

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

Balázs Jójárt^a, Tamás A. Martinek^b & Árpád Márki^{a,*}

^aDepartment of Pharmacodynamics and Biopharmacy, University of Szeged, Eötvös u. 6, H-6720, Szeged, Hungary; ^bDepartment of Pharmaceutical Chemistry, University of Szeged, Eötvös u. 6, H-6720, Szeged, Hungary

Received 29 July 2004; accepted in revised form 11 May 2005
© Springer 2005

Key words: AutoDock, benzoxazine, binding mode, binding pocket, oxytocin receptor, scoring function, 3D-QSAR

Summary

Molecular docking and 3D-QSAR studies were performed to determine the binding mode for a series of benzoxazine oxytocin antagonists taken from the literature. Structural hypotheses were generated by docking the most active molecule to the rigid receptor by means of AutoDock 3.05. The cluster analysis yielded seven possible binding conformations. These structures were refined by using constrained simulated annealing, and the further ligands were aligned in the refined receptor by molecular docking. A good correlation was found between the estimated ΔG_{bind} and the pK_i values for complex F. The Connolly-surface analysis, CoMFA and CoMSIA models ($q^2_{\text{CoMFA}} = 0.653$, $q^2_{\text{CoMSIA}} = 0.630$ and $r^2_{\text{pred,CoMFA}} = 0.852$, $r^2_{\text{pred,CoMSIA}} = 0.815$) confirmed the scoring function results. The structural features of the receptor–ligand complex and the CoMFA and CoMSIA fields are in closely connected. These results suggest that receptor–ligand complex F is the most likely binding hypothesis for the studied benzoxazine analogs.

Introduction

As a high proportion of neonatal mortalities can be attributed to premature delivery (birth before 37 completed weeks), drug treatment for the prevention of early uterine contractions is essential. The authorized medication for tocolytic purposes includes β_2 -sympathomimetics (e.g. fenoterol) and MgSO_4 .

Since the nonapeptide hormone oxytocin (OT) plays a key role during labor, other possibilities are OT antagonists. One such drug, the peptide Atosiban[®], is used in clinical practice, but its sensitivity to peptidases excludes its oral applica-

tion. In recent years, therefore, some research groups have concentrated on the development of selective orally active nonpeptide OT antagonist ligands, which can be divided into two groups: the benzodiazepine (BZD) and the benzoxazine (BZX) analogs. Of the BZD analogs, the sedatohypnotic diazepam, midazolam, nitrazepam and clonazepam exert tocolytic effects; this is related to the antagonist capacity of these compounds on the myometrial α -adrenergic receptors [1]. Another group of BZD analogs, the 5-phenyl-1,3-dihydro[e][1,4]diazepin-2-one derivatives, can inhibit the receptor binding of OT [2]. The BZX analogs were developed on the basis of the results at Otsuka [3]: OPC 21268 (Figure 1) has significant rat vasopressin 1a and 2 receptor ($V_{1a}R/V_2R$) selectivity and oral activity.

*To whom correspondence should be addressed. Phone: +36-62-545567; Fax: +36-62-545567; E-mail: marki@pharma.szote.u-szeged.hu

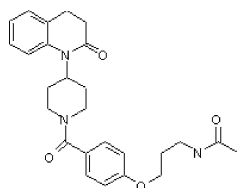


Figure 1. Structure of OPC21268.

Using human tissues, Williams et al. [4] demonstrated that OPC 21268 is a selective OT receptor (OTR) ligand, with an affinity for binding to uterine OTRs that is more than 60 times that to AVP-V_{1a}Rs in the platelets, uterus and liver. Various structural modifications of this compound have been prepared in order to increase the affinity toward human OTRs [5–9].

On the basis of the results of molecular modeling and mutagenesis studies, it has been clearly revealed that the native hormones OT and AVP and their agonist peptide analogs dock with their tocin/pressin ring completely nested in the receptor interior, while their C-terminal interacts with the first extracellular loop (Y115-V_{1a}R, D103-V₂R and F103-OTR) [10–12]. In contrast, the small nonpeptide ligands dock entirely in the hydrophobic core of the transmembrane region [13]. The residues Q2.57(92); V3.33(120), M3.36(123), T3.40(127); and A7.42(318), N7.45(321), S7.46(322) have been found to be responsible for the binding of YM087, a nonpeptide selective V_{1a}R and V₂R (but not OTR) antagonist [14]. Residues are numbered according to a consensus numbering scheme [15].

The exact binding site and binding mode of the nonpeptide benzoxazine OTR antagonists remain open questions. In the present work we have made an attempt to identify the key elements of the receptor–ligand interaction for 45 nonpeptide OTR BZX analogs taken from the literature [8, 9] (Tables 1a–f). The problem was addressed by using molecular docking incorporating receptor flexibility with constrained simulated annealing (CSA) in order to generate binding hypotheses. Quantitative structure–activity relationship analysis with scoring functions and CoMFA and CoMSIA computational techniques were deployed to identify the 3D structure of the binding pocket of the OTR and the binding conformations of the ligands, and hence to construct a 3D-QSAR model.

Table 1. The structure and measured pK_i of the OT antagonist (a–e: training set; f: test set).







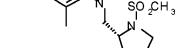
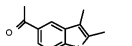
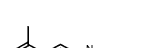
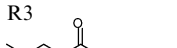
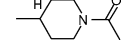
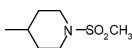
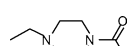
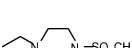

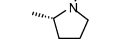
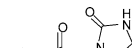
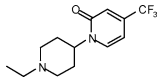
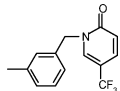
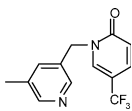
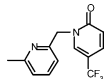









	Mol ID	R1	p <i>K</i> _i
(a)	01		7.4
	02		6.5
	03		6.4
	04		6.8
	05		6.9
	06		6.8
(b)	07	R2 	8.0
	08		6.4
	09		6.4
(c)	10	R3 	7.4
	11		7.2
	12		7.1
	13		7.1
	14		7.2
	15		7.6
	16		7.6
	17		7.6

Table 1. Continued

Mol ID	R1	p <i>K_i</i>
18		6.1
19		6.1
20		7.2
21		8.0
22	H	R4
23		7.2
24		7.8
25		7.6
26		8.2
27		8.5
28		7.8
29		8.4
30		8.6
31		8.7
32		8.2
33		8.9

(d)

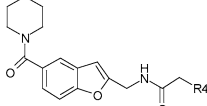
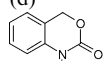


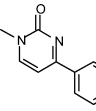
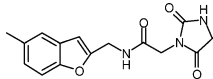
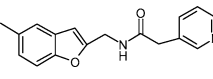
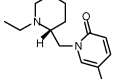
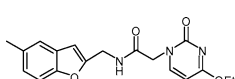
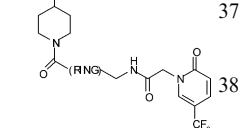


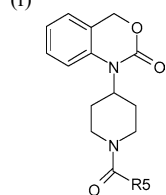
Table 1. Continued

Mol ID	R1	p <i>K_i</i>
33		8.7
34		9.3
35		9.0
36	R2	8.1
37		7.0
38		7.3
39	R5	7.2
40		6.9
41		6.7
42		8.2
43		7.6
44		8.0
45		8.7

(e)



(f)



The receptor binding studies were performed on the hOTR, using [³H]OT for displacement analysis. p*K_i*: negative logarithm of inhibition constant.

