Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 711–720
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Receptor-based QSAR studies of non-peptide human oxytocin receptor antagonists

Balázs Jójárt *, Árpád Márki

Department of Pharmacodynamics and Biopharmacy, University of Szeged, H-6720 Szeged, Eötvös u. 6, Hungary

Received 24 March 2006; received in revised form 30 May 2006; accepted 31 May 2006
Available online 7 June 2006

Abstract

In the present study, QSAR calculations were performed on the receptor-based alignment of 58 non-peptide human oxytocin receptor antagonists. With the aid of different scoring functions (AutoDock 3.05 built-in and X-Score 1.2) the evolved receptor–ligand complexes were characterized. By means of various datasets it was confirmed that the scoring functions were not capable to predict the biological activity correctly in compounds containing a rigid derivative in the variable region. To improve the pK_i prediction 3D-QSAR calculation was performed. The regions related to the biological activity were determined by using cross-validated $r^2(q^2)$-guided region selection (q^2-GRS) method. The predictive power of the CoMFA model [$r^2_{pred} = 0.89$, q^2(LMO, five groups) = 0.695 ± 0.034] allowed prediction of the biological activities of newly synthesized compounds and confirmed the receptor-based alignment.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Receptor-based QSAR; Scoring function; CoMFA with q^2-GRS; Human oxytocin receptor; Benzoxazinone antagonists

1. Introduction

Since its first application comparative molecular field analysis (CoMFA) [1] has been routinely used in 3D-QSAR studies and drug discovery [2]. The method is based on the assumption that changes in the binding affinities of compounds are related to the changes in the molecular interaction field (MIF, [3]) generated in the active conformation of the ligands. The crucial point of these calculations is the active conformation. For rigid molecules, such as the steroid dataset that has become the benchmark in 3D-QSAR calculations [4], the alignment of the molecules can be easy compared with that of flexible compounds, such as the angiotensin receptor antagonists [5]. In this case, different conformations of the most active compound were taken into consideration, and the 3D-QSAR model was built for all alignments. When the alignment of these flexible compounds is structure (protein/DNA)-based, the MIF calculation and accordingly the CoMFA calculation are matched to requirements.

In drug discovery the G-protein-coupled receptors (GPCRs) [6] play the major role as target molecules, because the primary target of the available drugs, either directly or indirectly, are the GPCRs. There are ~1000 genes encoding such receptors in the human genome [7], and these receptors regulate virtually all known physiological processes in mammals. There are several distinct modes of ligand (agonist or antagonist) binding [8]: (i) only transmembrane core; (ii) transmembrane core, extracellular loops and N-terminal part; (iii) only N-terminal segment; (iv) extracellular loops and N-terminal part. The cardinal preoccupation is that only the crystal structure of bovine rhodopsin (bRho) is available, therefore the homology modelling (at least 20% sequence identity is necessary) is the only way to model the 3D structure of these receptors. After the homology modelling the automated docking and virtual screening [9–12] is possible.

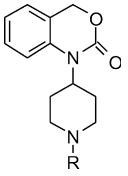
Oxytocin (OT) is a cyclic nonapeptide that binds to oxytocin receptor (OTR), which is a GPCR. The OT stimulates both the frequency and force of uterine contractions and plays an important physiological role in milk ejection. OT is the drug of choice for induction of labour. [13]. The molecules that competitively inhibit the interaction of OT with its membrane receptor have been of some interest because of their potential use in the treatment of preterm labour [14]. In case of the human

* Corresponding author. Tel.: +36 62 545567; fax: +36 62 545567.
E-mail address: jojartb@pharm.u-szeged.hu (B. Jójárt).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.05.010

oxytocin receptor and agonist/antagonist binding the peptide molecules bind to the core, exracellular loops and N-terminal regions of the human OTR (hOTR) [15,16]. In a previous publication [17], we performed QSAR studies on benzoxazinone (BZX) analogues. However, the structure-based alignment was performed by means of a hOTR model, which was built by using a δ-opiate template [18,19], and not the generally accepted bRho template [20]. In another study, the hOTR model was built with the bRho template [21]. In order to obtain relevant structure-activity relationships, we have now carried out QSAR calculations on the receptor-based alignment of 58 BZX hOTR antagonists (Tables 1 and 2, [22,23], compounds are provided in Supplementary Information in SMILES format), with the aims: (1) docking 57 non-peptide OT antagonists to the putative active site of the hOTR, determined earlier [20]; (2) characterization of the 58 receptor–ligand

Table 1
Structures and measured pK_i values of the OT antagonists (training set)



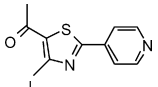
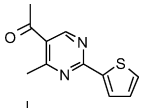
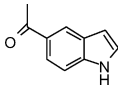
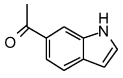
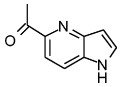
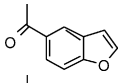
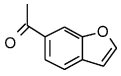
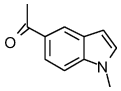
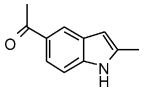
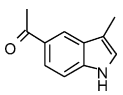
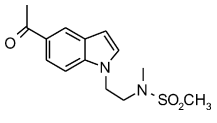
ID	R	pK_i
1		6.9
2		6.8
3		6.7
4		6.7
5		6.1
6		6.6
7		6.4
8		6.9
9		6.9
10		6.1
11		7.4

Table 1 (*Continued*)

