Design of anti-fungal agents by 3D-QSAR

Shreya S Sonak, Sandeep Pathare, Siddharth J Modi & Vithal M Kulkarni*

Department of Pharmaceutical Chemistry, Poona College of Pharmacy, Bharati Vidyapeeth (Deemed to be University), Pune - 411 038, Maharashtra, India

*E-mail: vmkulkarni60@gmail.com

An increase in the number of invasive fungal infections especially in immunocompromised patients is increasing the mortality rate worldwide. Due to the emergence of drug-resistant fungi, the currently available antifungal drugs have become ineffective. Because no alternative treatment is available, some existing drugs are still used. Therefore, there is a need to design and develop novel and effective anti-fungal drugs. Molecular docking and 3-dimensional quantitative structure-activity relationship (3D-QSAR) methods have been useful approaches for the design of novel molecules. A set of 30 molecules reported in the literature containing azoles and non-azoles have been used in this study to derive 3D-QSAR.CoMFA and CoMSIA models for the most active compound and least active compounds have been developed. The structural requirements were obtained by analysing the contour maps. The partial least square analysis for CoMFA and CoMSIA showed a significant cross-validated correlation coefficient of 0.625 and 0.67 and a non-cross validated correlation coefficient of 0.991 and 0.99, respectively. The model was validated by observing the predicted correlation for test molecules with the value of 0.699 and 0.659, respectively.

Keywords: Azoles, Anti-fungal, 3D-QSAR, CoMFA, CoMSIA

Fungi are eukaryotic organisms that include yeasts and moulds. There are about 600 species of fungi that are associated with human diseases of which 20 cause 99% of human fungal infections. Fungal reproduction forms spores which attack the human body either by inhalation or by direct contact causing fungal infection also called as mycosis. Fungal infections mostly affect skin, nails, or lungs, the most common fungal infections are athlete's foot, ringworm, yeast infection or fungal infection of the nail. The infections can be classified into (a) allergic reactions to fungal proteins, (b) toxic reactions to toxins present in certain fungi and (c) infections (mycosis). According to pathogens the fungal infection can be classified as endogenous (Candida infections) or acquired from the environment (Cryptococcus, Aspergillus infections).

The aerobic growth of fungal cell is carried out by ergosterol, the main sterol synthesized in the fungal cell. Ergosterol is responsible for maintaining the membrane fluidity, permeability and viability which helps in the growth of the cell. Synthesis of ergosterol is a multi-step process[1] that takes place in the cytoplasm of the fungal cell as shown in Fig. S1, Supplementary Data. Most of the anti-fungal agents, mainly azoles, interfere with the ergosterol biosynthesis by inhibiting enzymes required for its synthesis. The classification of antifungal agents according to their mechanism of action (MOA)[2] is shown in Fig. S2, Supplementary Data.

Azoles are the first-line drugs used for the treatment of fungal infections which act by inhibitinglanosterol14-α-demethylase, an enzyme required for the synthesis of ergosterol. Currently, azole containing drugs used for the treatment of fungal infections[2] are shown in Fig. S3, Supplementary Data. Due to increase in the number of fungal infections worldwide because of treatment of several diseases like diabetes, immunosuppressive viral infections, etc., has led to the development of multi-drug resistance (MDR) in humans. The generation of new potential antifungals is lagged behind compared to the pace of increasing number of fungal infections. Therefore, the designing of new antifungals is necessary[2]. In 2011, Sheng et al.,[3] reported several compounds containing azoles and non-azoles with their antifungal activity against *Candida Albicans* and it was observed that non-azole derivatives also showed fairly good antifungal activity.

Quantitative structure-activity relationship (QSAR) studies build mathematical models which find a correlation between structure and function using the