Materials and methods

Docking protocol and scoring functions

The 3D structure of the receptor model was obtained by courtesy of Arthur Gieldon et al. [14]. The model was constructed as follows: the incomplete (missing IL2/IL3 loops) δ -opiate receptor was used as template [16]. After the multiple sequence alignment, the initial V_{1a}R, V₂R and OTR models were obtained via mutations/insertions/deletions on the δ -opiate receptor. The IL2 and IL3 loops and the incomplete N and C terminals were added and the models were refined by a CSA cycle up to 1000 K *in vacuo*, with α C constrained atoms, and finally the obtained 3D structures were minimized in 2000 cycles *in vacuo*.

The most active compound, 34, was docked into the receptor by means of AutoDock 3.05 software [17]. Since the program AutoDock allows only one of the docking partners to be flexible, the receptor was kept rigid, and the ligand was allowed to be flexible. The grid box ($50 \times 70 \times 50$ Å) was centered on the putative active site, and the lattice point distance was set to 0.375 Å. Lennard-Jones parameters 12–10 and 12–6 were used for modeling H-bonds and van der Waals interactions, respectively. In calculations of the electrostatic grid map, the distance-dependent dielectric constant of Mehler and Solmayer was utilized.

During the docking procedure, the Lamarckian Genetic Algorithm with the pseudo-Solis and Wets method was used, with 250 individuals in the population. The stopping criterion was defined by the total number of energy evaluations, which was set to 2.5×10^7 . The translation step was set to 0.5 Å/step, and in both quaternion and torsion steps 5.0 degrees/step was used. The other parameters were set as previously published [18]. The docking procedure was performed 500 times and the obtained structures were subsequently subjected to cluster analysis. The ligand conformation with the lowest binding free energy was used as reference; the root-mean-square distance was calculated, and was used as descriptor for the cluster analysis.

In consequence of the rigid receptor model applied during the initial docking procedure, the resulting conformations were very rough, therefore

to remove the steric clashes in the obtained structures further refinement were performed by using the CSA protocol.

The initial structures for CSA were constructed by selecting the lowest-energy structures from the major clustered populations; all receptor residues closer than 10 Å to the ligands were retained, and the remainders of the residues were deleted from the models. The CSA protocol was set up as follows: (1) the α C atoms of the receptor residues were constrained to their initial positions; (2) 100 random structures were generated with Hoover-Nosé dynamics [19] at 900 K; (3) these structures were heated to 1000 K (5000 fs), equilibrated at 1000 K (20,000 fs), and cooled down to 50 K (10,000 fs); (4) after cooling, a final minimization was performed with the aid of the steepest descent, the conjugated gradient and finally the truncated Newton methods, in a consecutive manner, the gradients being 100, 10 and 0.001, respectively. The MMFF94 force field [20] was utilized, with a distance-dependent dielectric function ($\epsilon_r = 1$) and a 25 Å cut-off for the long-range non-bonding interactions. From the 100 receptor–ligand complexes obtained, the lowest-energy structure was selected as a binding hypothesis.

In the putative receptor–ligand complexes, the remaining ligands were superimposed with the docked, most active compound at the piperidinyl-BZX ring. The superimposing of the molecules was as follows: (1) for the piperidinyl-BZX ring, dummy atoms were placed at the positions of the heavy atoms in the most active compound; (2) a restraint of zero distance was applied between the dummy and the heavy atom; (3) finally, this structure, without the most active compound, was minimized as described previously. For the structures obtained, a final docking procedure with the previously described parameters (except for the docking runs, which were set to 200, and the energy evaluations, which were set to 250,000) was performed. With the binding free energy as descriptor, the resulting conformations were cluster-analyzed, using a root-mean-square tolerance of 0.3. From the largest population, the lowest-binding free-energy conformation was selected, and these receptor–ligand complexes were characterized via two different scoring functions: the built-in scoring function of AutoDock 3.05, and the empirical scoring function X-Score 1.2 developed by Wang et al. [21].

3D-QSAR parameters

The ligands aligned in the receptor by using AutoDock 3.05 were subjected to 3D-QSAR studies. The 45 BZX analogs were divided into two groups. The 3D-QSAR model was built up for 38 analogs (training set, Tables 1a–e), and the remaining 7 compounds were used as test set (Table 1f). For the 3D-QSAR study, the powerful and widely used methods CoMFA [22] and CoMSIA [23] were chosen. These methods are based on the assumption that changes in the binding affinities of ligands are related to changes in the molecular interaction field (MIF) [24] generated in the active conformation. In the CoMFA method, the two fields (steric and electrostatic) are calculated by using Lennard-Jones and Coulombic potentials, respectively. In CoMSIA, a distance-dependent Gaussian-type functional form has been introduced, which can avoid singularities at the atomic position and dramatic changes in potential energy for these grids in the proximity of the surface, while no arbitrary definition of cut-off limits is required. In CoMSIA, similarity is expressed in terms of five physicochemical properties: steric occupancy, partial atomic charges, local hydrophobicity, and H-bond donor and acceptor properties.

The CoMFA parameters were the default box size, i.e. the box automatically created by the software, with 2.0, 1.5 and 1.0 Å lattice point distances. For calculation of both the steric and electrostatic energy terms, an sp³ carbon atom with +1 charge, a distance-dependent dielectric constant and a 30 kcal/mol cut-off were used. In the CoMSIA method, the default box with lattice point distances of 2 Å and an attenuation factor of 0.3 was applied for the following combination of fields: steric-electrostatic, hydrophobic only, donor-acceptor, and all five fields. In order to choose the appropriate components and check the statistical significance of the models, leave-one-out (LOO) and leave-multiple-out (LMO) cross-validation schemes were used. In the LMO procedure, five groups were used, i.e. each property value was predicted with a model based on 80% of the available data. The maximum number of components (*c*) was set to 10, and appropriate column filtering was used to omit the lattice points with energy variance <2.0 kcal/mol. The optimum number of components and q^2 were determined by selecting the lowest PRESS. The final non-cross

validated prediction model was derived with the number of components obtained from the cross-validation analysis. The results were interpreted graphically by field contribution maps, using the field type “stdev × coeff”.

Hardware and software

We used AutoDock 3.05 to complete the docking procedure, and the further structure refinement with CSA was carried out with Molecular Operating Environment 2003.02 [25] on a 3.0 GHz 2× Xeon dual PC under Red Hat 9.0 with the OpenMosix operating system. One structure refinement with CSA took 16–20 h of processor time, depending on the size of the complex.

The CoMFA and CoMSIA calculations and the statistical analysis were performed on a Silicon Graphics Origin2000 with the Irix 6.4 operating system, using Sybyl 6.9.1 software [26].

Results

Molecular docking of the most active compound (34)

In order to generate multiple binding hypotheses, the docking of the most active compound (34) to the putative binding site of the OTR was performed 500 times in a rigid receptor model having the initial coordinates from homology modeling. The cluster analysis of the obtained conformations resulted in seven major populations. The lowest-energy molecules of these populations (denoted by A, B, C, D, E, F and G) were selected for further analysis (Figure 2).

The initial set of the possible receptor–ligand bioactive conformations were very crude, and required further refinement. The structural refinement of the possible 3D structures of the binding pocket was achieved by using CSA. For each receptor–ligand complex obtained from the rigid docking, 100 structures were generated, and the structures having the lowest potential energies were selected as the initial hypotheses for the receptor–ligand complex.

