Journal of Molecular Graphics and Modelling 27 (2008) 409–420

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

3D QSAR studies of 1,3,4-benzotriazepine derivatives as CCK_2 receptor antagonists

Kirandeep Kaur, Tanaji T. Talele *

Department of Pharmaceutical Sciences, College of Pharmacy and Allied Health Professions, St. John's University, Jamaica, NY 11439, United States

ARTICLE INFO

Article history:
Received 14 February 2008
Received in revised form 15 July 2008
Accepted 19 July 2008
Available online 30 July 2008

Keywords:
3D QSAR
CoMFA
CoMSIA
HQSAR
CCK_2 receptor
1,3,4-Benzotriazepine

ABSTRACT

A number of CCK_2 antagonists have been reported to play an important role in controlling gastric acid-related conditions, nervous system related disorders and certain types of cancer. To obtain the helpful information for designing potent antagonists with novel structures and to investigate the quantitative structure–activity relationship of a group of 62 different CCK_2 receptor antagonists with varying structures and potencies, CoMFA, CoMSIA, and HQSAR studies were carried out on a series of 1,3,4-benzotriazepine-based CCK_2 receptor antagonists. QSAR models were derived from a training set of 49 compounds. By applying leave-one-out (LOO) cross-validation study, cross-validated (r^2_{cv}) values of 0.673 and 0.608 and non-cross-validated (r^2_{ncv}) values of 0.966 and 0.969 were obtained for the CoMFA and CoMSIA models, respectively. The predictive ability of the CoMFA and CoMSIA models was determined using a test set of 13 compounds, which gave predictive correlation coefficients (r^2_{pred}) of 0.793 and 0.786, respectively. HQSAR was also carried out as a complementary study, and the best HQSAR model was generated using atoms, bonds, hydrogen atoms, and chirality as fragment distinction with fragment size (2–5) and six components showing r^2_{cv} and r^2_{ncv} values of 0.744 and 0.918, respectively. CoMFA steric and electrostatic, CoMSIA hydrophobic and hydrogen bond acceptor fields, and HQSAR atomic contribution maps were used to analyze the structural features of the datasets that govern their antagonistic potency.

1. Introduction

Cholecystokinin (CCK), gastrin, and related peptides include a family of peptide hormones and neuropeptides which exert a variety of physiological actions on the gastrointestinal tract and the central nervous system (CNS) [1,2]. Although both CCK and gastrin share an identical carboxyl-terminal pentapeptide sequence they differ in selectivity for the two known CCK receptor subtypes, the CCK_1 (CCK_1R) and the CCK_2 (CCK_2R) receptors as a result of tyrosine sulfation at the seventh position (CCK) or at the sixth position (gastrin) from the carboxyl terminus. Furthermore, while both receptors recognize sulfated CCK with comparable high affinity, the CCK_2R has high affinity for both sulfated and non-sulfated gastrin [3,4].

CCK_1R and CCK_2R are seven-transmembrane spanning receptors that belong to the superfamily of G-protein-coupled receptors (GPCR's) and have 50% homology [5–7]. The gastrointestinal polypeptide hormone gastrin has been known to stimulate gastric acid secretion and gastrointestinal cell growth in peripheral tissues. Hence, efforts have been made to develop agents that inhibit gastrin activity especially on gastric secretion, by acting as CCK_2R antagonists [8]. These compounds have been found to be beneficial in the treatment of gastric acid-related disorders, such as gastro esophageal reflux disease (GERD), proton pump inhibitor (PPI)-evoked rebound acid hypersecretion [9], and in certain types of cancer [10–12].

Furthermore, since activation of CCK_2R in the CNS by CCK led to the mediation of pain, panic, and anxiety, it is also possible that CCK_2 antagonists may have a role in controlling these disorders [13,14]. Prominent actions mediated by CCK_2R's within the digestive tract include stimulation of acid secretion from gastric parietal cells [15,16], the release of histamine from enterochromaffin-like cells [17], and the regulation of gut motor functions [18,19].

The development of selective CCK_2R antagonists encompassing a diverse range of chemical structures have been accomplished by one of two main approaches [20]. The peptoid-based compound (CI-988) [21] and the indole derivative (JB93182) [22] are the most significant examples to stem from using the native peptide hormone as the starting point but, unfortunately, these compounds exhibited low oral potency. Initial efforts aimed at creating chiral 1,4- and 1,5-benzodiazepine-based CCK_2 antagonists led to

* Corresponding author.
E-mail address: talelet@stjohns.edu (T.T. Talele).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.07.003

the development of compound (L-365,260) [23], a 1,4-benzodiazepine, whose biological testing has progressed to human studies that showed its capability of reversing the anxiogenic effects produced by tetragastrin [24], but its ineffectiveness in patients with panic disorder [25]. Other 1,4-benzodiazepine-based CCK_2 antagonists based on the paradigm of compound L-365,260 [23] were compound GR 199114X [26] which lacks a charged functionality, compounds L-736,380 [27] and Z-360 [28] which contain acidic substituents, and compounds L-740,093 [29] and YF476 [30], which contain basic substituents. Consequently, McDonald et al. [31] have reported the synthesis of series of 1,3,4-benzotriazepine-based CCK_2R antagonists which are achiral and maintained selectivity over CCK_1R. Further, they reported optimization of 1,3,4-benzotriazepine-based CCK_2R antagonists so as to obtain potent, orally active inhibitors of gastrin-mediated gastric acid secretion [32].

CCK_2R is a seven-transmembrane spanning receptor that belongs to the superfamily of GPCR's. At present, derivation of information on the three-dimensional (3D) structure of GPCR's has not been feasible through the application of standard structure determination techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy because of difficulties in receptor purification and of insolubility of the receptor in environments lacking phospholipids. This situation represents a major bottleneck for structure-based drug design. Therefore, a ligand-based approach, based on extensive experimental structure–activity relationships (SAR's), could be helpful in identifying the probable binding site of CCK_2R antagonists. Tokarski and Hopfinger [33] have carried out three-dimensional molecular shape analysis-QSAR (3D-MSA-QSAR) studies on 1,4-benzodiazepine derivatives as CCK_1R antagonists. From molecular shape analyses, these authors have found molecular shape, as represented by common overlap steric volume and nonoverlap steric volume, to be the major contributing factor to the affinity of the CCK_1R for benzodiazepine compounds. Sinha et al. [34] have performed traditional QSAR studies on a series of 1,4-benzodiazepine CCK_1R and CCK_2R antagonists and determined logP and an indicator variables to contribute to the antagonistic activity. Recently a new molecular modeling approach based on field points as a simple descriptor of electrostatic and van der Waals maxima and minima surrounding a structurally diverse series of nonpeptidic analogues of JB 93182 has been used to derive a pharmacophore model for CCK_2R antagonists [35]. Tairi-Kellou et al. [36] have described traditional QSAR equations correlating electronic properties of 1,4-benzodiazepines to CCK_1R and CCK_2R antagonism. An additional application of 2D QSAR techniques to 1,4-benzodiazepine analogues with CCK_1R antagonistic activity have been recently reviewed [37].

In the present study different 3D QSAR techniques such as comparative molecular field analysis (CoMFA) [38,39] and comparative molecular similarity indices analysis (CoMSIA) [40] were applied to a series of recently discovered achiral 1,3,4-benzotriazepine compounds [31,32] with potent CCK_2R antagonistic activity so as to correlate molecular property fields of aligned antagonists using the partial least squares (PLS) method [41,42]. Well established rational drug design methods such as CoMFA and CoMSIA have been successfully applied in our laboratory to malonyl CoA decarboxylase inhibitors [43], cytotoxic agents [44] and HCV NS5B polymerase inhibitors [45]. Molecular hologram QSAR, HQSAR (HQSAR™ Manual SYBYL 7.2) [46] study was also carried out to generate a molecular fingerprint of the 2D structure of the 1,3,4-benzotriazepines.

Each of the QSAR methods used in the present study possesses inherent merits and limitations. In the case of the CoMFA technique, strong features are: (a) the possibility of analyzing molecules with identical pharmacophores but differing atom connectivities together; (b) general applicability to series of molecules for which alignable models can be constructed and for which a desired property is believed to be the result of an alignment-dependent non-covalent molecular interaction; (c) the possibility of gaining an understanding of the biological properties of a set of ligand molecules from a suitable sampling of the steric and electrostatic fields used in the analysis; (d) the use of parameters that represent the interaction energy of an entire ligand, not just the interaction of a more or less arbitrarily selected substructure of the ligand [38,40].

Some of the limitations associated with the use of CoMFA technique include (a) the need to specify an initial alignment rule and an active conformation for each individual compound within a series of interest; (b) inherently underdetermined nature since the number of coefficients to be evaluated are many times higher than the number of compounds evaluated even though PLS provides a robust self-consistent QSAR, with cross-validation ensuring a high probability of predictive utility; (c) over interpretation of contour coefficient maps with which all possible relevant aspects of a ligand–receptor interaction cannot be explored with test results for a few dozen compounds; (d) likelihood of failure when there is a small group of molecules that are very dissimilar from the rest because it cannot predict the behavior of the dissimilar molecules and thus, to cross-validate the test; (e) sensitivity to changes in orientation of the superimposed molecules in the lattice; (f) very steep Lennard–Jones potential and Coulombic potential functions near the van der Waals surface of the molecule and which will require the use of strict cutoff values, thus making difficult to interpret fragmented and not contiguously connected contour maps; (g) generation of maps denoting regions apart from the molecules and where interactions with a putative environment are to be expected; (h) the molecular mechanics snapshot of steric and electrostatic nonbonded enthalpies, which produces the CoMFA parameters, does not include entropically based factors such as hydrophobicity, which also contribute to ligand-binding [38,40].

