Development of Pharmacophore Alignment Models as Input for Comparative Molecular Field Analysis of a Diverse Set of Azole Antifungal Agents

Tanaji T. Talele, Santosh S. Kulkarni, and Vithal M. Kulkarni*

Pharmaceutical Division, Department of Chemical Technology, University of Mumbai, Mumbai 400 019, India

Received March 10, 1999

Molecular modeling studies were performed on a diverse set of 24 cytochrome P-450$_{14\alpha DM}$ inhibiting azole antifungals that demonstrate different degrees of biological activity. The studied compounds, which have been reported to be active in vitro against *Candida albicans*, were divided into a training set of 20 compounds and a test set of 4 compounds. In an effort to develop a ligand-binding model for the cytochrome P-450$_{14\alpha DM}$ receptor, a pharmacophore mapping program (Apex-3D) was used to search structural features that are common to ligands that exhibit moderate to high antifungal activity. Apex-3D then was utilized to propose a common biophoric region that included one low-energy conformation of each compound in the training set. These aligned structures suggested a three-point pharmacophore map (two atom-centered descriptors and one aromatic ring centroid) for the azole antifungals. The resulting alignment was used in a comparative molecular field analysis (CoMFA) study in an attempt to correlate the steric and electrostatic fields of the compounds with the respective biological activity. The predictive ability of each resultant CoMFA model was evaluated using a test set consisting of four compounds. The best 3D quantitative structure−activity relationship model yielded cross-validated, conventional, and predictive r^2 values equal to 0.536, 0.998, and 0.778, respectively. A predictive model was obtained from the CoMFA analysis, which shall be useful for the development of new azole analogues as potential antifungals.

INTRODUCTION

Cytochrome P-450-dependent lanosterol 14α-demethylase (P-450$_{DM}$) catalyzes the transformation of lanosterol to ergosterol, an important constituent of the cell membrane in fungi, by causing the elimination of the 14α-methyl group of lanosterol to give the C14−C15 unsaturated sterol.[1,2] The accumulation of 14α-methylated sterols in azole-treated fungal cells affects membrane structure and function, resulting in an inhibition of the growth of fungi. Differential inhibition of this enzyme between pathogenic fungi and man is the basis for the clinically important activity of these azole antifungal agents. The specificity of the inhibitors is determined by the differential complementarity between the structure of the agent and the active sites of the fungal and host enzymes.[3] In fact, one of the reasons to pursue the search for superior antifungal agents is to increase their specificity toward fungal enzyme. This is crucial, especially under pathological cases where the immune system is compromised to a great extent (AIDS pathology) and the side effects (inhibition of the host P-450) due to overdosage of the azole compounds may eventually cause the death of patients.[4,5]

Since no structural information is available about the azole antifungal compounds in complex with cytochrome P-450$_{14\alpha DM}$ enzyme, a ligand-based approach was implemented in an attempt to understand the important interactions necessary for binding. In this paper, the combination of Apex-3D[6,7]-based alignment and comparative molecular field analysis (CoMFA)[8] techniques were employed to obtain a three-dimensional quantitative structure−activity relationship (3D-QSAR) for 20 diverse cytochrome P-450$_{14\alpha DM}$ inhibitors. The results of conformational analysis, pharmacophore identification, and comparative molecular field analysis studies for the azole antifungal agents that exhibit a range of in vitro inhibitory activity (MIC) against *Candida albicans* (*C. albicans*), are presented. The term "MIC" refers to log 1/MIC. The compounds, which have been studied, are shown in Table 1.[9−23]