Estimation of ΔG_{bind} for the active series

With seven binding models available for the most active compound and the OTR, the remainders of

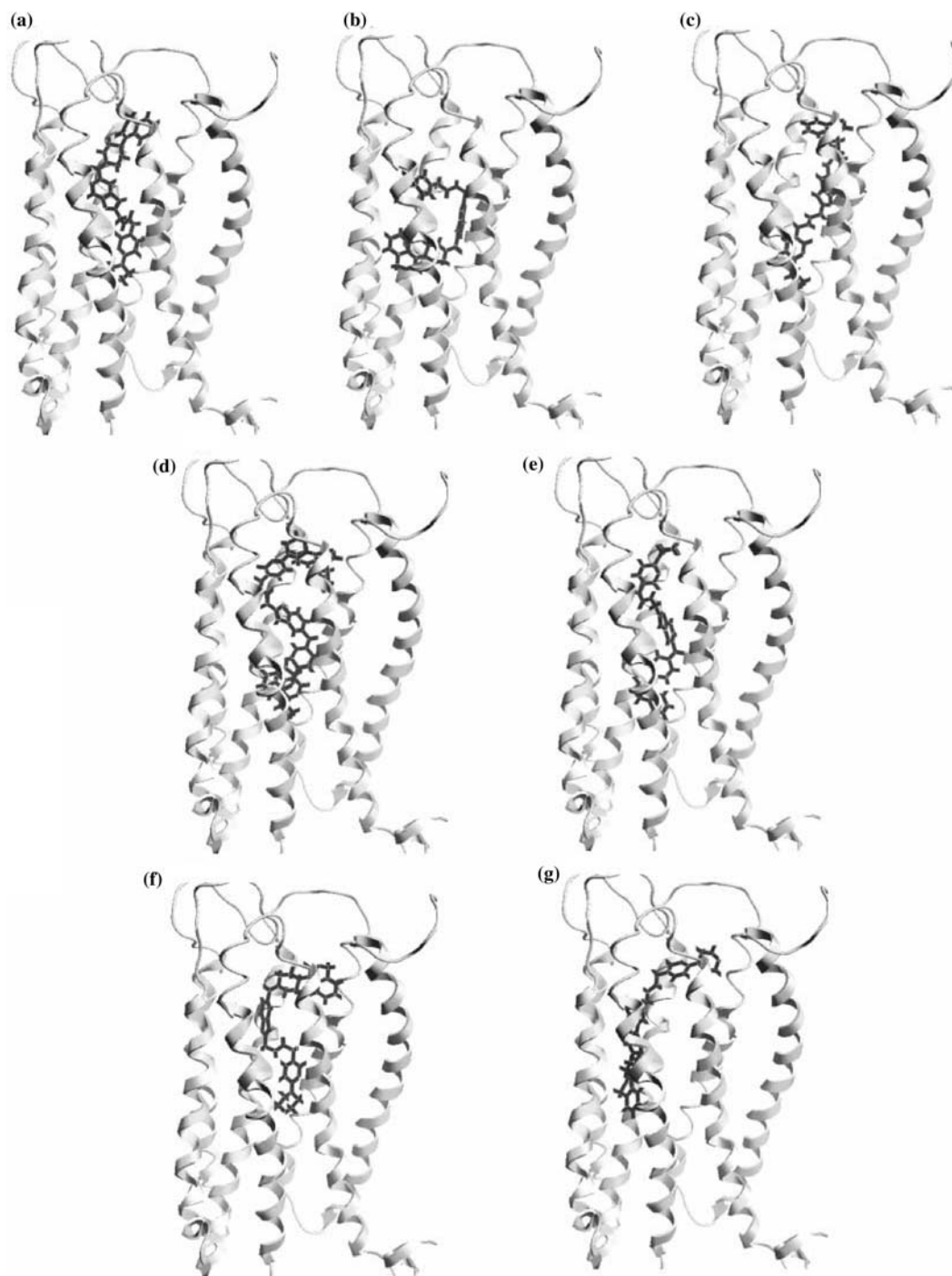


Figure 2. The input structures of the CSA protocol: (a) complex A, (b) complex B, (c) complex C, (d) complex D, (e) complex E, (f) complex F and (g) complex G.

the molecules in the series were incorporated in the calculations to find the model exhibiting the best correlation between the estimated free energies of

binding and the biological activities. The molecules were fitted to the most active ligand, subsequent rigid docking iterations were run, and

two different scoring functions were applied for the energy computations as described above.

The estimated ΔG_{bind} of complex F are presented in Table 2, and the rest of the data are available as supporting information. The binding energies exhibited negative values after the CSA receptor refinement supporting that the ligands are likely to bind into the optimized binding site. The regression coefficients (r^2) of the ΔG_{bind} with the biological activity are listed in Table 3.

With both scoring functions, the best linear correlation was obtained between the measured $\text{p}K_i$ and the estimated binding free energies for binding hypothesis F: $r_{\text{AutoDock}}^2 = 0.636$, $\text{p}K_i^{\text{AutoDock}} = -0.16 \times \Delta G_{\text{bind}}^{\text{Autodock}} + 5.31$ and $r_{\text{HSScore}}^2 = 0.445$, $\text{p}K_i^{\text{HSScore}} = -0.615 \times (\Delta G_{\text{bind}}^{\text{HSScore}} + 0.576)$ (Figure 3). The other binding models revealed rather low regression coefficients, pointing to model F being the possible binding mode.

The linear models could be improved by removing the two outlying compounds 18 and 19 from the HSScore dataset, which led to a regression coefficient of $r_{\text{HSScore}}^2 = 0.595$ ($\text{p}K_i^{\text{HSScore}} = -0.685 \times \Delta G_{\text{bind}}^{\text{HSScore}} + 0.474$). For the AutoDock dataset, the deletion of outliers 09 and 32 resulted in a regression coefficient of $r_{\text{AutoDock}}^2 = 0.702$ ($\text{p}K_i^{\text{Autodock}} = -0.159 \times \Delta G_{\text{bind}}^{\text{Autodock}} + 5.33$).

Receptor surface analysis and identification of pharmacophore points

The analytical Connolly surface [27] reveals the hydrophobic, hydrophilic and possible H-bond anchor points of the receptor around the ligand. The surface was calculated with the receptor atoms closer than 4.5 Å to the ligand, and a spacing of 0.5 Å was used (Figure 4).

The regular and the potential H-bonds and the residues responsible for the hydrophobic pockets around the ligand were determined by analyzing the best binding mode. We accepted any electrostatic interaction as a regular H-bond when the distance between the hydrogen atom and the acceptor atom ($d(\text{A} \cdots \text{H})$) was lower than 2.8 Å, and the angle formed by the acceptor, hydrogen and donor atoms ($\alpha(\text{A} \cdots \text{H} - \text{D})$) was greater than 120° (Table 4). Further inspection of the structure revealed potential H-bonds having longer distances and short angles between the pillar atoms, but whose formation could be facilitated by the dynamic nature of the receptor–ligand complexes.

Table 4 lists both the potential (denoted by asterisks) and the regular H-bonds.

The hydrophobic residues are P2.60(95), V3.28(115), I5.39(201), F6.44(284) and F7.35(311) (Figure 5).

3D-QSAR studies of the 45 BZX analogs

The cross-validated CoMFA and CoMSIA models
The molecular docking protocol successfully predicted a binding mode in the transmembrane region of the OTR, which was supported by the correlation between the estimated binding free energy and the biological activity. In order to reinforce the above results, 3D-QSAR studies were performed.

For the protein-based alignment of the compounds (Figure 6), CoMFA and CoMSIA analyses were performed by using various parameters (Tables 5–7).

The CoMFA analysis was run with three different lattice point distance setups: 2.0, 1.5 and 1.0 Å, denoted by models CoMFA_A, CoMFA_B and CoMFA_C, respectively. The maximum LOO q^2 value was obtained for complex F in each case; it was significantly different from the values calculated for the remainder of the binding hypotheses. q^2 varies within the range 0.61–0.65 (Table 5) which may indicate that the predictivity of the QSAR model is reasonably good. The highest q^2 was obtained for parameter set CoMFA_B, and this was subjected to further analysis. We constructed three other datasets, removing those outliers which were identified with the application of the scoring functions, and the outliers of the 3D-QSAR model. The model CoMFA_D was constructed by removing compounds 18 and 19, model CoMFA_E by deleting compounds 08, 18 and 19, and CoMFA_F by removing ligands 09 and 32 (AutoDock outliers). The resulting characteristics of the QSAR models are summarized in Table 6.