Some of the merits surrounding the CoMSIA technique are: (a) the use of a Gaussian-type distance dependence, and the absence of singularities at the atomic positions; features that will preclude the need for setting arbitrary definitions of cutoff limits and deficiencies arising from different slopes of the fields; (b) the generation of contour maps that are both superior to and easier to interpret than those derived by other techniques, and having those regions within the area occupied by the ligand skeletons that require a particular physicochemical property important for activity highlighted; (c) insensitivity to changes in orientation of the superimposed molecules in the lattice; (d) ability to produce contour maps that are easy to interpret through the use of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields and the introduction of partitioning of variance into different field types [40].

Major limitations associated with the use of the CoMSIA technique are: (a) the need to use more fields than with the CoMFA approach, a feature which, however, does not necessarily increases its accuracy or predictive ability because of the correlations between fields; (b) the use of Gaussian distributions to build similarity fields and, in this way, avoid strict cutoff values but at the cost of sacrificing accuracy with respect to steric fields [40].

The use of the HQSAR technique can offer advantages such as (a) the ability to rapidly and easily generate QSAR models of a high statistical quality and predictive value; (b) the use of an extended form of fingerprint known as molecular hologram, which encodes more information on, for example, branched and cyclic fragments than a traditional 2D fingerprint; (c) its predictive power in terms of r^2_{cv} values is comparable to that of CoMFA; (d) ability to generate QSAR models for both small and large data sets in a very rapid manner; (e) intrinsic simplicity, rapidity, ease of use, and accuracy in predicting activities; (f) accepts inputs that are just 2D

molecules and which do not require an alignment; (g) following its implementation within SYBYL, it will allow the coloring of the molecule based on the contribution of the various atoms to the model, i.e., whether they have a positive or detrimental effect on activity, so it is also easy to interpret [46].

As indicated for other QSAR techniques, the use of a HQSAR approach is susceptible to certain limitations, notably (a) its 2.5D, rather than 3D, nature since it is based on counts of fragment fingerprints; (b) it does not allow for accurately extrapolating biological activity predictions to molecules lacking fragments included in the original training data set and which were used to create the model; and (c) does not take into account stereochemical aspects thus making it impossible to differentiate between two stereoisomers or to predict which stereoisomer (or diastereoisomer) would be more active [46].

The resulting contour maps from the 3D QSAR models will help to better understand the role of steric, electrostatic, hydrophobic, and hydrogen bond descriptors in governing CCK_2R inhibition and will serve as a tool for developing more potent CCK_2R antagonists.

2. Materials and methods

2.1. Data sets and biological activity

The training and the test set used include a series of 1,3,4-benzotriazepine derivatives previously reported as CCK_2R antagonists [31,32]. The pK_i values ($-\log K_i$) were used as a dependent variable in the CoMFA, CoMSIA, and HQSAR analyses. As a rule of thumb, the pK_i values of the training data set should span approximately 3 log units. Accordingly, the pK_i values of the training set described in this manuscript span 4.37 log units. Selection of the training set (49 compounds) and the test set (13 compounds) was accomplished by maintaining the similar distribution of structural diversity and a range of biological activities. The structures of the compounds in the training and test sets are shown in Table 1.

2.2. Molecular modeling and alignment

All molecular modeling studies were performed using SYBYL version 7.2 [47] on a Dell Precision 470n workstation with the RHEL 4.0 operating system. The geometry of the most active compound **30**, built using the Sketch Molecule function in SYBYL, was fully optimized with the conjugate gradient method by using the Tripos force field [48] and Gasteiger–Marsili charges [49] with a convergence criterion of 0.001 kcal/mol Å. The conformation thus obtained for compound **30** was subjected to simulated annealing to identify presumably the global minimum energy conformation. The system was heated at 1000 K for 1 ps and then cooled at 200 K for 1 ps. The exponential annealing function was used and 10 such cycles were run. The least energy conformation was determined and subsequently subjected to minimization with the same criteria as mentioned above, and used as a template to build structures of all other training and test set compounds. The conformation of compound **30** was compared with structurally related truncated (7-bromo-5-phenyl-1,3,4-benzotriazepin-2-one) analog for which X-ray crystal data (CCDC Code: QEQXEP) is available [31]. The superimposition of all heavy atoms of the 1,3,4-benzotriazepine core of compound **30** to the corresponding heavy atoms of the truncated analog resulted in rmsd of 0.646, thus the conformation of compound **30** obtained from simulated annealing was a reasonable choice as a template molecule. The newly built compounds were minimized by the same criteria as mentioned above without subjecting them to simulated annealing protocol with objective criteria as maximum atomic overlap with most active compound **30**. In order to maintain the maximum overlap criteria we needed to modify the torsional angles of the anilide side chain of compounds **41**, **42**, **44**, **48**, **49**, and **62** while keeping the conformational energy within +10 kcal/mol of their lowest energy conformation. The most important requirement for the CoMFA and CoMSIA studies is that the 3D structures to be analyzed are aligned according to a suitable conformational template [38]. Since structural information on these antagonists–receptor complexes is presently not available; one of the most potent molecules (compound **30**) was chosen as a template to fit the rest of the training and test set compounds by using the SYBYL fit atoms function. The reference atoms N-1, N-2 and C-3 of the 1,3,4-benzotriazepine ring in compound **30** were used for alignment (Fig. 1a). The resulting representative alignment model (Fig. 1b) was subjected to CoMFA and CoMSIA study. The mol_2 file of the aligned entire data set is provided in Supplementary information (file name: entiredatasetalign.mol_2).

2.3. CoMFA and CoMSIA 3D QSAR models

To derive the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with grid spacing of 2 Å in *x*, *y* and *z* directions was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp^3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard–Jones 6–12 potential) and electrostatic (Coulombic potential) field energies with a distance-dependent dielectric at each lattice point. Values of steric and electrostatic energy were truncated at a default value of 30 kcal/mol.

CoMSIA descriptors were derived with the same lattice box as used for the CoMFA calculations. Five physicochemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. A probe atom sp3 carbon with a charge of +1, hydrophobicity of +1 [50], and H-bond donor and acceptor property of +1 [51,52] was placed at every grid point to measure the electrostatic, steric, hydrophobic, and H-bond donor or acceptor field. The default value of 0.3 was used as the attenuation factor (α).

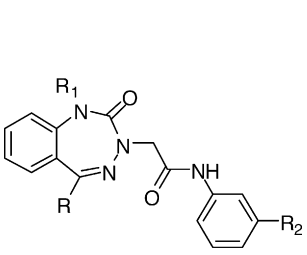
The CoMFA and CoMSIA descriptors were used as independent variables and the pK_i values were used as dependent variables in partial least square (PLS) [41,42] regression analyses to derive 3D QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-one-out (LOO) cross-validation [39,53]. The optimal number of components (ONC) obtained from the cross-validated PLS analysis were used to derive the final QSAR model using the compounds in the training set (without cross-validation).

The boot strapping analysis [54] for 100 runs and the number of cross-validations (e.g., two and five) were carried out and confirmed by the average value for 50 runs from each cross-validation. To test the utility of the model as a predictive tool, an external set of compounds (test set) with known activities, but not used in model generation, were used.

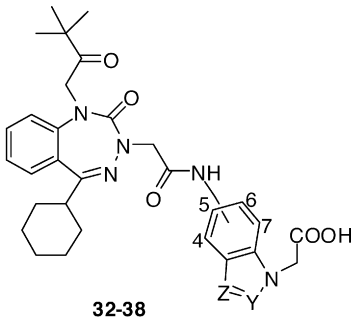
2.4. HQSAR model

In HQSAR, a molecule is described as a unique string of numbers or "bins" (molecular hologram). The bins represent all of the unique fragments included within a particular molecule and are assigned by a cyclic redundancy check (CRC) algorithm. All linear, branched and overlapping structure fragments were used to generate HQSAR descriptors. The structure fragments from each original molecule consisted of a user defined minimum and a maximum number of atoms, which are fragment-size parameters. The information in each fragment is defined by fragment distinction parameters, including atoms (A), bonds (B), connections (Con), hydrogen atoms (H), chirality (Ch), and H donor or

Table 1
Structures of the training and test set compounds[a]



