

Structural investigations of CXCR2 receptor antagonists by CoMFA, CoMSIA and flexible docking studies

SHRAVAN KUMAR GUNDA
ROHITH KUMAR ANUGOLU*
SRI RAMYA TATA
SAIKH MAHMOOD

Bioinformatics Division, Osmania
University, Hyderabad-500007
Andhra Pradesh, India

Three-dimensional quantitative structure activity relationship (3D QSAR) analysis was carried out on a set of 56 *N,N'*-diarylsquaramides, *N,N'*-diarylureas and diaminocyclobutenediones in order to understand their antagonistic activities against CXCR2. The studies included comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Models with good predictive abilities were generated with CoMFA $q^2 = 0.709$, r^2 (non-cross-validated square of correlation coefficient) = 0.951, F value = 139.903, r^2 bs = 0.978 with five components, standard error of estimate = 0.144 and the CoMSIA $q^2 = 0.592$, $r^2 = 0.955$, F value = 122.399, r^2 bs = 0.973 with six components, standard error of estimate = 0.141. In addition, a homology model of CXCR2 was used for docking based alignment of the compounds. The most active compound then served as a template for alignment of the remaining structures. Further, mapping of contours onto the active site validated each other in terms of residues involved with reference to the respective contours. This integrated molecular docking based alignment followed by 3D QSAR studies provided a further insight to support the structure-based design of CXCR2 antagonistic agents with improved activity profiles. Furthermore, *in silico* screening was adapted to the QSAR model in order to predict the structures of new, potentially active compounds.

Keywords: CXCR2, homology modeling, CoMFA, CoMSIA, PLS, docking

Accepted July 17, 2012

Chemokines are small, mainly basic proteins involved in host defense that summon immune cells in response to invasion by pathogenic organisms (1). They consist of 340–380 amino acid residues, which play key roles in immunomodulation and host defense mechanisms (2). Chemokine receptors are G protein-coupled receptors containing 7

* Correspondence; e-mail: rohit.casper@gmail.com

trans-membrane domains with three intracellular and three extracellular hydrophilic loops, and an intracellular C-terminus containing serine and threonine residues. Chemokines, an expanding family of small cytokines, have an important role in the chemotaxis of several leucocyte subsets. Chemokines comprise a large protein family that can be divided into subfamilies on the basis of structural motifs. These receptors are divided into four subgroups: CC (proteins have two adjacent cysteines), C (protein has only two cysteines), CX₃C (proteins have three amino acids between the two cysteines) and CXC (proteins have two N-terminal cysteines, which are separated by one amino acid) chemokine ligands, based upon the presence and positioning of the first two of the four conserved cysteine residues (3).

CXC chemokines are characteristically heparin binding proteins. The candidate CXC chemokine receptors that mediate angiogenic activity are CXCR1 and/or CXCR2. Only IL-8 and GCP-2 specifically bind to CXCR1, whereas all ELR⁺ CXC chemokines bind to CXCR2 (4). CXCR2 is expressed by a wide range of cell types, for example, neutrophils, mast cells, T cells, keratinocytes, and cerebellar neurons. Increased levels of CXCR2 and its ligand IL-8 have been observed in humans with diseases such as arthritis, asthma, rheumatoid arthritis, psoriasis, reperfusion injury, and chronic obstructive pulmonary disease (5). This suggests that the CXCR2 receptor and IL-8 may play a pivotal role in these inflammatory disorders. An immune response is characterized by the activation of leukocytes and migration of hematopoietic cells into secondary lymphoid organs or to the site of inflammation. Under inflammatory conditions, the expression of specific chemokines, a class of small proteins with potent chemotactic activity, is up-regulated (6). In the light of these findings, small molecule antagonists of the CXCR2 receptor are attractive biological targets for molecular drug discovery.

Quantitative structure-activity relationship (QSAR) enables the investigators to establish reliable quantitative structure-activity and structure-property relationships to derive *in silico* QSAR models to predict the activity of novel molecules prior to their synthesis. Comparative molecular field analysis (CoMFA) (7) and comparative molecular similarity indices analysis (CoMSIA) (8) have been performed to better understand the structure-activity relationship. In the present work, the ligand guided approach and the molecular modeling studies of CXCR2 antagonists were carried out using docking and three-dimensional quantitative structure-activity relationship (3D-QSAR) analyses.

EXPERIMENTAL

Data set

Biological and chemical data of 56 CXCR2 antagonists, taken from the literature (9, 10) were selected for the QSAR study. The half maximal inhibitory concentration is a measure of the effectiveness of a compound in inhibiting a biological or biochemical function, which is usually represented as IC_{50} in mmol L^{-1} . Six molecules, whose IC_{50} values were reported as $> 270 \text{ mmol L}^{-1}$ and $> 2000 \text{ mmol L}^{-1}$, respectively, were not considered in this study. Reported IC_{50} values of molecules were converted into pIC_{50} values (negative log of molar concentration value) and used as dependent variables in CoMFA and CoMSIA analyses. The 3D QSAR models were generated using a training

set of 42 molecules. Predictive power of the resulting models was evaluated using a test set of 14 molecules. The test set compounds were selected manually so that the structural diversity and wide range of activity in the data set were included.

Molecular modeling

Molecular modeling studies were performed using the SYBYL programming package, version 6.7, Tripos Associates (St. Louis, USA), on a Silicon graphics workstation. The geometry of ligand molecules was optimized using the Tripos force field with a distance-dependent dielectric and Powell's conjugate gradient method. Partial atomic charges were applied using Gasteiger-Hückel charges. The most active compound 65 was used for alignment and the rest of the molecules were aligned to it by using the common substructure (Fig. 1). All the molecules were aligned by the align database command available in SYBYL using maximum substructure. It adjusts the geometry of the molecules so that its steric and electrostatic fields match the fields of the template molecule.

