INTERNATIONAL JOURNAL OF INSTITUTIONAL PHARMACY AND LIFE SCIENCES

Life Sciences

Research Article……!!!

Received; accepted

3D-QSAR COMFA, COMSIA & DOCKING STUDIES ON ARYL SULFONYL PIPERAZINE DERIVATIVES AS 5-HT6 RECEPTOR LIGANDS

Uday Chandra Kumar* and Mahmood Shaik

Bioinformatics Division, Environmental Microbiology Lab, Department of Botany, Osmania University, Hyderabad 500 007, A.P., India

Keywords:

Comparative Molecular Field Analysis; Comparative Molecular Similarity Indices Analysis; 3D-QSAR; Aryl sulfonyl piperzine; 5-HT6 receptor

For Correspondence:

Uday Chandra Kumar

Bioinformatics Division, Environmental Microbiology Lab, Department of Botany, Osmania University, Hyderabad 500 007, A.P., India

E-mail:

dr.udaynair16@gmail.com

ABSTRACT

Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA), Docking studies were performed on a series of Arylsulfonylpiperzine derivatives as 5-HT_6 receptor derivatives. Ligand molecular superimposition on the template structure was performed by the atom/shape based root mean square fit and database alignment methods. Training set of 27 molecules improved which were validated by a test set of 9 compounds. The atom and shape based root mean square alignment yielded the best predictive CoMFA model q^2 cv = 0.783 , r^2 (non-cross-validated square of correlation coefficient) = 0.910, *F* value = 42.494, r^2 bs = 0.958 with five components, standard error of estimate = 0.051 and while the CoMSIA model yielded q^2 cv = 0.777, r^2 (non-cross- -validated square of correlation coefficient) = 0.996, *F* value = 1508.216, r^2 bs = 0.998 with six components, standard error of estimate = 0.050. The contour maps obtained from 3d-QSAR studies were appraised for activity trends for the molecules analyzed. Results that indicate steric, electrostatic, hydrophobic (lipophilic) and hydrogen bond donor substituents play a significant role in 5-HT_6 inhibitory activity and selectivity of the compounds. The data generated from the present study will further helpful for designing new novel, potent 5-HT_6 receptor derivatives.

INTRODUCTION

Serotonin (5-hydroxytryptamine, 5-HT) is a major neurotransmitter which interacts with 5-HT receptors to produce a lot of its effects. There are 15 different human serotonin receptors that have been cloned and divided into 7 subclasses (5-HT1-7). The 5-HT6 receptor is one of the latest subtypes to have been identified and belongs to the G protein coupled receptor (GPCR) superfamily, which is positively coupled to adenylate cyclase. 5-HT6 receptors are present in brain regions such as the cerebral cortex, nucleous accumbens, caudate-putamen and hippocampus with high densities and thalamus and substantia nigra with low densities, which are associated with learning and memory. Actually, compromised serotonergic function may have an important contribution to cognitive decline related to aging, Alzheimer's disease (AD) and schizophrenia. Thus serotonergic system became a potential target for the treatment of memory dysfunction. It is also suggested that the 5-HT6 receptor has a major role in obesity, based on the knockout mice study that 5-HT6 receptor-knockout mice are resistant to weight gain when exposed to a high-fat diet. The 5-HT6 receptor antagonists, designed, synthesized and biologically evaluated by Hun Yeong et.al as arylsulfonylpiperazine derivatives [1-3] as novel 5-HT6 receptor ligands.

The present study deals with these reported ligands which were used to extract descriptors and predict properties using 3d-QSAR, CoMFA, CoMSIA and then the most active molecule and also the least active molecule was docked to a protein of interest using SYBYL(6.7).

The binding mode of aryl sulfonyl piperazine was studied on human recombinant 5-HT6 serotonin receptor stably expressed by HEK293 cell line through [3H]-lysergic acid diethylamide (LSD) binding assay. This technique was experimented by Hun Yeong et.al. The results are summarized in Table 1.

The log values of IC50 values (% inhibition at 10 μM) were evaluated using equation (1) in MS excel sheet (MS OFFICE application):

$$\text{Log (IC50 values)} = 9 - \log (RC^*)$$

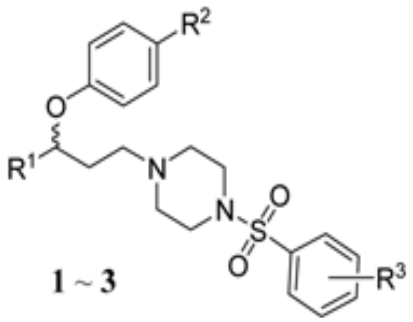
* RC represents location of Rows & Columns of excel spreadsheet.

The –log(IC50 values) obtained were used for predicting properties of the compounds using COMFA AND COMSIA. R1, R2, R3 represents the side chains of the compound that is synthesized as aryl sulfonyl piperazine. The derivatives such as phenyl, ethyl, H etc., mentioned in Table 1 were assigned as side-chains. The derivatives were assigned to the aryl sulfonylpiperazine compound. These structures were sketched, cleaned up and were minimized in a molecular modeling software called as SYBYL (6.7)

which runs exclusively on Linux OS. For minimization, force field was set as "TRIPOS", charges as "Gasteiger-Huckel" and with 10000 iterations. The minimized molecules were saved in a database.

The objective of this research study is to predict and analyze the activity of 5-HT6 receptor antagonists using 3D-QSAR, CoMFA and CoMSIA using SYBYL (6.7). To predict and validate CoMFA and CoMSIA models by observing contour maps and to analyze and predict the binding mode of most active, moderate active and least active ligand to a protein of interest whose PDB ID is 1UNQ and present the importance scoring functions.

Table 1. IC50 values of the aryl sulfonyl piperazine derivatives against 5-HT6 receptor



1 ~ 3

COMPOUND	R1	R2	R3	% inhibition at 10 µM	IC50 (µM)
1a	phenyl	CF3	H	24	>10
1b	Phenyl	CF3	o-Cl	32	>10
1c	phenyl	CF3	m-Cl	30	>10
1d	Phenyl	CF3	p-Cl	23	>10
1e	Phenyl	CF3	o-F	52	6.0
1f	Phenyl	CF3	m-F	32	>10
1g	Phenyl	CF3	p-F	41	>10
1h	Phenyl	CF3	o-Me	41	>10

1i	phenyl	CF3	m-Me	34	>10
1j	Phenyl	CF3	p-Me	33	>10
1k	Phenyl	CF3	m-OMe	30	>10
1l	Phenyl	CF3	3,4-Me2	20	>10
2a	ethyl	CF3	H	58	4.2
2b	ethyl	CF3	o-Cl	32	>10
2c	Ethyl	CF3	m-Cl	59	4.3
2d	Ethyl	CF3	p-Cl	54	7.8
2e	Ethyl	CF3	o-F	43	>10
2f	Ethyl	CF3	m-F	55	4.7
2g	Ethyl	CF3	p-F	44	>10
2h	Ethyl	CF3	o-Me	60	1.5
2i	ethyl	CF3	m-Me	50	7.1
2j	Ethyl	CF3	p-Me	58	6.2
2k	Ethyl	CF3	m-OMe	50	9.6
2l	Ethyl	CF3	3,4-Me2	47	>10
3a	Ethyl	CN	H	29	>10
3b	Ethyl	CN	o-Cl	37	>10
3c	Ethyl	CN	m-Cl	43	>10
3d	Ethyl	CN	p-Cl	37	>10
3e	Ethyl	CN	o-F	60	3.6