COMPUTATIONAL METHODS

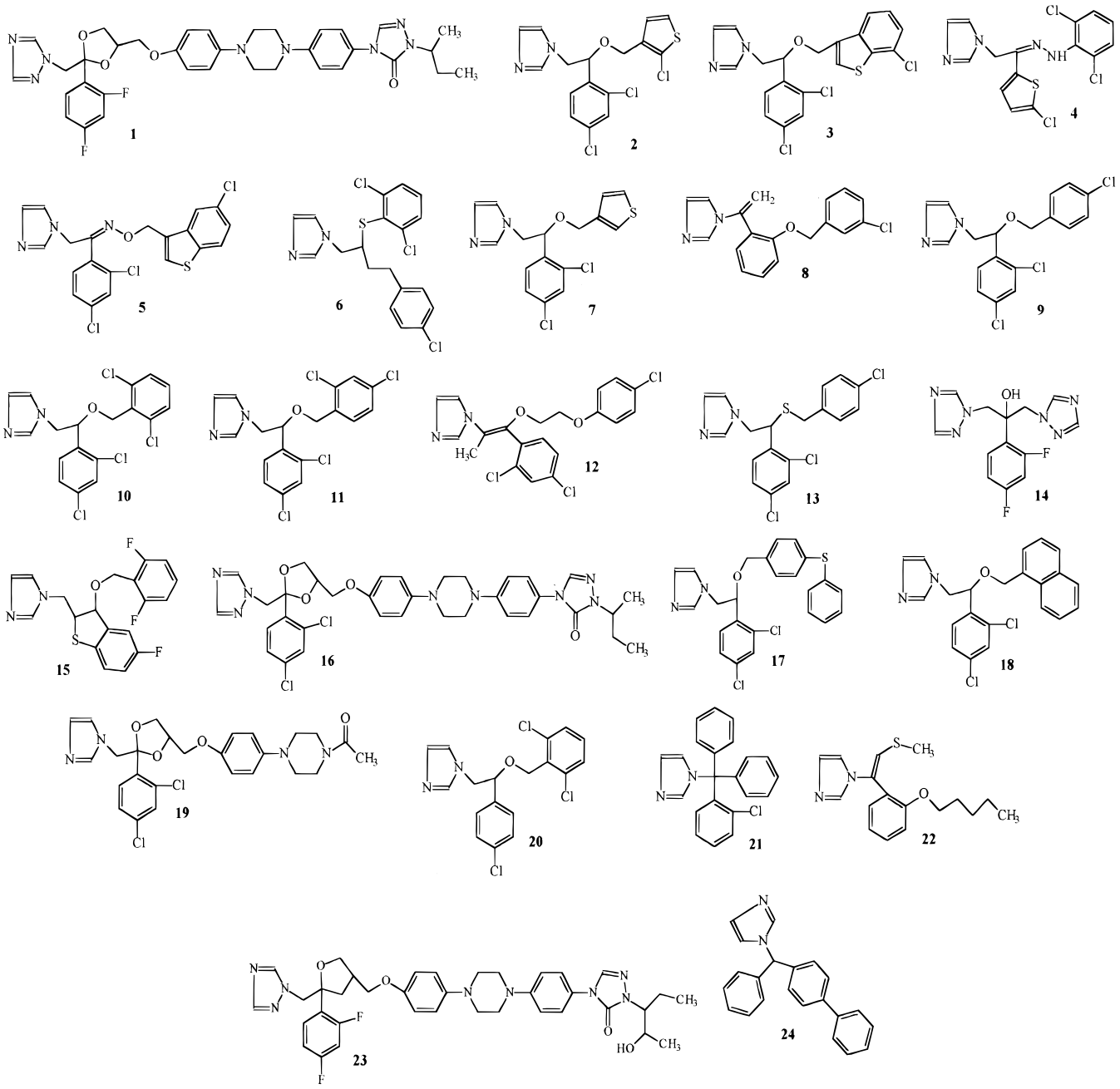
Modeling techniques described herein were performed on Silicon Graphics INDIGO^2 Solid Impact R4400 workstation (SGI) using the INSIGHT II 97.0 (Molecular Simulations Inc.),[24] and Sybyl 6.22 (TRIPOS Associates Inc., St. Louis, MO)[25] software.