1-31, 39-62



32-38

Compound	R	R_1	R_2	pK_i
Training set				
1	*c*-C_6H_{11}	CH_3	COOH	5.92
2	*c*-C_6H_{11}	*t*-$BuCOCH_2$	COOH	8.29
3	Me_2CHCH_2	*t*-$BuCOCH_2$	COOH	7.05
4	*c*-C_5H_9	*t*-$BuCOCH_2$	COOH	7.62
5	*c*-C_7H_{13}	*t*-$BuCOCH_2$	COOH	8.32
6	1-Ad	*t*-$BuCOCH_2$	COOH	7.62
7	*c*-C_6H_{11}	$Me_2CH(CH_2)_2$	COOH	7.69
8	*c*-C_6H_{11}	Pyrrolidin-1-yl-$COCH_2$	COOH	7.79
9	*c*-C_6H_{11}	$EtO(CH_2)_2$	COOH	6.31
10	*c*-C_6H_{11}	1-Ad$COCH_2$	COOH	8.25
11	*c*-C_6H_{11}	2-Me$C_6H_4COCH_2$	COOH	8.12
12	*c*-C_6H_{11}	*c*-$C_5H_9COCH_2$	COOH	9.00
13	*c*-C_6H_{11}	*c*-$C_6H_{11}COCH_2$	COOH	9.10
14	*c*-C_6H_{11}	1-Me-*c*-$C_5H_9COCH_2$	COOH	8.92
15	*c*-C_6H_{11}	*t*-$BuCOCH_2$	CH_2COOH	8.67
16	*c*-C_6H_{11}	*t*-$BuCOCH_2$	SCH_2COOH	9.19
17	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$N(Me)CH_2COOH$	8.97
18	*c*-C_6H_{11}	*t*-$BuCOCH_2$	OCH_2COOH	8.26
19	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$NHCH_2COOH$	8.21
20	*c*-C_6H_{11}	*t*-$BuCOCH_2$	1(2)*H*-tetrazol-5-yl	9.00
21	*c*-C_6H_{11}	*t*-$BuCOCH_2$	1,2,4-Oxadiazol-3-yl-5(2*H*)-one	9.70
22	*c*-C_6H_{11}	*t*-$BuCOCH_2$	N(Me)-1(2)*H*-tetrazol-5-yl	8.71
23	*c*-C_6H_{11}	*t*-$BuCOCH_2$	SCH_2-1(2)*H*-tetrazol-5-yl	9.15
24	*c*-C_6H_{11}	*t*-$BuCOCH_2$	3-C_6H_4COOH	8.80
25	*c*-C_6H_{11}	*t*-$BuCOCH_2$	2-Thiazol-4-yl-COOH	9.18
26	*c*-C_6H_{11}	*t*-$BuCOCH_2$	4-Oxazol-2-yl-COOH	9.44
27	*c*-C_6H_{11}	*t*-$BuCOCH_2$	1-Imidazol-4-yl-(1*H*)-CH_2COOH	8.39
28	*c*-C_6H_{11}	*t*-$BuCOCH_2$	1-Pyrrol-2-yl-CH_2COOH	9.04
29	*c*-C_6H_{11}	*c*-$C_5H_9COCH_2$	CH_2COOH	9.55
30	*c*-C_6H_{11}	*c*-$C_5H_9COCH_2$	1,2,4-Oxadiazol-3-yl-5(2*H*)-one	9.97
31	*c*-C_6H_{11}	*c*-$C_5H_9COCH_2$	SCH_2COOH	9.48

Compound	Position of attachment	Y	Z	pK_i
Training set				
32	6	CH	CH	9.53
33	6	CH	N	7.77
34	6	N	CH	9.08
35	5	CH	CH	9.11
36	5	N	CH	8.03
37	4	CH	CH	8.22
38	4	N	CH	8.56

Compound	R	R_1	R_2	pK_i
Training set				
39	2-Py	Pyrrolidin-1-yl-$COCH_2$	NHMe	5.60
40	Ph	*t*-$BuCOCH_2$	NHMe	7.13
41	*c*-C_6H_{11}	*t*-$BuCOCH_2$	NMe_2	8.22
42	*c*-C_6H_{11}	*t*-$BuCOCH_2$	NH_2	7.57
43	*c*-C_6H_{11}	*t*-$BuCOCH_2$	Morpholin-1-yl	7.78
44	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$NH(CH_2)_2OCH_2CH_3$	7.35
45	*c*-C_6H_{11}	*t*-$BuCOCH_2$	Imidazol-1-yl	8.04
46	*c*-C_6H_{11}	*t*-$BuCOCH_2$	Me	8.05
47	*c*-C_6H_{11}	*t*-$BuCOCH_2$	OMe	8.26
48	*c*-C_6H_{11}	*t*-$BuCOCH_2$	CH_2OH	7.88
49	*c*-C_6H_{11}	*t*-$BuCOCH_2$	COOH	8.29

Table 1 (*Continued*)

Compound	R	R_1	R_2	pK_i
Test set				
50	*c*-C_6H_{11}	$MeCOCH_2$	COOH	6.48
51	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$(CH_2)_2COOH$	8.73
52	*c*-C_6H_{11}	*t*-$BuCOCH_2$	SO_2CH_2COOH	8.37
53	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$CONHSO_2Me$	8.41
54	*c*-C_6H_{11}	*t*-$BuCOCH_2$	CH_2-1(2)*H*-tetrazol-5-yl	8.33
55	*c*-C_6H_{11}	*t*-$BuCOCH_2$	5-furan-2-yl-COOH	9.61
56	*c*-C_6H_{11}	*c*-$C_5H_9COCH_2$	$(CH_2)_2COOH$	9.47
57	2-Py	*t*-$BuCOCH_2$	NHMe	6.71
58	2-Py	2-$CH_3C_6H_4COCH_2$	NHMe	6.54
59	*c*-C_6H_{11}	*t*-$BuCOCH_2$	NHMe	8.10
60	*c*-C_6H_{11}	*t*-$BuCOCH_2$	Pyrrolidin-1-yl	7.80
61	*c*-C_6H_{11}	*t*-$BuCOCH_2$	$NCH_3(CH_2)_2NHCH_3$	7.53
62	*c*-C_6H_{11}	*t*-$BuCOCH_2$	OH	7.88

[a] Compounds **1**, **32**, and **34** were outliers in the CoMFA and CoMSIA models whereas HQSAR model had only one outlier, i.e., compound **32**.

acceptor (DA). The generated fragments were then hashed into a fixed length array to produce a molecular hologram. The fixed length was defined as hologram length parameters. The HQSAR module provides 12 default hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401), which are prime numbers, to minimize the possibility of fragment collision [46]. The particular nature of substructure fragments generated by HQASR and, consequently, the information contained in the resultant molecular holograms was altered by adjusting these parameters. For hologram generation process, different combinations of these parameters were considered using the fragment-size default (4–7). The HQSAR analysis was performed by screening 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts were then related to the measured biological activity of the training set. The generated models were investigated using r^2_{cv} and the predictive ability of the models was assessed by their r^2_{ncv} values.

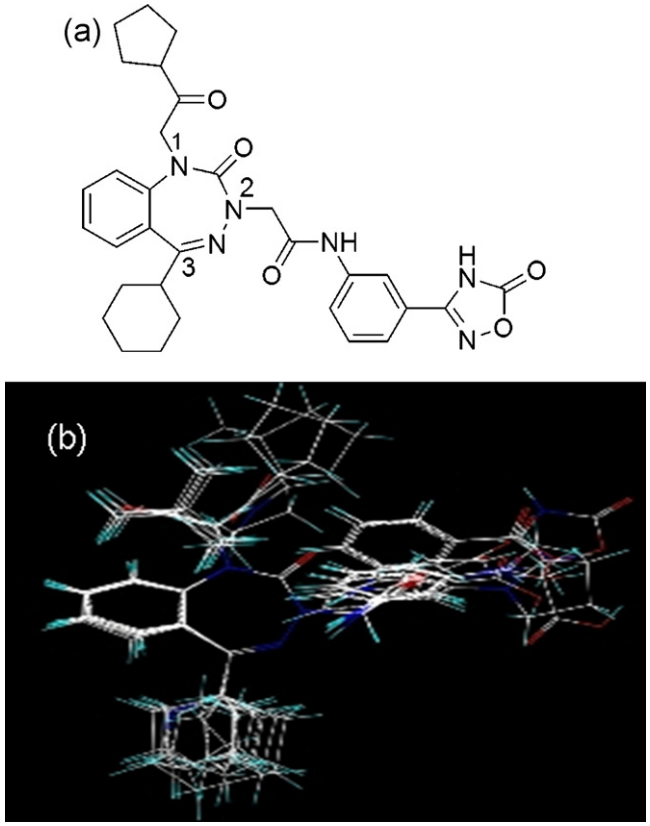


Fig. 1. (a) Compound **30** used as a template for atom based alignment. The atoms (N-1, N-2, and C-3) were used for the alignment. Atom numbering does not follow IUPAC rules. (b) For clarity only representative compounds are shown in alignment model.

3. Results and discussion

3.1. CoMFA and CoMSIA statistical results

The 3D QSAR models for 1,3,4-benzotriazepine-based CCK_2R antagonists were derived using CoMFA and CoMSIA techniques.

Table 2
Summary of CoMFA and CoMSIA results

PLS statistics	49-Compound model		46-Compound model	
	CoMFA	CoMSIA	CoMFA	CoMSIA
r^2_{ncv} [a]	0.940	0.638	0.966	0.969
SEE[b]	0.241	0.572	0.170	0.161
F_{test} [c]	134.312	40.587	224.615	250.855
r^2_{cv} [d]	0.282	0.336	0.673	0.608
SEP[e]	0.824	0.801	0.584	0.610
r^2_{pred} [f]	0.830	0.687	0.793	0.786
PLS components[g]	5	2	5	5
Contribution				
Steric	0.507	–	0.486	–
Electrostatic	0.493	0.363	0.514	0.419
Hydrophobic	–	0.234	–	0.239
H-bond acceptor	–	0.403	–	0.342
r^2_{boot} [h]	–	–	0.983	0.981
SEE_{boot} [i]	–	–	0.112	0.123
r^2_{LHO} [j]	–	–	0.484	0.444
SD_{LHO} [k]	–	–	0.683	0.675
r^2_{5cv} [l]	–	–	0.627	0.558
$S.D._{5cv}$ [m]	–	–	0.622	0.633

Bold values indicate the best model.
[a] Correlation coefficient.
[b] Standard error of estimate.
[c] Ratio of r^2 explained to unexplained = $r^2/(1-r^2)$.
[d] Cross-validated correlation coefficient after leave-one-out procedure.
[e] Standard error of prediction.
[f] Predicted correlation coefficient for test set of compounds.
[g] Optimal number of principal components.
[h] Average of correlation coefficient for 100 samplings using bootstrapped procedure.
[i] Average standard error of estimate for 100 samplings using bootstrapped procedure.
[j] Average cross-validated correlation coefficient for 50 runs using leave-half-out (LHO) group.
[k] Standard deviation of average cross-validated correlation coefficient for 50 runs.
[l] Average cross-validated correlation coefficient for 50 runs using five cross-validation group.
[m] Standard deviation of average cross-validated correlation coefficient for 50 runs.