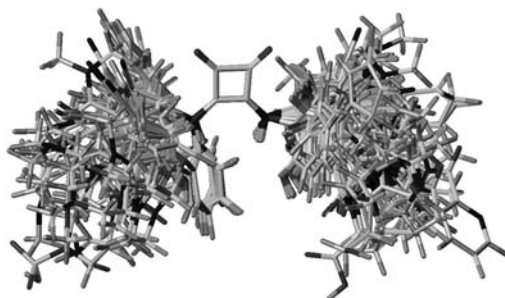


Fig. 1. Alignment structure of all 56 molecules using the most active molecule as the template.

Homology modeling

Comparative modeling was used to build the model. The 3-D structure of the receptor is not available by either X-ray crystallography or nuclear magnetic resonance spectroscopy. Therefore, homology modeling was done for a CXCR2 protein with unknown structure, which was retrieved from the Uniprot database (<http://www.uniprot.org>; accession number: P41597). Crystal structure of the FimXEAL domain from *Pseudomonas aeruginosa* (PDB code: 3HV8) served as a template to construct 3-D models for the CXCR2 receptor protein. Softwares like Clustal-X and MODELLER 9v7 (11) were used for alignment and structure building. Accuracy of the protein model generated was judged by the validity report generated by PROCHECK (12). Parameter comparisons of these proteins were made with well-refined structures having similar resolution. The main chain parameters plotted were Ramachandran Plot (Institute of Molecular Function, Saitama, Japan) quality, peptide bond planarity, bad non-bonded interactions, main chain hydrogen bond energy, C-alpha chirality and the overall G factor. In the Ramachandran plot analysis, the residues were classified according to their regions in the quadrangle.

Docking studies

Molecular docking studies were carried out using the FlexX (13) program interfaced with SYBYL 6.7. In this automated docking program, the flexibility of ligands is considered while the protein or biomolecule is considered as a rigid structure. The ligand is built in an incremental fashion, where each new fragment is added in all possible positions and conformations to a pre-placed base fragment inside the active site. All the molecules for docking were sketched in the SYBYL and all the charges were removed. The 3D coordinates of active sites were taken from the X-ray crystal structure of the FimXEAL domain protein for docking. All water molecules were removed and the protein was used to dock the inhibitor (14). The ligand was pre-processed before docking calculations by giving charges according to the Gasteiger-Hückel method, followed by energy minimization with 10,000 iterations of conjugate gradient algorithm using the Tripos force field. The compounds were docked into the binding site using FlexX, which uses a robust incremental construction algorithm in which the ligand is decomposed into pieces and then flexibly is built up in the active site, using a variety of placement strategies. The active site is defined within the 6.5×10^{-10} m radius around the co-crystallized ligand. This ligand-based study was performed using the pharmacophore based program in SYBYL 6.7, and the receptor-guided study was performed using molecular docking with FlexX. The receptor structure of CXCR2 was modeled by using the crystal structure of the FimXEAL domain protein (PDB code: 3HV8) as a template. A linear method such as PLS is used to find quantitative relationships between the structures of CXCR2 antagonists and their biological activities, and the results obtained by this method are compared.

CoMFA studies

Steric and electrostatic interactions were calculated using the Tripos force field with a distance-dependent dielectric constant in all interactions in a regularly spaced (2×10^{-10} m) grid taking a sp^3 carbon atom as steric probe and a +1 charge as electrostatic probe. The cut-off was set to 30 kcal mol⁻¹. With standard options for the scaling of variables, the regression analysis was carried out using the full cross-validated partial least squares (PLS) method (leave-one-out) (15). The minimum sigma (column filtering) was set to 2.0 kcal mol⁻¹ to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. The final model, non-cross-validated conventional analysis, was developed with the optimum number of components to yield a non-cross-validated r^2 value.

CoMSIA studies

In CoMSIA, a distance-dependent Gaussian-type physicochemical property was adopted to avoid singularities at the atomic positions and dramatic changes to potential energy for those grids in the proximity of the surface. With the standard parameters and no arbitrary cut-off limits, five physicochemical properties, namely, steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, were calculated. The properties were calculated using a C+ probe atom with a radius of 1×10^{-10} m placed at a regular grid spacing of 2×10^{-10} m to enclose all the binding conformations

of the inhibitors. In general, the similarity index $A_{F,K}$ between the compounds of interest was computed by placing a probe atom at the intersections of lattice points using Eq. (1):

$$A_{F,K}^q(j) = - \sum_{i=1}^n W_{\text{probe},k} W_{ik} e^{-\alpha r_{iq}^2} \quad (1)$$

where q represents a grid point, i is the summation index over all atoms of molecule j under computation, W_{ik} is the actual value of the physicochemical property k of atom i , and $W_{\text{probe},k}$ is the value of the probe atom. In the present study, we used a probe atom ($W_{\text{probe},k}$) with charge +1, radius 10^{-10} m, hydrophobicity +1, and attenuation factor (α) of 0.3 for the Gaussian type distance (16). Statistical evaluation for the CoMSIA analysis was performed in the same way as described for CoMFA.

Partial least squares analysis

In 3D-QSAR, the CoMFA and CoMSIA descriptors were used as independent variables and pIC_{50} as the dependent variable. To quantify the relationship between these structural parameters (CoMFA and CoMSIA interaction energies) and biological activities, the PLS algorithm was used. The CoMFA descriptors were used as independent variables, and pIC_{50} values as dependent variables in partial least square regression analysis. Cross-validation partial least square method of leave-one-out (LOO) was performed to obtain the optimal number of components used in subsequent analyses. The minimum sigma (column filtering) was set to 2.0 kcal mol⁻¹ to improve the signal-to-noise ratio. The optimum number of principle components in the final non-cross-validated QSAR equations was determined to be that leading to the highest coefficient (r^2) and the lowest standard error in the LOO cross-validated predictions. The non-cross-validation was used in the analysis of CoMFA results and prediction of the model. The same method was also used for CoMSIA; thereafter a full PLS was run using column filtering of 1.0 kcal mol⁻¹ (17). Auto scaling was applied to all CoMSIA analysis. The cross-validated correlation coefficient (q^2), which resulted in optimum number of components and lowest standard error of prediction, was considered for further analysis and calculated using Eq. (2):

$$q^2 = 1 - \frac{\sum_{\gamma} (\gamma_{\text{pred}} - \gamma_{\text{actual}})^2}{\sum_{\gamma} (\gamma_{\text{actual}} - \gamma_{\text{mean}})^2} \quad (2)$$

where γ_{pred} , γ_{actual} , and γ_{mean} are the predicted, actual, and mean values of the target property (pIC_{50}), respectively.