chemometric technique. In simple terms, the physicochemical properties like affinity of binding of ligands to the receptor, inhibition constants, rate constants, and other biological endpoints, with an atom, group or molecular properties such as lipophilicity, polarizability, electronic and steric properties are correlated[4]. In 3D-QSAR, models are developed by correlating activity with non-covalent interaction fields surrounding the molecules by chemometric technique like partial-least-square (PLS) analysis[4,5]. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Analysis (CoMSIA) are the two approaches among others of 3D-QSAR studies. CoMFA describes the molecular properties by fields and CoMSIA calculates similarity fields to probe atoms and groups, these fields are Gaussian potentials[5]. Kulkarni et al. have reported the application of these methods to several classes of drugs and successfully designed new therapeutic molecule[6-8]. Though there are innumerable reports on these methods, they still find useful application for drug design. In the present study, a 3D-QSAR model is developed for a series of 30 compounds containing azoles and non-azoles using these two methods.
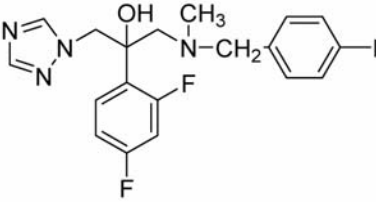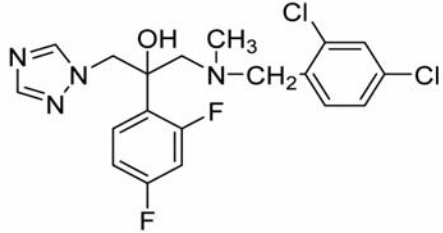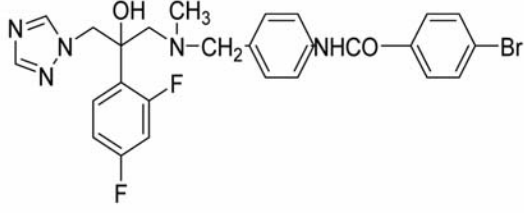
Experimental Details

3D-QSAR Study

Datasets

A series of 30 compounds containing azoles and non-azoles reported by Sheng and Zhang are used for the present 3D-QSAR studies[3]. Authors report in this paper that structure-based methods can significantly improve the efficiency for discovering new azoles with good activities. Non-azoles included in this study also have a similar mechanism of action acting on Cyp51. Our paper is aimed to address these. The activity of the compounds determined as IC_{50} was converted to pIC_{50}. These compounds were divided into a training set (70%) and test set (30%), as the sets are required to generate 3D-QSAR models, the selection of compounds takes place randomly. These sets undergo CoMFA and CoMSIA analysis. To create the training set and test set, an option in the software Sybyl-Xver1.3 'create and random method' is used. The symmetrical distribution of data can be done by PLS regression analysis, where the logarithmic affiliation helps in distributing the data[9]. The structures of compounds, IC_{50} and pIC_{50} data are given in Table 1. Test molecules are marked with hashtag #.

Table 1 — Structures, experimental and predicted IC50 values

Compound No.*	Structure	IC_{50} µg/ml *C.Albicans*	pIC_{50} µg/ml *C.Albicans*
13#		0.032	6.494
14.		0.016	7.795
15#		0.008	8.090

(Contd.)

Table 1 — Structures, experimental and predicted IC50 values (*Contd.*)

Compound No.*	Structure	IC$_{50}$ µg/ml *C.Albicans*	pIC$_{50}$ µg/ml *C.Albicans*
16.		0.001	9.000
17.		0.002	8.690
18.		0.023	7.630
19#		0.027	7.568
20.		0.004	8.400
21#		0.004	8.400
22.		0.004	8.400

(*Contd*

Table 1 — Structures, experimental and predicted IC50 values (*Contd.*)

Compound No.*	Structure	IC$_{50}$ µg/ml C.Albicans	pIC$_{50}$ µg/ml C.Albicans
23.		0.063	7.204
24.		0.016	7.806
25.		0.004	8.390
26.		0.001	9.000
27.		0.016	7.806
28#		0.001	9.000
29.		0.016	7.280
30#		0.016	7.280

Table 1 — Structures, experimental and predicted IC50 values (*Contd.*)

Compound No.*	Structure	IC_{50} μg/ml C.Albicans	pIC_{50} μg/ml C.Albicans
39.		160.000	3.795
40.		40.000	4.397
41.		160.000	3.795
42.		40.000	4.390
43.		100.000	4.000
44#		50.000	4.301
45.		0.063	7.204
46#		0.031	7.508
47.		0.031	7.508
48.		0.063	7.204
49.		8.000	5.096
50#		8.000	5.096

Alignment of database

Alignment is an important requirement for CoMFA and CoMSIA studies, where the orientation of a molecule and its conformation in space is determined. Sketch software was used to convert 2D structures into 3D, and then Gasteiger-Huckel charges were applied to all compounds. The energy minimization of the compounds was done by the Standard Tripos Molecular Mechanism force field. The distil alignment function was also performed[10]. The alignment of molecules is shown in Fig. 1.