Conformational Analyses and Apex-3D Model Development. The molecular structures of azole antifungals were drawn in 2D and converted to 3D using the *Sketch* and *Converter* module in INSIGHT 97.0 and then energy minimized to generate a reasonable standard geometry. A systematic conformational search in the *Search*/*Compare* module identified the sterically accessible conformations for these structures using a rotation increment of 15° for all the torsional angles. These conformational results were confirmed by means of the molecular mechanics method using the conjugate gradient algorithm of the Discover 97.0 program (Molecular Simulations, Inc.). The list of allowed conformations was too large, so only the lowest energy conformers for each compound having potential energies below 10 kcal mol^{-1} of the current global minimum were saved.[26,27]

The "alignment rule", that is, the superimposition of the molecular models within a three-dimensional fixed lattice,

10.1021/ci990020o CCC: $18.00
Published on Web 10/08/1999

Table 1. Structures of the Training (**1**−**20**) and Test (**21**−**24**) Set of Compounds



is one of the most important requirements in CoMFA studies. To define the alignment rules, one can use a variety of methods that are generally dependent on whether crystallographic data are available. Since no X-ray structural information for any of the receptor−azole complexes was available, we studied two types of alignment strategies, viz., rms (root-mean-square) pharmacophoric atom fit and automated Apex-3D pharmacophore alignment (guided by conformational flexibility and biological activity). The pharmacophoric atom fit alignment is achieved by rms fitting the corresponding key substructures. The key substructures in azole antifungals identified by Apex-3D, which are essential in the interaction with the binding site of cytochrome P-450$_{14\alpha DM}$, are (*i*) N3 and N4 of the azole ring, (*ii*) etherial oxygen atom, and (*iii*) aromatic ring centroid.[26] Thus, it was reasonable to align our compounds by following a pharmacophoric scheme mentioned above. Additionally, the CoMFA field fit alignment technique was applied to each alignment separately, leading to a total of four alignments for the training set of 20 molecules.

Apex-3D was used for automated pharmacophore identification and 3D-QSAR model building. To identify statistically significant biophore for these flexible compounds, a set of conformers were generated by a systematic conformational search method using molecular mechanics CVFF (consistent valence force field) calculations as described in Computational Methods. The low-energy bioactive conformers were identified that represent the pharmacophore. Due to the high flexibility of these compounds, the number of conformers generated were too large, a *set_energy_params* command was used to limit the maximum number of conformers to 100 with a molecular mechanics energy

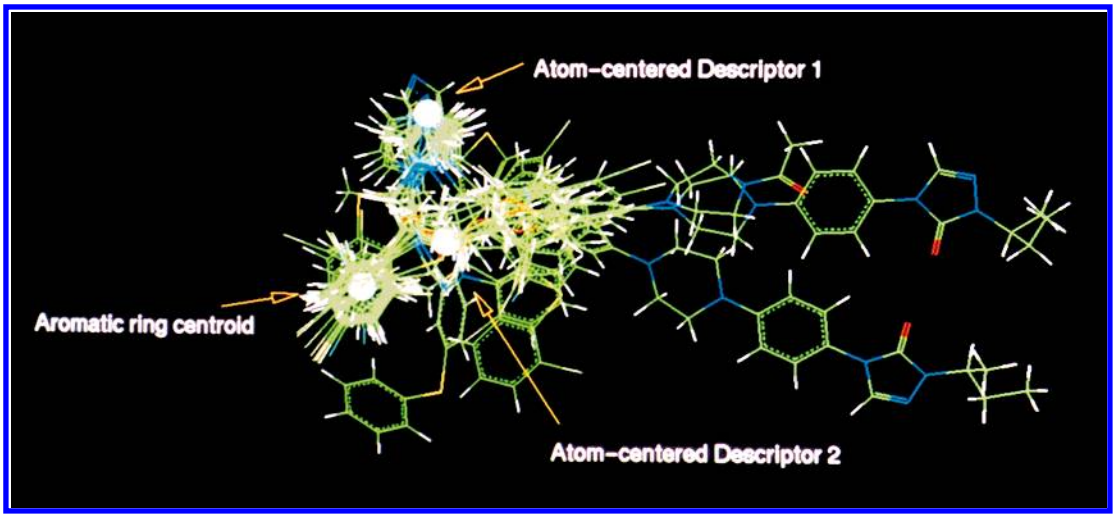


Figure 1. Pharmacophore alignment and biophoric site points of model I (20 compounds) identified by Apex-3D. Circles denote two atom-centered descriptors and one ring center.

threshold set to 10 kcal mol^{-1} to reduce the number of conformers to only the energetically most stable. This process effectively excluded those conformers that were not unique. To filter out conformers that did not add new information on biophore geometry and resultant superimpositioning of flexible ligands, Apex-3D was then used for internal conformer clusterization. The sample of conformers for each compound were partitioned into a number of clusters. After the conformer clustering, the lowest energy conformer of each cluster was selected as representative of that cluster and exported to Apex-3D, which is within INSIGHT II (97.0), for automated biophore identification and 3D-QSAR model building. The best fit pharmacophoric conformers of all the 20 compounds of model I are shown superimposed in Figure 1.