For the 49 compounds in the training set, the model yielded ($r^2_{cv} = 0.282$ and $r^2_{ncv} = 0.940$) and ($r^2_{cv} = 0.336$ and $r^2_{ncv} = 0.638$) values for the CoMFA and CoMSIA models, respectively (Table 2). Various 3D QSAR tools available in SYBYL such as region focusing, inertial alignment, different charge calculation method, field fit, and steric/electrostatic cutoffs were used to enhance r^2_{cv} with least standard deviation. The region focusing approach yielded only slightly better results with [($r^2_{cv} = 0.415$, SEP = 0.728, $r^2_{ncv} = 0.803$ and SEE = 0.398) and ($r^2_{cv} = 0.645$, SEP = 0.588, $r^2_{ncv} = 0.937$ and SEE = 0.231)] for the CoMFA and CoMSIA, respectively. Inertial alignment model also led to no improvement in statistical data [($r^2_{cv} = 0.341$, SEP = 0.816, $r^2_{ncv} = 0.930$ and SEE = 0.259) and ($r^2_{cv} = 0.360$, SEP = 0.839, $r^2_{ncv} = 0.669$ and SEE = 0.547)] for the CoMFA and CoMSIA, respectively. Furthermore we have computed Gasteiger–Huckel charges for molecules in the datasets to be able to evaluate the effect of the charge calculation method on the statistical results. In this instance, the results were poorer than those obtained by the Gasteiger–Marsili method for the CoMFA and CoMSIA models (data not shown). On the other hand, the field fit approach yielded [($r^2_{cv} = 0.362$, SEP = 0.745, $r^2_{ncv} = 0.787$ and SEE = 0.414) and ($r^2_{cv} = 0.469$, SEP = 0.694, $r^2_{ncv} = 0.779$ and SEE = 0.421)] for CoMFA and CoMSIA, respectively. Even though the effect of steric/electrostatic cutoffs was studied, the corresponding statistical results were inferior to the default values (data not shown).

The training set was also examined for outliers. There are several reasons for the existence of outliers, including unique structural differences, the insignificant mathematical value in defining the biological activity, and a higher residual between the observed and the predicted biological activity of an antagonist. According to this criterion, compounds **1**, **32**, and **34** were found to be outliers. Structurally compound **1** is unique since it is the only one in the training set possessing a N_1-methyl substituent. According to McDonald et al. [32], high receptor affinity will be achieved in the presence of a ketone carbonyl group and a bulky hydrophobic *tert*-butyl group at the N_1-position, features that are missing in compound **1**. A notable underprediction of pK_i values for compounds **32** and **34** stems from the undesirable occupancy of the sterically unfavorable large yellow contour by a 6-substituted indole- and indazole-N_1-acetic acid functional group, respectively, according to CoMFA steric contour map discussed below. Even though compound **33** is structurally similar to compounds **32** and **34**; it was predicted quite closer to the observed pK_i value as its pK_i value itself is significantly lower than that of compounds **32** and **34**. Interestingly, the antagonistic activity of compounds **35**–**38** with either 4- or 5-substituted bicyclic heterocycles was accurately predicted since their N_1-acetic acid functional group did not occupy this large yellow contour. Thus, after dropping compounds **1**, **32**, and **34** from the training set, the 3D QSAR models were re-performed on the remaining 46 compounds and led to a substantial enhancement in the r^2_{cv} value (0.673), with the optimal number of principal components (ONC) being 5 for CoMFA and 0.608 for CoMSIA again with (ONC) 5 (Table 2). Since statistically significant CoMFA and CoMSIA models were derived from the compounds remaining in the training set, these compounds served as the basis for further assessment and discussion. The observed and predicted pK_i values derived from the training set by the best CoMFA model are presented in Supplementary Information, Table S1 and Fig. 2a.

It has been established that the five different descriptor fields are not totally independent of each other and such dependencies of individual fields usually decrease the statistical significance of the models [55,56]. All possible combinations of fields (Supplementary Information, Table S2) were evaluated to determine which of the five CoMSIA fields are actually needed for the generation of a predictive model. The hydrophobic and hydrogen bond donor fields yielded the best individual field models with r^2_{cv} of 0.483 and 0.484, respectively. In the combined models, the best r^2_{cv}, r^2_{ncv}, and least standard error of estimation were obtained by the combination of electrostatic, hydrophobic, and hydrogen bond acceptor fields ($r^2_{cv} = 0.608$, $r^2_{ncv} = 0.969$ and SEE = 0.161). Thus, the model with electrostatic, hydrophobic, and hydrogen bond acceptor fields appeared to be the superior among all the derived models. The detailed observed and predicted pK_i values based on the selected CoMSIA model for the training set are shown in Supplementary Information, Table S1 and Fig. 2b.

To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis for 100 runs was performed (Table 2). Bootstrapping involves the generation of many new datasets from the original datasets, after randomly choosing samples from the original dataset. The bootstrapped r^2 of 0.983 for CoMFA and 0.981 for CoMSIA suggest that a good internal consistency exists within the underlying dataset.

Next, a cross-validation analysis was applied to the set of compounds in the training set to investigate the stability of the CoMFA and CoMSIA models. The training set model was cross-validated using two (leave-half-out) and five (leave-20%-out) cross-validation groups, 50 times each. The average and standard deviation values of r^2_{cv} are shown in Table 2. When

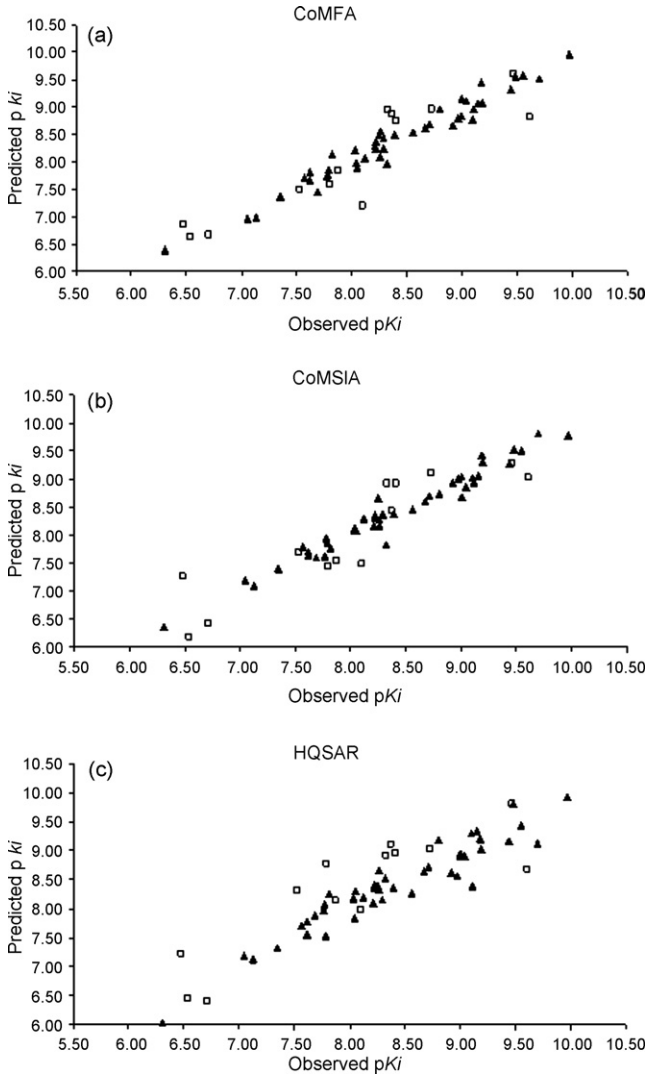


Fig. 2. Plot of observed versus predicted pK_i values for the training set (▲) and the test set (□) compounds based on the CoMFA (a), CoMSIA (b) and HQSAR (c) model.

two cross-validation groups were used, the average r^2_{cv} values for CoMFA and CoMSIA were 0.484 (standard deviation = 0.683) and 0.444 (standard deviation = 0.675), respectively. By using five cross-validation groups, the average r^2_{cv} and S.D. values were 0.627 and 0.622, respectively, for CoMFA. For CoMSIA, these values were 0.558 and 0.633, respectively. Thus, two and five cross-validation analyses of the training set composition, for each run, were consistent for the CoMFA and CoMSIA models.

3.2. Predictive power of the CoMFA and CoMSIA models

The predictive abilities of the CoMFA and CoMSIA models were determined from a set of 13 test compounds not included in the model generation study. The predicted r^2 values from the CoMFA and CoMSIA models were found to be 0.793 and 0.786, respectively (Table 2). The detailed observed and predicted pK_i values based on the selected CoMFA and CoMSIA models for the test set compounds are listed in Supplementary Information, Table S1 and graphically shown in Fig. 2a and b. Statistically significant predictions support the validity of the derived models in predicting the antagonistic activity of the untested compounds.

3.3. CoMFA contour maps

In the CoMFA steric field, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% contributions, respectively. Similarly the red (electronegative charge favorable) and blue (electropositive charge favorable) contours in the CoMFA electrostatic field represent 80% and 20% contributions, respectively.