By means of four different field combinations, steric and electrostatic (CoMSIA_A), hydrophobic only (CoMSIA_B), H-bond donor and acceptor (CoMSIA_C), and all five fields (CoMSIA_D), CoMSIA analyses were run (Table 7). We obtained the maximum LOO q^2 for complex F, but complex D exhibited similar LOO predictivity, which prevented an unambiguous selection of the best model purely on the basis of the CoMSIA calculations.

Table 2. The values of the estimated ΔG_{bind} in kcal/mol using complex F.

Mol_id	pK _i	$\Delta G_{\text{bind}}^{\text{Autodock}}$	$\Delta G_{\text{bind}}^{\text{Hpscore}}$	$\Delta G_{\text{bind}}^{\text{Hmscore}}$	$\Delta G_{\text{bind}}^{\text{Hsscore}}$	$\Delta G_{\text{bind}}^{\text{AVEscore}}$
01	7.40	-14.30	-11.24	-11.48	-10.48	-11.07
02	6.50	-3.06	-9.85	-10.18	-9.19	-9.74
03	6.40	-6.52	-10.30	-10.63	-9.73	-10.22
04	6.80	-6.61	-10.28	-10.76	-9.40	-10.15
05	7.70	-11.80	-11.42	-11.24	-10.83	-11.16
06	6.80	-13.40	-11.55	-12.02	-11.20	-11.58
07	8.00	-13.00	-11.72	-11.80	-10.86	-11.45
08	6.40	-4.90	-11.14	-10.97	-10.60	-10.91
09	6.40	-15.00	-11.80	-11.61	-11.01	-11.47
10	7.40	-14.00	-11.48	-11.65	-11.21	-11.45
11	7.20	-13.80	-11.84	-12.10	-11.40	-11.78
12	7.10	-13.40	-11.63	-11.87	-10.71	-11.40
13	7.10	-14.10	-11.51	-11.92	-11.18	-11.54
14	7.20	-10.40	-10.93	-11.24	-10.18	-10.78
15	7.60	-13.30	-11.61	-12.06	-11.36	-11.68
16	7.60	-16.10	-12.07	-12.40	-11.35	-11.94
17	7.60	-16.70	-12.90	-13.04	-12.33	-12.76
18	6.10	-7.17	-11.89	-11.94	-11.39	-11.74
19	6.10	-7.31	-13.20	-13.42	-11.96	-12.86
20	7.20	-14.30	-13.11	-13.37	-12.07	-12.85
21	8.00	-16.30	-12.40	-12.39	-11.80	-12.19
22	7.20	-14.40	-11.35	-11.77	-10.91	-11.34
23	7.80	-15.70	-11.80	-12.06	-11.50	-11.78
24	7.60	-17.30	-12.22	-12.56	-11.87	-12.22
25	8.20	-16.80	-12.32	-12.43	-11.89	-12.21
26	8.50	-17.10	-12.22	-12.49	-11.55	-12.09
27	7.80	-16.60	-12.18	-12.88	-12.04	-12.36
28	8.40	-17.80	-12.37	-12.60	-11.88	-12.29
29	8.60	-16.60	-12.17	-12.33	-11.61	-12.03
30	8.70	-17.50	-12.71	-12.84	-12.17	-12.57
31	8.20	-18.20	-13.27	-13.26	-12.67	-13.07
32	8.90	-15.80	-12.60	-12.97	-12.13	-12.57
33	8.70	-18.00	-12.70	-12.85	-12.15	-12.57
34	9.30	-21.30	-13.26	-13.60	-12.62	-13.16
35	9.00	-19.20	-13.67	-13.27	-13.15	-13.36
36	8.10	-16.10	-12.49	-12.49	-11.92	-12.31
37	7.00	-9.70	-11.51	-11.65	-10.84	-11.34
38	7.30	-10.90	-11.98	-12.06	-11.13	-11.72
39	7.20	-14.20	-11.29	-11.31	-10.80	-11.14
40	6.90	-14.40	-11.58	-11.59	-10.79	-11.32
41	6.70	-14.50	-10.98	-11.28	-10.31	-10.86
42	8.20	-15.90	-11.96	-12.25	-11.32	-11.84
43	7.60	-16.80	-12.40	-12.73	-11.74	-12.29
44	8.00	-17.60	-13.31	-13.53	-12.74	-13.19
45	8.40	-18.00	-12.70	-12.93	-12.09	-12.57

All the field combinations provided acceptable LOO predictivities. In order to retain the highest amount of chemical information for back-projec-

tion purposes, we used the all-fields model (CoM-SIA_D) for further analysis despite its somewhat lower q^2 value. In a search for possible outliers,

Table 3. Correlation coefficients using the built-in scoring function of AutoDock 3.05, and X-Score 1.2.

Complex ID	r^2				
	$\Delta G_{\text{bind}}^{\text{Autodock}}$	$\Delta G_{\text{bind}}^{\text{Hpscore}}$	$\Delta G_{\text{bind}}^{\text{Hmscore}}$	$\Delta G_{\text{bind}}^{\text{Hsscore}}$	$\Delta G_{\text{bind}}^{\text{AVEScore}}$
A	0.248	0.251	0.221	0.305	0.239
B	0.105	0.105	0.082	0.138	0.111
C	0.083	0.199	0.223	0.176	0.204
D	0.376	0.164	0.103	0.198	0.156
E	0.388	0.112	0.087	0.225	0.150
F	0.635	0.382	0.385	0.450	0.414
G	0.035	0.066	0.071	0.090	0.077

HMScore, HPscore, and HSScore are appropriate for three different hydrophobic functions; AVEScore was calculated as (Hmscore + Hpscore + HSScore)/3 (see Ref. [21]).

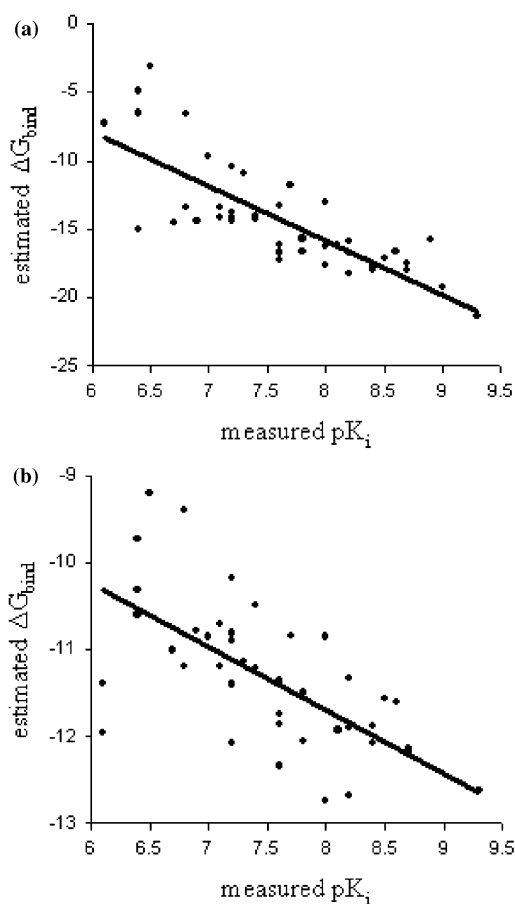


Figure 3. Plot of measured pK_i versus $\Delta G_{\text{bind}}^{\text{Autodock}}$ (a) and $\Delta G_{\text{bind}}^{\text{Hsscore}}$ (b).

further datasets were created. On removal of compounds 18 and 19 (CoMSIA_E), which were the pK_i outliers from the CoMSIA_D models too,

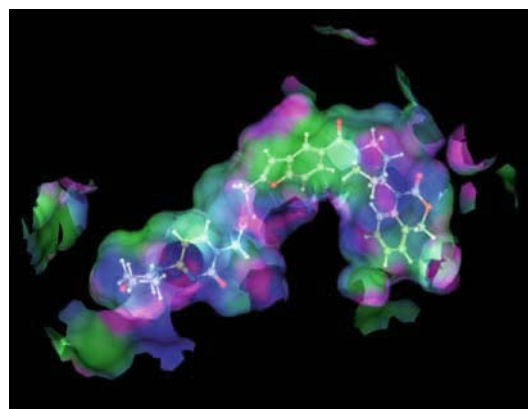


Figure 4. Analytical Connolly surface of the receptor in receptor–ligand complex F. The green regions are hydrophobic, and the blue ones are hydrophilic, while the purple regions are those through which H-bonds are likely to form.

the parameters changed as follows: $q^2 = 0.630$, $n = 2$, PRESS = 0.481. On removal of compounds 09 and 32 (CoMSIA_F), the q^2 values decreased significantly ($q^2 = 0.462$, $n = 1$, PRESS = 0.599).