The predicted ability of generated 3D-QSAR models was determined using a test set of 14 molecules that were separated during model generation. Test set molecules were chosen randomly considering a wide range of activity. Energy minimization and alignment of these 14 molecules were the same as the training set molecules described previously, and their activities were predicted using the model generated from the training set. The predicted correlation coefficient (r^2_{pred}) based on test set molecules, was computed using Eq. (3).

$$r^2_{\text{pred}} = \frac{SD - \text{PRESS}}{SD} \quad (3)$$

where SD is the sum of squared deviations between the biological activities of the test set and mean activities of the training molecules and $PRESS$ is the sum of squared deviations between the predicted and actual activities of the test set molecules.

RESULTS AND DISCUSSION

Homology modeling and molecular dynamic simulation

The sequence was aligned based on the structure. Based on sequence similarity analysis between the target and other proteins, known structures showed that the template (3HV8) had 63 % of amino acid sequence identity with our predicted model C-C chemokine receptor type 2 protein. Practically, at this level of sequence identity, it is satisfactory enough to use the crystallographic structure of 3HV8 as a template in order to obtain high quality alignment for structure prediction by homology modeling. In this study, ϕ and ψ torsion angles were also checked using the Ramachandran plots. Comparison of the results shows that one of the models generated by the MODELLER9v7 program maintained by Ben Webb (San Francisco, USA) is more acceptable. The best model predicted by MODELLER was used for further analysis by PROCHECK (European Bioinformatics Institute, Cambridge, UK). The overall general similarities and subtle differences between the 3D structure of template 3HV8 and the predicted model can be seen from the backbone superposition. PROCHECK statistics indicated that approximately 98 % of residues in the CXCR2 model were either in the most favored or in the additionally allowed regions; 1.5 % of residues were in the generously allowed regions, and 0.3 % of the residues were in the disallowed regions. As evident from superposition, the general folding topology of the structure is similar; however, some structural differences appear between the predicted model and the template. These differences are mainly due to insertions and deletions in different loop regions (18). The CXCR2 modeled protein interacting with the most active compound **65** is shown in Fig. 2.

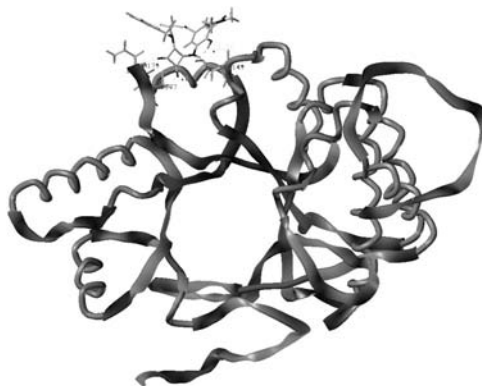


Fig. 2. The CXCR2 homology modeled protein interacting with the most active compound **65**.

Molecular docking

The most active compound **65** was docked into the receptor site along with the remaining fifty five molecules by using FlexX. The crystal structure (PDB ID: 3HV8) was used for docking. The ligand with all water molecules was deleted and Gasteiger-Hückel charges were assigned. The structure was then minimized using the conjugate gradient algorithm for 5,000 steps with no initial optimization, using the Tripos force field. The non-bonding cut-off was set to 15×10^{-10} m and a distance dependent dielectric constant was applied. All atoms of the protein were treated as aggregates, with the exception of those within the 15×10^{-10} m radius of the bound ligand. Based on rigid-body docking by FlexX, protein and ligand were analyzed for shape complementarity, hydrophobic effects resulting from a decrease in the solvent accessible surface, and electrostatic interactions. Docking studies yielded crucial information concerning the orientation of inhibitors in the binding pocket of the receptor. The key amino acid residues within the docking complex model involved in the interaction between the predicted CXCR2 and the three compounds (most active, least active and highest dock score) were THR149, VAL161, ASN179 and ASN182. Out of all the compounds, docking interactions with three compounds, **65** (most active compound), **14** (least active compound) and **24** (highest dock score compound) are shown in the Figs. 3a–c, respectively.