The CoMFA contour map of steric contribution in the presence of most potent compound **30** is depicted in Fig. 3a. From this figure, it is evident that different physicochemical field properties are predominantly distributed in the regions around the N_1- and N_3-positions of the 1,3,4-benzotriazepine ring system, suggesting their relevance to the activity of these compounds. A large green contour (G1) around the R_1 substituent at the N_1-position of the 1,3,4-benzotriazepine ring indicates that a sterically bulky group is favored in this region. The R_1 substituent of the most potent compounds **21** (pK_i = 9.70), **29** (pK_i = 9.55), and **30** (pK_i = 9.97) touches this green contour while the terminal methyl group of the ethoxyethyl moiety, the C-3 carbon atom of the pyrrolidyl moiety, and the 2-methylbenzoyl moiety of the R_1 substituents in least potent compounds **9** (pK_i = 6.31), **39** (pK_i = 5.60), and **58** (pK_i = 6.54), respectively, were found in the vicinity of the two small sterically unfavorable yellow contours (Y1 and Y2). Compounds **10** (pK_i = 8.25), and **11** (pK_i = 8.12) with the adamantyl and 2-methylphenyl R_1 substituents, respectively, displayed comparatively less activity compared to most potent compounds suggesting the requirement of the optimum bulk at this position. An R_2 substituent in the form of carboxylic acid group attached to the anilide moiety by a wide variety of lipophilic linkers such as –SCH_2 in **16** (pK_i = 9.19), –$N(CH_3)CH_2$ in **17** (pK_i = 8.97) and –CH_2CH_2 in **51** (pK_i = 8.73) is located near the small green contour (G2) explains their high antagonistic activity compared to those compounds where a –COOH group is directly attached to the *meta*-position of the anilide moiety. A green contour (G3) near the R substituent signifies that a bulky group needs to be present at this position to attain CCK_2R antagonistic activity. The steric bulk of a phenyl or pyridyl ring is small compared to that of a cyclohexyl ring, a finding that explains why compounds with a cyclohexyl substituent at the R-position (e.g., compounds **21** and **30**) are more potent than compounds with an aryl substituent as is the case with compounds **39** (pK_i = 5.60), **40** (pK_i = 7.13), **57** (pK_i = 6.71), and **58** (pK_i = 6.54).

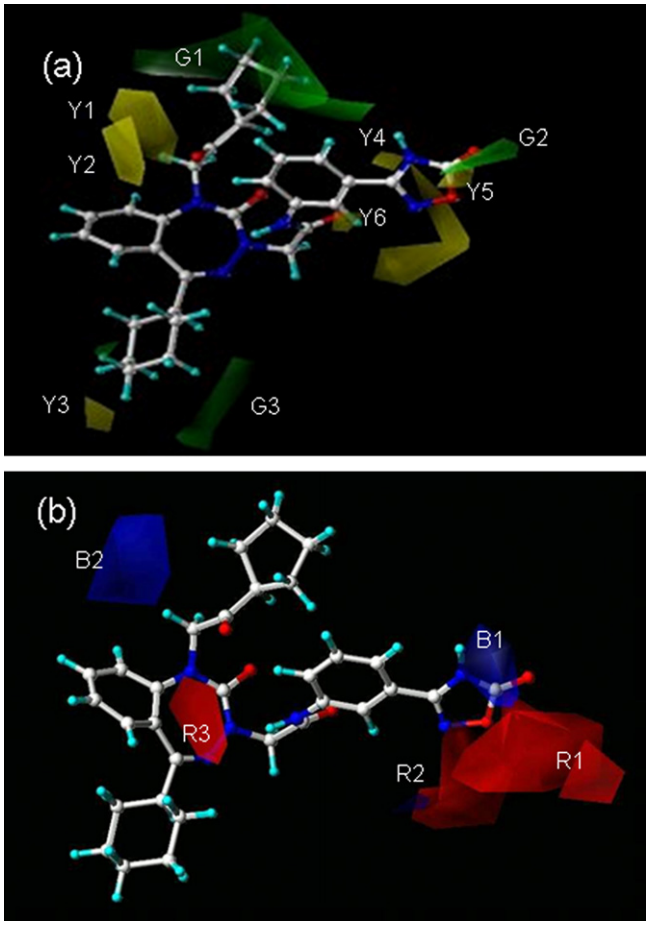


Fig. 3. (a) CoMFA stdev*coeff steric contour map; green contours (G1–G3) indicate regions where bulky groups increase activity, whereas yellow contours (Y1–Y6) indicate regions where bulky groups decrease activity. (b) CoMFA stdev*coeff electrostatic contour map; blue (B1–B2) contours indicate regions where electropositive groups increase activity, whereas red (R1–R3) contours indicate regions where electronegative groups increase activity. The most potent compound **30** is displayed as a reference.

For compounds **2**–**6**, structural differences in their R substituent are reflected in their pK_i values. For example, *c*-C_6H_{11} in **2** (pK_i = 8.29), –Me_2CHCH_2 in **3** (pK_i = 7.05), *c*-C_5H_9 in **4** (pK_i = 7.62), *c*-C_7H_{13} in **5** (pK_i = 8.32), and 1-Ad in **6** (pK_i = 7.62). It is obvious that increasing the bulk of the alkyl group at the R-position increases the CCK_2R antagonistic activity. Unexpectedly, however, the bulky adamantyl group in compound **6** showed decreased antagonistic activity. This finding suggests that this site of the CCK_2R has steric boundaries that preclude the bulky adamantyl group from optimum binding. Furthermore, a small yellow contour (Y3) is mapped near the R substituent which seems to be very close to the bulky adamantyl group and depicts that steric bulk larger than the cyclohexyl or cycloheptyl may lead to poor CCK_2R antagonistic activity. Two small yellow contours (Y4 and Y5) one at each side of the C-3 carbon atom of the oxadiazole ring suggest that steric bulk at this position would be unfavorable for CCK_2R antagonistic activity. These two yellow contours were found to be occupied by –$NHCH_3$ group at R_2-position of the less active compounds **39**, **40**, **57**, and **58**. Further, carboxylic acid group (R_2 substituent) is attached to the anilide moiety by a range of lipophilic linkers such as –CH_2 in **15**, –SCH_2 in **16**, –$N(CH_3)CH_2$ in **17**, and –CH_2CH_2 in **51** showed high activities, is surrounded by three yellow contours (Y4, Y5, and Y6) depicting that lipophilic linker with less steric bulk at R_2-position is favorable for CCK_2R

antagonistic activity. The polar linkers at R_2-position such as ether, amine, and sulfonyl, respectively, in compounds **18** (pK_i = 8.26), **19** (pK_i = 8.21), and **52** (pK_i = 8.37) were found to move rather close to the sterically unfavorable yellow contours and, hence, not to show marked activities in comparison to the lipophilic linker containing counterparts.

The electrostatic contour map of a CoMFA model is shown in Fig. 3b. To facilitate visualization, the most potent compound **30** is overlaid on the map. The electronegative ring oxygen and carbonyl oxygen atoms of the oxadiazole ring at R_2-position in the most active compounds **21** and **30** are seen surrounded by a large red contour (R1). The least potent compound **39** (R_2 = –$NHCH_3$) lacks the electronegative group at this position. The medium size red contour (R_2) adjacent to the large red contour encompasses the acetic acid function of the highly active 5-substituted indole (compound **35**, pK_i = 9.11) and indazole (compound **36**, pK_i = 8.03) N_1-atom. The third small red contour (R3) seen in the vicinity of the 3-position of the indole ring of compound **37** (pK_i = 8.22) and 3-position of the indazole ring of compound **38** (pK_i = 8.56), suggests that antagonistic activity will increase by the presence of an electronegative substituent at this position. A medium blue contour map (B1) localized around the carbonyl carbon atom of the oxadiazole ring of compounds **21** and **30** indicates that the electropositive group at this position would increase the CCK_2R antagonistic activity. The electropositive contour may have resulted from the electron deficient carbonyl carbon atom which is directly attached to the O and N atoms of compounds **21** and **30**. A large blue contour map (B2) near the R_1 substituent indicates the presence of electropositive groups such as alkyl chains which, as exemplified by the most potent compounds having either a bulky *tert*-butyl group or *c*-C_5H_9 group at this position, would increase the CCK_2R antagonistic activity.

3.4. CoMSIA contour maps

In the CoMSIA hydrophobic field, the orange (hydrophobic favorable) and white (hydrophobic unfavorable or hydrophilic favorable) contours represent 80% and 20% contributions, respectively. In the CoMSIA hydrogen bond acceptor field, the magenta (hydrogen bond acceptor favorable) and cyan (hydrogen bond acceptor unfavorable) contours represent 80% and 20% contributions, respectively.

The electrostatic contour maps of CoMSIA model are similar to those obtained from CoMFA, and their interpretation is as described for the CoMFA electrostatic contour map. The hydrophobic contour map of the CoMSIA model in the presence of most potent compound **30** is displayed in Fig. 4a. The large orange contour map (O1) located near the R_1 substituent suggests that bulky hydrophobic groups at this position would favor CCK_2R antagonistic activity. The significantly lower activity observed with compounds **1** (pK_i = 5.92), **7** (pK_i = 7.69), and **50** (pK_i = 6.48) is probably the result of their having relatively small hydrophobic substituents such as methyl, isoamyl, and acetylmethyl at R_1-position; whereas the high activity of compounds **21** and **30** is associated with the presence of a large hydrophobic substituents at R_1-position. This information is also consistent with the fact that the simultaneous presence of a large hydrophobic group and a ketone carbonyl function at R_1-position is necessary for a potent CCK_2R antagonistic activity (e.g., compounds **2** vs **7** and **2** vs **50**). A large white contour map (W1) encompassing positions 1 and 5 of the oxadiazole ring of the most potent compounds **21** and **30** suggest the importance of hydrophilic groups for high CCK_2R antagonistic activity. Most of the compounds with –COOH group attached directly to the anilide ring showed moderate activity as their –COOH group was found to be somewhat away from the large white contour. The highly active compounds **16** (pK_i = 9.19), **17** (pK_i = 8.97), **25** (pK_i = 9.18), and **26** (pK_i = 9.44), possessing a –COOH group linked through a variety of linkers to the anilide ring, were found to be in contact with the large white contour. McDonald et al. [31] have pointed out that the anilide binding site might be exposed to the solvent, as a result, be able to tolerate substituents with varying bulkiness as well as electronic properties. The presence of a white (hydrophobic unfavorable or alternatively hydrophilic favorable) contour map at R_2-position, further reinforces the suggestion of McDonald et al. [31] that this site is probably solvent-exposed. Such a possibility can be verified either by site-directed mutagenesis combined with photoaffinity labeling or by X-ray crystallographic study of the CCK_2R–antagonist complex. Whereas the former approach has been found feasible [57–59], the latter one is surrounded by difficulties associated with the isolation of the membrane bound GPCR's in a pure form. An alternative way to gaining an understanding of the molecular mechanisms associated with CCK_2R–antagonist interactions is based exclusively on a theoretical (homology modeling) concept. The least potent compound **39**, possessing a hydrophobic –$NHCH_3$ group, was seen in the