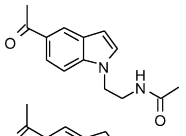
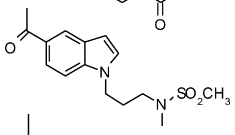
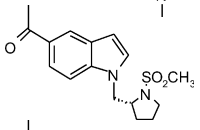
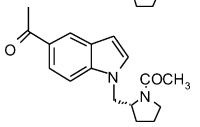
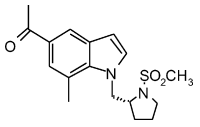
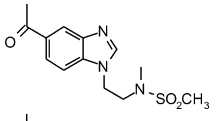
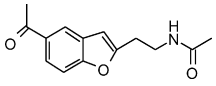
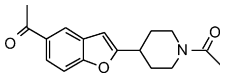
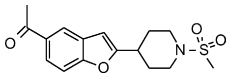
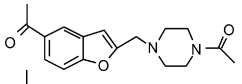
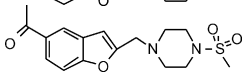
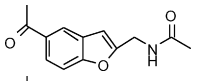
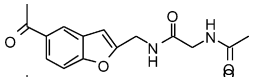
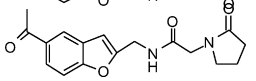
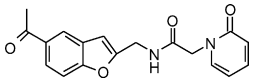
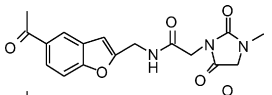
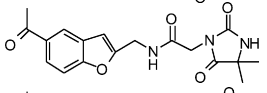
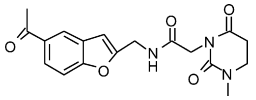
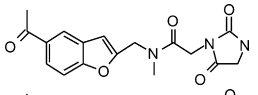
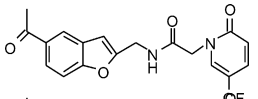
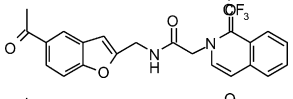
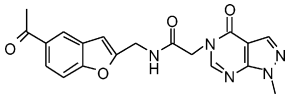
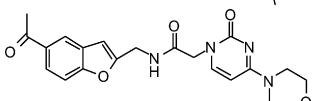
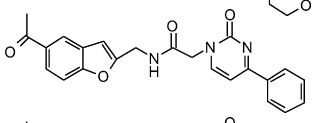
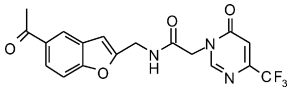
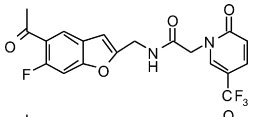
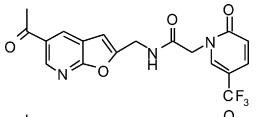
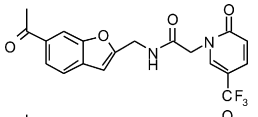
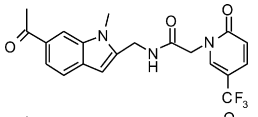
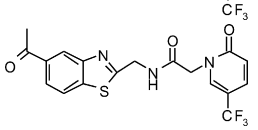
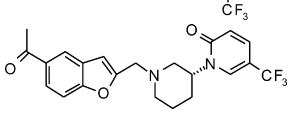
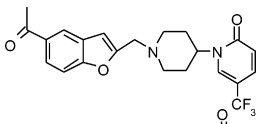
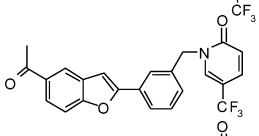
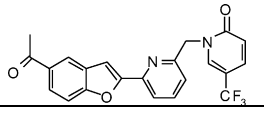
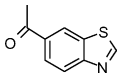
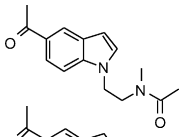
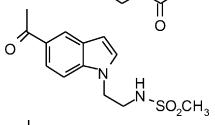
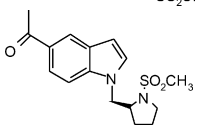
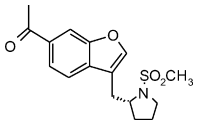
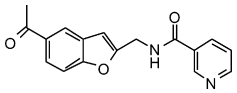
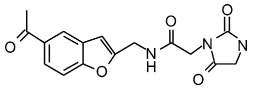
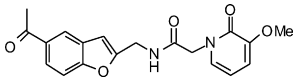
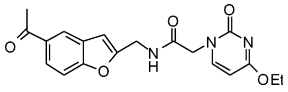
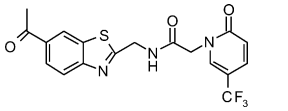
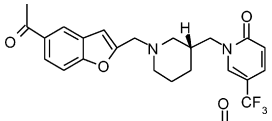
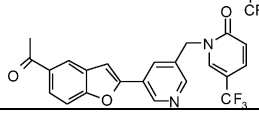
ID	R	pK_i
12		6.4
13		6.8
14		7.7
15		6.8
16		8.0
17		6.4
18		7.4
19		7.2
20		7.1
21		7.1
22		7.2
23		7.2
24		7.8
25		7.6
26		8.2
27		8.5
28		7.8
29		8.4
30		7.6

Table 1 (*Continued*)

ID	R	pK_i
31		7.6
32		8.7
33		8.2
34		8.9
35		9.3
36		9.0
37		8.4
38		8.1
39		8.5
40		7.0
41		7.3
42		7.1
43		7.6
44		6.1
45		6.1
46		8.0

The receptor-binding studies were performed on the hOTR, with [^{3}H]OT for displacement analysis. pK_i: negative logarithm of inhibition constant [22,23].

Table 2
Structures and measured pK_i values of the OT antagonists (test set)

ID	R	pK_i
47		6.1
48		7.2
49		6.5
50		6.9
51		6.7
52		7.6
53		8.2
54		8.6
55		8.7
56		8.0
57		8.0
58		7.2
		

The receptor-binding studies were performed on the hOTR, using [^{3}H]OT for displacement analysis. pK_i: negative logarithm of inhibition constant [22,23].

complexes via scoring functions; (3) construction of an orientation-independent CoMFA model.

2. Materials and methods

2.1. Receptor-based alignment of the investigated compounds

In a previous paper [20], we described the potential dynamic conformation of the hOTR–BZX complex. The evolved conformation was used as the starting point in this study to perform QSAR calculations. The compounds were super-

imposed with the most active ligand, and docking calculations were performed to obtain the receptor-based alignment. The docking procedures were carried out with the AutoDock 3.05 Software [24] with the following parameters: the grid box ($40 \times 40 \times 60$ grid point, 0.375 Å grid point distance) was centred on the ligand; in the calculation of the electrostatic grid map, the distance-dependent dielectric constant of Mehler and Solmayer [25] was applied, for the van der Waals interaction the 12–6, and for the H-bond the 12–10 Lennard–Jones potential was utilized. The docking was performed 50 times for each compound; the translational, quaternion and torsional step sizes were set to 0.5 Å, 5.0° and 5.0°, respectively. The possible active conformation of the ligand was selected by means of cluster analysis, carried out with 0.5 cluster tolerance, using the free energy of binding (ΔG_{bind}) as descriptor. From the most loaded cluster, the lowest energy structure was selected. After the docking, all receptor–ligand complexes were minimized in a multistep protocol by means of Molecular Operating Environment 2005.06 Software [26], using the MMFF94s force field [27]. The minimization was as follows: (i) the receptor atoms were restrained to their initial position with a 1000 kcal/mol Å^2 force constant until the gradient attained a value of 10; (ii) in the second stage, the force constant was decreased to 100 kcal/mol Å^2 until the gradient was less than 1; (iii) in the final stage, all receptor residues closer than 4.5 Å to the ligands were allowed move freely, while the remainder of the protein was restrained to its position with a 100 kcal/mol Å^2 force constant [28]. During the minimization process, a distance-dependent dielectric constant ($\varepsilon_r = 4$) and a 10 Å cut-off were used for the long-range non-bonding interactions.