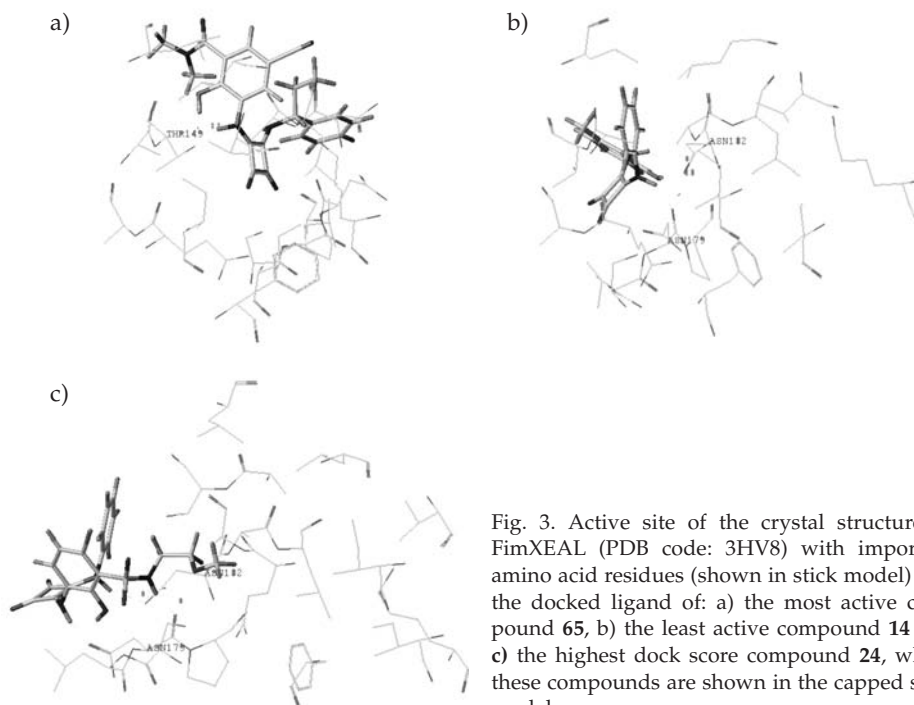


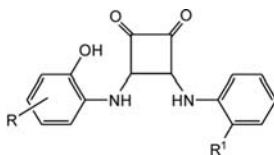
Fig. 3. Active site of the crystal structure of FimXEAL (PDB code: 3HV8) with important amino acid residues (shown in stick model) and the docked ligand of: a) the most active compound **65**, b) the least active compound **14** and c) the highest dock score compound **24**, where these compounds are shown in the capped stick model.

3D-QSAR analysis

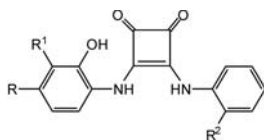
CoMFA analysis. – Forty-two compounds out of fifty six derivatives were used in the training set and fourteen compounds were used as the test set. PLS analysis was carried out for the training set and a cross-validated q^2 of 0.709 for five components was obtained. The non-cross-validated PLS analysis with the optimum components revealed $r^2 = 0.951$, F value = 139.903 and standard error of estimate (SEE) 0.144. Bootstrap analysis for 100 runs was then carried out for further validation of the model by statistical sampling of the original dataset to create new datasets.

Thus, the difference in the parameters calculated from the original data and the average of the parameters calculated from the N (=100) runs of bootstrapping sampling was a measure of the bias of the original calculation. This yielded higher $r^2_{\text{bootstrap}}$ value 0.978 for CoMFA with a standard error value of 0.096. This further supports the statistical validity of the developed models. The structures, predicted activities for the derivatives *vs.* their experimental values are listed in Table I (N,N' -diarylsquaramide derivatives) and Table II (3,4-diaminocyclobutene derivatives) and the correlation between the predicted activities and experimental values is depicted in Fig. 4a. Illustrated predicted activities using the CoMFA model are in good agreement with the experimental data, suggesting that the CoMFA model should have a satisfactory predictive ability.

Table I. Structure, experimental and predicted activity along with the docking scores of N,N' -diarylsquaramides derivatives (series-1)



Compd. No.	R	R ¹	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
1	H	H	235	6.63	6.67	-0.04	6.70	-0.07	-23.8
2a	3-CN	H	32	7.50	6.94	0.56	7.21	0.29	-23.2
4a	4-CN	H	15	7.82	6.98	0.84	7.18	0.64	-22.3
6	4-Me	H	205	6.69	7.09	-0.40	7.09	-0.40	-24.7
8	4-NO ₂	H	22	7.66	7.27	0.39	7.16	0.50	-24.8
10	4-NO ₂ , 5-Cl	H	52	7.28	7.50	-0.21	7.42	-0.14	-25.0
12	5-Me	H	308	6.51	6.79	-0.27	7.01	-0.50	-24.4
13	5-Cl	H	54	7.27	6.88	0.39	6.59	0.68	-23.6
14	6-Me	H	1720	5.76	6.68	-0.92	6.86	-1.10	-22.4
15	6-NO ₂	H	358	6.45	5.86	0.58	6.04	0.41	-22.5
16	H	Br	51	7.29	6.96	0.33	7.34	-0.05	-22.0
17	4-CN	Br	36	7.44	7.47	-0.03	7.20	0.22	-21.2
19	4-NO ₂	Br	37	7.43	7.60	-0.17	7.98	-0.51	-24.2



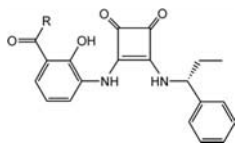
Compd. No.	R	R ¹	R ²	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
21	Cl	H	Br	32	7.45	7.56	-0.06	7.37	0.13	-20.7
22a	CN	H	Br	15	7.82	7.46	0.36	7.39	0.43	-27.8
23	Me	SO ₂ NMe ₂	Me	15	7.83	7.45	0.37	7.28	0.54	-21.9
24	Cl	SO ₂ NH(CH ₂) ₂ OMe	H	8	8.10	8.20	-0.10	8.22	-0.12	-29.3
25	F		Br	13	7.89	7.79	0.09	8.02	-0.13	-22.5
26	Cl		<i>n</i> -Pr	8	8.10	7.87	0.23	7.99	0.11	-20.7
28	Cl		Br	10	8.00	7.76	0.24	7.90	0.103	-24.1
29a	Cl		Cl	20	7.70	8.01	-0.31	8.13	-0.43	-25.8
30a	Cl		H	18	7.74	7.23	0.51	7.35	0.39	-22.5
31	Cl	CONH ₂	H	14	7.85	7.78	0.08	7.72	0.13	-26.1
35	Cl		Br	28	7.55	7.42	0.13	7.57	-0.02	-28.0
36a	H		H	27	7.57	8.08	-0.51	7.95	-0.38	-24.9
37	H		H	44	7.36	7.51	-0.16	7.24	0.11	-27.2

a – Test set compounds.