3f	Ethyl	CN	m-F	67	2.4
3g	Ethyl	CN	p-F	48	>10
3h	Ethyl	CN	o-Me	55	3.7
3i	Ethyl	CN	m-Me	68	2.4
3j	Ethyl	CN	p-Me	52	7.0
3k	ethyl	CN	m-OMe	58	6.7
3l	ethyl	CN	3,4-Me2	57	5.8

MATERIALS & METHODOLOGY

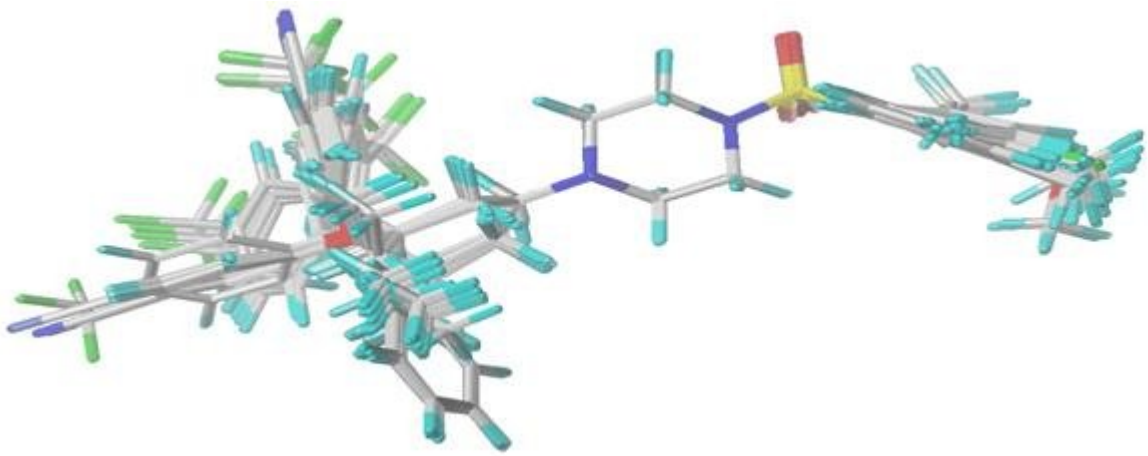
Data Set: In the present work, a total of 36 compounds with their 5-HT6 receptor inhibitory activity from the reported work [4] were selected. The reported compounds showed wide variations in their structures and potency profile. Considering a high deviation in the biological activity and structural variations among the compounds of the series it was considered as an ideal series for performing QSAR analysis. Biological data with negative logarithm of % inhibition expressed in µm was used as a dependent variable in the 3D-QSAR study, thus correlating the data linearly to the free energy change.

Selection of Training and Test Sets: In view of the finding that q2 appears to be a necessary but not a sufficient condition for a model to have high predictive power, [5] an emphasis has been give in the present study for validation of the developed model using an external test set. The whole set of 36 compounds was divided into training set (27) and test set (9) compounds. In the training set, most potent, moderately active an low active compounds were included to spread the activity range. The test set compounds were selected in such a manner that at least one structural analog of the training set was chosen for the test set.

Computational Details: A Silicon Graph is Fuel workstation was IRIX 6.5 operating system running SYBYL 6.7 (Sybyl 7.0; Truois Ubc. 1699 south Hanley Road, St. Louis, Missouri 63144, U.S.A) was used for three-dimensional structure building and molecular modeling studies. Initial optimization of the structures was carried out using TRIPOS force field with Gasteiger-Huckel charges, and repeated minimization was performed using steepest-descent and conjugate gradient methods until the root-mean-square (rms) deviation of 0.001 kcal/mol was achieved. Conformational energies were computed with

electrostatic terms; the lowest energy structures finally minimized were used in superimposition. The partial atomic charges required for the electrostatic interactions were computed by the semi empirical molecular orbital methods using Molecular Orbital package (MOPAC) with Austin Model 1 (AM1) Hamiltonian. [6]

Selection of Template and Molecular Alignment: In the development of 3D-QSAR models, the choice of the template conformation is the most important factor to provide a reliable pharmacophore model. This renders the spatial alignment of molecules under study as one of the most sensitive and determining factors in obtaining a robust and meaningful model. CoMFA and CoMSIA results may be extremely sensitive to a number of factors such as alignment rules, over all orientation of the aligned compounds, lattice shifting step size and probe atom. The accuracy of the prediction of CoMFA and CoMSIA models and the reliability of the contour models depend strongly on the structural alignment of the molecules [7] and thus we applied molecular alignment to align all the molecules used in present study. The molecular alignment was achieved by the SYBYL routine database align. The most active compound (12) was used as an alignment template and the rest of the molecules were aligned to it by using the common substructure as shown in below fig.



Molecular alignment of 36 arylsulphonyl piperazine derivatives.

3D QSAR studies: In order to have better understanding and explore the contributions of electrostatic, steric, and hydrophobic fields of the data set, and to build predictive 3D-QSAR models, CoMFA and CoMSIA studies were performed based on the molecular alignment as described. CoMFA calculates steric and electrostatic properties according to Lennard-Jones and coulomb potentials, respectively, whereas CoMSIA calculates similarity indices in the space surrounding each of the molecules in the dataset.

CoMFA Studies: To derive the CoMFA 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 sp3 carbon probe atom with a van der Walls radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard-Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. The steric and electrostatic energy values were truncated at a default value of 30 kcal/mol. The CoMFA steric and electrostatic fields generated were scaled by the COMFA standard option available in SYBYL.

CoMSIA Studies: Comparative molecular similarity indices analysis (CoMSIA) was performed to evaluate steric, electrostatic, Hydrophobic, hydrogen bond donor, and hydrogen bond acceptor properties of molecules by employing the standard options in SYBYL. The steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor fields were calculated separately using the sp3 carbon atom probe with a charge of +1 provided in SYBYL 6.7. Similar to CoMFA, a data table has been constructed from similarity indices calculated at the intersections of regularly spaced lattice (2 Å spacing). Similarity indices AF,K between the compounds of interest and a probe atom have been calculated according to eq. 3

$$A^{q}_{F,K}(j) = -\sum_{i=1}^{n} \omega_{probe,k}\omega_{ik}e^{-\alpha r^{2}_{iq}}$$

where q is the grid point for molecule j; ωik is the actual value of the physicochemical property k of atom I; ωprobe, k indicates probe atom with charge +1, radius 1 Å, hydrophobicity +1, H-bond donor and acceptor property +1; α is an attenuation factor; and riq is the mutual distance between the probe atom and grid point q and atom I of the test molecule. The default value of α is 0.3.

