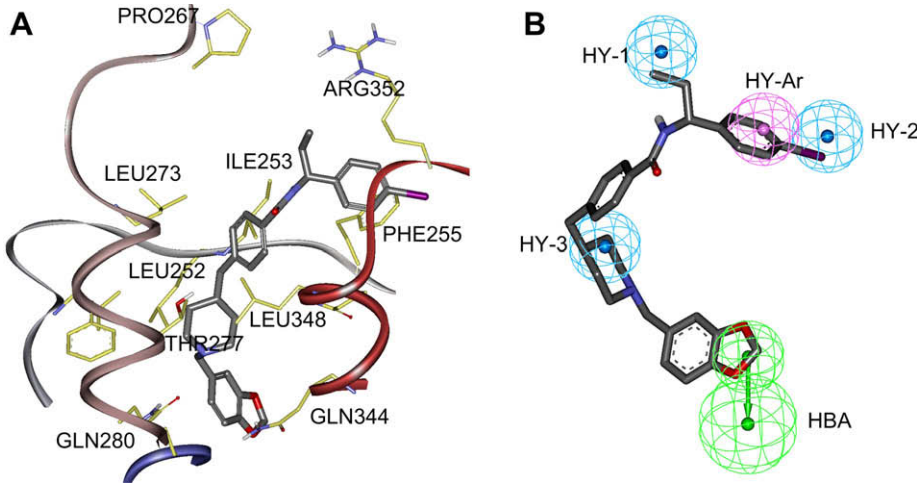


Fig. 7. Comparison between the highest-ranking pharmacophore model and the best performing docking/scoring/CoMFA combination. (A) The pose and conformation of inhibitor **28** as it docks into site **A1** of the homology model of MCH1R via JAIN scoring function. (B) The highest-ranking pharmacophore model mapped to inhibitor **28** (Ki = 20.0 nM). HY-1, hydrophobic group 1; HY-2, hydrophobic group 2; HY3, hydrophobic group 3; HBA, hydrogen-bond acceptor; HY-Ar, hydrophobic aromatic group.

from the top-scoring conformers/poses according to each considered scoring function and each particular binding site. For each alignment, the interaction fields between the ligands and proton (positively charged), hydrogen-bond donor/acceptor, and methyl (neutral) probes were calculated employing a regularly spaced rectangular grid of 2.0 Å spacing. The spatial limits of the molecular field were defined automatically and were extended past the van der Waals volume of all the molecules in the *X*, *Y*, and *Z* directions. The ligands were assigned partial charges using the Gasteiger method implemented within CERIUS2. The energy fields were calculated employing the default UNIVERSAL force field (version 1.02) implemented within CERIUS2 [50] and were truncated to ±50 kcal/mol. The calculation gave around 3460 variables for each compound (*ca.* 1150 variable/probe).