CoMSIA analysis. – CoMSIA analyses were performed using five descriptor fields: steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor, where the best q^2 was found by using all five different descriptor variables. This demonstrates that these variables are necessary to describe the interaction mode of the *N,N'*-diarylsquaramides, *N,N'*-diarylureas and diaminocyclobutenediones inhibitors with CXCR2, as well as the field properties around the inhibitors. The predicted binding affinities derived from Co-

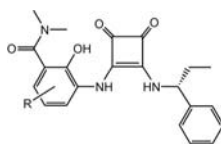
Table II. Structure, experimental and predicted activity along with the docking scores of 3,4-diaminocyclobutenediones (series-2)

Compd. No.	Structure	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
38		5	8.30	8.30	0.04	8.2	0.08	-21.1
39		12	7.92	7.91	0.007	8.16	-0.24	-21.8
40a		25	7.60	7.72	-0.12	8.11	-0.51	-22.4
42		22	7.66	7.63	0.027	7.50	0.20	-19.0

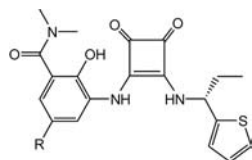


Compd. No.	R	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
46a		28	7.56	7.72	-0.29	7.95	-0.39	-21.3
47		219	6.66	6.82	-0.16	6.45	0.21	-21.3
48		34	7.47	7.60	-0.17	7.77	-0.30	-20.9
49a		4.4	8.36	7.91	0.45	7.50	0.86	-20.8
50		11	7.96	7.77	0.19	7.80	0.16	-21.5
51		9	8.04	7.90	0.14	8.21	-0.16	-18.4
52a		10	8.00	7.64	0.36	7.74	0.26	-21.1
53		12	7.92	7.65	0.27	7.44	0.45	-21.7
54		20	7.70	7.61	0.09	7.81	-0.11	-21.7

55		18	7.74	7.76	-0.02	7.90	-0.19	-20.5
56a		20	7.70	7.20	0.50	7.31	0.39	-18.7
57a		5	8.30	7.78	0.52	7.91	0.39	-25.0
58		27	7.59	7.72	-0.15	7.76	-0.19	-21.5
59		20	7.70	7.77	-0.07	7.67	0.03	-22.0



Compd. No.	R	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
60a	H	4.4	8.36	7.88	0.48	8.19	0.55	-21.1
61	4-bromo	4	8.40	7.98	0.42	7.77	0.63	-23.4
62	5-methyl	203	6.70	6.90	-0.22	6.90	-0.20	-21.3
63a	4-phenyl	17	7.77	8.03	-0.26	8.19	-0.42	-20.0
64	4-(2-pyridyl)	15	7.80	7.89	-0.06	7.94	-0.11	-22.5
65	5-cyano	3	8.52	7.92	0.60	7.95	0.57	-18.7
67	6-methyl	4.5	8.35	8.04	0.31	8.19	0.16	-20.1



Compd. No.	R	IC ₅₀ (mmol L ⁻¹)	pIC ₅₀	CoMFA predicted	CoMFA residual	CoMSIA predicted	CoMSIA residual	Docking score
68	H	5	8.30	7.87	0.43	8.08	0.23	-20.2
69	cyano	5	8.30	8.06	0.24	7.96	0.34	-17.3
70	chloro	50	7.30	7.80	-0.55	7.61	-0.30	-21.1
71	bromo	61	7.20	7.80	-0.60	7.86	-0.64	-20.3
72	methyl	76	7.12	7.48	-0.36	7.07	-0.58	-20.6

a – Test set compounds.

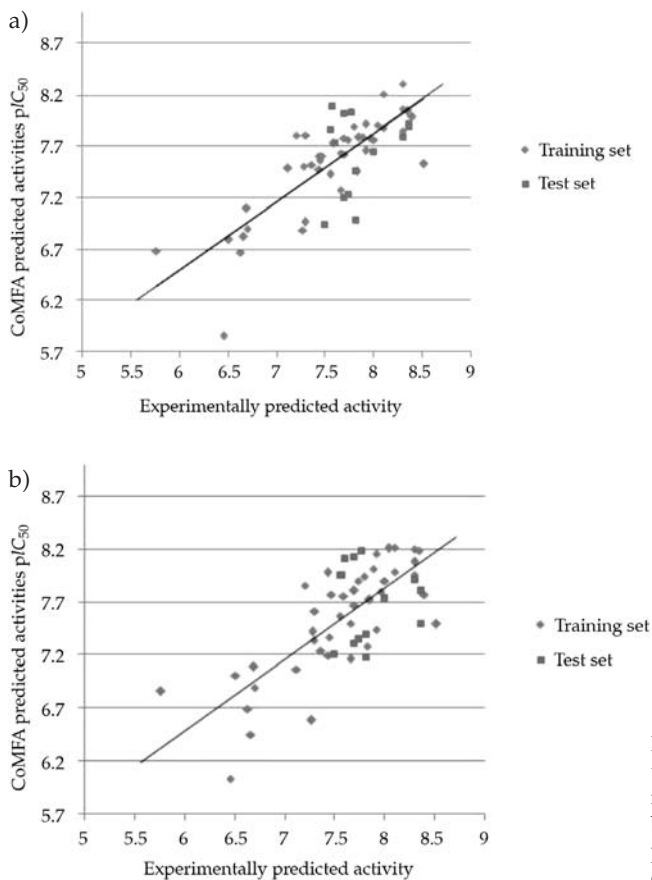


Fig. 4. a) Plot of predicted vs. actual pIC_{50} values of training set molecules for the CoMFA model, b) plot of predicted vs. actual pIC_{50} values of training set molecules for the CoMSIA model.

Table III. Statistical results of CoMFA and CoMSIA models

Component	CoMFA		CoMSIA	
q^2	0.709		0.592	
r^2	0.951		0.955	
N	5		6	
F -value	139.903		122.399	
SEE	0.144		0.141	
CV	0.681		0.641	
Boot-strap	mean	std. dev.	mean	std. dev.
SEE	0.096	0.063	0.0102	0.062
r^2	0.978	0.011	0.973	0.010

q^2 – LOO-cross-validated correlation coefficient; r^2 – non-cross-validated correlation coefficient; N – number of components used in the PLS analysis; SEE – standard error of estimation; F -value – F -statistic for the analysis

MSIA analysis are also listed in Tables I and II and the correlation between the predicted activities and experimental values is depicted in Fig. 4b. The CoMSIA study revealed leave-one-out $q^2 = 0.592$, non-cross-validated $r^2 = 0.955$, SEE 0.141 and $F = 122.399$. The PLS analysis results of CoMFA and CoMSIA are summarized in Table III.