2.2. Estimation of the free energy of binding

The evolved receptor–ligand complexes were characterized with two empirical scoring functions: the built-in AutoDock 3.05 (BIND) and X-Score 1.2 [29]. The empirical scoring function of AutoDock 3.05 includes five terms: van der Waals, H-bond, electrostatic, solvation and ligand deformation terms. The coefficients of the five terms were determined by linear regression, using 30 protein–ligand complexes [24]. The scoring function of X-Score 1.2 also contains five terms: van der Waals, H-bond, hydrophobic (which could be derived in four ways: hydrophobic matching (HM), hydrophobic contact potential (HP), hydrophobic molecular surface (HS), and the average of the three terms (AVE)), deformation and the regression constant, which implicitly include the entropy loss during the binding. The first version of this scoring function was calibrated by using 200 protein–ligand complexes [29], which was subsequently extended to 800.

2.3. 3D-QSAR studies

In the 3D-QSAR calculations, the powerful and widely used CoMFA method was utilized, with the cross-validated $r^2(q^2)$-guided region selection method [30], with which it is easy to eliminate the orientation-dependent properties of 3D-QSAR studies. The protocol was as follows: (i) the conventional CoMFA calculation with an automated generated box was performed; (ii) the initial box was divided into 125 equal small boxes; (iii) for each small box, CoMFA calculations were performed with the C.3 (+1) atom type, and 1.0 Å step size; (iv) a master region file was generated through the use those boxes where q^2 was greater than a special cut-off (in this study, we used cut-offs of 0.1, 0.2, 0.3, 0.4 and 0.5).

With five master regions obtained, CoMFA calculations were performed. In the CoMFA electrostatic field calculation, a distance-dependent dielectric form was utilized, with a 30 kcal/mol cut-off for both fields. To choose the appropriate parameters and check the statistical significance of the models, leave-one-out (LOO) and leave-multiple-out (LMO, five groups) cross-validation schemes were used. During the PLS calculations, the maximum number of components was set to 10, and the minimum sigma value to 2 kcal/mol. The final non-cross-validated models were constructed with the number of components obtained from the cross-validated calculations. The results were interpreted graphically by means of field contribution maps (stdev × coeff.).The prediction ability of the QSAR models was also investigated. Models were built by using 46 compounds (training set) and the biological activity was estimated on 12 compounds (test set). In 3D-QSAR calculations, Sybyl 7.1 Software was used [31].

3. Results and discussion

In order to support the docking results and establish a simple model with which the order of magnitude of the biological activity could be estimated, we characterized the evolved receptor–ligand complexes with the estimated ΔG_{bind}, calculated by using the AutoDock 3.05 (BIND) and X-Score 1.2 scoring functions (HP, HM, HS, AVE). The calculated ΔG_{bind} values are available as Supplementary Information (Table S1). The plots of the pK_i versus ΔG_{bind} are depicted in Fig. 1.

Linear regression with cross-validation was performed to determine the statistical significance of the models. Linear regression was performed first, including all 58 compounds; the parameters obtained are listed in Table 3.

We investigated the influence of compounds **44** and **45** (Table 4).

On removal of the two outliers, the prediction ability ($q^2 = \sim 0.5$) and the goodness of fit ($r^2 = \sim 0.55$) increased significantly. To test the predictive power of the models, linear regression was also performed for the training set (without the outliers), and tested on the test set. The parameters obtained are listed in Table 5.

By means of the two sets, the predictive ability of the models were also acceptable ($q^2 > 0.5$, $r^2_{pred} > 0.4$). The predicted pK_i values for the test set are listed in Table 6.

After the removal of compound 58 (residual = −1.3) from the dataset, the r^2_{pred} values increased significantly: 0.66, 0.69, 0.76, 0.71 and 0.73 for HP, HM, HS, AVE and BIND, respectively. The common structural element in compounds **44**, **45** and **58** is a rigid derivative at the position of the first amide group after the benzofuran ring. Based on these results, we concluded that the models obtained from scoring calculations

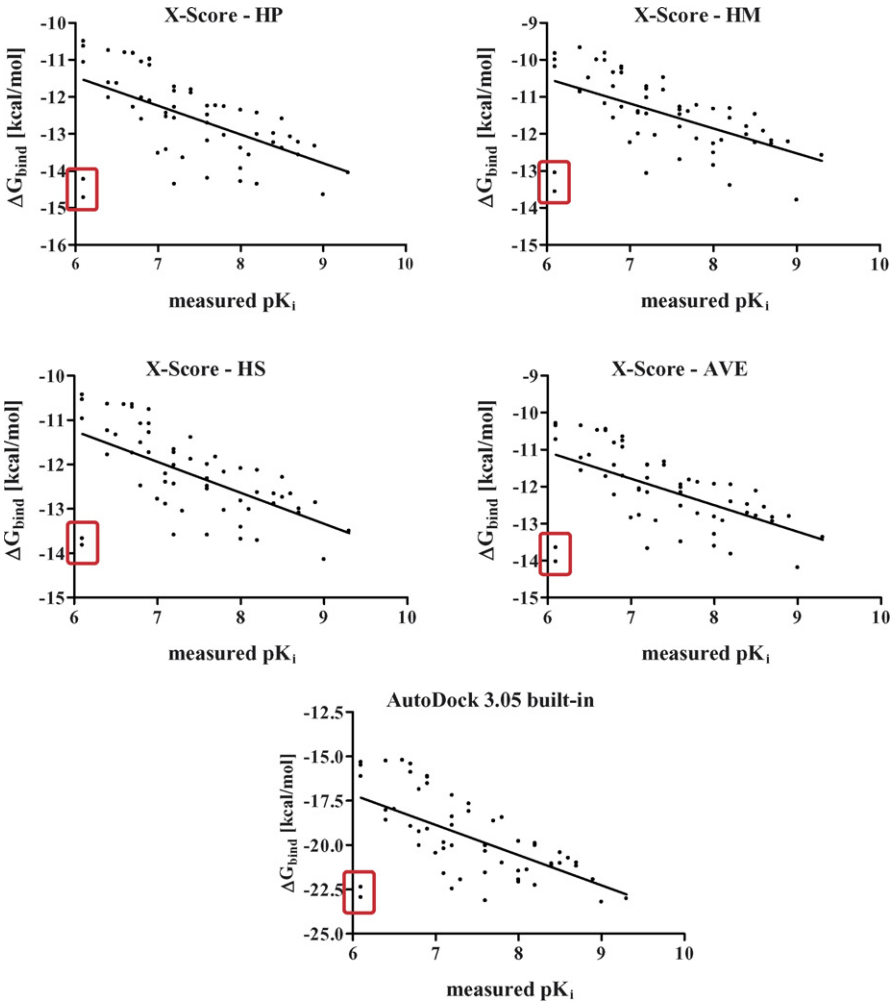


Fig. 1. Plots of the estimated ΔG_{bind} vs. measured pK_i using various scoring functions including all **58** compounds. The determined outliers (compound **44** and **45**) are depicted with a red box.

are not capable to estimate the biological activity, when there is a rigid derivative at the position of the amide group.