CoMFA and CoMSIA models

For each alignment produced, the steric and electrostatic fields were calculated by each interaction in the grid of a 3D cubic lattice with 2 Å grid space which was extended for the aligned dataset in all three directions X, Y and Z. Lennard-jones and Columbic potentials were used at lattice interaction[7]. CoMFA gives an interaction between an inhibitor and its molecular target by considering only two properties i.e., steric and electronic fields, while CoMSIA considers five different properties: steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor fields. van der Waals potential and Columbic term represents electrostatic and steric fields, respectively and were calculated by Tripos force field[10,11]. For CoMFA, the descriptors were derived by considering a sp^3 carbon atom with radius 1.52 Å with +1 charge which is a probe atom and for the derivation of both the fields interaction energy was calculated using force field, ±30 kcal/mol was the cut-off value[7]. The CoMFA and CoMSIA studies are developed on the assumption that the binding affinities of ligand change with the changes in molecular properties which are shown by the fields[9]. For CoMSIA studies a Gaussian function was established to calculate the distance between the probe atom (molecular atom) at all grid points; similarity indices at the molecular surface are calculated by the CoMSIA equation[7]
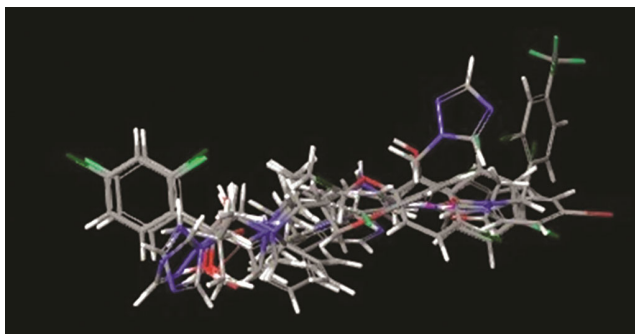
$$A_{F,k\,(j)}^{q} = \sum W_{probe,k}\, W_{ik}\, e^{-\alpha r^2 iq}$$

where, A is the similarity index at a grid point, q are the atoms of molecule j, W_{probe} where k is the probe atom having radius 1 Å, W_{ik} is the actual value of physicochemical property k of atom e and α is the attenuation factor[7]. The common fragment of the azoles is shown in Fig. 2.

Partial least square (PLS) analysis and model validation

PLS is an approach of multiple regression analysis, which is carried out for the development of 3D-QSAR. The data set of definite molecules were treated using PLS analysis. 3D contour maps were developed with an optimum number of components say, 6. The generated PLS algorithm is used to develop a correlation between structural property and pharmacological activity[9]. Leave one out (LOO) and cross-validation analysis is performed using PLS. q^2 represents the cross-validation and r^2 represents the conventional correlation coefficient, if $q^2 \geq 0.5$ and $r^2 \geq 0.64$ the models are accepted[7]. q^2 is calculated by the equation:

$$q^2 = 1 - \sum \left(Y_{predicted} - Y_{observed}\right) / \sum \left(Y_{observed} - Y_{mean}\right)$$

The non-cross validation method calculates the validation of conventional correlation r^2_{ncv}, standard error of estimate (SEE) and F (Fischer's) values. If the QSAR models are developed with an optimum number of components (ONC) with higher LOO and q^2 value and lower SEE value, then the probabilities of over-fitted models are negligible. A bootstrap analysis is carried out which checks the robustness of the generated model. This method is performed several times and arbitrarily selects the original set of an object. For our study, we performed 10 cycles for 100 times[9].
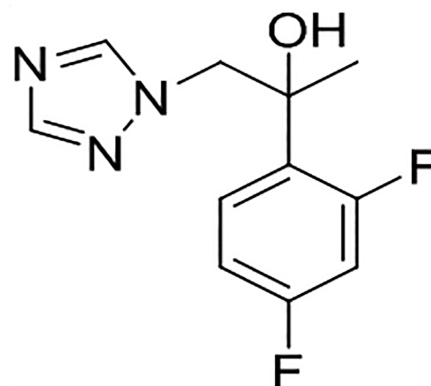

Fig. 1 — Aligned dataset of azoles

Fig. 2 — Common Fragment of azoles

Predictive correlation coefficient (r^2_{pred})

The predictive power of the 3D-QSAR model is generated by the test set. The r^2_{pred} value is determined with the help of test set compounds using the equation:

$$r^2 = \frac{(SD - PRESS)}{SD}$$

where, SD = total squared deviation between biological activities of test set compound and mean activities of training set compounds, and PRESS = total squared deviation between experimental and predicted activity values for each compound in test set[10].