CoMFA and CoMSIA contour analysis

The CoMFA steric and electrostatic fields from the final non-cross-validated analysis were plotted as 3-D contour maps with the field energies at each lattice point calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table ($SD \cdot \text{coeff}$), always plotted as the percentages of the CoMFA equation contribution. These maps show regions where differences in molecular fields are associated with the differences in biological activity. The CoMFA contours for steric and electrostatic fields are shown in Figs. 5a,b, while those of CoMSIA steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields are shown in Figs. 6a–e, respectively.

The most active compound **65** is shown superimposed with CoMFA steric and electrostatic contour maps in Figs. 5a,b. Two colored isopleths were present, in which a large region was present near the R substituted phenyl ring. Here the steric counter is present around the C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃ carbons, indicating the favorability of the sterically bulky group for activity and accounting for the better activity of compounds **61**, **67**, **68** and **69** compared to that of the other molecules. On the other hand, regions indicated by isopleths correspond to regions for unfavorable presence of steric groups and should be avoided for better activity of the molecule. Four isopleths were observed around the compound, indicating the removal of bulky groups for enhancing the activity of this compound. For the electrostatic contour map, a few isopleths were observed near the R substituted phenyl ring indicating that the electronegative substitution was favorable for activity. This is consistent with that of the steric contour, providing confirmation that a bulky electronegative substituent is highly favorable. The contour around the phenyl ring indicates that the electropositive environment is desirable at this position and hence the electropositive environment -NH group occupying this position is present in all potent compounds.

The CoMSIA steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor contour map for the highly active compound **65** is shown in Fig. 6. The CoMSIA steric field is more or less similar to that of the corresponding CoMFA fields. In Fig. 6a, large

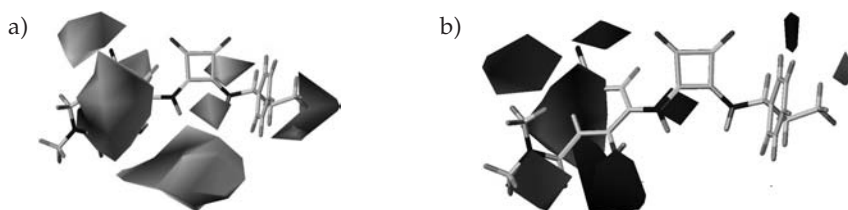


Fig. 5. CoMFA contour maps: a) steric field distribution for the highly active compound, b) electrostatic field distribution for the highly active compound.

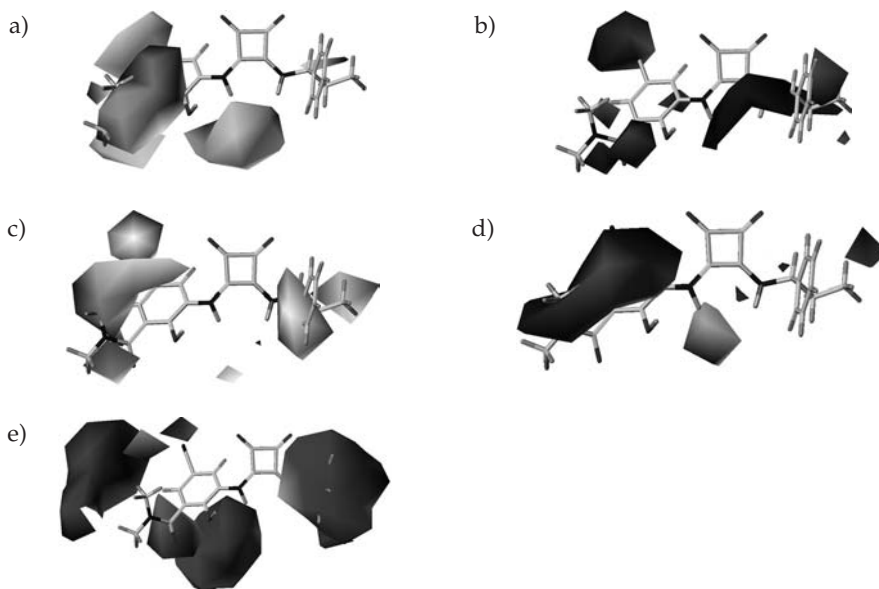


Fig. 6. CoMSIA contour maps: a) steric field distribution for the highly active compound, b) electrostatic field distribution for the highly active compound, c) hydrophobic field distribution for the highly active compound, d) hydrogen bond donor field distribution for the highly active compound, e) hydrogen bond acceptor field distribution for the highly active compound.

unfavorable isopleths and large favorable isopleths are present near the R substituted phenyl ring, indicating the presence of both unfavorable and favorable conditions for bulky groups at this position. Here, the steric counter is present at C₉, C₁₀, C₁₁, C₁₂, C₁₈, N₂₈ and another counter is present near N₇, N₂₁, C₂₀ atoms. Placement of bulkier and lighter group counters might increase the activity of the molecule. Similar variation can be also observed from the CoMSIA plot (Fig. 6b), where many small regions of two different contour regions are present on the most active compound 65. Here, the prominent negative counter is present at C₂₀, N₂₁ and the positive counter is present at N₁₄. Adding more negative groups (like Cl⁻, Br⁻) and positive R groups (like NH₃⁺) in their respective counter regions might increase the activity of the molecule.