Henceforth, to confirm the previous statement, we performed the linear regressions on the dataset after removing all those compounds, which contain rigid derivatives: **14**, **15**, **16**, **19**, **20**, **21**, **22**, **30**, **43**, **44**, **45**, **46**, **50**, **51**, **57** and **58** (Tables 7–9; Fig. 2).

We obtained significantly increased r^2_{pred} values (∼0.8) in each case. Based on these results we can conclude that the prediction ability of these scoring functions is poor if there is a rigid derivative at the position of the amide group and has a good predictive power in pK_i range 6.0–7.5. Accordingly, we suggest that the simple linear models could be useful to exclude compounds without rigid derivatives and with low biological activity.

To improve the biological estimation of newly synthetized compounds, we performed 3D-QSAR calculations. To determine the important regions around the molecules, and to

Table 3
The parameters of the cross-validated and non-cross-validated models, obtained by using different scoring functions, all 58 compounds were included in the calculations

Scoring function	Cross-validated model parameters		Non-cross-validated model parameters			
	q^2	SDEP	r^2	s	$F(1, 56)$	Linear model
HP	0.255	0.722	0.320	0.690	26.313	pK_i = 2.261–0.411 × ΔG_{bind} (HP)
HM	0.235	0.732	0.307	0.696	24.847	pK_i = 2.182–0.457 × ΔG_{bind} (HM)
HS	0.301	0.699	0.358	0.670	31.201	pK_i = 1.151–0.512 × ΔG_{bind} (HS)
AVE	0.266	0.717	0.331	0.684	27.691	pK_i = 1.835–0.462 × ΔG_{bind} (AVE)
BIND	0.322	0.689	0.371	0.663	33.079	pK_i = 3.153–0.218 × ΔG_{bind} (BIND)

q^2: cross-validated r^2; SDEP: standard error of prediction, s: standard error of estimate; r^2: conventional r^2; F: Fisher (1, 56) value.

Table 4
The parameters of the cross-validated and non-cross-validated models, obtained by using different scoring functions, with removal of compounds **44** and **45** from the dataset

Scoring function	Cross-validated model parameters		Non-cross-validated model parameters			
	q^2	SDEP	r^2	s	$F(1, 54)$	Linear model
HP	0.503	0.572	0.538	0.552	62.970	$pK_i = 0.766–0.536 \times \Delta G_{bind}$ (HP)
HM	0.514	0.566	0.548	0.546	65.503	$pK_i = 0.386–0.621 \times \Delta G_{bind}$ (HM)
HS	0.542	0.549	0.573	0.531	72.410	$pK_i = -0.416–0.647 \times \Delta G_{bind}$ (HS)
AVE	0.526	0.559	0.559	0.539	68.420	$pK_i = 0.188–0.605 \times \Delta G_{bind}$ (AVE)
BIND	0.523	0.561	0.555	0.542	67.236	$pK_i = 2.358–0.262 \times \Delta G_{bind}$ (BIND)

q^2: cross-validated r^2; SDEP: standard error of prediction; s: standard error of estimate; r^2: conventional r^2; F: Fisher (1, 54) value.

Table 5
The parameters of the cross-validated and non-cross-validated models, obtained by using different scoring functions for the training set without compounds **44** and **45**

Scoring function	Cross-validated model parameters		Non-cross-validated model parameters				
	q^2	SDEP	r^2	s	$F(7, 42)$	Linear model	r^2_{pred}
HP	0.538	0.553	0.529	0.578	57.432	$pK_i = 0.673–0.546 \times \Delta G_{bind}$ (HP)	0.42
HM	0.554	0.543	0.521	0.590	60.461	$pK_i = 0.311–0.630 \times \Delta G_{bind}$ (HM)	0.43
HS	0.551	0.545	0.523	0.587	59.649	$pK_i = -0.387–0.646 \times \Delta G_{bind}$ (HS)	0.53
AVE	0.555	0.543	0.520	0.592	60.838	$pK_i = 0.134–0.612 \times \Delta G_{bind}$ (AVE)	0.46
BIND	0.532	0.557	0.533	0.571	55.965	$pK_i = 2.505–0.256 \times \Delta G_{bind}$ (BIND)	0.52

q^2: cross-validated r^2; SDEP: standard error of prediction; s: standard error of estimate; r^2: conventional r^2; F: Fisher (1, 47) value; r^2_{pred}: predictive r^2 for the test set.

Table 6
The predicted pK_i values obtained by using different scoring functions without compounds **44** and **45**

ID	Measured pK_i	Predicted pK_i				
		HP	HM	HS	AVE	BIND
47	6.1	6.47	6.61	6.35	6.46	6.47
48	7.2	7.13	7.06	7.14	7.11	7.21
49	6.5	7.02	6.91	6.93	6.95	7.10
50	6.9	7.28	7.41	7.19	7.29	7.39
51	6.7	7.37	7.35	7.19	7.31	7.35
52	7.6	7.86	7.75	7.71	7.78	7.61
53	8.2	7.45	7.43	7.44	7.44	7.59
54	8.6	7.81	7.82	7.78	7.81	7.81
55	8.7	7.88	7.98	8.07	7.98	7.88
56	8.0	7.97	8.03	7.89	7.97	7.99
57	8.0	8.27	8.19	8.27	8.25	8.15
58	7.2	8.50	8.54	8.38	8.49	8.25

exclude the influence of the orientation, we applied the q^2-GRS method, developed by Tropsha and co-workers [30].

Use of the protein-based alignment of the 58 compounds (Fig. 3), the conventional CoMFA calculation provided the following parameters: $q^2 = 0.723$, SDEP = 0.452 and $n = 4$.

After the conventional CoMFA calculation, the q^2-GRS method was applied with the following q^2 cut-offs: 0.1, 0.2, 0.3, 0.4 and 0.5 (Table 10, Fig. 4).