Results and Discussion

Results of CoMFA and CoMSIA models

The CoMFA model is developed using the steric and electrostatic parameters, initially, PLS analysis was taken into consideration using leave-one-out, the cross-validated correlation q^2 was observed 0.625 (ONC=6). The non-cross validation correlation was performed by column filtering set to 2.0 with same ONC, r^2_{ncv}=0.991: F-value = 435.063: SEE = 0.180. The bootstrap analysis r^2_{bs} = 0.998, SEE$_{bs}$= 0.087 and cross-validation coefficient r^2_{cv} = 0.614 supports the uniformity of the developed QSAR model. The statistical parameters are reported in Table 2 for CoMFA and CoMSIA models.

The experimental IC50, predicted pIC50 and residual values of the training set of derivatives for

CoMFA and CoMSIA models are given in Table 3 and for the test set of derivatives in Table 4. CoMSIA model was developed using the same data set, the q^2 was found 0.670 and non-cross validated r^2_{ncv} was found 0.990 with a F-value of 368.25 having the same ONC of 6. Bootstrap analysis r^2_{bs} = 0.996 and SEE$_{bs}$= 0.120 which supports the model performed and confirms the internal reliability of data set by cross-validation, r^2_{cv}= 0.637. The model was developed using 5 descriptors:steric, electrostatic, H-bond donor, H-bond acceptor and hydrophobicity with values of 0.389, 0.748, 0.835, 0.757 and 0.984, respectively. The prediction of correlation coefficient

Table 2 — Statistical parameters from CoMFA and CoMSIA models

PLS analysis parameter	CoMFA	CoMSIA
r^2 loo (q^2)	0.625	0.670
ONC	6.000	6.000
SEE	0.180	-
r^2_{ncv}	0.991	0.990
F value	435.063	368.250
Steric Field Contribution	-	0.389
Electrostatic Field Contribution	-	0.748
Donor contribution	-	0.835
Acceptor contribution	-	0.757
Hydrophobic contribution	-	0.984
r^2_{bs}	0.998	0.996
SEE$_{bs}$	0.087	0.120
r^2_{cv}	0.614	0.637
Test set r^2_{pred}	0.699	0.659

Table 3 — Experimental IC50, predicted pIC50 and residual values of training set by CoMFA and CoMSIA analysis
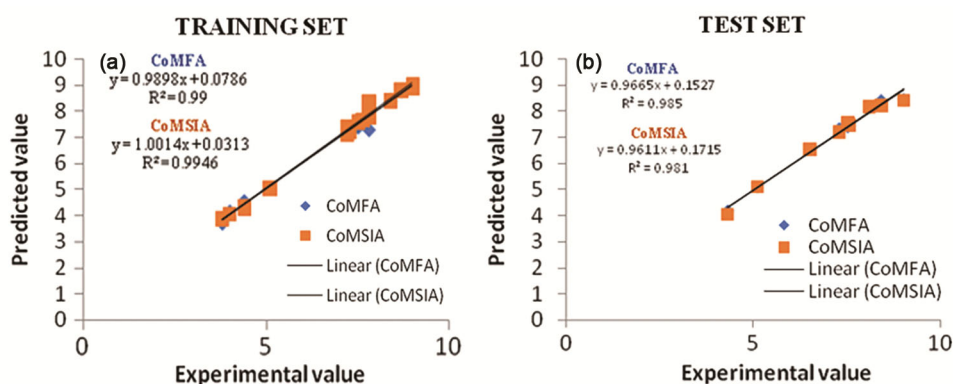
Compound No.*	Experimental value	CoMFA		CoMSIA	
		Predicted	Residual	Predicted	Residual
14	7.795	7.801	-0.006	7.859	-0.064
16	9.000	8.968	-0.032	9.004	-0.004
17	8.690	8.768	-0.078	8.799	-0.109
18	7.630	7.573	0.057	7.649	-0.019
20	8.400	8.378	0.022	8.390	0.010
22	8.400	8.441	-0.041	8.400	0.000
23	7.204	7.161	-0.043	7.372	-0.168
24	7.806	7.289	-0.023	7.744	0.062
25	8.390	8.393	-0.003	8.418	-0.028
26	9.000	8.911	0.089	8.900	0.100
27	7.806	8.338	-0.532	8.327	-0.521
29	7.280	7.319	-0.039	7.204	0.076
39	3.795	3.751	0.044	3.882	-0.087
40	4.397	4.398	-0.001	4.297	0.100
41	3.795	3.722	0.073	3.923	-0.128
42	4.390	4.561	-0.071	4.367	0.023
43	4.000	4.163	-0.163	4.068	0.233
45	7.204	7.227	-0.023	7.283	-0.079
47	7.508	7.393	0.115	7.577	-0.069
48	7.204	7.317	-0.113	7.135	0.069
49	5.096	5.096	0.076	5.056	0.040