Molecular docking: The docking studies were carried out using the FlexX program [5] interfaced with SYBYL 6.7. In this automated docking program, the flexibility of the 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 minimized using PM3 method and all the charges were removed. The 3D coordinates of the active sites were taken from the X-ray crystal structures of the 5-HT6 were obtained from protein databank (PDB code 1UNQ). The PDB file obtained from protein data bank was used as a receptor site. All water molecules were removed and the protein was modified to dock inhibitor and also hydrogens were added.

The active site was defined with a distance of 6.5 Å around the co-crystallized ligand. Formal charges were assigned to all the molecules and FlexX run was submitted.

QSAR coefficient contour maps: The visualization of the results of the best CoMFA and CoMSIA models have been performed using the "StDev*Coeff" mapping option contoured by contribution. Favored and disfavored levels fixed at 80% and 20%, respectively. The contours of the CoMFA and CoMSIA steric maps are shown in green (more bulk is favored) and yellow (less bulk is favored). The electrostatic fields of both CoMFA and CoMSIA contours are colored blue (positive charge is favored) and red (negative charge is favored). The contours of the CoMSIA hydrophobic fields are colored yellow (hydrophobic groups enhance activity) and white (hydrophilic groups enhance activity). The hydrogen bond field contours show regions where hydrogen bond acceptors (magenta) on the receptor enhance the activity and hydrogen bond donor (cyan) increase the activity.

RESULTS AND DISCUSSION

A data set of 27 analogues was selected as a training set and to derive the conventional CoMFA, CoMSIA models. CoMFA and CoMSIA were carried out using the QSAR options of SYBYL. Steric and electrostatic fields of CoMFA were calculated using Lennard - Jones and Coulombic potential, respectively. The five physicochemical properties for CoMSIA (steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor) were evaluated in the QSAR options of SYBYL. To test the statistical significance of the models, cross- validations were done by means of the "leave – one – out" (LOO) procedure using the enhanced version of "partial least square" PLS, the SAMPLS method. Based on the optimal number of components, the final model was built using the result of non - cross - validation to predict the affinities of the compounds in the training set and test set.

CoMFA analysis: Twenty seven compounds out of the total 36 5-HT6 receptor were used as training set and nine compounds were used as test set. The test set compounds were selected manually so that the structural diversity and wide range of activity in the dataset were included. PLS analysis was carried out for the training set and a cross-validated q2 of 0.783 . The non cross-validated PLS analysis with the optimum components revealed a conventional r2 value of 0.910, F = 42.494 and an standard error of estimation = 0.051. Bootstrap analysis for 10 runs was then carried out for further validation of the model by statistical sampling of the original dataset to create new datasets.

CoMSIA analysis: The CoMSIA analyses were performed using five descriptor fields: steric, electrostatic, hydrophobic and hydrogen bond donor and acceptor. The CoMSIA study revealed a cross validated q2 of 0.770, a conventional r2 of 0.917 with a standard error of estimation = 0.050 and F =

36.942 for training set. Bootstrap analysis for 10 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 N(=10) runs of bootstrapping sampling is a measure of the bias of the original calculation.

Summary of 3D-QSAR analyses:

	CoMFA		CoMSIA	
q2	0.783		0.770	
r2	0.910		0.917	
SEE	0.051		0.050	
Boot Strap	Mean	Stddev	Mean	Stddev
SEE	0.035	0.029	0.039	0.030
r2	0.959	0.025	0.947	0.031
F value	42.494		36.942	

Contour analysis: The visualization of the results of the CoMFA and CoMSIA models have been performed using the StDev*Coeff mapping option contoured by contribution. The default level of contour with contribution, 80% for favored region and 20% for disfavored region was set during contour analysis.

CoMFA contour maps: In CoMFA the steric interaction is represented by green and yellow contours, while electrostatic interaction is denoted by red and blue contours. Favorable steric interactions are shown in green; sterically unfavorable regions are shown in yellow. Blue contours indicate regions where hydrophobic interactions enhance binding; red contours show regions where hydrophobic properties decrease affinity.

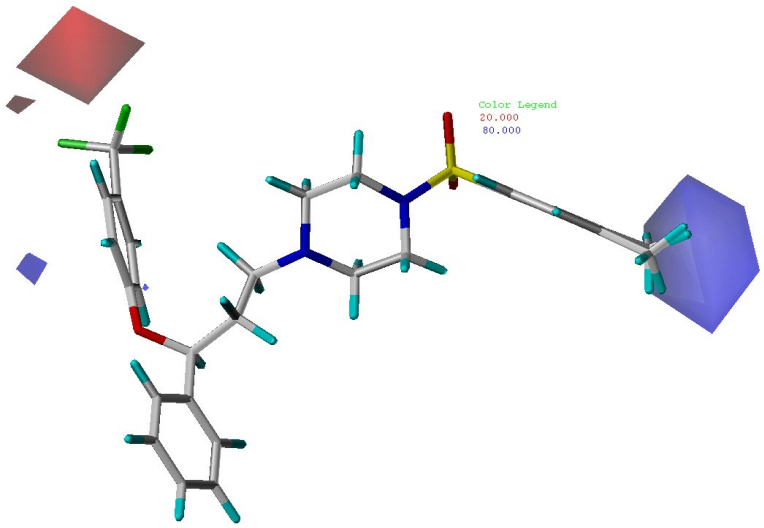


Figure 1. CoMFA electrostatic contour map of 5HT6 receptor (1l) having highest activity. In which Red: negative potential favorable; blue: positive potential favorable.

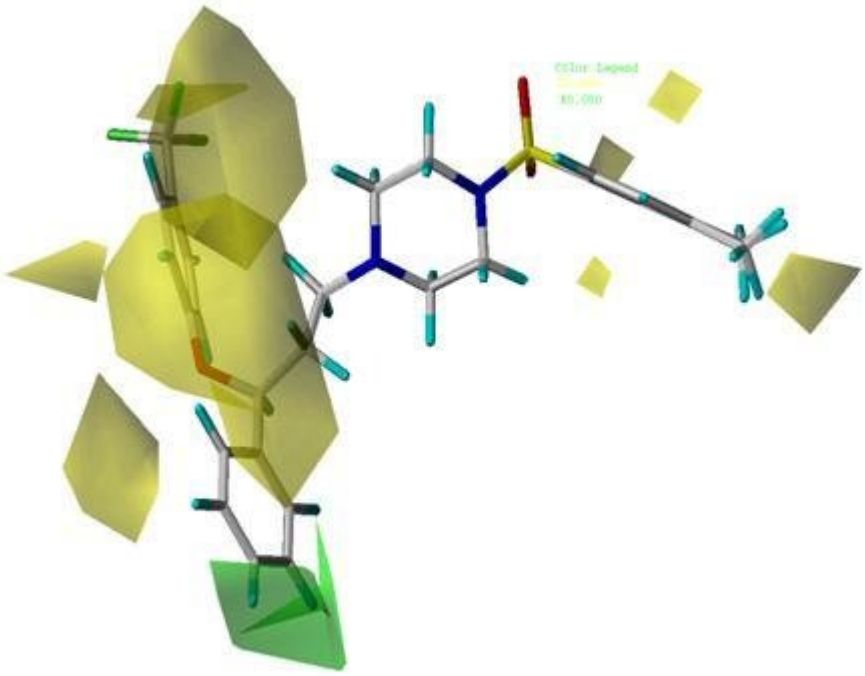


Figure 2. CoMFA steric contour map of 5HT6 receptor (1l) having highest activity. In which Green: sterically favorable regions; yellow: sterically unfavorable regions.