With the various q^2 cut-offs, the parameters of the models did not change significantly. This indicated that, by increasing the cut-off value, we could exclude the sub regions which are not related to the biological activity. On the other hand, these results suggest that we could extract the maximum chemical information from the dataset. It is a well-known fact that a high q^2 does not guarantee predictive power for external datasets [32,33]. To test the predictive power of the CoMFA models, further calculations were performed on the training set, and the

Table 7
The parameters of the cross-validated and non-cross-validated models, obtained by using different scoring functions, without rigid derivatives containing compounds (**14**, **15**, **16**, **19**, **20**, **21**, **22**, **30**, **43**, **44**, **45**, **46**, **50**, **51**, **57** and **58**)

Scoring function	Cross-validated model parameters		Non-cross-validated model parameters			
	q^2	SDEP	r^2	s	$F(1, 40)$	Linear model
HP	0.635	0.547	0.667	0.522	80.009	$pK_i = -0.380–0.637 \times \Delta G_{bind}$ (HP)
HM	0.636	0.546	0.671	0.519	81.404	$pK_i = -0.669–0.725 \times \Delta G_{bind}$ (HM)
HS	0.679	0.513	0.706	0.491	96.069	$pK_i = -1.724–0.765 \times \Delta G_{bind}$ (HS)
AVE	0.658	0.529	0.688	0.505	88.242	$pK_i = -0.990–0.714 \times \Delta G_{bind}$ (AVE)
BIND	0.655	0.531	0.684	0.509	86.675	$pK_i = 1.642–0.306 \times \Delta G_{bind}$ (BIND)

q^2: cross-validated r^2; SDEP: standard error of prediction; s: standard error of estimate; r^2: conventional r^2; F: Fisher (1, 40) value.

Table 8
The parameters of the cross-validated and non-cross-validated models, obtained by using different scoring functions for the training set, without rigid derivatives containing compounds (**14**, **15**, **16**, **19**, **20**, **21**, **22**, **30**, **43**, **44**, **45**, **46** and **50**)

Scoring function	Cross-validated model parameters		Non-cross-validated model parameters				
	q^2	SDEP	r^2	s	$F(1, 32)$	Linear model	r^2_{pred}
HP	0.616	0.562	0.657	0.531	61.371	$pK_i = -0.017–0.607 \times \Delta G_{bind}$ (HP)	0.76
HM	0.613	0.563	0.657	0.531	61.211	$pK_i = -0.240–0.685 \times \Delta G_{bind}$ (HM)	0.80
HS	0.649	0.537	0.686	0.508	69.760	$pK_i = -1.307–0.729 \times \Delta G_{bind}$ (HS)	0.85
AVE	0.635	0.548	0.673	0.518	65.998	$pK_i = -0.582–0.678 \times \Delta G_{bind}$ (AVE)	0.81
BIND	0.629	0.552	0.668	0.522	64.450	$pK_i = 1.971–0.289 \times \Delta G_{bind}$ (BIND)	0.83

q^2: cross-validated r^2; SDEP: standard error of prediction; s: standard error of estimate; r^2: conventional r^2; F: Fisher (1, 32) value; r^2_{pred}: predictive r^2 for the test set.

Table 9
The predicted pK_i values obtained by using different scoring functions

ID	Measured pK_i	Predicted pK_i				
		HP	HM	HS	AVE	BIND
47	6.1	6.43	6.61	6.24	6.44	6.45
48	7.2	7.17	7.10	7.13	7.15	7.29
49	6.5	7.05	6.95	6.89	6.98	7.16
52	7.6	7.99	7.85	7.78	7.90	7.74
53	8.2	7.53	7.51	7.47	7.52	7.72
54	8.6	7.92	7.93	7.86	7.93	7.96
55	8.7	8.01	8.10	8.18	8.12	8.04
56	8.0	8.11	8.16	7.98	8.11	8.17

In the and test sets the rigid derivatives containing compounds were not included (compounds **51**, **57** and **58**).

model obtained was tested on the test set (Tables 11 and 12 and Fig. 5). The predictive power of the models indicates that compounds containing rigid derivatives have not been proved to be outliers and the non-cross-validated model could be used in biological activity prediction.

In the selection of the test set we have to pay attention that the selected activities (6.1–8.7) must cover the range of activities (6.1–9.3), and the structures were selected randomly.

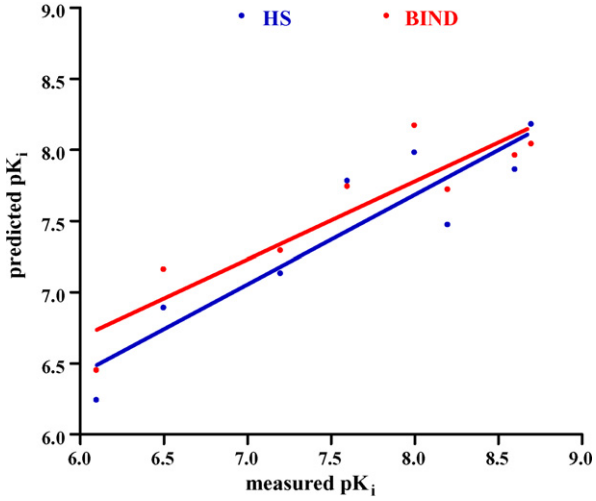


Fig. 2. Plot of the predicted vs. the measured pK_i for the test set obtained by using the HS (blue) and BIND (red) scoring functions. In the training and test sets the rigid derivatives containing compounds were not included (compounds **14**, **15**, **16**, **19**, **20**, **21**, **22**, **30**, **43**, **44**, **45**, **46**, **50**, **51**, **57** and **58**).

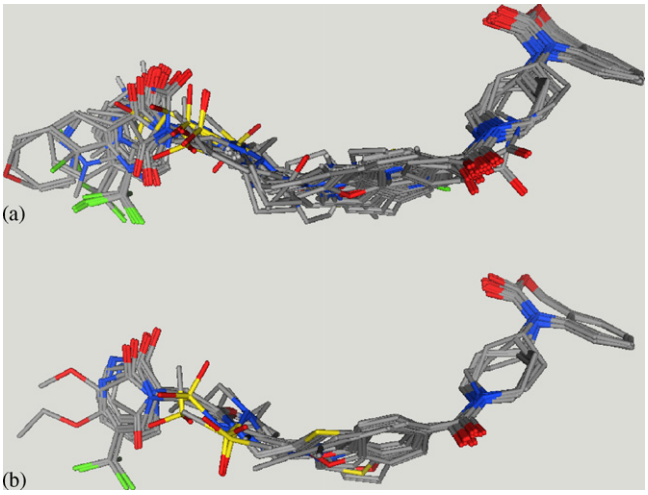


Fig. 3. The receptor-based alignment of the training set (a: 46 compounds) and the test set (b: 12 compounds).