*As reported [3]

Table 4 — Experimental IC50, predicted pIC50 and residual values of test set by CoMFA and CoMSIA analysis

Compound No.*	Experimental value	CoMFA		CoMSIA	
		Predicted	Residual	Predicted	Residual
13	6.494	6.509	-0.015	6.554	-0.060
15	8.090	8.171	-0.081	8.177	-0.087
19	7.568	7.463	0.105	7.478	0.090
21	8.400	8.416	-0.016	8.219	0.181
28	9.000	8.451	0.549	8.411	0.589
30	7.280	7.328	-0.048	7.205	0.075
44	4.301	4.163	0.138	4.068	0.233
46	7.508	7.393	0.115	7.577	-0.069
50	5.096	5.080	0.016	5.115	-0.019

*As reported [3]

Fig. 3 — (a) The plot of experimental and predicted IC$_{50}$ values of the training set of compounds of CoMFA and CoMSIA models and (b) The plot of experimental and predicted IC$_{50}$values of the test set of compounds of CoMFA and CoMSIA models

(r^2_{pred}) of the test set of molecules for CoMFA and CoMSIA was found 0.699 and 0.659, respectively.

QSAR Visualization

An important feature of the 3D-QSAR model development of CoMFA and CoMSIA is visualizing the 3D contour maps which are generated when there are variations in the molecular fields. On the basis of the contour maps, modifications can be made in the structure to improve the activity of the compounds. The most active compound 26 (IC$_{50}$=0.001 µg/ml) and the least active compound 39 (IC$_{50}$=160 µg/ml) were chosen for generating the contour maps. A graph of linear correlation for the training set of compounds for CoMFA and CoMSIA experimental and predicted values is shown in Fig. 3a. A graph of linear correlation for the test set of compounds for CoMFA and CoMSIA experimental and predicted values is shown in Fig. 3b.

CoMFA contour maps

The contour maps are created around the molecule and the style was set to transparent for better visualization. Green colour represents the favourable region for steric group substitutions while yellow represents unfavourable region for a steric group (80:20 contribution) respectively. Substitution of bulky groups in the green area enhances the inhibitory activity, whereas substitution in the yellow area decreases the inhibitory activity. The region where positively charged groups are substituted is represented by blue colour for electrostatic field and the red colour represents the region where negatively charged groups are substituted (80:20 contribution) respectively. The contour maps for steric and electrostatic fields for compounds 26 and 39 are shown in Fig. 4 (a), (b) and (c), (d), respectively.