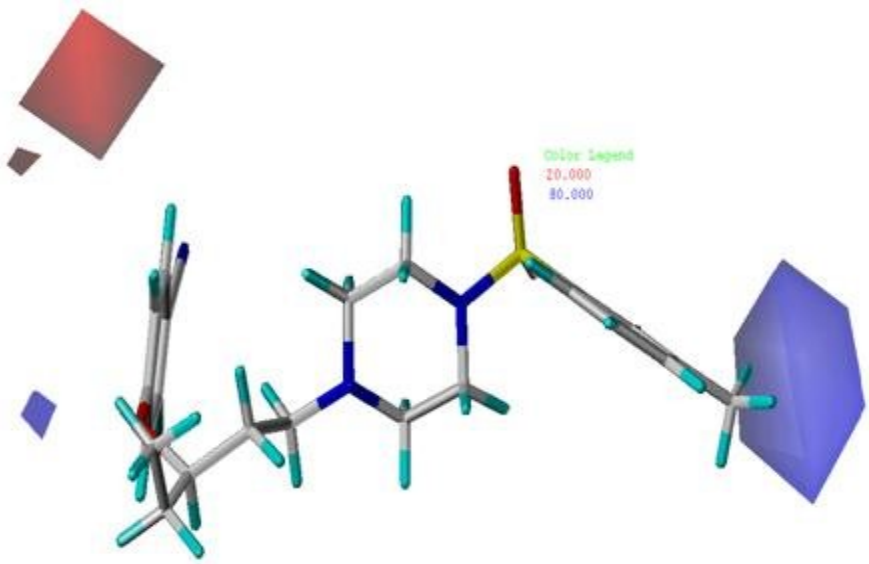


Figure 3. CoMFA electrostatic contour map of 5HT6 receptor (3i) having lowest activity. In which Red: negative potential favorable; blue: positive potential favorable.

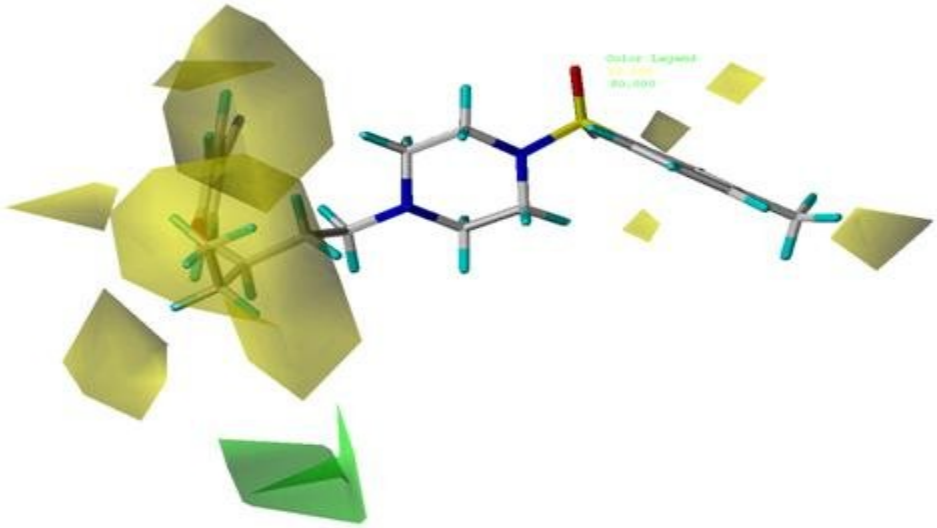


Figure 4. CoMFA steric contour map of 5HT6 receptor (3i) having lowest activity. In which Red: negative potential favorable; blue: positive potential favorable.

In Figure 1 and 3, regions where increased negative charge is associated with enhanced activity are indicated in red while regions where positive charge is associated with enhanced activity are indicated in blue. In Fig 2 and 4, the steric map shows that more bulk is favorable near green area, and unfavorable near yellow area.

CoMSIA contour maps: The contours maps of CoMSIA were derived using steric, electrostatic, and hydrophobic and hydrogen bond donor and acceptor fields. CoMSIA steric and electrostatic are more or less similar to those of CoMFA steric and electrostatic contour maps. Hydrophobic contour map of CoMSIA model, in which yellow indicates regions where hydrophobic groups increase activity and white indicates regions where hydrophilic groups increase activity. Hydrogen-bond donor contour map of CoMSIA model, in which cyan indicates regions where hydrogen-bond donor groups increase activity and purple indicates regions of unfavorable contributions from hydrogen-bond donor. Hydrogen-bond acceptor contour map of CoMSIA model, in which magenta indicates regions where hydrogen-bond acceptor groups increase activity and red indicates regions where they make unfavorable contribution to the activity.

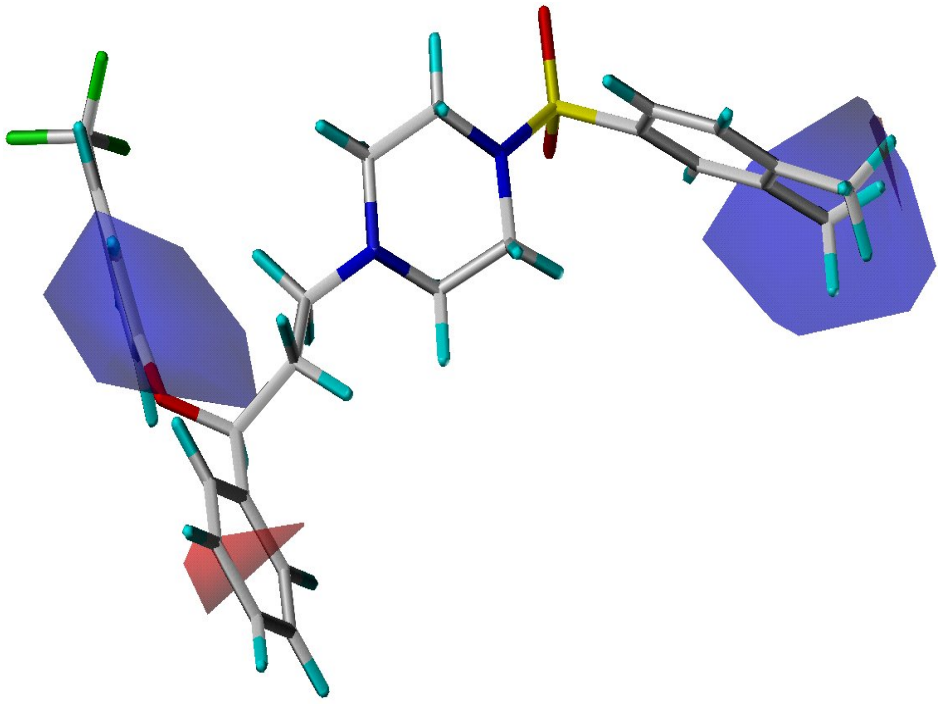


Figure 5. CoMSIA electrostatic contour map of 5HT6 receptor (1l) having highest activity. In which Red: negative potential favorable; blue: positive potential favorable.