To derive the best possible 3D-QSAR statistical model for each alignment, we used genetic partial least squares (G/PLS) analysis to search for optimal regression equations capable of correlating the variations in biological activities of the training compounds with variations in the corresponding interaction fields [52]. Our preliminary diagnostic trials suggested the following optimal G/PLS parameters: explore linear equations at mating and mutation probabilities of 50%; population size = 500; number of generations

```
MCH1R   105                                          SYI NIIMPSVFGT ICLLG----I IGNSTVIFAV VKKSKLHWCN
1hzxA   1      MNGTEGPNF YVPFSNKTGV VRSPFEAPQY YLAEP--WQF SMLAAYMFLL I-MLG----F PINFLTLYVT VQHKKLR---
1l9hA   1      MNGTEGPNF YVPFSNKTGV VRSPFEAPQY YLAEP--WQF SMLAAYMFLL I-MLG----F PINFLTLYVT VQHKKLR---
1f88A   1     M-NGTEGPNF YVPFSNKTGV VRSPFEAPQY YLAEP--WQF SMLAAYMFLL I-MLG----F PINFLTLYVT VQHKKLR---
1ln6A   40                                                LAAYMFLL I-MLGFPINF LTLYVTVQHK ---KLRT---
1u19A   1      MNGTEGPNF YVPFSNKTGV VRSPFEAPQY YLAEP--WQF SMLAAYMFLL I-MLG----F PINFLTLYVT VQHKKLR---

MCH1R   144    NVPDIFIINL S--VVDLLFL LGMPFMIHQL MG-NGVWHFG ETMCTLITAM DA-NSQFTST YILTAMAIDR YLATVHPISS
1hzxA   70     TPLNYILLNL A--VADLFMV FG-GFTTTLY TSLHGYFVFG PTGCNLEGFF AT-LGGEIAL WSLVVLAIER YVVVCKPMSN
1l9hA   70     TPLNYILLNL A--VADLFMV FG-GFTTTLY TSLHGYFVFG PTGCNLEGFF AT-LGGEIAL WSLVVLAIER YVVVCKPMSN
1f88A   70     TPLNYILLNL A--VADLFMV FG-GFTTTLY TSLHGYFVFG PTGCNLEGFF AT-LGGEIAL WSLVVLAIER YVVVCKPMSN
1ln6A   71     PL--NYILLN LAVADLFMVF GG-FTTTLYT --SLHGYFVF GPTGCNLEGF FATLGGEIAL WSLVVLAIER YVVVCKPMSN
1u19A   70     TPLNYILLNL A--VADLFMV FG-GFTTTLY TSLHGYFVFG PTGCNLEGFF AT-LGGEIAL WSLVVLAIER YVVVCKPMSN

MCH1R   220    TKFRKPSVAT LVICLLWALS FISITPVWLY -ARLIPFPGG AVGCGIR--L PNPDTDLYWF TLYQFFLAFA LPFVVITAAY
1hzxA   146    FRF-GENHAI MGVAFTWVMA LACAAPPLVG WSRYIP-EGM QCSCGIDYYT PHEETNNESF VIYMFVVHFI IPLIVIFFCY
1l9hA   146    FRF-GENHAI MGVAFTWVMA LACAAPPLVG WSRYIP-EGM QCSCGIDYYT PHEETNNESF VIYMFVVHFI IPLIVIFFCY
1f88A   146    FRF-GENHAI MGVAFTWVMA LACAAPPLVG WSRYIP-EGM QCSCGIDYYT PHEETNNESF VIYMFVVHFI IPLIVIFFCY
1ln6A   146    FRF-GENHAI MGVAFTWVMA LACAAPPLVG WSRYIP-EGM QCSCGIDYYT PHEETNNESF VIYMFVVHFI IPLIVIFFCY
1u19A   146    FRF-GENHAI MGVAFTWVMA LACAAPPLVG WSRYIP-EGM QCSCGIDYYT PHEETNNESF VIYMFVVHFI IPLIVIFFCY

MCH1R   297    VRILQ---RM TSSV-----A PASQRSIRLR TKRVTRTAIA ICLVFFVCWA PYYVLQLTQL SISRFTLTFV YLYNAAISLG
1hzxA   224    GQLVF---TV KEAAA----- ATTQKA---- EKEVTRMVII MVIAFLICWL PYAGVAFYIF THQGSDFGPI FMTIPAF-FA
1l9hA   224    GQLVF---TV KEAAA----- ATTQKA---- EKEVTRMVII MVIAFLICWL PYAGVAFYIF THQGSDFGPI FMTIPAF-FA
1f88A   224    GQLVF---TV KEAAAS---- ATTQKA---- EKEVTRMVII MVIAFLICWL PYAGVAFYIF THQGSDFGPI FMTIPAF-FA
1ln6A   224    GQLVFTVKEA AAQQQESATT QKAEKE---- VTRMVIIMVI AFL---ICWL PYAGVAFYIF THQGSDFGPI FMTIPAF-FA
1u19A   224    GQLVF---TV KEAAAQQQES ATTQKA---- EKEVTRMVII MVIAFLICWL PYAGVAFYIF THQGSDFGPI FMTIPAF-FA

MCH1R   369    YANSCLNPFV -YIVLCETFR KRLVLSVKPA AQ
1hzxA   296    KTSAVYNPVI -YIMMNKQFR NCMVTTLCCG KNP-LGD--S TTVSKTETSQ VAPA
1l9hA   296    KTSAVYNPVI -YIMMNKQFR NCMVTTLCCG KNP-LGD--S TTVSKTETSQ VAPA
1f88A   296    KTSAVYNPVI -YIMMNKQFR NCMVTTLCCG KN------PS TTVSKTETSQ VAPA
1ln6A   296    KTSA--VYNP VIYIMMNKQF RNCMVTTLCC GKNPLGDDEA STTVSKTETS QVAP
1u19A   296    KTSAVYNPVI -YIMMNKQFR NCMVTTLCCG KNPLGDDEAS TTVSKTETSQ VAPA
```