Comparative Molecular Field Analysis. Each model obtained from the Apex-3D analysis provided the alignments for the CoMFA study which correlated the steric and electrostatic fields of each molecule with its MIC against *C. albicans*. CoMFA uses a partial least squares (PLS)[28] algorithm to perform the correlation and was set to develop a five component model with five cross-validation (cv) groups. Subsequent CoMFA analyses with the best model were performed using the number of cross-validation groups set to the number of observations. The best analysis was then repeated without cross-validation (with an optimum number of components obtained during cross-validation), and the resulting equation was used for the internal prediction of MIC activities.

To assist selection among various 3D-QSAR calibration equations (models) and to test their utility as predictive tools, an external set (the so-called test set) of compounds with known activities but not used in model generation was predicted. The predictive r^2 based only on molecules from the test set was reported as the most appropriate parameter to evaluate the predictive power of a CoMFA model.

Predictive r^2 is calculated as[8]

$$\text{predictive } r^2 = 1 - (\text{“press”}/\text{SD})$$

where SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the training set compounds and “press” is the sum of the squared deviations between predicted and actual activities for every compound in the test set. A predictive r^2 value of 1 means that the CoMFA model is perfectly predictive for the test set, while prediction of a mean value of the training set for every member in the test set yields a predictive $r^2 = 0$. This procedure was used to evaluate our CoMFA models.

As a further attempt to validate the best CoMFA model obtained, the bootstrap validation method was used as implemented in Sybyl.[29] This technique is used to estimate the stability of the parameters (mean and standard deviation) associated with the statistical models.

CoMFA fields were calculated within the QSAR module of SYBYL 6.22. The steric (van der Waals interaction) and electrostatic (Coulombic values with a $1/r$ distance-dependent dielectric function) fields were calculated at each lattice intersection on a regularly spaced grid. The grid spacing was 2.0 Å in each direction. We used two different probe atoms in the study, an sp^3 hybridized carbon probe atom with a van der Waals radius of 1.52 Å and a charge +1 (default probe atom in Sybyl) and an sp^3 oxygen with a charge of −1.0. Two different energy truncation values of 30/30 and 5/30 kcal/mol were set, respectively, for the steric and electrostatic interactions. In case of 5/30 kcal/mol energy truncation as used by other authors,[30,31] the steric energy was truncated at lower values to avoid unjustified large parametric variance due to the steep increase of the steric field contribution at lattice points close to the molecules. The electrostatic interactions experienced at lattice points where steric energy exceeded the truncation value were set to the mean of all the electrostatic values at the same locations. However, the truncation of the steric energy at a low cutoff value can imply that some lattice points where the electrostatic field displays a significant value are not appropriately considered in the analysis. Column filtering was equal to 2.0 kcal/mol. QSAR tables were generated from the training set with the compounds as rows and the target biological data as a column. Partial atomic charges required for calculation of the electrostatic interaction were computed by a semiempirical molecular orbital method[32] using the MOPAC program. The charges were computed using the AM1 method.[33]

Table 2. Conformational Data and MIC of Training (**1**−**20**) and Test (**21**−**24**) Set of Azole Antifungals

compd	conformers[a]	ΔE[b]	MIC[c]
1	16	8.72	1.1740
2	57	2.13	−1.0969
3	17	6.31	0.6776
4	26	1.52	−0.7930
5	11	2.27	−1.0409
6	22	4.17	−1.3010
7	10	0.30	−0.7930
8	46	3.27	−1.7770
9	13	3.46	−2.0000
10	49	3.06	−2.0000
11	67	4.10	−1.0000
12	55	3.70	0.4318
13	30	3.52	−0.7121
14	38	8.63	0.5312
15	25	4.29	0.4424
16	30	0.30	−0.7258
17	29	7.79	−1.3010
18	13	3.95	−1.0000
19	52	5.94	2.0000
20	7	3.51	−1.0000
21			−0.8696
22			−0.8096
23			0.9935
24			−0.3979

[a] Within 10 kcal mol^{-1} above that of the lowest energy conformation found. [b] Energy difference between the Apex-3D conformer and the lowest energy conformer. [c] Biological activity (MIC in micrograms per milliliter) is expressed as (log 1/MIC).