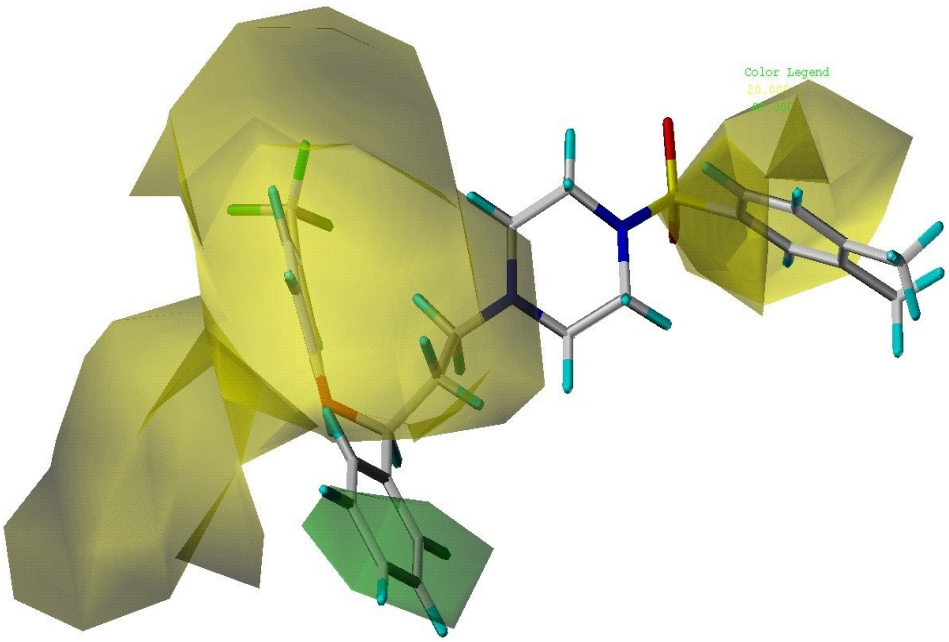


Figure 6. CoMSIA steric contour map of 5HT6 receptor (1l) having highest activity. In which Green: sterically favorable regions; yellow: sterically unfavorable regions.

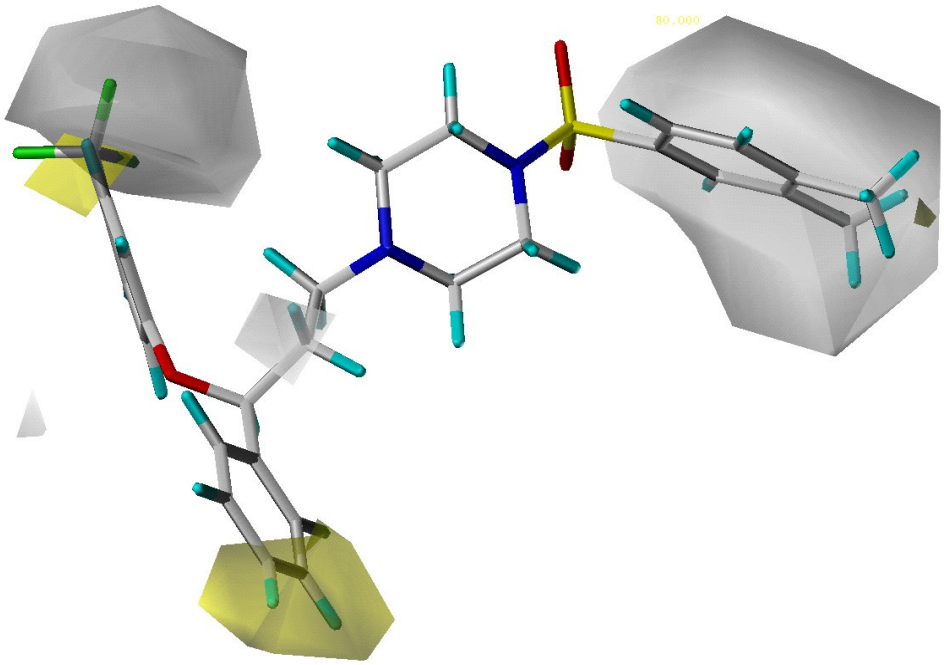


Figure 7. COMSIA hydrophobic contour map of 5 HT6 receptor (1l) having highest activity. Yellow indicates regions where hydrophobic groups increase activity and white indicates regions where hydrophilic groups increase activity.

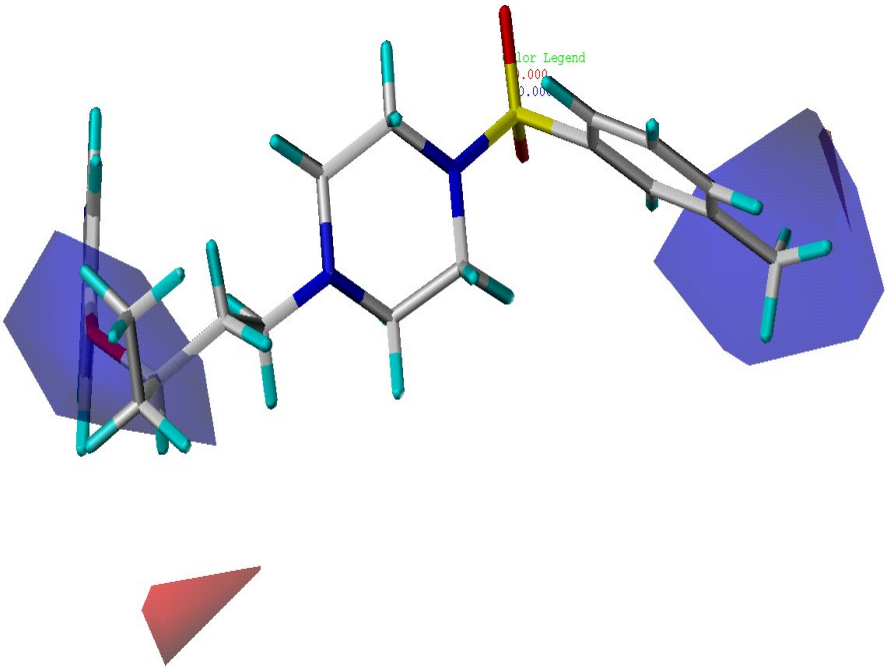


Figure 8. CoMSIA electrostatic contour map of 5HT6 receptor (3i) having lowest activity. In which Red: negative potential favorable; blue: positive potential favorable.