Prediction for external dataset by 3D-QSAR models.

To test the predictive power of the models CoMFA_B, CoMFA_D, CoMFA_E, and CoMSIA_E, we determined the biological activity of the test set (Table 1f), which was not included in the model building. The regression coefficients between the measured and predicted pK_i values were changed as follows: 0.848, 0.709 and 0.709 and 0.815 for the CoMFA_B, CoMFA_D, CoMFA_E and CoMSIA_E models, respectively. The measured and the predicted pK_i values are given in Table 8, and the measured versus predicted pK_i

Table 4. Parameters of H-bonds between ligand and receptor in complex F.

Receptor–Ligand	d (A···H) (Å)	α (A···H–D) (°)
N7.49(325)_HD21–L_O78	1.54	165.5
N7.49(325)_HD21–L_O71*	3.17	44.6
S7.46(322)_HG–L_O65	2.72	136.7
S7.46(322)_HG–L_N61	2.27	136.8
Q3.32(119)_CO–L_H24	1.75	152.4
Q3.32(119)_HE21–L_O26*	2.97	90.4
Q3.32(119)_HE22–L_O26*	2.79	100.8
W6.48(288)_HE1–L_O30*	3.25	95.6

d (A···H) is the distance between the hydrogen atom and the acceptor atom, α (A···H–D) is the angle formed by the acceptor, hydrogen and donor atoms. L denotes the ligand. The potential H-bonds not meeting the detection criteria are indicated by asterisks.

values for CoMFA_B and CoMSIA_E are depicted in Figure 7.

The non-cross validated parameters of the CoMFA_B and CoMSIA_E QSAR models are shown in Tables 9 and 10, and the predicted versus measured pK_i values are depicted in Figure 8. The contour maps of the CoMFA_B and CoMSIA_E models are depicted in Figures 9–13.

Testing the robustness of models and construction of final models

LMO cross-validation was performed for the two 3D-QSAR models. The values obtained were

$q^2 = 0.592 \pm 0.007$, $c = 2.71 \pm 0.998$ PRESS = 0.544 ± 0.041 for the CoMFA_B model, and $q^2 = 0.589 \pm 0.045$, $c = 2 \pm 0.133$ PRESS = 0.504 ± 0.025 for the CoMSIA_E model.

Finally, we constructed the CoMFA and CoMSIA models with all 45 compounds. The resulting models had a predictive power of $q^2 = 0.694$, $c = 6$, PRESS = 0.466, and $r^2 = 0.997$, $s = 0.047$, $F(6,38) = 2390.257$, and $q^2 = 0.602$, $c = 2$, PRESS = 0.535, and $r^2 = 0.81$, $s = 0.358$, $F(2,42) = 89.354$.

Discussion

Without prior knowledge of the binding mode of compound 34, we generated 500 structures at the putative active site of the OTR. Cluster analysis of the obtained conformations identified seven active conformation and binding mode hypotheses. The docking procedure performed on the rigid receptor model having the initial atomic coordinates of the homology model led to two consequences: (1) the evolved structures are very crude; (2) the 3D structure of the receptor could not be changed, and therefore the conformation of the binding pocket and the ligand could not mutually accommodate. To solve these problems, we used CSA for the seven complexes to sample the conformational space, and, due to the complexity of the system, we used a truncated receptor model.

In order to identify the most likely active receptor–ligand complex, we applied two scoring functions to estimate the free energy of binding for the studied series of ligands. The negative binding energies obtained for all the scoring functions

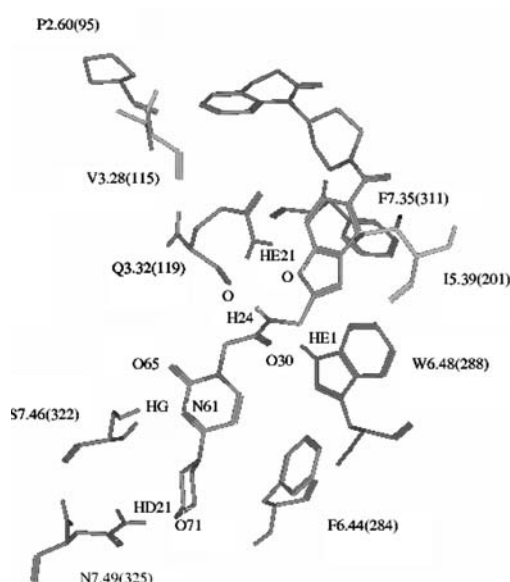


Figure 5. Electrostatic, H-bond and hydrophobic contacts between receptor and ligand in complex F.



Figure 6. As an example, the protein-based alignment of the 45 BZX analogs in receptor complex F.

(Table 2) indicate, that the receptor–ligand complex refinement was appropriate. For complex F, acceptably predictive relationships for both scoring functions ($r^2 = 0.450$, and $r^2 = 0.635$, for HScore and AutoDock, respectively) were generated between the measured pK_i and the estimated ΔG_{bind} . The degree of goodness of the models was increased significantly by removing certain outliers. These results suggest that complex F could be the active conformation.

The calculated analytical Connolly surface is in accordance with the structural features of the ligand. There are nonpolar residues (P2.60(95), V3.28(115), I5.39(201), F6.44(284) and F7.35

(311)) in the proximity of the hydrophobic parts of the ligand, and accordingly hydrophobic interactions can be formed advantageously. This finding is in good agreement with the fact that small molecular weight ligands bind to sites within the hydrophobic core formed by the transmembrane α -helices [28]. Further important interactions are the H-bonds (N7.49(325)_HD21-L_O71, S7.46(322)_HG-L_O65, S7.46(322)_HG-L_N61 and Q3.32(119)_CO-L_H24), which play key roles during the binding. Although no regular H-bonds were observed between W6.48(288)_HE1 and L_O30, or Q3.32(119)_HE21/22 and L_O26, this can change during the binding because of the dynamic nature of the receptor–ligand complexes. From these results, the explanation of the low biological activity ($pK_i = 6.4$) of 08 as an example is clear. The hydrophobic part of the molecule, which can interact with F6.44(284), is missing and it can not form H-bonds (or electrostatic interactions) with the residues N7.49(325), S7.46(322) and Q3.32(119). The above structural observations strongly suggest that complex F as a likely binding conformation forms a stable complex due to

Table 6. QSAR model parameters of complex F obtained for various datasets.

QSAR model	q^2	c	PRESS
CoMFA_D	0.660	3	0.468
CoMFA_E	0.673	3	0.448
CoMFA_F	0.513	1	0.562

For the removed outliers, see the text.