A possible explanation for the good predictive power is that the chemical variability of the training set explains the chemical variability of the test set. Accordingly, we performed LMO cross-validation for the whole dataset, where the software

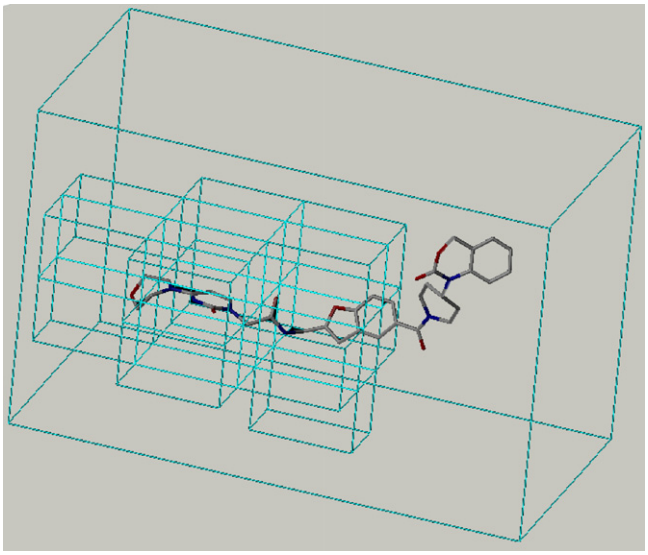


Fig. 4. The locations of the regions used in the conventional CoMFA calculations (cut-off $q^2 > 0.5$).

Table 10
The parameters of the various CoMFA calculations on 58 compounds

Model	q^2 cut-off	Number of boxes	Cross-validated parameters			Non-cross-validated parameters			Norm. coeff. (frac.)	
			q^2	SDEP	n	s	r^2	F	Ste	Ele
Conventional	–	125	0.723	0.452	4	0.244	0.920	151.872[a]	1.669 (0.452)	2.027 (0.548)
Model A	0.1	34	0.711	0.462	4	0.249	0.916	144.270[a]	1.614 (0.443)	2.030 (0.557)
Model B	0.2	31	0.709	0.464	4	0.248	0.917	146.328[a]	1.641 (0.447)	2.029 (0.553)
Model C	0.3	26	0.706	0.466	4	0.249	0.916	144.665[a]	1.644 (0.4489	2.022 (0.552)
Model D	0.4	23	0.710	0.464	6	0.177	0.959	200.803[b]	2.550 (0.461)	2.636 (0.539)
Model E	0.5	17	0.706	0.476	6	0.183	0.957	187.184[b]	2.214 (0.458)	2.624 (0.542)

q^2: cross-validated r^2; SDEP: standard error of prediction; n: optimum number of components; s: standard error of estimate; r^2: conventional r^2; F: Fisher value [[a](4, 53), [b](6, 51)].

Table 11
The parameters of the various CoMFA calculations, using the training set (46 compounds)

Model	q^2 cut-off	Cross-validated parameters			Non-cross-validated parameters					r^2_{pred}
		q^2	SDEP	n	r^2	s	$F(4, 41)$	Norm. coeff. (frac.)		
								Ste	Ele	
Model A	0.1	0.649	0.518	4	0.247	0.920	118.567	1.657 (0.436)	2.143 (0.564)	0.90
Model B	0.2	0.645	0.521	4	0.245	0.921	119.944	1.667 (0.436)	2.160(0.564)	0.91
Model C	0.3	0.641	0.524	4	0.247	0.920	118.625	1.670 (0.436)	2.160 (0.564)	0.90
Model D	0.4	0.647	0.519	4	0.246	0.921	118.860	1.664 (0.431)	2.192 (0.569)	0.89
Model E	0.5	0.645	0.521	4	0.251	0.918	114.656	1.632 (0.427)	2.189 (0.573)	0.89

q^2: cross-validated r^2; SDEP: standard error of prediction; n: optimum number of components; s: standard error of estimate; r^2: conventional r^2; $F(4, 41)$: Fisher value, r^2_{pred}: predictive r^2 on test set.

selects the test sets randomly. The LMO cross-validation was performed for 500 times with five groups, which means that the model was built by using 80% of the available data (46 compounds) and the model obtained was tested on 12 compounds (Table 13). Accordingly, the partition of the data was same as used before.

Among the 500 LMO cross-validation runs, there were 6 calculations in which q^2 was higher than 0.7, and 6 in which the q^2 was lower than 0.6.

The results confirmed that the model is likely to be of certain predictive power when extended to other ligands.

The field contribution maps obtained from the non-cross-validated model E are depicted in Figs. 6 and 7.

Table 12
The predicted biological activities for the test set, using various models

ID	Measured pK_i	Predicted pK_i				
		Model A	Model B	Model C	Model D	Model E
47	6.1	6.50	6.49	6.50	6.51	6.49
48	7.2	6.78	6.77	6.76	6.72	6.79
49	6.5	6.74	6.74	6.74	6.78	6.79
50	6.9	6.93	7.02	7.00	7.08	7.14
51	6.7	6.95	6.87	6.90	6.82	6.87
52	7.6	8.19	8.19	8.20	8.19	8.19
53	8.2	8.25	8.23	8.21	8.21	8.20
54	8.6	8.54	8.53	8.58	8.58	8.63
55	8.7	8.69	8.69	8.72	8.72	8.62
56	8.0	8.00	8.00	8.00	8.00	8.00
57	8.0	7.85	7.86	7.84	7.83	7.75
58	7.2	7.29	7.29	7.28	7.30	7.26

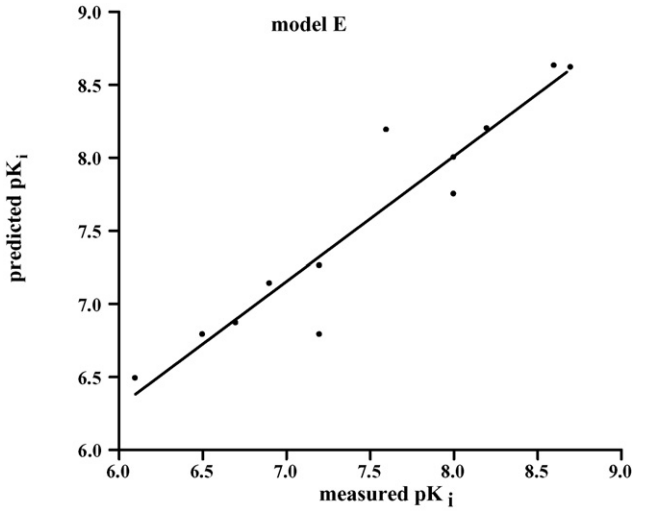


Fig. 5. The plot of measured vs. predicted pK_i using the non-cross-validated model of model E.