The CoMSIA hydrophobic contour maps are shown in Fig. 6c, which are similar to that of CoMSIA electrostatic isopleths; representing the presence of favored isopleths near the R substituted phenyl ring and -NH position indicates that any bulk group present at this position will represent hydrophobically favored regions. Three small regions on the benzene ring indicate that hydrophilic groups will be favorable, thus compounds that have the NH group at that position will have much greater potency than compounds without the NH group. Hydrogen bond donor interactions are represented in Fig. 6d. The contours represent regions that favor hydrogen bond donors, as well as contours with disfavored regions. The presence of a large contour isopleth near the R substituted phenyl ring indicates the presence of an unfavorable region for hydrogen bond donor

groups. There is only a small contour isopleth present near the –NH group, favorable for the donor group around this region. Hydrogen-bond donor and acceptor contour maps from CoMSIA are shown in Fig. 6e. The contour isopleths favor the positioning of the hydrogen bond acceptor group in these regions whereas a few isopleths disfavored the presence of H-bond acceptor groups. The hydrogen-bond acceptor contour map is shown with three large contours in which the hydrogen-bond acceptor is not favored. Contour regions are mainly positioned at the –NH group and also near the R substituted phenyl ring. Only a small contour region is present near the phenyl ring, which favors the H-bond acceptor group around this region.

In silico screening

In silico methods can help identify drug targets *via* bioinformatics tools. They can be also used to analyze the target structures for possible binding active sites, generate candidate molecules, check their drug likeness, dock these molecules with the target, rank them according to their binding affinities, further optimize molecules to improve their binding characteristics. Use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. Use of complementary experimental and informatics techniques increases the chance of success in many stages of the discovery process, from the identification of novel targets and elucidation of their functions to the discovery and development of lead compounds with desired properties. Thus, the *in silico* procedure can be applied as a physicochemical filter to reduce the number of compounds to be tested experimentally for hit/lead generation. In other words, the *in silico* procedure minimizes the time and cost associated with identifying new leads. Virtual screening was performed by insertion, deletion and substitution of different substituents on the original molecules and the effects of structural modifications on the biological activity were investigated (19). Then, the domain of QSAR model application was defined for using the model for screening new compounds. The newly designed molecule showed higher activity than the most active molecule, as shown in Fig. 7. The designed molecule contains a negatively charged Br⁻ at C₂₃ and a bulky pentane ring instead of two methyl groups near N₂₈. Adding these two groups to the most active molecule helped in increasing the antagonistic activity of the molecule.

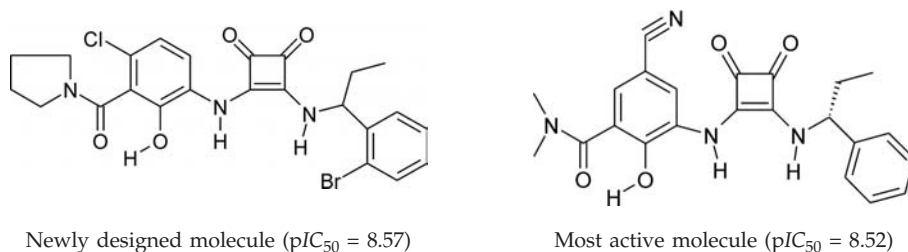


Fig. 7. Comparison of the newly designed molecule with the most active molecule.

CONCLUSIONS

A receptor independent 3D-QSAR has been established for CXCR2 antagonistic molecules employing the most widely used techniques CoMFA and CoMSIA. The present studies highlight the importance of ligand orientation and selection of the training set molecules in the development of statistically significant QSAR models. CoMSIA models provided better statistical models than CoMFA, which points to the significance of the hydrogen bond donor and hydrophobic fields in the selectivity and activity of these ligands in addition to steric and electrostatic fields. Statistical significance and robustness of the generated 3D-QSAR models were confirmed using an external set of molecules. The structural requirements identified in the present study can be utilized strategically in designing novel, potent CXCR2 antagonistic molecules.

REFERENCES

1. A. Zlotnik and O. Yoshie, Chemokines: a new classification system and their role in immunity, *Immunity* **12** (2000) 121–127; DOI: 10.1016/S1074-7613(00)80165-X.
2. L. Yang, C. Zhou, L. Guo, G. Morriello, G. Butora, A. Pasternak, W. H. Parsons, S. G. Mills, M. MacCoss, P. P. Vicario, H. Zweerink, J. M. Avala, S. Goval, W. A. Hanlon, M. A. Cascieri and M. S. Springer, Discovery of 3,5-bis(trifluoromethyl) benzyl L-arylglycinamide based potent CCR2 antagonists, *Bioorg. Med. Chem. Lett.* **16** (2006) 3735–3739; DOI: 10.1016/j.bmcl.2006.04.045.
3. J. A. Belperio, M. P. Keane, D. A. Arenberg, C. L. Addison, J. E. Ehlert, M. D. Burdick and R. M. Strieter, CXC chemokines in angiogenesis, *J. Leukocyte Biol.* **68** (2000) 1–8.
4. P. Loetscher, M. Seitz, I. Clark-Lewis, M. Baggiolini and B. Moser, Both interleukin-8 receptors independently mediate chemotaxis: Jurkat cells transfected with IL-8R1 or IL-8R2 migrate in response to IL-8, GRO α and NAP-2, *FEBS Lett.* **341** (1994) 187–192; DOI: 10.1016/0014-5793(94)80454-0.
5. J. C. Acosta, A. O’Loughlen, A. Banito, M. V. Guijarro, A. Augert, S. Raguz, M. Fumagalli, M. Da Costa, C. Brown, N. Popov, Y. Takatsu, J. Melamed, F. d’Adda di Fagagna, D. Bernard, E. Hernandez and J. Gil, Chemokine signaling via the CXCR2 receptor reinforces senescence, *Cell* **133** (2008) 1006–1018; DOI: 10.1016/j.cell.2008.03.038.
6. J. Busch-Petersen, Small molecule antagonists of the CXCR2 and CXCR1 chemokine receptors as therapeutic agents for the treatment of inflammatory diseases, *Curr. Med. Chem.* **6** (2006) 1345–1352.
7. R. D. Cramer III, D. E. Patterson and J. D. Bunce, Comparative molecular-field analysis (COMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.* **110** (1988) 5959–5967; DOI: 10.1021/ja00226a005.
8. G. Klebe, U. Abraham and T. Mietzner, Molecular similarity indexes in a comparative-analysis (COMSIA) of drug molecules to correlate and predict their biological-activity, *J. Med. Chem.* **37** (1994) 4130–4146; DOI: 10.1021/jm00050a010.
9. B. W. McClelland, R. S. Davis, M. R. Palovich, K. L. Widdowson, M. L. Werner, M. Burman, J. J. Foley, D. B. Schmidt, H. M. Sarau, M. Rogers, K. L. Salyers, P. D. Gorycki, T. J. Roethke, G. J. Stelman, L. M. Azzarano, K. W. Ward and J. Busch-Petersen, Comparison of *N,N'*-diarylsquaramides and *N,N'*-diarylureas as antagonists of the CXCR2 chemokine receptor, *Bioorg. Med. Chem. Lett.* **17** (2007) 1713–1717; DOI: 10.1016/j.bmcl.2006.12.067.