The most active compound 26 is divided into four different parts (A), (B), (C) and (D) shown in Fig. 5. The green contour on the C segment suggests that bulky groups are favourable in that region for better activity, the green contour on the alkyl chain of the B part suggest that increase in the number of carbon chain (n= ≤ 4) can increase the inhibitory action. Triazole ring, A part is not favoured for bulky groups as substituents represented by the yellow contour. The least active compound 39 is a non-azole molecule that is divided into three parts (a), (b) and (c) as shown in Fig. 5. The green contour map suggests that the b part
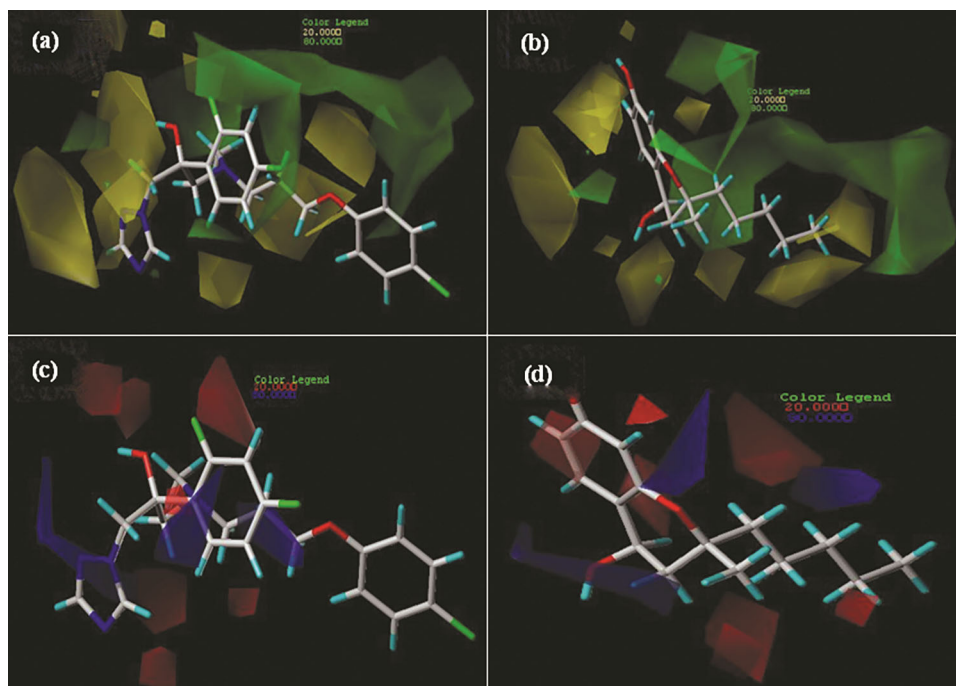
Fig. 4 — CoMFA contour maps of steric field (a, b) and electrostatic field (c, d) for compound 26 and compound 39
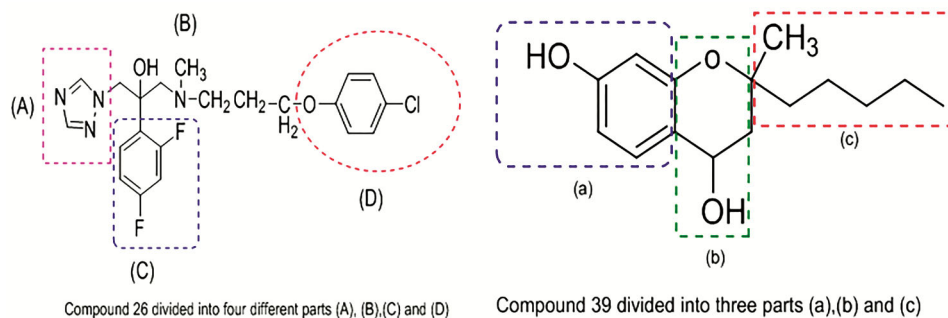

Compound 26 divided into four different parts (A), (B),(C) and (D)   Compound 39 divided into three parts (a),(b) and (c)

Fig. 5 — Fragmentation of Compound 26 and Compound 39

is cyclic and the cpart is carbon chain, which are favoured areas for bulky group substitutions.

The blue and red contour maps represent the electrostatic field, in the most active compound 26. The blue contour around the propyl chain of B part and triazole ring A part suggest that positively charged groups can be substituted in these regions which will lead to an increase in activity of the compounds. The blue contour region around the hydroxyl group of b part of the least active compound 39 indicates that it is the most favoured region for substitution of the positively charged groups.

CoMSIA contour maps

The most active compound 26 was considered for generating the CoMSIA contour maps. The steric and

electrostatic contour maps are almost similar to those observed in CoMFA studies. The contour maps of steric, electrostatic, H-bond acceptor, H-bond Donor and Hydrophobic fields are shown in Fig. 6 (a), (b), (c), (d) and (e), respectively. The green contour area in Fig. 6a indicates that part B and part C of the compound can be substituted with bulky groups. It is a sterically favoured region that can improve biological activity. The blue contour area in Fig. 6b represents that positively charged groups can be substituted in part C which is 2,4 diflurobenzene.