Table 13
The results of LMO cross-validation (500×) using five groups for model E

Model E	Mean ± S.D.	Minimum	Maximum
q^2	0.695 ± 0.034	0.589	0.766
SDEP	0.477 ± 0.0326	0.424	0.562
n	4.74 ± 0.88	2	7

q^2: cross-validated r^2; SDEP: standard error of prediction; n: optimum number of components.

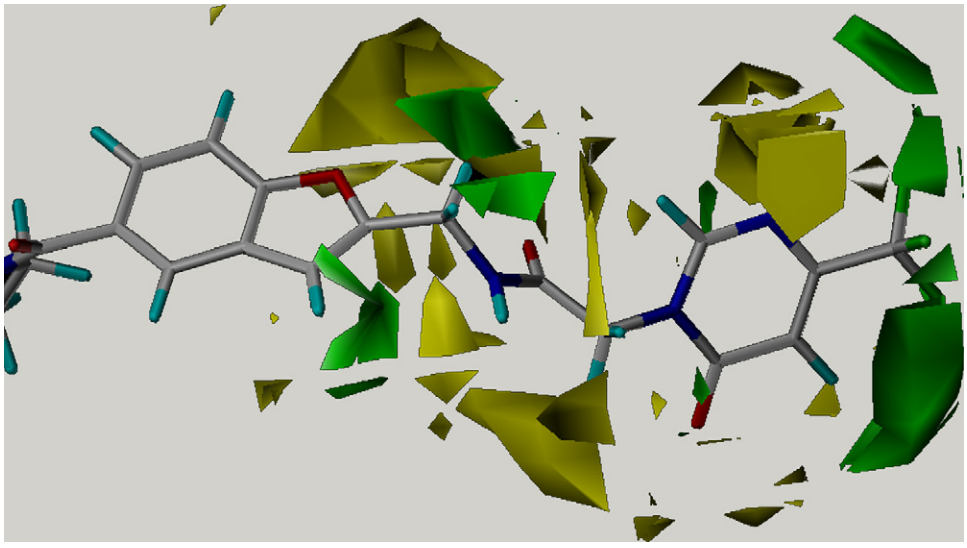


Fig. 6. Contour plot of steric field contribution maps (S.D. × coeff.). Larger pK_i values correlated with more bulk near green and less bulk near yellow. Compound **37** (pK_i = 8.4) is depicted as reference molecule.

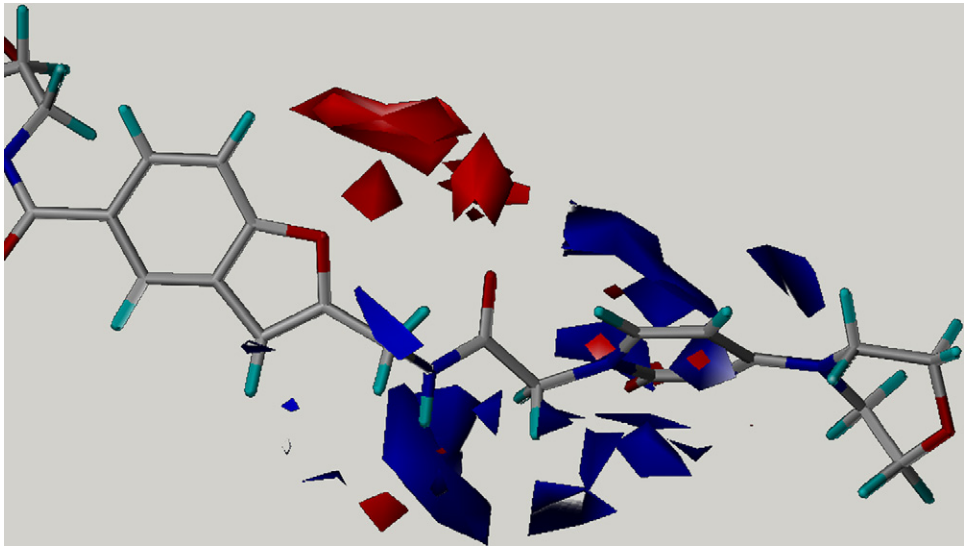


Fig. 7. Contour plot of electrostatic field contribution maps (S.D. × coeff.). Larger pK_i values correlated with more positive charge near blue, and with more negative charge near red. The most active compound (**35**, pK_i = 9.3) is depicted as reference molecule.

The important regions in the ligand are: (1) the bulky substituent nearby the CF_3 moiety; (2) the negative charge on the benzofuran ring at the position of the O atom and at the position of the amide O; (3) the positive charge at the position of the amide H. The steric unfavourable regions are the followings: bulky substituent near the furan ring, the amide group and nearby the N atom in the pyrimidin ring.

4. Summary

In the present study, QSAR calculations were performed on 58 benzoxazinone analogues taken from the literature. The final linear model of the scoring functions was developed by using only 42 compounds, because compounds containing rigid derivative were removed from the dataset. The simple linear model of scoring functions (Table 7) could be useful in preliminary research in the pK_i range 6.0–7.5. By means of the q^2-GRS method, the important CoMFA regions (17) were identified, and the high predictive ability of model E [q^2(LOO) = 0.706 (all 58 compounds), q^2(LOO) = 0.645 (the training set, 46 compounds), q^2(LMO) = 0.695 ± 0.034 (58 compounds) and $r^2_{pred} = 0.89$ (the test set, 12 compounds)] confirmed the receptor-based alignment of the compounds and therefore the most likely active conformation. As concerns only those parts of the receptor cavity which was determined with the q^2-GRS method, the small molecule fragment docking is easier, and therefore the determination of new variable part of a potent non-peptide oxytocin is possible. Thereafter, by connecting these small fragments to the benzoxazinone-benzofuran derivative,

biological prediction is possible by using the final CoMFA model from model E.

Acknowledgements

The work was supported by the Centenarium Foundation of Gedeon Richter Ltd., and the National Council for Research and Technology (NKFP) Budapest (RET 08/2004).

Computational support (Sybyl 7.1) was provided by the Hungarian Academy of Sciences Biological Research Centre, Szeged (OTKA M 045378).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.05.010.