RESULTS AND DISCUSSION

Conformational Analysis. To identify a statistically significant biophore for these flexible compounds, a set of conformers was generated by a systematic conformational search method using molecular mechanics calculations as described in Computational Methods. The number of conformers generated were restricted to 100 by applying molecular mechanics energy threshold to 10 kcal mol^{-1}. This process effectively excluded those conformers that were not unique. These conformers were further partitioned into a number of clusters within the Apex-3D program. Having determined the lowest energy state of each of the compounds, the energy of each compound in the pharmacophoric configuration was calculated and compared to determine whether the pharmacophoric conformation was a viable entity or a modeling artifact and shown in Table 2.[26] The best fit pharmacophoric conformers of all 20 compounds of model I are shown superimposed in Figure 1. Most of the compounds in the training set were able to attain a common pharmacophoric configuration with a moderate energy penalty (Table 2, ΔE < 10 kcal/mol) that represents the difference in energy of the pharmacophoric and the lowest energy conformations.

Apex-3D Analysis. Apex-3D (automated pharmacophore alignment method) uses a logicostructural approach to identify molecular features common to a set of diverse chemical structures. Advanced statistical techniques and 3D pattern-matching algorithms are used to assign probabilities for the identified biophores that are contributing to the biological activity of these compounds. These biophores are composed of various descriptor centers. Calculations with this method were performed on these sets of compounds. The automated method identified the following features important for activity from the training set (Table 1) superimposed on three distinct pharmacophoric site points: two atom-centered descriptors and one aromatic ring center. Initially, 10 models were selected with one model associated with each conformer of the training set. The selection of an appropriate model is based on the minimal deviation of the superimposed biophoric centers, the best possible fit of each of the superimposed ensembles, and the results obtained from 3D-QSAR analysis. These 10 models served as trial alignments for CoMFA studies. The CoMFA models were derived for each of these alignments. Only alignment model I−III and VIII show some correlation to the CoMFA fields and biological activity (Table 3). Close inspection revealed that model I seems to represent the most rational alignment when the structural diversity of all 20 compounds in the training set is considered. Biophores that have representative compounds (model I) with a high match were subsequently used in CoMFA studies.

Table 3. Apex-3D Model Evaluation

	r^2 with CoMFA		
model	5 cv groups[a]	2 cv groups[b]	20 cv groups
I	0.209	0.126	0.207
II	0.212	0.119	0.226
III	0.235	0.089	0.271
IV	0.082	−0.220	0.117
V	0.063	−0.510	0.173
VI	0.111	0.051	0.129
VII	0.129	0.000	0.165
VIII	0.204	0.138	0.206
IX	0.101	−0.015	0.101
X	0.147	0.046	0.139
XI[c]	−0.024	−0.082	−0.024
XII[d]	0.037	−0.101	−0.026
XIII[e]	0.061	−0.014	0.074

[a] Average of 50 cross-validated runs. [b] Average of 50 cross-validated runs. [c] Atom fit alignment using pharmacophoric atoms of the molecules. [d] Atom fit alignment model XI refined by field fit using compound 1 as template. [e] Model I refined by field fit using compound 1 as template.

CoMFA Analysis. To decide on the best alignment rule, both techniques, Apex-3D derived automated alignment and pharmacophore atom fit (rms fitting) alignment and subsequent field fitting, were studied. Field fit alignment is considered to be more relaxed in comparison with atom pair fit techniques. It is based on fitting the steric and/or electrostatic fields of the template and target molecules by global rotation/translation in order to minimize the differences between the fields of both molecules. If the aligned molecules are focused to share a common global shape and location in the 3D lattice, the entropic contributions to the free energy are expected to be similar.[34] The results of different alignment techniques are given in Table 3. The results demonstrate that the field fit and pharmacophore atom fit alignments do not improve the significance of CoMFA models and show the adequacy and applicability of the Apex-3D based automated alignment rule to the present problem. The structural diversity of the studied compounds made it difficult to align them by any other methods.