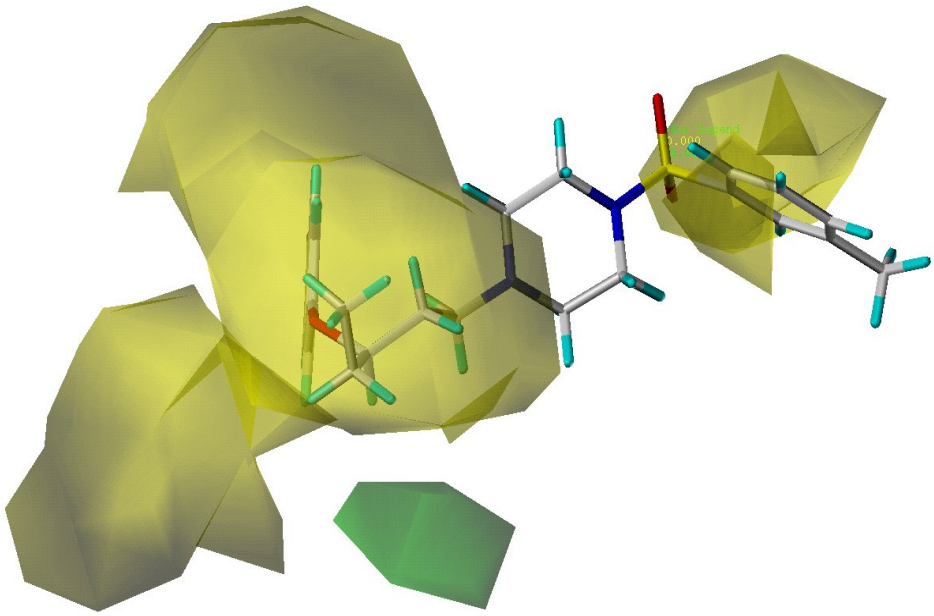


Figure 9. CoMSIA steric contour map of 5HT6 receptor (3i) having lowest activity. In which Green: sterically favorable regions; yellow: sterically unfavorable regions.

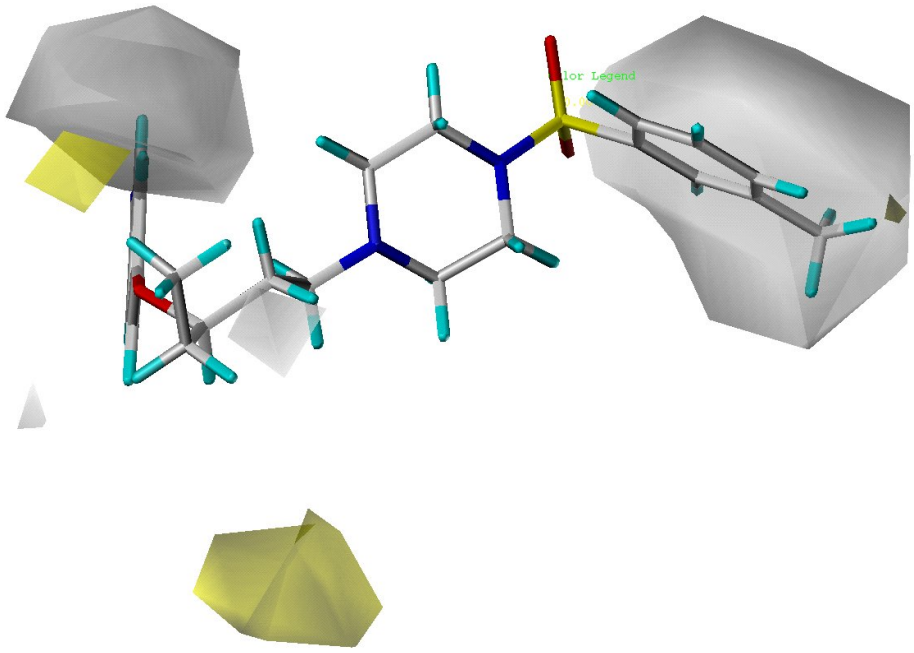


Figure 10. COMSIA hydrophobic contour map of 5 HT6 receptor (1l) having highest activity. Yellow indicates regions where hydrophobic groups increase activity and white indicates regions where hydrophilic groups increase activity.

TRAINING SET VALUES:

TABLE-2

		CoMFA		CoMSIA	
COMPOUNDS	% INHIBITION	PREDICTED VALUES	RESIDUAL VALUES	PREDICTED VALUES	RESIDUAL VALUES
1a	4.619	4.529	0.09	4.514	0.106
1j	4.481	4.643	-0.161	4.6	-0.118
1k	4.522	4.534	-0.011	4.583	-0.06

1l	4.698	4.526	0.173	4.501	0.198
2a	4.236	4.197	0.04	4.245	-0.009
2c	4.229	4.283	-0.054	4.232	-0.003
2d	4.267	4.356	-0.089	4.391	-0.124
2f	4.259	4.26	-0.001	4.206	0.054
2g	4.356	4.278	0.078	4.3	0.056
1b	4.494	4.534	-0.04	4.518	-0.023
2h	4.221	4.274	-0.053	4.296	-0.074
2i	4.301	4.223	0.078	4.243	0.058
2j	4.236	4.345	-0.108	4.316	-0.079
2k	4.301	4.239	0.062	4.247	0.054
2l	4.327	4.267	0.061	4.275	0.053
3e	4.221	4.262	-0.04	4.218	0.004
1c	4.522	4.547	-0.024	4.498	0.025
3f	4.173	4.268	-0.094	4.261	-0.087
3g	4.318	4.245	0.074	4.273	0.046
3h	4.259	4.242	0.018	4.226	0.034
3i	4.167	4.243	-0.075	4.247	-0.08
3j	4.283	4.231	0.053	4.218	0.066
3k	4.236	4.178	0.059	4.168	0.069
3l	4.244	4.264	-0.02	4.308	-0.063

1d	4.638	4.544	0.094	4.564	0.074
1f	4.494	4.496	-0.001	4.499	-0.004
1i	4.468	4.572	-0.103	4.643	-0.174

TEST SET VALUES:

TABLE-3

		CoMFA		CoMSIA	
COMPOUNDS	% INHIBITION	PREDICTED VALUES	RESIDUAL VALUES	PREDICTED VALUES	RESIDUAL VALUES
2b	4.494	4.2009	0.289	4.2009	0.289
2e	4.366	4.22573	0.144	4.22573	0.144
3a	4.537	4.25688	0.283	4.25688	0.283
3b	4.431	4.21274	0.217	4.21274	0.217
3c	4.366	4.20629	0.163	4.20629	0.163
3d	4.431	4.303	0.127	4.303	0.127
1e	4.283	4.51265	-0.232	4.51265	-0.232
1g	4.387	4.60926	-0.219	4.60926	-0.219
1h	4.387	4.53322	-0.143	4.53322	-0.143