Fig. 8. The alignment of MCH1R and bovine rhodopsin templates (PDB codes: 1hzxA, 1l9hA, 1f88A, 1ln6A and 1u19A) that yielded the homology model **A**. Transmembrane domains are bolded, gaps are shown as (–), identical and similar residues are high-lighted in yellow and green, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

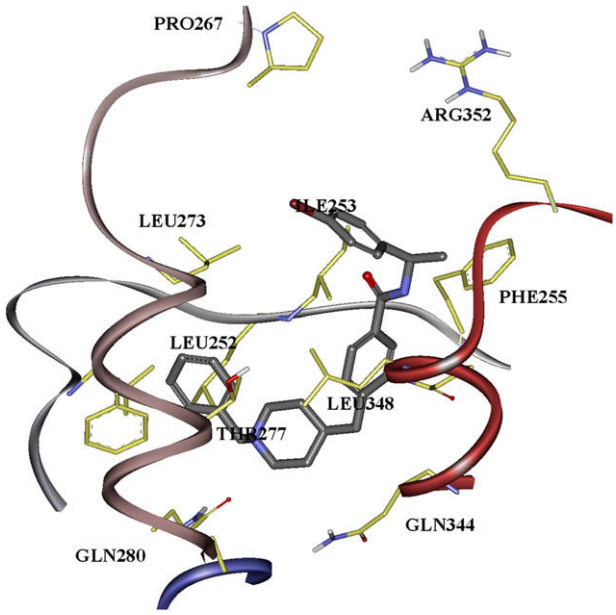


Fig. 9. The docked conformer/pose of inhibitor **10** (Ki = 2584 nM) as it docks into site **A1** of the homology model of MCH1R via JAIN scoring function.

(iterations) = 30 000 and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (i.e., CoMFA descriptors) and PLS latent variables (i.e., principle components), it was decided to scan and evaluate all possible CoMFA models resulting from all combinations of 1–5 principal components and 5–9 CoMFA descriptors.

All CoMFA models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}) [52] and predictive r^2 (r^2_{PRESS}) calculated from the test subsets.

A particular CoMFA model is defined as successful if it exhibits r^2_{LOO} and r^2_{PRESS} values above 0.5. However, CoMFA models corresponding to the best r^2_{PRESS} values in a particular alignment were further cross-validated through leave-20%-out validation. This process was repeated 10 times, and the average predictive r^2 ($r^2_{L\text{-}20\%\text{-}O}$) is determined. This cross-validation technique has been shown to yield better indices for the robustness of a model than the normal LOO procedure [53]. An additional validation was also performed for high-ranking 3D-QSAR models to rule out the possibility of chance correlations. All biological activities were randomized 99 times (confidence level 99%) and were subjected to regression analysis, and the mean randomization r^2 was calculated [54]. The statistical performance of the different docking/scoring/CoMFA models is shown in Tables 5 and 6.