The hydrogen bond acceptor fields are represented by magenta and red contour maps (80:20 contribution). The magenta contour area of Fig. 6c suggests that part B of the compound is the favoured region for H-bond acceptor substituents and the red colour contour at
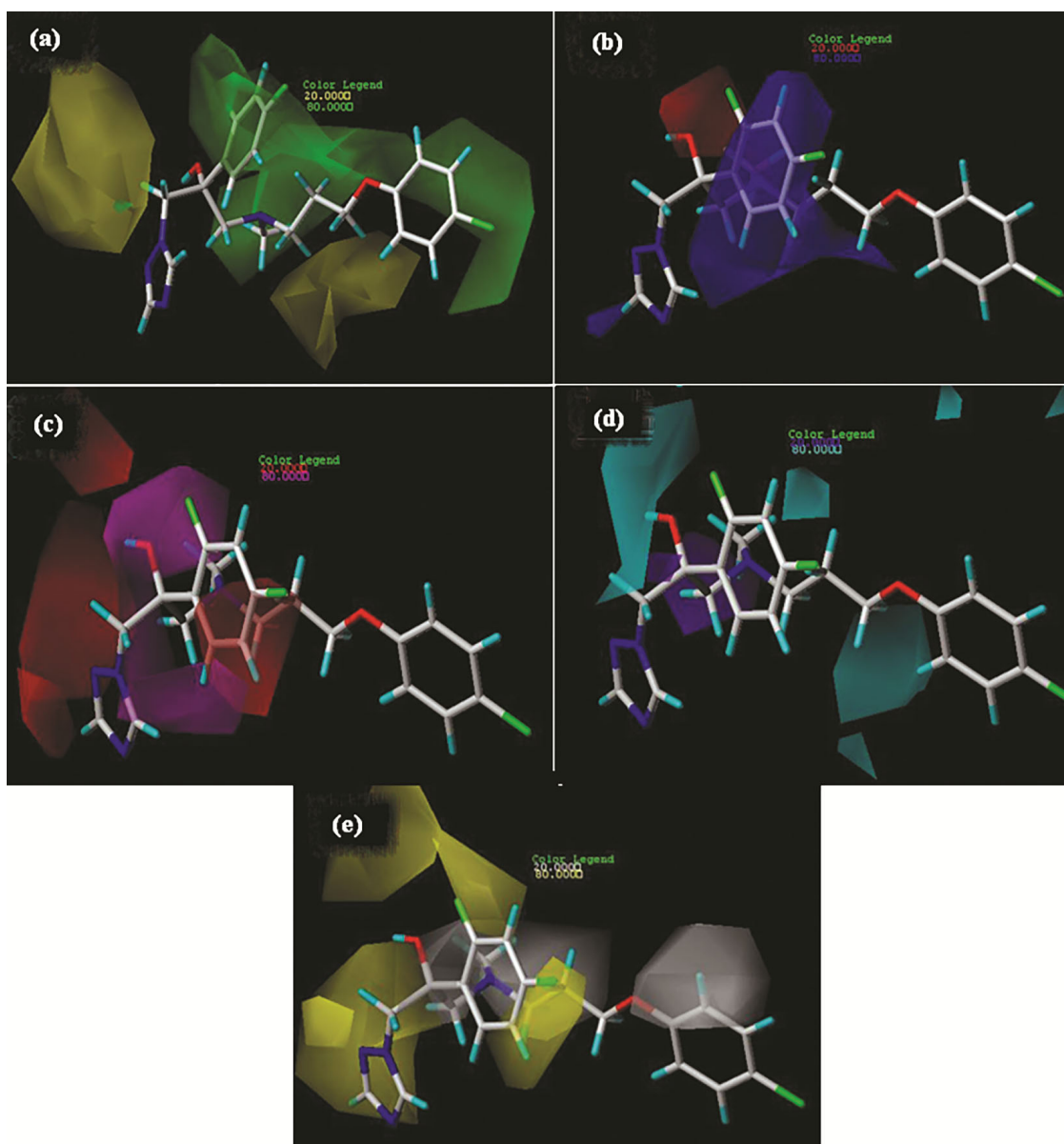
Fig. 6 — CoMSIA contour maps of most active compound 26 (a) steric, (b) electrostatic, (c) Hydrogen bond acceptor, (d) Hydrogen bond donor and (e) Hydrophobic fields

2,4diflurobenzene ring shows that the ring is unfavourable for H-bond acceptor substituents. Cyan colour in contour maps of Fig. 6d depicts the favoured region of the H-bond donor field and the purple colour depicts the unfavoured region. The carbon chain in part B of the compound shows a small part of cyan contour which indicates that it is the favourable region for H-bond donating groups and will help in increasing the biological activity.