DOCKING RESULTS:

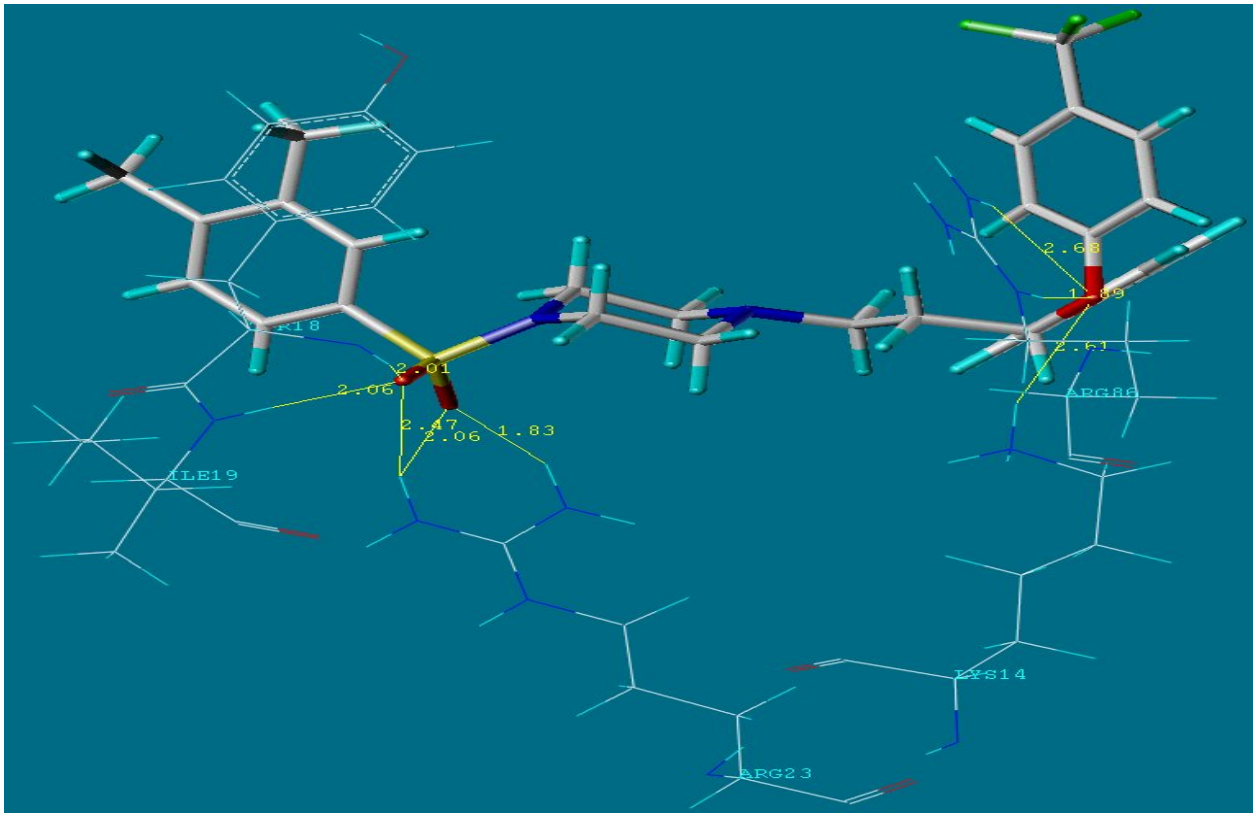


FIG 1

ACTIVE SITE	MOST ACTIVE LIGAND	DOCK SCORE
LYS 14	11	-20.2
ILE 19	11	-20.2
ARG 23	11	-20.2
ARG 86	11	-20.2
TYR 18	11	-20.2