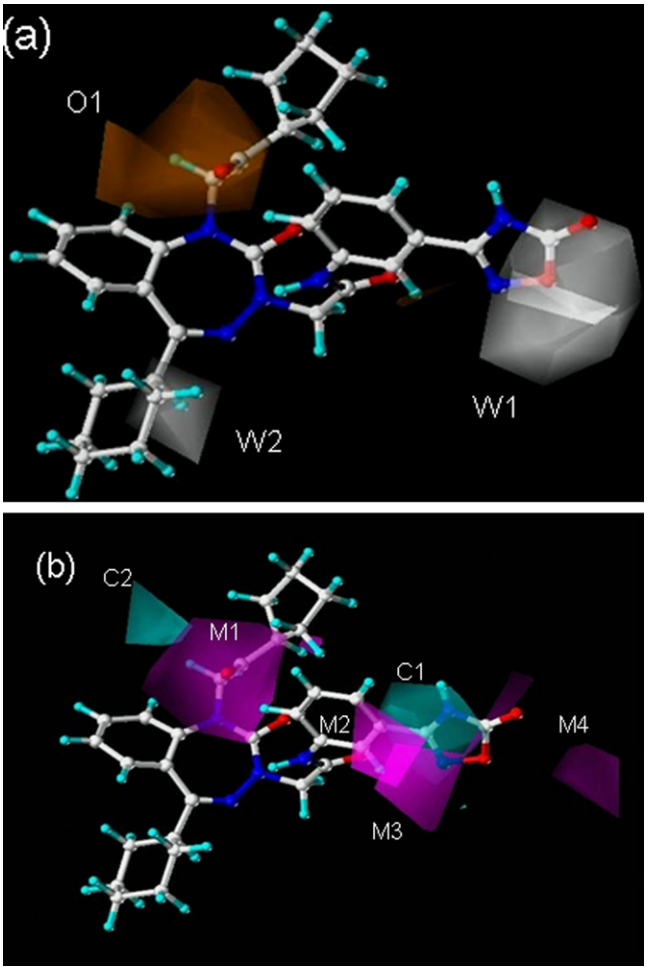


Fig. 4. (a) CoMSIA stdev*coeff hydrophobic contour maps; orange (O1) and white (W1–W2) contours indicate favorable and unfavorable hydrophobic groups, respectively, for CCK2 antagonistic activity. (b) CoMSIA stdev*coeff hydrogen bond acceptor contour maps; magenta (M1–M4) and cyan (C1–C2) contours indicate favorable and unfavorable hydrogen bond acceptor groups, respectively, for CCK2 antagonistic activity. The most potent compound **30** is displayed as a reference.

Table 3
Results of HQSAR analysis for various fragment distinctions on the key statistical parameters using fragment-size default (4–7)[a]

Model	Fragment distinction	r^2_{cv}	r^2_{ncv}	SEP	SEE	HL	*N*
1	A/B	0.634	0.879	0.534	0.320	53	5
2	A/B/C	0.672	0.922	0.526	0.256	257	5
3	A/B/C/H	0.650	0.862	0.543	0.341	257	5
4	A/B/C/H/Ch	0.657	0.867	0.537	0.335	257	5
5	A/B/C/H/Ch/DA	0.627	0.901	0.568	0.293	83	6
6	A/B/H	0.658	0.922	0.544	0.260	307	6
7	A/B/C/Ch	0.647	0.912	0.552	0.275	59	6
8	A/B/DA	0.626	0.928	0.569	0.250	151	6
9	A/B/C/DA	0.574	0.876	0.599	0.324	307	5
10	A/B/H/DA	0.676	0.912	0.529	0.277	257	6
11	A/B/C/Ch/DA	0.577	0.897	0.597	0.295	353	5
12	A/B/C/H/DA	0.680	0.907	0.526	0.284	53	6
13	A/B/H/Ch/DA	0.693	0.877	0.515	0.326	61	6
14	A/B/Ch	0.645	0.875	0.547	0.324	71	5
15	**A/B/H/Ch**	**0.692**	**0.914**	**0.516**	**0.273**	**151**	**6**

Bold values indicate the best model.
[a] r^2_{cv}, cross-validated correlation coefficient; SEP, cross-validated standard error; r^2_{ncv}, non-cross-validated correlation coefficient; SEE, non-cross-validated standard error; HL, hologram length; *N*, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

vicinity of the large white contour. A white contour (W2) near the R substituent suggests that small hydrophilic groups can be tolerated on the cyclohexyl/aryl rings.

The hydrogen bond acceptor contour map of the CoMSIA model in the presence of the most potent compound **30** is depicted in Fig. 4b. A magenta contour (M1) around the R_1 substituent points to the higher activity of compounds having a hydrogen bond acceptor group at this position. The magenta contour around the hydrogen bond acceptor ketone group of the most potent compounds **21**, **29**, and **30** suggests the importance of this group for potent antagonistic activity. Interestingly, the lack of marked activity for compounds exhibiting a hydrogen bond acceptor oxygen atom at the R_1-position in the form of an ether (compound **9**, $pK_i = 6.31$) or amide (compounds **8**, $pK_i = 7.79$ and **39**, $pK_i = 5.60$) is taken as an indication that the oxygen atom needs to be present as a ketone function. The two magenta contours (M2 and M3) surrounding the R_2 substituent indicate the higher activity of compounds having hydrogen bond acceptor group at this position. Since the most potent compounds **21** and **30** possess a hydrogen bond acceptor group at positions O-1, N-2 and C-5 of their oxadiazole ring and the least potent compound **39** does not; this group must be required for potent antagonistic activity. An additional magenta contour (M4) at the R_2-position resulted from the carbonyl oxygen of the indole and indazole N_1-acetic acid group in the highly active compounds **35**–**38**. A large cyan contour map (C1) seen at the hydrogen bond donating ring –NH of the oxadiazole moiety (R_2 substituent) indicates that hydrogen bond acceptor group is unfavorable. Another small cyan contour (C2) is located near the R_1 substituent appears to improve the antagonistic activity of a hydrogen donor group present at this position.

3.5. HQSAR model

The HQSAR model was first generated with a default fragment size (4–7), combined with various fragment distinctions and hologram lengths. Table 3 summarizes the results generated for different fragment distinctions and hologram lengths. The best model was obtained using a hologram length of 151 and A + B + H + Ch as the fragment distinctions ($r^2_{cv} = 0.692$, $r^2_{ncv} = 0.914$, ONC = 6). The information obtained from fragment distinctions such as atoms (A), bonds (B), hydrogen atoms (H), and chirality (Ch), made an important contribution to the biological activity. With the best fragment distinctions parameters, PLS analysis was performed to investigate if different fragment sizes could improve the statistical parameters. The HQSAR results for the different fragment sizes used are summarized in Table 4. These results show that variation in fragment size can lead to improvement of the HQSAR model ($r^2_{cv} = 0.744$, $r^2_{ncv} = 0.918$, ONC = 6) using a hologram length of 97. The observed versus predicted pK_i values for the training set based on the best HQSAR model are listed in Supplementary Information, Table S1 and graphically shown in Fig. 2c. As the structure encoded within a 2D fingerprint is directly related to biological activity of molecules within the training set, the HQSAR model is able to predict the activity of a test set of structurally related molecules from its fingerprint. The predictive power of the best HQSAR model derived using the 48 training set molecules was assessed by predicting the pK_i values for 13 test molecules not included in the training set. The observed and predicted pK_i values for molecules in the test set are depicted in Supplementary Information, Table S1 and graphically shown in Fig. 2c.

3.6. HQSAR contribution maps

HQSAR enables us to divide a molecule into every atom contributing to the biological activity of the molecule. Individual atomic contributions to the bioactivity of compounds **30**, **39**, and **41** are depicted in Fig. 5a–c, respectively. The colors at the red end of the spectrum (i.e., red, red orange, and orange) reflect unfavorable (negative) contributions, and colors at the green end (i.e., yellow, green blue and blue) indicate favorable (positive) contributions. Atoms with intermediate contributions are colored in white. From Fig. 5a, it can be inferred that the 1,3,4-benzotriazepine ring, the cyclohexyl ring, the anilide moiety, and the first two carbon fragments of the N_1-substituent are strong contributors and the oxadiazole ring a moderate contributor to the

Table 4
HQSAR analysis for the influence of various fragment sizes on the key statistical parameters using the best fragment distinction (atoms, bonds and chirality)

Fragment size	r^2_{cv}	r^2_{ncv}	SEP	SEE	HL	*N*
2–5	**0.744**	**0.918**	**0.470**	**0.266**	**97**	**6**
3–6	0.725	0.927	0.487	0.251	257	6
5–8	0.665	0.871	0.539	0.335	83	6
6–9	0.592	0.769	0.586	0.441	53	5
7–10	0.599	0.816	0.589	0.399	53	6
3–10	0.570	0.717	0.595	0.482	53	4