Several cross-validation groups were used to assess the predictive power of the various models. Cross-validated r^2 from 20, 5, and 2 groups are shown in Table 3. The analyses with two and five cross-validation groups were performed;

Table 4. CoMFA Results from Apex-3D Alignment Model I

analysis	observation deleted	no. of cv groups	no. of components	(r^2/R^2)	SEP or SEE	*F* value	probability of $R^2 = 0$
1	none	20	5	0.207	0.995		
2	none	0	5	0.997	0.070	912.491	0.0
3	16	19	4	0.465	0.875		
4	16	0	4	0.995	0.086	744.203	0.0
5	17	19	1	0.249	0.979		
6	17	0	1	0.698	0.621	39.317	0.0
7	20	19	5	0.192	1.024		
8	20	0	5	0.997	0.070	901.333	0.0
9	7	19	5	0.235	1.025		
10	7	0	5	0.997	0.068	962.059	0.0
11	6	19	5	0.227	0.997		
12	6	0	5	0.997	0.070	875.059	
13	16, 17	18	3	0.506	0.836		
14	16, 17	0	3	0.989	0.133	404.851	0.0
15	16, 6	18	5	*0.536*	0.845		
16	16, 6	0	5	*0.998*	0.054	1492.580	
17	16, 20	18	5	0.454	0.897		
18	16, 20	0	5	0.998	0.054	1482.664	0.0
19	16, 7	18	4	0.489	0.905		
20	16, 7	0	4	0.996	0.085	758.132	0.0
21	16, 15	18	4	0.488	0.882		
22	16, 15	0	4	0.995	0.089	660.257	0.0
23	16, 15, 17	17	3	0.549	0.843		
24	16, 15, 17	0	3	0.987	0.144	329.736	0.0
25	16, 17, 7	17	5	0.529	0.871		
26	16, 17, 7	0	5	0.998	0.059	1251.766	0.0
27	16, 17, 20	17	3	0.494	0.864		
28	16, 17, 20	0	3	0.989	0.137	377.309	0.0
29	16, 15, 7	17	4	0.497	0.917		
30	16, 15, 7	0	4	0.996	0.082	790.560	0.0
31	16, 15, 20	17	4	0.474	0.908		
32	16, 15, 20	0	4	0.995	0.094	586.834	0.0

each of the respective submodels consisted of 50 and 80% of the compounds (randomly selected), and the remaining ones were predicted. Since the random formation of cross-validation groups might have an impact on the results, this analysis was repeated 50 times.

The model I from Apex-3D analysis was chosen for detailed CoMFA analysis. A summary of the numerical results from these CoMFA studies is listed in Table 4. Models II and III (Table 3) with better statistical data were not chosen because some of the training set compounds were not aligned properly within the pharmacophore. For example, in model III structurally similar compounds **1** and **16** are not aligned in a similar fashion. There is no reason these compounds should have different modes of binding to the receptor. These alignment models were compared with the orientations of the azole antifungals in the active site of the receptor model obtained from our docking studies.[35] The model I alignment had all 20 compounds in the same orientation as that observed from the docking studies. Hence, models II and III were not considered further for detailed CoMFA studies.

The process of developing a suitable CoMFA model required the evaluation of various training sets (Table 4) utilizing cross-validation methods. In certain CoMFA training sets, one or more compounds were eliminated from the analysis with subsequent analysis of the correlation. The "leave-one-out" cross-validation method was used so that each compound would be systematically excluded and its activity predicted. Several cross-validated analyses in which two or more compounds were dropped (analyses 12, 15, and 23) showed significant increases in r^2 to a value of 0.536 when two compounds **6** and **16** were excluded from the analysis. At present it is difficult to rationalize the exclusion of these two compounds. We felt that the best CoMFA results were obtained from analysis 15 that gave a $r^2 = 0.536$ with an optimal number of five PLS components. Analysis 15 was chosen since it accommodated the largest number of compounds (18). Comparison to results obtained using Gasteiger−Marsilli charges revealed a decrease in $r^2 = 0.449$ from the analysis 15. The numerical results of non-cross-validated PLS analyses of 1 and 15 are listed in Table 5 and a plot of the fitted versus actual MIC is shown in Figure 2. The results from this analysis indicated a significant correlation ($R^2 = 0.998$), a small standard error (0.054), and a statistically insignificant probability of a spurious correlation ($p = 0.000$). The steric contribution was 56.1%, while the electrostatic contribution was 43.9%.