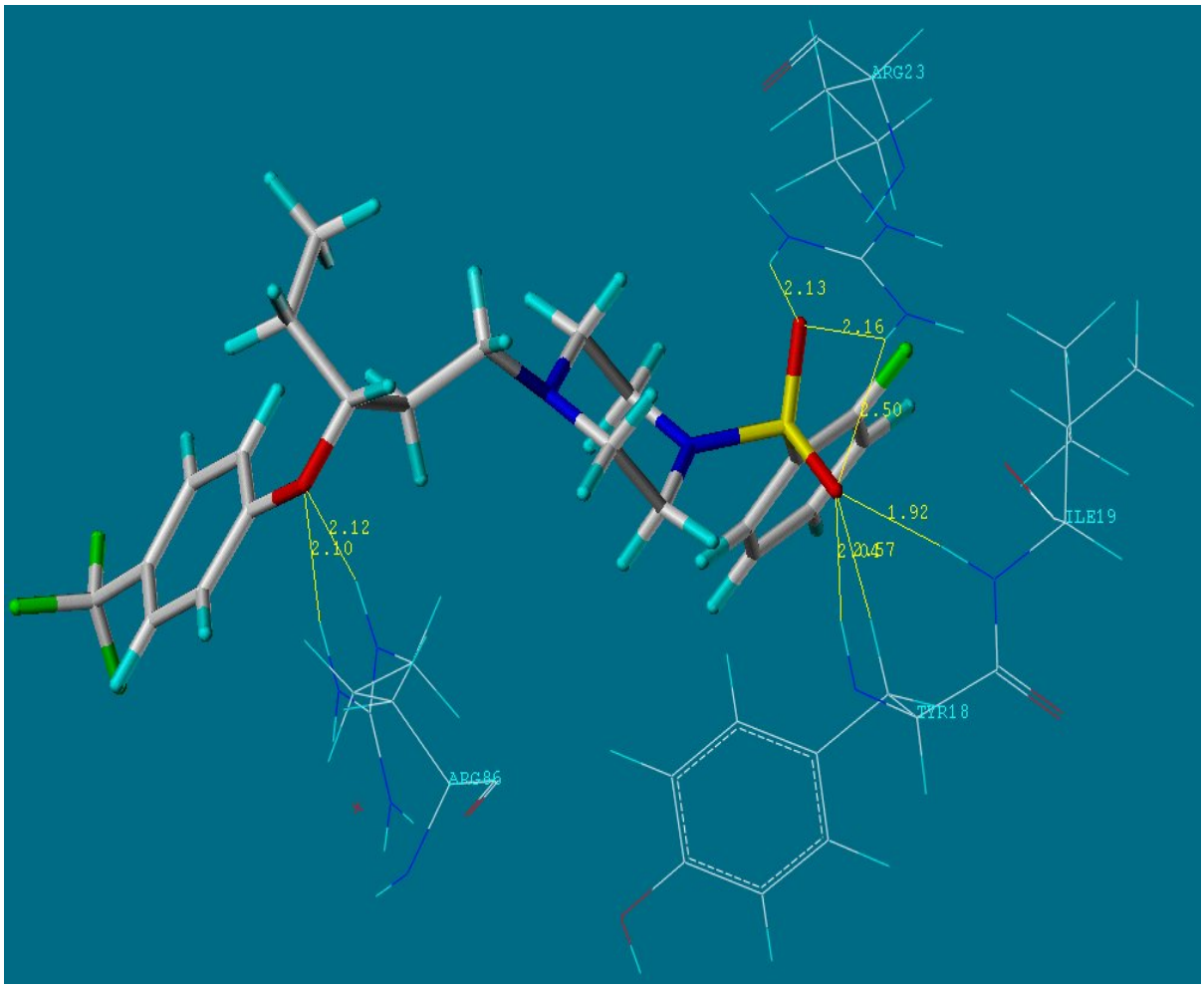
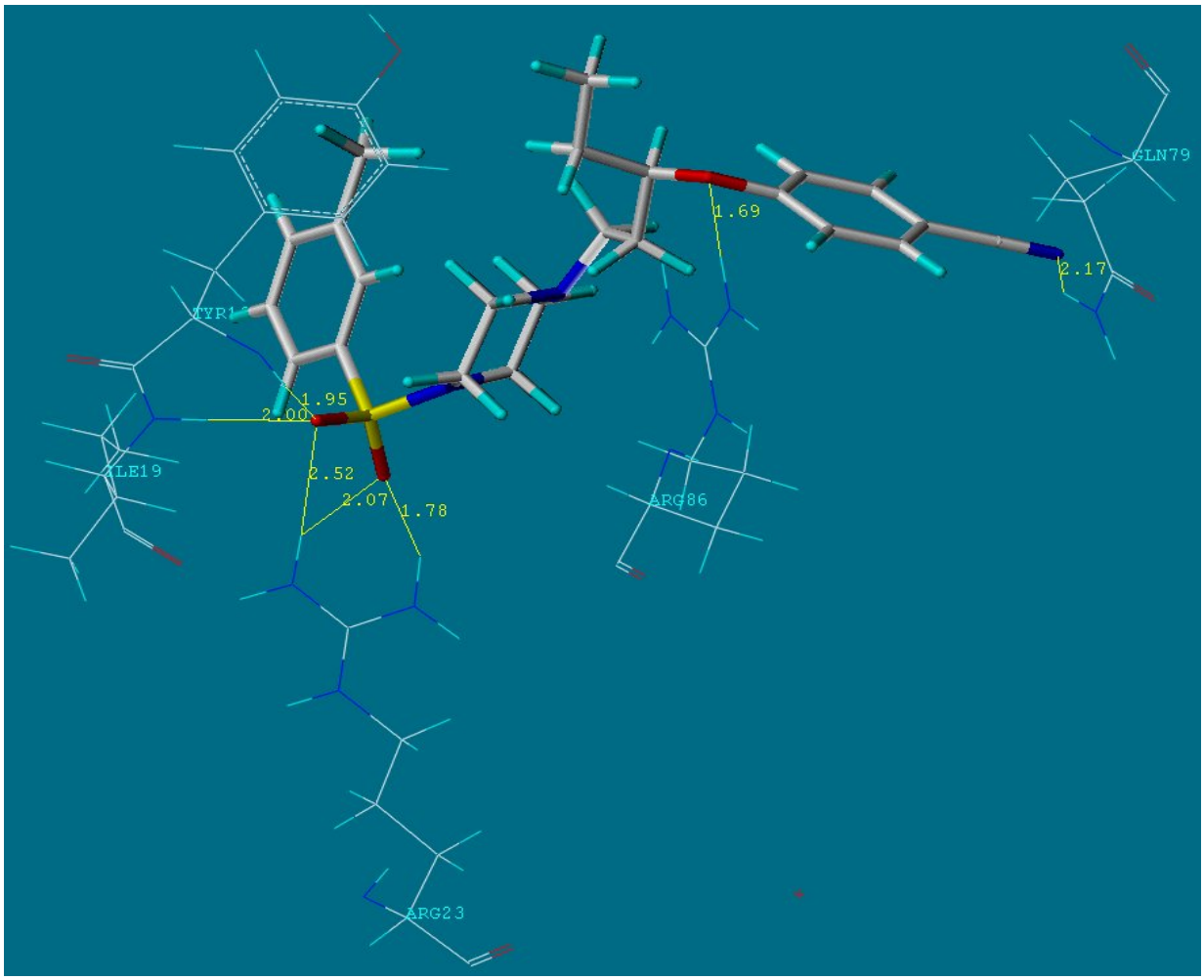


FIG 2

ACTIVE SITE	MODERATE ACTIVE LIGAND	DOCKSCORE
TYR 18	2e	-22.5
ILE 19	2e	-22.5
ARG 86	2e	-22.5
ARG 23	2e	-22.5
GLN 79	2e	-22.5
ARG 23	2e	-20.4



ACTIVE SITE	LEAST ACTIVE LIGAND	DOCKSCORE
TYR 18	3i	-20.4
ILE 19	3i	-20.4
ARG 23	3i	-20.4
ARG 86	3i	-20.4
GLN 79	3i	-20.4

CONCLUSION

The use of 3D structural information is useful in drug design and the CoMFA and CoMSIA models generated in our studies are clearly stable and robust, exhibiting both good internal and external consistency. The ability of the QSAR models to accurately predict the property value, along with the important information gathered from 3D contour maps appear to be valuable tools for the design of new compounds having improved receptor activity. Docking of most active (1L), moderate active (2e) and least active (3i) reveals different modes of binding affinity with the amino acids. The binding modes were evaluated and validated using scoring function. All the three categories of active ligand have high affinity with ARG 23, ILE 19 and TYR 18, this reveals that modification of ligand at those specific sites can improve the inhibitory activity of the receptor. Hence the process of docking can be regarded as a key aspect in reforming the correlation between calculated and observed binding affinities in effect to develop an effective novel compound.

REFERENCES

1. “The role of in silico approaches in modern drug discovery”, by G. Narahari Sastry Molecular Modelling Division, India Institute of Chemical Technology.

2. Eriksson L, Johansson E, Muller M, Wold S, On the selection of the training set in environmental QSAR analysis when compounds are clustered" J. Chemometrics, 2007, Vol. 14, 599-616.

3. 3D Qsar in Drug Design (2008) by Hugo Kubinyi.

4. Sun Ah Jeon, Hyunah Choo, Woo-Kyu Park,Hyewhon Rhim, Soo Y. Ko, Yong Seo Cho, Hun Yeong Koh, Ae Nim Pae, Synthesis and Biological Evaluation of Aryl sulfonyl piperazine Derivatives as 5-HT6 Receptor Ligands, Bull. Korean Chem. Soc., 2007; Vol. 28(2): 285-291.

5. Clark M, RD Cramer, The probability of chance correlation using partial least squares (PLS). III, Quant. Struct.-Act. Relationships, 1993; 12: 137-145.

6. Dewar M, Zoebisch E, Healy E, Stewart J, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 1985, Vol. 107(13), 3902-3909.

7. S.J. Cho, A. Tropsha, Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results, J. Med. Chem, (1995) Vol. 38(7), 1060-1066.