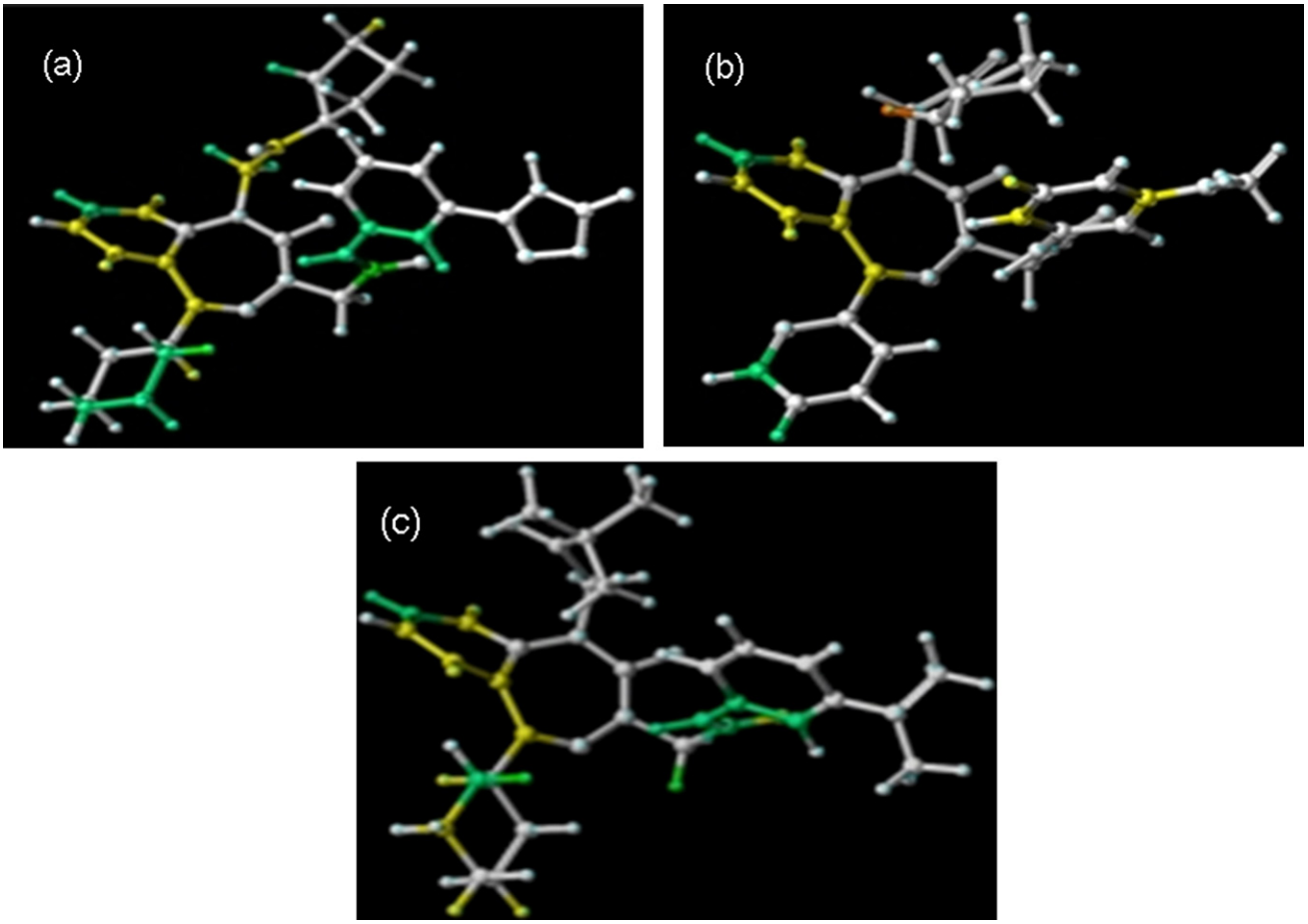


Fig. 5. Individual atomic contributions for the activity of the most potent compound **30** (panel a), least active compound **39** (panel b), and moderately active compound **41** (panel c).

CCK_2R antagonistic activity. Likewise, it is apparent that the hydrogen atoms attached to the first carbon atom at the N_1-position also make a positive contribution. Interestingly, the replacement of these hydrogen atoms by a bulky group may potentially lead to sterically unfavorable interactions as reflected by CoMFA sterically unfavorable yellow (Y1–Y2) contours around these hydrogen atoms. In Fig. 5b the benzotriazepine ring of compound **39** captured the green, yellow, and white colors in a manner similar to that of highly and moderately active compounds. However, the anilide moiety was colored in yellow and white rather than in green and white seen in potent compounds. In addition, the C-2 hydrogen atom of the pyrrolidine moiety was colored red due to its negative contribution. Not surprisingly the C-2 position of the pyrrolidine moiety was found

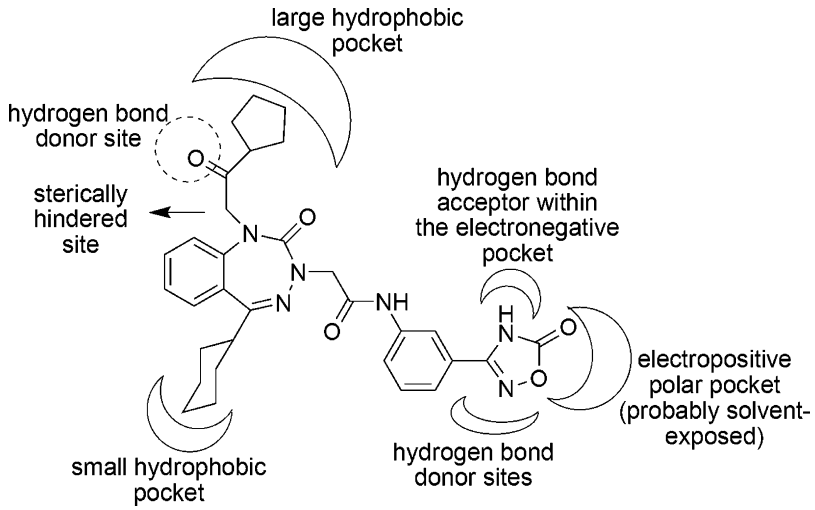


Fig. 6. Proposed hypothetical CCK_2 receptor active site model for 1,3,4-benzotriazepine series of compounds.

to be surrounded by sterically unfavorable yellow (Y1–Y2) contours. In Fig. 5c, the benzotriazepine, cyclohexyl, and anilide moieties were colored in green, yellow, and white, a situation that suggests their significant role in the antagonistic activity of compound **41**. The *tert*-butylcarbonylmethyl group and the *meta*-$N(CH_3)_2$ moieties at the N-1 and anilide positions, respectively, are shown colored in white color, thus suggesting their moderate contribution to antagonistic activity. The regions with intermediate contributions can be identified as potential targets for future molecular modification and experimental SAR studies.

In summary, the CoMFA steric/electrostatic and CoMSIA hydrophobic/hydrogen bond acceptor contour maps associated with differences in antagonistic potencies demonstrate that the structural variations are dominated by chemical features at the N_1 and N_3-positions of the 1,3,4-benzotriazepine ring. The common steric, electrostatic, hydrophobic, and hydrogen bonding properties of the hypothetical CCK_2R active site for the 1,3,4-benzotriazepine class of CCK_2R antagonists are displayed in Fig. 6. The novel ligand–receptor binding interactions determined through this study are: (1) the –R and $–R_1$ substituent binding sites of the CCK_2R appear to be quite stringent in terms of optimum size/volume of the hydrophobic groups; (2) the binding site for the –R substituent is relatively smaller than the corresponding binding site for the $–R_1$ substituent; (3) a ketone function acts as hydrogen bond acceptor needs to be present at the R_1-position for potent CCK_2R antagonistic activity to exist; (4) the branching at the carbon next to the N_1 atom of the 1,3,4-benzotriazepine ring may lead to steric hindrance; (5) the oxadiazole present as a R_2 substituent serves as hydrogen bond acceptor moiety and may be responsible for the high antagonistic activity exhibited by compounds in this group; (6) the *meta*-substituent of the anilide moiety acts both as a mixed hydrogen bond donor/acceptor and as a electropositive polar functional group; and (7) the anilide binding site of CCK_2R seems to be quite accommodating for large groups with differing structural features. For example, the 3-COOH group can be replaced by a variety of bioisosteric 5-membered heterocyclic rings, separation of the 3-COOH group by a variety of linkers, and replacement of the anilide moiety by aminoindole/aminoindazole/aminobenzimidazole rings led to roughly similar potencies. The absence of any contour map at the 1,3,4-benzotriazepine ring could be due to the absence of structural variation in this region.

4. Conclusions

Three QSAR methods, CoMFA, CoMSIA, and HQSAR are used to investigate the relationship between the 1,3,4-benzotriazepine structures and CCK_2R antagonist activity. The high r^2_{cv} obtained from these different QSAR methods suggest that all of these models possess excellent internal predictivity. Excellent correlations between the observed and predicted antagonistic activities for 13 test compounds further verified the reliability of the constructed QSAR models. The 3D QSAR approach has revealed the importance of the shapes of different groups in influencing CCK_2R antagonist property to be determined. The generated 3D QSAR and HQSAR models are valuable computational tools for both the rational design of potent CCK_2R antagonist and the prediction of antagonist property of these agents prior to their synthesis and bioevaluation.

Acknowledgements

Support to TT in the form of start-up funds and resources from the College of Pharmacy of St. John's University are gratefully acknowledged. We thank Maulik Patel, Pallav Patel, Shridhar Kulkarni, and Dr. Cesar Lau-Cam for critical readings of the manuscript and for helpful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.07.003.