References

[1] R.D. Cramer III, D.E. Patterson, D.B. Jeffrey, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1998) 5959–5967.
[2] H. Kubinyi, G. Folkers, Y.C. Martin, 3D QSAR in Drug Design—Recent Advances, vol. 3, Kluwer Escom, London, 1998, pp. 317–338.
[3] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[4] H. Kubinyi, G. Folkers, Y.C. Martin, 3D QSAR in Drug Design—Recent Advances, vol. 3, Kluwer Escom, London, 1998, pp. 199–214.
[5] P. Datar, P. Desai, E. Coutinho, K. Iyer, CoMFA and CoMSIA studies of angiotensin (AT1) receptor antagonists, J. Mol. Model. 8 (2002) 290–301.
[6] U. Gether, Uncovering molecular mechanisms involved in activation of g protein-coupled receptors, Endocr. Rev. 21 (2000) 90–113.
[7] R. Fredriksson, M.C. Lagerström, L.-G. Lundin, H.B. Schiöth, The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints, Mol. Pharmacol. 63 (2003) 1256–1272.
[8] T.H. Ji, M. Grossmann, H. Ji, G protein-coupled receptors. I. Diversity of receptor–ligand interactions, J. Biol. Chem. 273 (1998) 17299–17302.
[9] M. Ashton, M.H. Charlton, M.K. Schwarz, R.J. Thomas, M. Whittaker, The selection and design of GPCR ligands: from concept to the clinic, Comb. Chem. HTS 7 (2004) 441–452.
[10] A. Evers, G. Hessler, H. Matter, T. Klabunde, Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols, J. Med. Chem. 48 (2005) 5448–5465.
[11] A. Evers, T. Klabunde, Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor, J. Med. Chem. 48 (2005) 1088–1097.
[12] J.R. Bock, D.A. Gough, Virtual screen for ligands of orphan G protein-coupled receptors, J. Chem. Inf. Model. 45 (2005) 1402–1414.
[13] J. Kruse, Oxytocin: pharmacology and clinical application, J. Family Pract. 23 (1986) 473–479.
[14] M.K. Schwarz, P. Page, Preterm labour: an overview of current and emerging therapeutics, Curr. Med. Chem. 10 (2003) 1441–1468.
[15] R. Postina, E. Kojro, F. Fahrenholz, Separate agonist and peptide antagonist binding sites of the oxytocin receptor defined by their transfer into the V_2 vasopressin receptor, J. Biol. Chem. 271 (1996) 31593–31601.
[16] S.R. Hawtin, H.C. Howard, M. Whetaley, Identification of an extracellular segment of the oxytocin receptor providing agonist-specific binding epitopes, Biochem. J. 354 (2001) 465–472.
[17] B. Jójárt, T.A. Martinek, Á. Márki, The 3D structure of the binding pocket of the human oxytocin receptor for benzoxazine antagonists, determined by molecular docking, scoring functions and 3D-QSAR methods, J. Comput.-Aided Mol. Des. 19 (2005) 341–356.
[18] A. Giełdoñ, R. Kaźmierkiewicz, R. Ślusarz, J. Ciarkowski, Molecular modeling of interactions of the non-peptide antagonist YM087 with the human vasopressin V1a, V2 receptors and with oxytocin receptors, J. Comput.-Aided. Mol. Des. 15 (2002) 1085–1104.
[19] I.D. Pogozheva, A.L. Lomize, H.I. Mosberg, Opioid receptor three-dimensional structures from distance geometry calculations with hydrogen bonding constraints, Biophys. J. 75 (1998) 612–634.
[20] 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, Crystal structure of rhodopsin: a G protein-coupled receptor, Science 289 (2000) 739–745.
[21] B. Jójárt, A. Márki, Possible dynamic anchor points in a benzoxazinone derivative-human oxytocin receptor system—a molecular docking and dynamics calculation, and the possible electrostatic and hydrophobic anchor points were determined by means of structural and energetical filters of the dynamic complex, J. Mol. Model. 13 (2007) 1–10.
[22] P.G. Wyatt, M.J. Allen, J. Chilcott, A. Foster, D.G. Livermore, J.E. Mordaunt, J. Scicinski, P. Woollard, Identification of potent and selective oxytocin antagonists. Part 1. Indole and benzofuran derivatives, Bioorg. Med. Chem. Lett. 12 (2002) 1399–1404.
[23] P.G. Wyatt, M.J. Allen, J. Chilcott, C.J. Gardner, D.G. Livermore, J.E. Mordaunt, F. Nerozzi, M. Patel, M.J. Perren, G.G. Weingarten, S. Shabbir, P.M. Wollard, P. Zhou, Identification of potent and selective oxytocin antagonists. Part 2. Further investigation of benzofuran derivatives, Bioorg. Med. Chem. Lett. 12 (2002) 1405–1411.
[24] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.
[25] E.L. Mehler, T. Solmajer, Electrostatic effects in proteins: comparison of dielectric and charge models, Protein Eng. 4 (1991) 903–910.
[26] Molecular Operating Environment, Version 2005.06, CCG Inc., Montreal, Quebec, Canada, 2005.
[27] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parametrization, and performance of MMFF94, J. Comput. Chem. 17 (1996) 490–519;
T.A. Halgren, Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions, J. Comput. Chem. 17 (1996) 520–552;
T.A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, J. Comput. Chem. 17 (1996) 553–586;
T.A. Halgren, Merck molecular force field. IV. Conformational energies and geometries for MMFF94, J. Comput. Chem. 17 (1996) 587–615;
T.A. Halgren, Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules, J. Comput. Chem. 17 (1996) 616–641.
[28] A.D. Mackerrel Jr, Empirical force fields for biological macromolecules: overview and issues, J. Comput. Chem. 25 (2004) 1584–1604.
[29] R. Wang, L. Lai, S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction, J. Comput.-Aided Mol. Des. 16 (2002) 11–26.
[30] S.J. Cho, A. Tropsha, Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results, J. Med. Chem. 38 (1995) 1060–1066.
[31] Sybyl Molecular Modeling System, Version 7.1, Tripos Associates, St. Louis, MO, 2005
[32] E. Novellino, C. Fattorusso, G. Greco, Use of comparative molecular field analysis and cluster analysis in series design, Pharm. Acta Helv. 70 (1995) 149–154.
[33] H. Kubinyi, F.A. Hamprecht, T. Mietzner, Three-dimensional quantitative similarity–activity relationships (3D QSiAR) from SEAL similarity matrices, J. Med. Chem. 41 (1998) 2553–2564.