5.3.3. Pharmacophore modeling

5.3.3.1. Conformational analysis. The molecular flexibilities of the collected inhibitors (**1**–**35**, Fig. 1 and Table 1) were taken into account by considering each training compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor was explored adopting the "best conformer generation" option within CATALYST. Default parameters were employed in the conformation generation procedure, i.e., conformational ensembles were generated with energy threshold of 20 kcal/mol from the local minimized structure and a maximum limit of 250 conformers per molecule [55–58]. Conformational sampling was performed utilizing the poling algorithm implemented within CATALYST. Poling promotes conformational variation via employing molecular mechanical force field algorithm that penalizes similar conformers [55,56]. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling within the training compounds [71].

5.3.3.2. Pharmacophore exploration. Generally, default HYPOGEN settings were employed in exploring the pharmacophoric space of MCH1R inhibitors. However, the following parameters were modified from their default values to optimize the resulting pharmacophoric models: (i) Bioactivity spread (GenerateHypo.Inactive.Spread parameter) = 3.2, (ii) Spacing = 300 pm, (iii) constrains on number of features: hydrogen-bond acceptors (HBA) = 0–3, hydrogen-bond donors (HBD) = 0–3, hydrophobic-aliphatic = 0–3, hydrophobic aromatic = 0–3, positive ionizable = 0–5, (iv) constrains on number of total features = 4–5.

5.3.3.3. Pharmacophore validation. When generating hypotheses, HYPGEN tries to minimize a cost function consisting of three terms: weight cost, error cost and configuration cost [55–58]. Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2. The deviation between the estimated

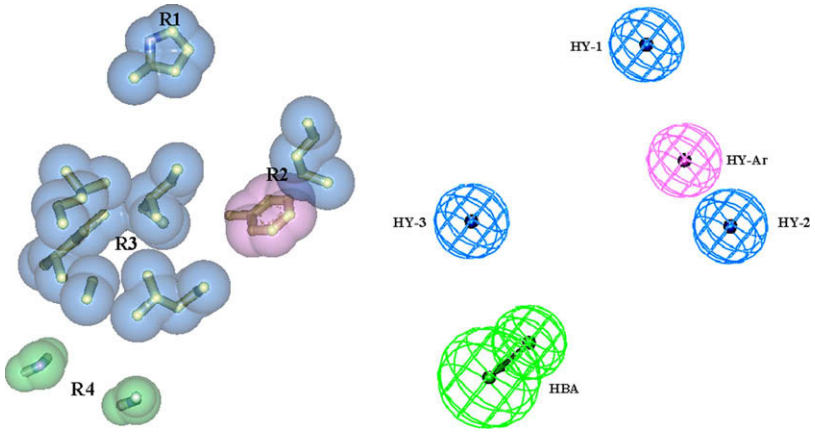


Fig. 10. A presentation of binding pocket of MCH1R showing the different binding regions (denoted from **R1** to **R4**). The three hydrophobic regions (**R1**, **R2**, and **R3**) are shown in blue (the aromatic hydrophobic in **R2** shown in pink), the hydrogen-bonding area (**R4**) is shown in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

activities of the training set and their experimentally determined values adds to the error cost. The activity of any compound can be estimated from a particular hypothesis through Eq. (1) [58].

$$\log(\text{Estimated activity}) = I + \text{Fit} \quad (1)$$

where, I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit values of the training compounds. The Fit value for any compound is obtained automatically employing Eq. (2) [58].

$$\text{Fit} = \sum \text{mapped hypothesis features} \times W\left[1 - \sum (\text{disp}/\text{tol})^2\right] \quad (2)$$

where, ∼ mapped hypothesis features represent the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, W is the weight of the corresponding hypothesis feature spheres. disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as Tolerance, equals to 1.6 Å by default). $\sim(\text{disp}/\text{tol})^2$ is the summation of $(\text{disp}/\text{tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound [58].