In Fig. 6e, the yellow contour area suggests favourable hydrophobic groups and the white contour indicate about the region for favourable hydrophilic groups (80:20 contribution), respectively. A small yellow contour in part B of the compound 26 indicates that the propyl chain is favourable for hydrophobic substitutions and is similar to the steric favoured region and can help to increase in the activity. In fact, some steric groups are hydrophobic. The white contour is also observed in the same region which indicates that the N-methyl is favourable for hydrophilic substitutions. Part D of the compound is also favourable for hydrophilic substitutions. The favourable and unfavourable regions generated by contour maps are depicted in Fig. 7.
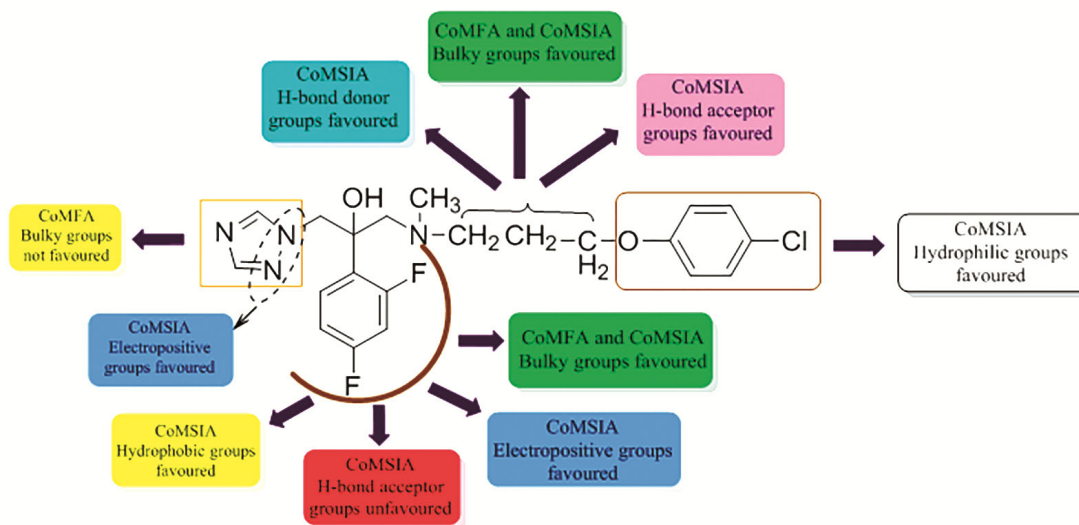
Fig. 7 — Structural Interpretation from QSAR model study

Conclusion

3D-QSAR models for a reported series of azoles and non-azoles are discussed. 3D-QSAR models were generated using two different analyses CoMFA and CoMSIA. To analyse these models, PLS was taken into consideration to correlate the descriptors with antifungal activity. The cross-validation coefficient (q^2) was observed 0.625 and 0.670 for CoMFA and CoMSIA, respectively. The validation of other statistical parameters of CoMFA and CoMSIA models shows values of r^2_{ncv} = 0.991 and 0.990, respectively with the same ONC value of 6. To confirm the robustness of the model, bootstrap analysis was performed resulting in values r^2_{bs}= 0.998 and 0.996 for CoMFA and CoMSIA, respectively. These statistical results and the contour maps generated explain that the model developed is satisfactory. From the contour maps, it is suggested that the substitution of bulky groups and electropositive groups on part C of compound 26 and also elongating the carbon chain in the B part will increase the activity. These models lead to the understanding that antifungal activity is dependent on steric, electrostatic, and hydrophobic fields. The study also signifies that the non-azole molecules are less effective than the azoles and for future development,

it is appropriate to consider azole scaffold more prominently. The results of 3D-QSAR models discussed in this study can be used for designing novel molecules as antifungal agents.

Supplementary Information

Supplementary information is available in the website http://nopr.niscair.res.in/handle/123456789/58776.

References

1   M Hata, Y Ishii, E Watanabe, K Uoto, S Kobayashi, K-I Yoshida, T Otani, A Ando, *Med Mycol*, 48 (2010) 613
2   Ghannoum M A & Rice L B, *Clin Microbiol Rev*, 12 (1999) 501
3   Sheng C & Zhang W, *Curr Med Chem*, 18 (2011) 733
4   Verma J, Khedkar V M & Coutinho E C, *Curr Top Med Chem*, 10 (2010) 95.
5   Kubinyi H, *Drug Discov Today*, 2 (1997) 457.
6   Gokhale V M & Kulkarni V M, *J Med Chem*, 42 (1999) 5348.
7   Modi S J, Modh D H & Kulkarni V M, *J Appl Pharm Sci*, 10 (2020) 1.
8   Bhansali S G & Kulkarni V M, *Der Pharma Chem*, 7 (2015) 156.
9   Shinde M G, Modi S J & Kulkarni V M, *J Appl Pharm Sci*, 7 (2017) 181.
10 Modi S J & Kulkarni V M, *J Appl Pharm Sci*, 8 (2018) 15.
11 Purushottamachar P & Kulkarni V M, *Bio & Med Chem.*, 11 (2003) 3487.