Table 5. Parameters (q^2 = cross-validated r^2 , c = number of components, PRESS = standard error of prediction) of the three CoMFA models.

Complex ID	CoMFA_A			CoMFA_B			CoMFA_C		
	q^2	c	PRESS	q^2	c	PRESS	q^2	c	PRESS
A	0.360	3	0.688	0.339	3	0.699	0.260	1	0.719
B	0.042	1	0.818	0.046	1	0.817	0.049	1	0.815
C	0.385	4	0.685	0.396	4	0.679	0.389	4	0.683
D	0.385	1	0.656	0.366	1	0.665	0.404	1	0.646
E	0.257	1	0.721	0.223	1	0.737	0.245	1	.0727
F	0.611	2	0.529	0.653	5	0.522	0.622	3	0.529
G	0.328	2	0.699	0.394	3	0.670	0.357	2	0.680

CoMFA_A – lattice point distance (d) 2.0 Å, CoMFA_B – lattice point distance (d) 1.5 Å, CoMFA_C – lattice point distance (d) 1.0 Å. For calculation of both the steric and electrostatic energy terms, an sp^3 carbon atom with charge of +1, a distance-dependent dielectric constant and a 30 kcal/mol cut-off were used.

Table 7. The parameters (q^2 = cross validated r^2 , c = number of components, PRESS = standard error of prediction) of the four CoMSIA models.

Complex ID	CoMSIA_A			CoMSIA_B			CoMSIA_C			CoMSIA_D		
	q^2	C	PRESS	q^2	c	PRESS	q^2	c	PRESS	q^2	c	PRESS
A	0.263	1	0.718	0.215	1	0.741	0.263	1	0.718	0.299	1	.0700
B	0.120	1	0.784	0.287	1	0.706	0.248	1	0.725	0.289	1	0.765
C	0.120	2	0.795	0.285	1	0.707	0.276	2	0.705	0.281	1	0.709
D	0.580	5	0.575	0.365	1	0.666	0.401	3	0.666	0.506	4	0.614
E	0.341	3	0.698	0.299	1	0.700	0.334	2	0.692	0.398	2	0.658
F	0.589	4	0.560	0.399	1	0.648	0.592	2	0.542	0.568	2	0.557
G	0.242	1	0.728	0.327	1	0.686	0.372	2	0.672	0.412	1	0.641

CoMSIA_A = using the steric and electrostatic fields, CoMSIA_B = using only the hydrophobic field, CoMSIA_C = using H-bond donor and acceptor fields, CoMSIA_D = using all five fields. The attenuation factor was set to 0.3, and the lattice point distance to 2.0 Å.

Table 8. Predicted pK_i values for seven compounds.

Mol ID	Measured pK_i	CoMFA_B predicted pK_i	CoMFA_D predicted pK_i	CoMFA_E predicted pK_i	CoMSIA_E predicted pK_i
39	7.2	7.53	7.63	7.66	7.37
40	6.9	7.50	7.52	7.52	7.16
41	6.7	7.36	7.60	7.69	7.26
42	7.6	8.09	8.14	8.13	8.05
43	8.2	7.91	7.82	7.86	8.12
44	8.0	8.18	8.07	8.12	7.61
45	8.7	8.79	8.84	8.83	8.48

hydrophobic and electrostatic (H-bond) interactions.

On the basis of the assumption that changes in the binding affinities of ligands are related to changes in the molecular interaction field generated in the active conformation, the protein-based alignments of the compounds were subjected to CoMFA and CoMSIA analyses. By means of 3D-QSAR methods, we obtained several models with acceptably good predictive power for complex F, which indicate that this complex is the most likely active conformation.

We removed some outliers of the training sets, and the results obtained allowed the following conclusions. On removal of the outliers 09 and 32, the predictive power of the models decreased significantly (q^2 = 0.513 and 0.462), indicating that the chemical diversity represented by these outliers is necessary. When other (08, 18 and 19) outliers were removed the q^2 values were increased, but there were no significant differences between

the values obtained: 0.653, 0.660, 0.673 and 0.630 for the CoMFA_B, CoMFA_D, CoMFA_E and CoMSIA_E models, respectively. Therefore, we constructed the non-cross validated models for each and these non-cross validated models were used to predict the biological activities of seven compounds which were not included in the model building. The best regression coefficients, $r^2_{\text{pred}} = 0.852$ and $r^2_{\text{pred}} = 0.815$, were obtained for CoMFA_B and CoMSIA_E, respectively, although the other two r^2_{pred} (0.709 for both CoMFA_D, and CoMFA_E) were also very good.

In the evaluation of the 3D-QSAR model, we applied an LMO cross-validation scheme for the training set, using the CoMFA_B, and CoMSIA_E models. The results confirmed that the model is likely to be of certain predictive value when extended to other related compounds. From these results, new CoMFA and CoMSIA models were elaborated, including all 45 compounds. The good results of this model validated the protein-

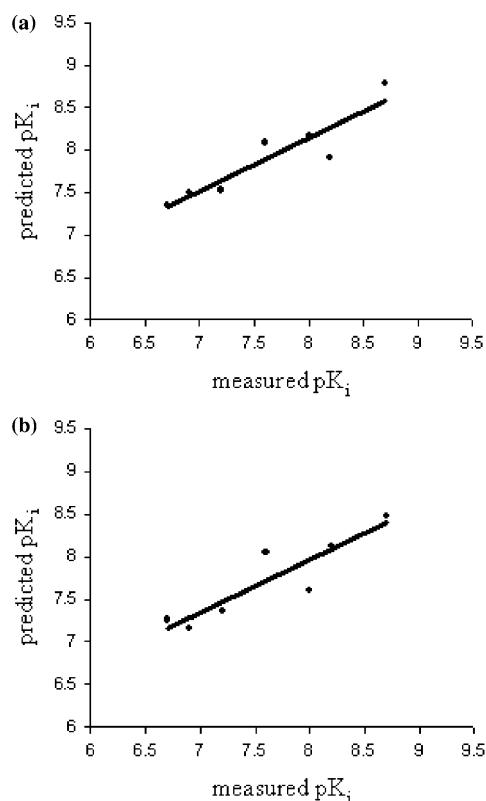


Figure 7. Plot of measured pK_i versus predicted pK_i for the test set (Table 1f) using the CoMFA_B (a) and CoMSIA_E (b) non-cross validated models.

Table 9. Final non-cross validated CoMFA model parameters using all compounds of the training set.

Complex F				
CoMFA_B			Field	Fraction (norm. coeff.)
	r^2	0.991	Steric	0.489 (2.057)
	s	0.082	Electrostatic	0.511 (2.150)
	$F(5,32)$	741.849		

$d = 1.5 \text{ \AA}$. (r^2 = conventional r^2 , s = standard error of estimate, F = Fisher value).

based alignment of the OT antagonists, and strongly supported the hypothesis that complex F is the most likely active conformation.

The contour maps of the non-cross validated models of CoMFA_B, and CoMSIA_E allowed the following conclusions: (1) Figures 9 and 11 depicts compound 08, with low biological activity $pK_i = 6.4$. The variable region of the ligand fills

Table 10. Parameters (r^2 = conventional r^2 , s = standard error of estimate, F = Fisher value) of non-cross validated model without outliers 18 and 19, using 2 components.