References

[1] J.N. Crawley, R.L. Corwin, Peptides 15 (1994) 731.
[2] K. Miyasaka, A. Funakoshi, J. Gastroenterol. 38 (2003) 1.
[3] S. Silvente-Poirot, M. Dufresne, N. Vaysse, D. Fourmy, Eur. J. Biochem. 215 (1993) 513.
[4] F. Noble, S.A. Wank, J.N. Crawley, J. Bradwejn, K.B. Seroogy, M. Hamon, B.P. Roques, Pharmacol. Rev. 51 (1999) 745.
[5] A. De Weerth, J.R. Pisegna, K. Huppi, S.A. Wank, Biochem. Biophys. Res. Commun. 194 (1993) 811.
[6] S.A. Wank, J.R. Pisegna, A. De Weerth, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 8691.
[7] Y.M. Lee, M. Beinborn, E.W. McBride, M. Lu, L.F. Kolakowski, A.S. Kopin, J. Biol. Chem. 268 (1993) 8164.
[8] I.M. McDonald, Exp. Opin. Ther. Patents 11 (2001) 445.
[9] G.J. Dockray, Best Pract. Res. Clin. Endocrinol. Metab. 18 (2004) 555.
[10] A. Aly, A. Shulkes, G.S. Baldwin, Biochim. Biophys. Acta 1 (2004) 1.
[11] C.R. Haigh, S.E. Attwood, D.G. Thompson, J.A. Jankowski, C.M. Kirton, D.M. Pritchard, A. Varro, R. Dimaline, Gastroenterology 124 (2003) 615.
[12] I. Chau, D. Cunningham, C. Russell, A.R. Norman, T. Kurzawinski, P. Harper, P. Harrison, G. Middleton, F. Daniels, T. Hickish, J. Prendeville, P.J. Ross, B. Theis, R. Hull, M. Walker, N. Shankley, B. Kalindjian, G. Murray, A. Gillbanks, J. Black, Br. J. Cancer 94 (2006) 1107.
[13] V. Dauge, I. Lena, Neurosci. Biobehav. Rev. 22 (1998) 815.
[14] Z. Wiesenfeld-Hallin, X.J. Xu, Regul. Pept. 65 (1996) 23.
[15] H. Kulaksiz, R. Arnold, B. Goke, E. Maronde, M. Meyer, F. Fahrenholz, W.G. Forssmann, R. Eissele, Cell Tissue Res. 299 (2000) 289.
[16] Y. Ochi, S. Horie, T. Maruyama, K. Watanabe, S. Yano, Life Sci. 77 (2005) 2040.
[17] H.L. Waldum, P.M. Kleveland, A.K. Sandvik, E. Brenna, U. Syversen, I. Bakke, K. Tommeras, Pharmacol. Toxicol. 91 (2002) 359.
[18] M. Giralt, P. Vergara, Regul. Pept. 81 (1999) 73.
[19] G. Dal Forno, C. Pietra, M.V. Urciuoli, F.T. Amsterdam, G. Toson, G. Gaviraghi, D. Trist, J. Pharmacol. Exp. Ther. 261 (1992) 1056.
[20] A.M. Capelli, D. Donati, F. Micheli, Curr. Med. Chem. 2 (2002) 287.
[21] D.C. Horwell, J. Hughes, J.C. Hunter, M.C. Pritchard, R.S. Richardson, E. Roberts, G.N. Woodruff, J. Med. Chem. 34 (1991) 404.
[22] S.B. Kalindjian, I.M. Buck, J.M. Davies, D.J. Dunstone, M.L. Hudson, C.M. Low, I.M. McDonald, M.J. Pether, K.I. Steel, M.J. Tozer, J.G. Vinter, J. Med. Chem. 39 (1996) 1806.
[23] M.G. Bock, R.M. DiPardo, B.E. Evans, K.E. Rittle, W.L. Whitter, D.E. Veber, P.S. Anderson, R.M. Freidinger, J. Med. Chem. 32 (1989) 13.
[24] J. Bradwejn, D. Koszycki, A. Couetoux du Tertre, H. Van Megen, J. Den Boer, H. Westenberg, Arch. Gen. Psychiatry 51 (1994) 486.
[25] M.S. Kramer, N.R. Cutler, J.C. Ballenger, W.M. Patterson, J. Mendels, A. Chenault, R. Shrivastava, D. Matzura-Wolfe, C. Lines, S. Reines, Biol. Psychiatry 37 (1995) 462.
[26] N. Bailey, P.C. Box, R.A.E. Carr, J.W.B. Cooke, B. Evans, H. Finch, J.E. Head, M. Pass, P. Shah, J.R. Wheatcroft, Bioorg. Med. Chem. Lett. 7 (1997) 281.
[27] J.L. Castro, R.G. Ball, H.B. Broughton, M.G. Russell, D. Rathbone, A.P. Watt, R. Baker, K.L. Chapman, A.E. Fletcher, S. Patel, A.J. Smith, G.R. Marshall, W. Ryecroft, V.G. Matassa, J. Med. Chem. 39 (1996) 842.
[28] N. Miura, T. Yoneta, H. Ukawa, Y. Fukuda, R. Eta, Y. Mera, T. Omata, T. Kinomoto, T. Kurimoto, Z. Itoh, Gastroenterology 120 (2001) A311.
[29] S. Patel, A.J. Smith, K.L. Chapman, A.E. Fletcher, J.A. Kemp, G.R. Marshall, R.J. Hargreaves, W. Ryecroft, L.L. Iversen, S. Iversen, R. Baker, G. Showell, S. Bourrain, J. Neduvelil, V. Matassa, S. Freedman, Mol. Pharmacol. 46 (1994) 943.
[30] Y. Takinami, H. Yuki, A. Nishida, S. Akuzawa, A. Uchida, Y. Takemoto, M. Ohta, M. Satoh, G. Semple, K. Miyata, Ailment. Pharmacol. Ther. 11 (1997) 113.
[31] I.M. McDonald, C. Austin, I.M. Buck, D.J. Dunstone, E.P. Griffin, E.A. Harper, R.A.D. Hull, S.B. Kalindjian, I.D. Linney, C.M.R. Low, M.J. Pether, J. Spencer, P.T. Wright, T. Adatia, A. Bashall, J. Med. Chem. 49 (2006) 2253.
[32] I.M. McDonald, J.W. Black, I.M. Buck, D.J. Dunstone, E.P. Griffin, E.A. Harper, R.A.D. Hull, S.B. Kalindjian, E.J. Lilley, I.D. Linney, M.J. Pether, S.P. Roberts, M.E. Shaxted, J. Spencer, K.I.M. Steel, D.A. Sykes, M.K. Walker, G.F. Watt, L. Wright, W. Xun, J. Med. Chem. 50 (2007) 3101.
[33] J.S. Tokarski, A.J. Hopfinger, J. Med. Chem. 37 (1994) 3639.
[34] J. Sinha, A. Kurup, A. Paleti, S.P. Gupta, Bioorg. Med. Chem. 7 (1999) 1127.
[35] C.M.R. Low, I.M. Buck, T. Cooke, J.R. Cushnir, S.B. Kalindjian, A. Kotecha, M.J. Pether, N.P. Shankley, J.G. Vinter, L. Wright, J. Med. Chem. 48 (2005) 6790.
[36] S. Tairi-Kellou, A. Cartier, B. Maouche, B. Maigret, J. Mol. Struct. (Theochem) 571 (2001) 207.
[37] S.P. Gupta, Curr. Pharm. Des. 8 (2002) 111.
[38] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
[39] R.D. Cramer, J.D. Bunce, D.E. Patterson, Quant. Struct. Act. Relat. 7 (1988) 18.

[40] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130.
[41] S. Wold, C. Albano, W.J. Dunn, U. Edlund, K. Esbenson, P. Geladi, S. Hellberg, W. Lindburg, M. Sjostrom, in: B. Kowalski (Ed.), Chemometrics: Mathematics and Statistics in Chemistry, Dordrecht, Reidel, The Netherlands, 1984, pp. 17–95.
[42] L. Staahle, S. Wold, J. Chemom. 1 (1987) 185.
[43] M.R. Patel, T.T. Talele, Bioorg. Med. Chem. 15 (2007) 4470.
[44] M.R. Patel, J.R. Dimmock, T.T. Talele, J. Chem. Inf. Model. 47 (2007) 2110.
[45] P.D. Patel, M.R. Patel, N. Kaushik-Basu, T.T. Talele, J. Chem. Inf. Model. 48 (2008) 42.
[46] *HQSAR*™ Manual *SYBYL 7.2*, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144.
[47] *SYBYL 7.2*, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144.
[48] M. Clark, R.D. Cramer, N. Van Opdenbosch, J. Comput. Chem. 10 (1989) 982.
[49] J. Gasteiger, M. Marsili, Tetrahedron 36 (1980) 3219.
[50] V.N. Viswanadhan, A.K. Ghose, G.R. Revenkar, R.K. Robins, J. Chem. Inf. Comput. Sci. 29 (1989) 163.
[51] G. Klebe, J. Mol. Biol. 237 (1994) 212.
[52] G. Klebe, T. Mietzner, F. Weber, J. Comput. Aided Mol. Des. 13 (1999) 35.
[53] S. Wold, Technometrics 4 (1978) 397.
[54] R. Vong, P. Geladi, S. Wold, K. Esbensen, J. Chemom. 2 (1988) 281.
[55] G. Bringmann, C. Rummey, J. Chem. Inf. Comput. Sci. 43 (2003) 304.
[56] M. Bohm, J. Stürzebecher, G. Klebe, J. Med. Chem. 42 (1999) 458.
[57] E.M. Hadac, E.S. Dawson, J.W. Darrow, E.E. Sugg, T.P. Lybrand, L.J. Miller, J. Med. Chem. 49 (2006) 850.
[58] C. Galés, M. Poirot, J. Taillefer, B. Maigret, J. Martinez, L. Moroder, C. Escrieut, L. Pradayrol, D. Fourmy, S. Silvente-Poirot, Mol. Pharmacol. 63 (2003) 973.
[59] C. Giragossian, D.F. Mierke, Biochemistry 41 (2002) 4560.