The third term penalizes the complexity of the hypothesis, i.e. the configuration cost. The overall cost (total cost) of a hypothesis is calculated by summing over the three cost factors. HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost, the more likely that the hypothesis does not reflect a chance correlation.

Another approach to assess the quality of HYPOGEN pharmacophores is to apply cross-validation using the Cat-Scramble program implemented in CATALYST. This validation procedure is based on Fischer's randomization test [59]. The goal of this type of validation is to check whether there is a strong correlation between the chemical structures and the biological activity. In this validation test, we selected a 95% confidence level, which generates 19 random spreadsheets by the Cat-Scramble command. These random spreadsheets were used to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses [60]. Table 7 shows the different success criteria of the top 10 scoring hypotheses. Table 8 shows the X, Y, Z coordinates of the highest-ranking hypothesis.

Acknowledgements

This project was sponsored by the Hamdi-Mango Centre for Scientific Research (Grant Number 04200010). The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Centre for Scientific Research at the University of Jordan for their generous funds. M.O.T. thanks the University of Jordan for funding him during his sabbatical leave.

References

[1] T. Schwede, A. Diemand, N. Guex, M.C. Peitsch, Res. Microbiol. 151 (2000) 107–112.
[2] R.A. Abagyan, S.J. Batalov, Mol. Biol. 273 (1997) 355–368.
[3] D.M. Standley, V.A. Eyrich, Y. An, D.L. Pincus, J.R. Gunn, R.A. Friesner, Proteins 1 (Suppl. 5) (2001) 133–139.
[4] M.G. Williams, H. Shirai, J. Shi, H.G. Nagendra, J. Mueller, K. Mizuguchi, R.N. Miguel, S.C. Lovell, C.A. Innis, C.M. Deane, et al., Proteins(Suppl. 5) (2001) 92–97.
[5] M.A. Marti-Renom, A. Fiser, M.S. Madhusudhan, B. John, A. Stuart, N. Eswar, P. Pieper, M.-Y. Shen, A. Sali, Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 2003, V. 5, p. 5.1.1.
[6] R. Sanchez, A. Sali, Curr. Opin. Struct. Biol. 7 (1997) 206–214.
[7] H. Fan, A.E. Mark, Protein Sci. 13 (2004) 211–220.
[8] J. Schonbrun, W.J. Wedemeyer, D. Baker, Curr. Opin. Struct. Biol. 12 (2002) 348–354.
[9] C. Venclovas, A. Zemla, K. Fidelis, J. Moult, Proteins(Suppl. 5) (2001) 163–170.
[10] D. Baker, A. Sali, Science 294 (2001) 93–96.
[11] S. Moro, F. Deflorian, M. Bacilieri, G. Spalluto, Curr. Pharm. Des. 12 (2006) 2175–2185.
[12] A. Evers, T. Klabunde, J. Med. Chem. 48 (2005) 1088–1097.
[13] D. Massotte, B.L. Kieffer, Nat. Struct. Mol. Biol. 12 (2005) 287.
[14] J. Drews, Science 287 (2000) 1960.
[15] A. Patny, P.V. Desai, M.A. Avery, Curr. Med. Chem. 13 (2006) 1667–1691.
[16] T.M. Frimurer, R.P. Bywater, Proteins 35 (1999) 375.
[17] U. Gether, Endocr. Rev. 21 (2000) 90.
[18] S.U. Miedlich, L. Gama, K. Seuwen, R.M. Wolf, G.E. Breitwieser, J. Biol. Chem. 279 (2004) 7254.
[19] X.Q. Xie, J.Z. Chen, E.M. Billings, Proteins 53 (2003) 307.
[20] J. Chamber, R.S. Ames, D. Bergsma, A. Muir, L.R. Fitzgerald, G. Hervieu, G.M. Dytko, J.J. Foley, J. Martin, W.S. Liu, J. Park, et al., Nature 400 (1999) 261.
[21] Y. Saito, H.P. Nothacker, Z. Wang, S.H. Lin, F. Leslie, O. Civelli, Nature 400 (1999) 265.
[22] Y. Chen, C. Hu, C.K. Hsu, Q. Zhang, C. Bi, M. Asnicar, H.M. Hsiung, N. Fox, L.J. Slieker, D.D. Yang, M.L. Heiman, Y. Shi, Endocrinology 143 (2002) 2469.
[23] D.J. Marsh, D.T. Weingarth, D.E. Novi, H.Y. Chen, M.E. Trumbauer, A.S. Chen, X.M. Guan, M.M. Jiang, Y. Feng, R.E. Camacho, et al., Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 3240.