Complex F				
CoMSIA_E			Field	Fraction (norm. coeff.)
	r^2	0.849	Steric	0.112 (0.254)
	s	0.305	Electrostatic	0.208 (0.471)
	$F(2,32)$	96.655	Hydrophobe	0.175 (0.396)
			H-bond donor	0.267 (0.604)
			H-bond acceptor	0.238 (0.538)

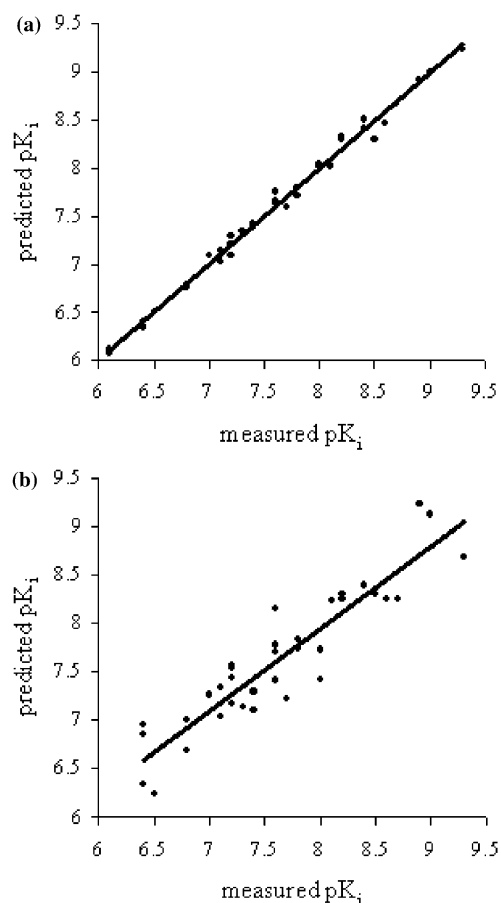


Figure 8. Plot of measured pK_i versus, predicted pK_i from final non-cross validated model, using CoMFA_B and CoMSIA_E models.

the sterically disfavored (represented by the yellow contour) space, and thus the methyl group of the ligand comes into van der Waals contact with αC of M7.39(315) (see the yellow dashed line). This

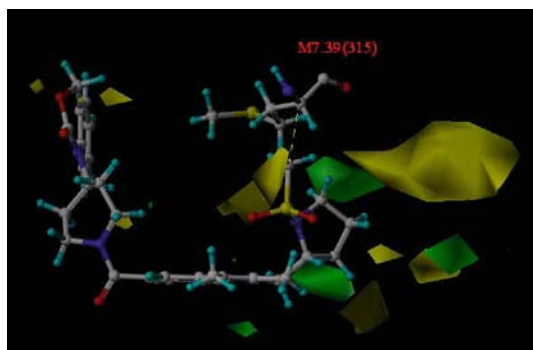


Figure 9. Contour plot of CoMFA_B steric fields (stdev \times coeff). Larger values of pK_i correlated with more bulk near green, and with less bulk near yellow. Compound 08 is shown as reference molecule.

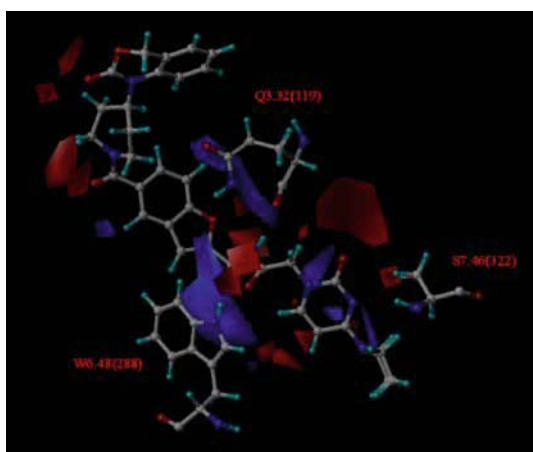


Figure 10. Contour plot of CoMFA_B electrostatic fields (stdev \times coeff). Larger values of pK_i correlated with more positive charge near blue, and with more negative charge near red. The most active compound 34 is shown as reference molecule.

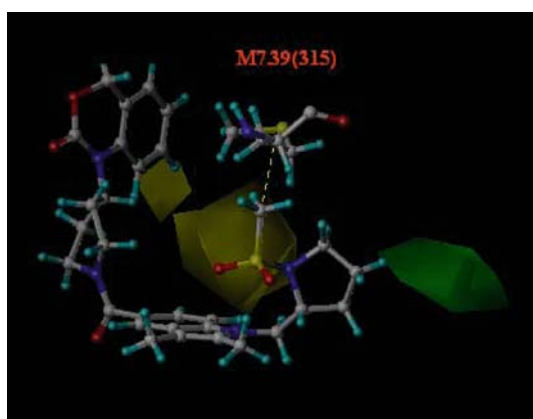


Figure 11. Contour plot of CoMSIA_E steric fields (stdev \times coeff). Favorable steric areas with more bulk are indicated in green, and disfavored steric areas by yellow isopleths. Compound 08 is shown as reference molecule.

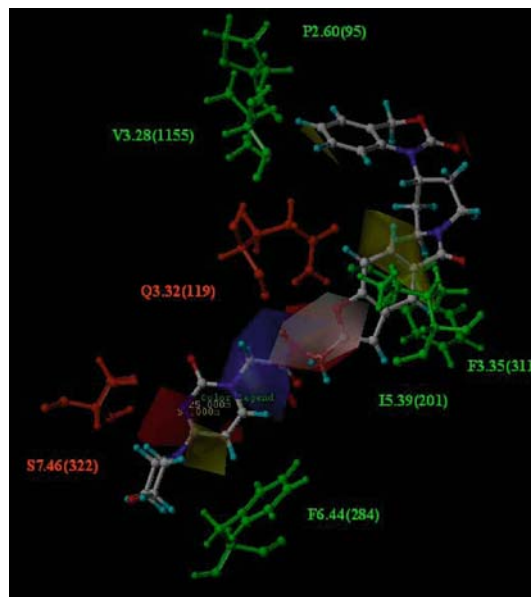


Figure 12. Contour plot of CoMSIA_E electrostatic and hydrophobic fields (stdev \times coeff). Favorable electrostatic areas with more positive charge are indicated in blue, and those with more negative charge by red isopleths. Favorable hydrophobic sites are shown in yellow, and disfavored ones by white isopleths. Compound 34 is shown as reference molecule.

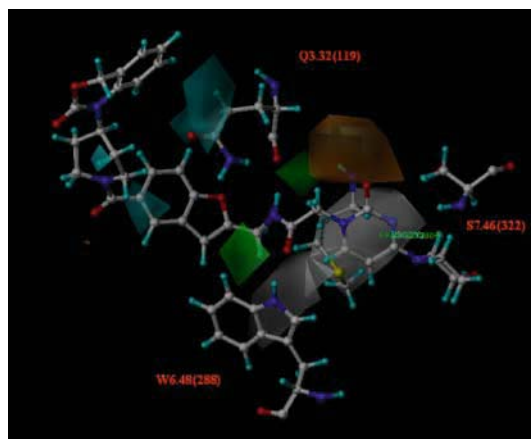


Figure 13. Contour plot of CoMSIA_E H-bond donor and acceptor fields (stdev \times coeff). Green isopleths contour maps beyond the ligands, where H-bond acceptor groups in the receptor will be favorable, and white shows those unfavorable for biological activity. Orange isopleths contour maps beyond the ligands, where H-bond donor groups in the receptor will be favorable, and cyan denotes those unfavorable for biological activity. Compound 34 is shown as reference molecule.

observation contributes to the explanation of the low biological activity of compound 08; (2) there are red and blue isopleths near the residue W6.48(288) (Figure 10), which indicates that both positive and negative charge would increase the biological activity near W6.48(288). The explanation of this conflict is that the position of the amide group is not constant in the active conformation along the series of molecules. In certain cases, the CO points toward the W6.48(288)_HE1, thereby forming an electrostatic interaction, while in other molecules, the amide group is translated and rotated, which facilitates other advantageous interactions. This also holds for the CoMSIA H-bond donor and acceptor fields (Figure 13).