10. C. Aki, J. Chao, J. A. Ferreira, M. P. Dwyer, Y. Yu, J. Chao, R. J. Merritt, G. Lai, M. Wu, R. W. Hipkin, X. Fan, W. Gonsiorek, J. Fosseta, D. Rindgen, J. Fine, D. Lundell, A. G. Taveras and P. Biju, Diaminocyclobutenediones as potent and orally bioavailable CXCR2 receptor antagonists: SAR in the phenolic amide region, *Bioorg. Med. Chem. Lett.* **19** (2009) 4446–4449; DOI: 10.1016/j.bmcl.2009.05.049.
11. A. Fiser, R. K. Do and A. Sali, Modeling of loops in protein structures, *Protein Sci.* **9** (2000) 1753–1773; DOI: 10.1110/ps.9.9.1753.
12. R. A. Laskowski, J. A. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR* **8** (1996) 477–486; DOI: 10.1007/BF00228148.
13. M. Rarey, B. Kramer, G. Metz, T. Lengauer and G. Klebe, A fast flexible docking method using an incremental construction algorithm, *J. Mol. Biol.* **261** (1996) 470–489; DOI: 10.1006/jmbi.1996.0477.
14. K. M. Gilbert, W. J. Skawinski, M. Misra, K. A. Paris, R. A. Buono, H. M. Deutsch and C. A. Venanzi, Conformational analysis of methylphenidate: comparison of molecular orbital and molecular mechanics methods, *J. Comput. Aided Mol. Des.* **18** (2004) 719–738; DOI: 10.1007/s10822-004-7610-1.
15. S. Wold, Cross-validated estimation of the number of components in factor and principal components models, *Technometrics* **20** (1978) 397–405.
16. G. Klebe, The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands, *J. Mol. Biol.* **237** (1994) 212–235; DOI: 10.1006/jmbi.1994.1223.
17. R. D. Cramer III, J. D. Bunce, D. E. Patterson and I. S. Frank, Crossvalidation, bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies, *Quant. Struct.-Act. Rel.* **7** (1988) 18–25; DOI: 10.1002/qsar.19880070105.
18. Y. Zhang and J. Skolnick, Scoring function for automated assessment of protein structure template quality, *Proteins* **57** (2004) 702–710; DOI: 10.1002/prot.20264.
19. G. Melagraki, A. Afantitis, H. Sarimveis, P. A. Koutentis, J. Markopoulos and O. Igglessi-Markopoulou, Optimization of biaryl piperidine and 4-amino-2-biarylurea MCH1 receptor antagonists using QSAR modeling, classification techniques and virtual screening, *J. Comput. Aided Mol. Des.* **21** (2007) 251–267; DOI: 10.1007/s10822-007-9112-4.

S A Ž E T A K

Strukturno istraživanje antagonista CXCR2 receptora pomoću CoMFA, CoMSIA i fleksibilnih doking studija

SHRAVAN KUMAR GUNDA, ROHITH KUMAR ANUGOLU, SRI RAMYA TATA i SAIKH MAHMOOD

U radu je opisano trodimenzijsko ispitivanje kvantitativnog odnosa strukture i antagonističkog djelovanja na CXCR2 receptore (3D QSAR) u seriji od 56 *N,N'*-diarilskvaramida, *N,N'*-diarilurea i diaminociklobutendiona. Provedene studije uključuju komparativnu analizu molekulskih polja (CoMFA) i komparativnu analizu indeksa molekulske sličnosti (CoMSIA). Modeli s dobrom predvidljivošću izvedeni su pomoću CoMFA $q^2 = 0,709$, r^2 (neunakrsno-validirani kvadrat koeficijenta korelacije) = 0,951, $F = 139,903$, r^2 bs = 0,978 s pet komponenata, standardnom pogreškom procjene 0,144 i CoMSIA $q^2 = 0,592$, $r^2 = 0,955$, $F = 122,399$, r^2 bs = 0,973 sa šest komponenata, standardnom pogreškom pro-

cjene 0,141. Osim toga, model homologije CXCR2 receptora upotrijebljen je za svrstavanje spojeva doking metodom. Najaktivniji spoj poslužio je kao predložak za svrstavanje preostalih struktura. Nadalje, mapiranjem kontura na aktivno mjesto spojevi su se uzajamno vrednovali kroz ostatke s obzirom na svoje konture. Ovo integrirano svrstavanje na temelju molekuskog dokinga i naknadnih 3D QSAR studija pomoglo je dizajniranju CXCR2 antagonista s poboljšanom aktivnosti. *In silico* pretraživanje prilagođeno je QSAR modelu sa svrhom predviđanja strukture novih, potencijalno aktivnih spojeva.

Ključne riječi: CXCR2, model homologije, CoMFA, CoMSIA, PLS, doking

Bioinformatics Division, Osmania University, Hyderabad-500007, Andhra Pradesh, India