[24] J.C. Bittencourt, F. Presse, C. Arias, et al., J. Comp. Neurol. 319 (1992) 218–245.
[25] D.S. Ludwig, N.A. Tritos, J.W. Mastaitis, et al., J. Clin. Invest. 107 (2001) 379–386.
[26] A.R. Kennedy, J.F. Todd, S.A. Stanley, et al., Endocrinology 142 (2001) 3265–3268.
[27] S.R. Chiocchio, M.G. Gallardo, P. Louzan, V. Gutnisky, J.H. Tramezzani, Biol. Reprod. 64 (2001) 1466–1472.
[28] K.M. Knigge, J.E. Wagner, Peptides 18 (1997) 1095–1097.
[29] M.E. Monzon, M.M. de-Souza, L.A. Izquierdo, I. Izquierdo, D.M. Barros, S.R. de-Barioglio, Peptides 20 (1999) 1517–1519.
[30] L. Verret, R. Goutagny, P. Fort, et al., BMC Neurosci. 4 (2003) 19.
[31] J. Kela, P. Salmi, R. Rimondini-Giorgini, M. Heilig, C. Wahlestedt, Regul. Pept. 114 (2002) 109–114.
[32] B. Borowsky, M.M. Durkin, K. Ogozalek, M.R. Marzabadi, J. DeLeon, R. Heurich, H. Lichtblau, Z. Shaposhnik, I. Daniewska, T.P. Blackburn, T.A. Branchek, C. Gerald, P.J. Vaysse, C. Forray, Nat. Med. 8 (2002) 825–830.
[33] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, Nucleic Acids Res. 31 (13) (2003) 3381–3385.
[34] M. Akamatsu, Curr. Top. Med. Chem. 12 (2002) 1381–1394.
[35] M.O. Taha, M.A. AlDamen, J. Med. Chem. 10 (2005) 8016–8034.
[36] T. Tuccinardi, V. Calderone, S. Rapposelli, A. Martinelli, J. Med. Chem. 49 (2006) 4305–4316.
[37] A.M. Abu-Hammad, F.U. Afifi, M.O. Taha, J. Mol. Graph. Model. 26 (2007) 443.
[38] K.S. Wilson, S. Butterworth, Z. Dauter, V.S. Lamzin, M. Walsh, S. Wodak, J. Pontius, J. Richelle, A. Vaguine, C. Sander, R.W. Hooft, G. Vriend, J.M. Thornton, R.A. Laskowski, M.W. MacArthur, E.J. Dodson, G. Murshudov, T.J. Old, R. Kaptein, J.C. Rullmann, J. Mol. Biol. 276 (1998) 417–436.
[39] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, G. Vriend, Bioinformatics 14 (6) (1998) 523–528.
[40] F. Horn, G. Vriend, F.E. Cohen, Nucleic Acids Res. 29 (2001) 346–349.
[41] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[42] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.
[43] J. Su, B.A. McKittrick, H. Tang, M. Czarniecki, W.J. Greenlee, B.E. Hawesb, K. O'Neill, Bioorg. Med. Chem. 13 (2005) 1829–1836.
[44] CERIUS2 LigandFit User Manual (Version 4.10), Accelrys Inc., San Diego, CA, 2005, pp. 3–48.
[45] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, J. Mol. Graph. Model. 21 (2003) 289–307.
[46] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, J. Mol. Graphics Model. 23 (2005) 395–407.
[47] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Chem. Biol. 2 (1995) 317–324.
[48] I. Muegge, Y.C. Martin, J. Med. Chem. 42 (1999) 791–804.
[49] A.N. Jain, J. Comput.-Aided Mol. Des. 10 (1996) 427–440.
[50] CERIUS2 OFF, Accelrys Inc., San Diego, CA, 1997, pp. 5–109.
[51] CERIUS2 4.10 QSAR, Accelrys Inc., San Diego, CA, 2005, pp. 177–187.
[52] CERIUS2 4.10 QSAR, Accelrys Inc., San Diego, CA, 2005, pp. 237–250.
[53] W. Sippl, J. Comput.-Aided Mol. Des. 14 (2000) 559–572.
[54] M. Clark, R.D. Cramer, Quant. Struct.-Act. Relat. 12 (1993) 137–145.
[55] J. Sutter, O. Güner, R. Hoffmann, H. Li, M. Waldman, in: O.F. Güner (Ed.), Pharmacophore Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 501–511.
[56] H. Li, J. Sutter, R. Hoffmann, in: O.F. Güner (Ed.), Pharmacophore Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 173–189.
[57] O.O. Clement, A.T. Mehl, in: F.O. Guner (Ed.), Pharmacophore Perception, Development, and Use in Drug Design, IUL Biotechnology Series, International University Line, La Jolla, California, 2000, pp. 71–84.
[58] Catalyst User Guide, Accelrys Software Inc., San Diego, 2005.