It is interesting to note that the weight of the hydrophobic field in the final CoMSIA_E model is relatively low, while the ligands bind into the hydrophobic transmembrane region of the receptor. The explanation for this apparent controversy may be that the hydrophobic parts of the compounds are conserved in the studied series, and therefore the variance in the hydrophobic properties is rather low. This finding, however, does not rule out the possibility that the hydrophobic interactions may make a constant and high contribution to the binding energy.

Summary

The major goal of this study was to determine the 3D structure of the receptor–ligand complex formed between the OTR and a BZX antagonist. The results can be summarized as follows: (1) using AutoDock 3.05 and CSA, 1200 possible active site conformations were generated; (2) with the help of two scoring functions and the analyses of the receptor surface, we could identify a probable conformation: the lowest-energy structures of complex F; (3) we determined the possible pharmacophore points in the receptor–ligand complex: the important hydrophobic residues are P2.60(95), V3.28(115), I5.39(201), F6.44(284) and F7.35(311); for the H-bonds: Q3.32(119), S7.46(322) and N7.49(325); and for the electrostatic interactions: Q3.32(119) and W6.48(288); (4) with the CoMFA, and CoMSIA methods, using the protein-based alignment of the ligands, 3D-QSAR models were constructed with a predictive power of $q^2 = 0.653$ and $q^2 = 0.630$ for the

training set, and $r^2 = 0.852$, $r^2 = 0.815$ for CoMFA_B and CoMSIA_E models, respectively, for the test set; (5) the positions of the contour plots of both 3D-QSAR models were in good agreement with the previously determined interaction sites. On the basis of these results, we chose complex F as the most likely conformation of the receptor–ligand complex.

Here we note that the approximate receptor model applied here is a homology model adopted from opioid receptor, and might raise concerns about the accuracy of the docking results. The good correlation between the binding energy and the experimental bioactivity for complex F, however, supports the view that the resulting model has certain predictive ability. Nevertheless, this interaction model could be used as a good working hypothesis for experimental testing using the site-directed mutagenesis method, and in the knowledge of the 3D structures of the binding pockets, it is possible to design new compounds via *de novo* drug design methods.

Acknowledgements

We especially thank A. Gieldon and his group for the 3D structure of the OTR. The help of András Mikó is gratefully acknowledged for the code in Java, which was used in the evaluation of the docked structures. Computational support (Sybyl 6.9) was provided by the Hungarian Academy of Sciences Biological Research Center, Szeged (OTKA M 045378).

References

1. Zupkó, I., Jánossy, K., Maul, K., Márki, Á. and Falkay, G., *Life Sci.*, 72 (2003) 1093.
2. Wyatt, P.G., Allen, M.J., Chilcott, J., Hickin, G., Miller, N.D. and Woollard, P.M., *Bioorg. Med. Chem. Lett.*, 11 (2001) 1301.
3. Yamamura, Y., Ogawa, H., Chihara, T., Kondo, K., Onogawa, T., Nakamura, S., Mori, T., Tominaga, M. and Yabuuchi, Y., *Science*, 252 (1991) 572.
4. Williams, P.D., Clineschmidt, B.V., Erb, J.M., Freidinger, R.M., Guidotti, M.T., Lis, E.V., Pawluczyk, J.M., Pettibone, D.J., Reiss, D.R., Veber, D.F. and Woyden, C.J., *J. Med. Chem.*, 38 (1995) 4634.
5. Bell, I.B., Erb, J.M., Freidinger, R.M., Gallicchio, S.N., Guare, J.P., Guidotti, M.T., Halpin, R.A., Hobbs, D.W., Homnick, C.F., Kuo, M.S., Lis, E.V., Mathre, D.J., Michelson, S.R., Pawluczyk, J.M., Pettibone, D.J., Reiss,

- D.R., Vickers, S., Williams, P.D. and Woyden, S., *J. Med. Chem.*, 41 (1998) 2146.
6. Kuo, M.S., Bock, M.G., Freidinger, R.M., Guidotti, M.T., Lis, E.V., Pawluczyk, J.M., Perlow, D.S., Pettibone, D.J., Quigley, A.G., Reiss, D.R., Williams, P.D. and Woyden, C.J., *Bioorg. Med. Chem. Lett.*, 8 (1998) 3081.
7. Williams, P.D., Bock, M.G., Evans, B.E., Freidinger, R.M., Gallichio, S.N., Guidotti, M.T., Jacobson, M.A., Kuo, M.S., Levy, M.R., Lis, E.V., Michelson, S.R., Pawluczyk, J.M., Perlow, D.S., Pettibone, D.J., Quigley, A.G., Reiss, D.R., Salvatore, C., Stauffer, K.J. and Woyden, C.J., *Bioorg. Med. Chem. Lett.*, 9 (1999) 1311.
8. Wyatt, P.G., Allen, M.J., Chilcott, J., Foster, A., Livermore, D.G., Mordaunt, J.E., Scicinski, J. and Woolard, P., *Bioorg. Med. Chem. Lett.*, 12 (2002) 1399.
9. Wyatt, P.G., Allen, M.J., Chilcott, J., Gardner, C.J., Livermore, D.G., Mordaunt, J.E., Nerozzi, F., Patel, M., Perren, M.J., Weingarten, G.G., Shabbir, S., Wollard, P.M. and Zhou, P., *Bioorg. Med. Chem. Lett.*, 12 (2002) 1405.
10. Mouillac, B., Chini, B., Balestre, M.-N., Elands, J., Trumpp-Kallmeyer, S., Hoflack, J., Hibert, M., Jard, S. and Barberis, C., *J. Biol. Chem.*, 43 (1995) 25771.
11. Hibert, M., Hoflack, J., Trumpp-Kallmeyer, S., Mouillac, B., Chini, B., Mahe, E., Cotte, N., Jard, S., Manning, M. and Barberis, C., *J. Recept. Signal Transduct. Res.*, 19 (1999) 589.
12. Chini, B., Mouillac, B., Ala, Y., Balestre, M.N., Trumpp-Kallmeyer, S., Hoflack, J., Elands, J., Hibert, M., Manning, M. and Jard, S., *EMBO J.*, 14 (1995) 2176.
13. Thibonnier, M., Coles, P., Conarty, D.M., Plesnicher, C.L. and Shoham, M., *J. Pharmacol. Exp. Ther.*, 294 (2000) 195.
14. Geldon, A., Kazmierkiewicz, R., Slusarz, R. and Ciarkowski, J., *J. Comput Aided Mol. Des.*, 15 (2001) 1085.
15. Ballesteros, J.A. and Weinstein, H., *Methods Neurosci.*, 25 (1995) 366.
16. Pogozheva, I.D., Lomize, A.L. and Mosberg, H.I., *Biophys. J.*, 75 (1998) 612.
17. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., *J. Comput. Chem.*, 19 (1998) 1639.
18. Hetenyi, C. and van derSpoel, D., *Protein Sci.*, 11 (2002) 1729.
19. Nosé, S., *Mol. Phys.*, 81 (1984) 511.
20. Halgren, T.A., *J. Comput. Chem.*, 17 (1996) 490.
21. Wang, R., Lai, L. and Wang, S., *J. Comput Aided Mol. Des.*, 16 (2002) 11.
22. Kramer, R.D. III, Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
23. Klebe, G., Abraham, U. and Mietzner, T., *J. Med. Chem.*, 37 (1994) 4130.
24. Goodford, P.J., *J. Med. Chem.*, 28 (1985) 849.
25. Molecular Operating Environment (Version 2003.02), Chemical Computing Group Inc.
26. SYBYL Molecular Modeling System (Version 6.9), Tripos Ass., St. Louis, MO.
27. Connolly, M.L., *J. Appl. Crystallogr.*, 16 (1983) 548.
28. Gether, U. and Kobilka, B.K., *J. Biol. Chem.*, 273 (1998) 17979.