[59] R. Fischer, The Principle of Experimentation Illustrated by a PsychoPhysical Experiment, eighth ed. Hafner Publishing, New York, 1966, (Chapter II).
[60] E.M. Krovat, T. Langer, J. Med. Chem. 46 (2003) 716–726.
[61] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Nature 381 (1996) 272.
[62] S. Schlyer, R. Horuk, Drug Discov. Today 11 (2006) 481–493.
[63] D. Macdonald, N. Murgolo, R. Zhang, J.P. Durkin, X. Yao, C.D. Strader, M.P. Graziano, Mol. Pharmcol. 58 (2000) 217–225.
[64] R.D. Cramer, J.D. Bunce, D.E. Patterson, Quant. Struct.-Act. Relat. 7 (1988) 18–28.
[65] M.G.B. Drew, N.R. Lumley, N.R. Price, R.W. Watkins, in: K. Gundertofte, F.S. Jorgensen (Eds.), Proceedings of the 12th European Symposium on Quantitative Structure–Activity Relationships: Molecular Modeling and Prediction of Bioactivity, Kluwer Academic/Olenum Publishers, New York, 1998, pp. 453–454.
[66] A. Tropsha, P. Gramatica, V.K. Gombar, Quant. Struct.-Act. Relat. Comb. Sci. 22 (2003) 69–77.
[67] L.F. Ramsey, W.D. Schafer, The Statistical Sleuth, first ed. Wadesworth Publishing Company, USA, 1997.
[68] R.G. Karki, V.M. Kulkarni, Eur. J. Med. Chem. 36 (2001) 147–163.
[69] M.O. Taha, A.G. Al-Bakri, W.A. Zalloum, Bioorg. Med. Chem. Lett. 16 (2006) 5902–5906.
[70] M.O. Taha, Y. Bustanji, A.G. Al-Bakri, A.M. Yousef, W.A. Zalloum, I.M. Al-Masri, N. Atallah, J. Mol. Graphics Model. 25 (2007) 870–884.
[71] R.P. Sheridan, S.K. Kearsley, Drug Discov. Today 7 (2002) 903.
[72] R.A. Silverman, The Organic Chemistry of Drug Design and Drug Action, Academic Press, San Diego, CA, USA, 1991, pp. 62–65.