To find the optimum number of PLS components, several cross-validated runs were performed with an increment of 1−10 PLS components on analysis 15. The cross-validated r^2 initially increases with the number of components and then reaches an almost constant value after the optimum number of components is attained. This is shown in Figure 3. The cross-validated r^2 increases from a value of 0.472 for one component and reaches nearly a plateau at a value of 0.546 for seven components. As the number of components increases, the predictivity of the models increases. But this also increases the complexity of the models. To have a reasonable balance in the predictivity and complexity of the derived models a five-component model was selected.

The scrambling experiments were performed on this analysis to detect chance correlations. In this experiment, the biological activity of the training set compounds was randomly interchanged and a cross-validation model was

Table 5. Predicted MIC's and Residuals of a Training Set of Compounds (Model I) from CoMFA Analysis[a]

	analysis 2			analysis 16	
compd	actual	predicted	residual	predicted	residual
1	1.1740	1.1692	0.0048	1.1682	0.0058
2	−1.0969	−1.2345	0.1376	−1.2055	0.1086
3	0.6776	0.6306	0.0470	0.7064	−0.0288
4	−0.7930	−0.7729	−0.0201	−0.7453	−0.0477
5	−1.0409	−1.0558	0.0149	−1.0597	0.0188
6	−1.3010	−1.3201	0.0192	*	*
7	−0.7930	−0.7284	−0.0646	−0.7525	−0.0405
8	−1.7770	−1.8428	0.0658	−1.7748	−0.0022
9	−2.0000	−2.0013	0.0013	−2.0544	0.0544
10	−2.0000	−1.8364	−0.1636	−1.9337	−0.0663
11	−1.0000	−1.0156	0.0156	−1.0016	0.0016
12	0.4318	0.4043	0.0275	0.4382	−0.0064
13	−0.7121	−0.7751	0.0430	−0.6958	−0.0163
14	0.5312	0.5460	−0.0148	0.5363	−0.0051
15	0.4424	0.4575	−0.0151	0.4600	−0.0176
16	−0.7258	−0.7361	0.0103	*	*
17	−1.3010	−1.2620	−0.0389	−1.2719	−0.0291
18	−1.0000	−1.0278	0.0279	−1.0729	0.0729
19	2.0000	2.0389	−0.0389	1.9585	0.0415
20	−1.0000	−0.9413	−0.0587	−0.9563	−0.0437

[a] Asterisks indicate compounds compounds excluded from the analysis.

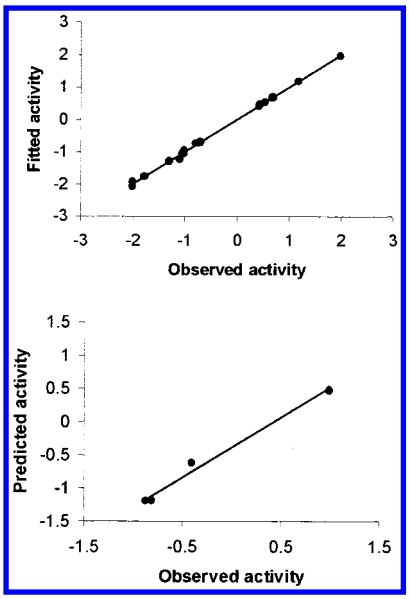


Figure 2. Fitted predictions versus measured (top) MIC values for the CoMFA analysis 16 of 18 compounds of the training set. The model was derived using five principal components yielding a cross-validated $r^2 = 0.536$. Predicted versus measured (bottom) MIC values for the test set of four compounds using the model I (analysis 16) to generate predictions. Predictive $r^2 = 0.778$.

derived using the optimum number of components. This analysis was repeated 50 times, and the mean cv r^2 obtained was −0.213 (standard deviation, 0.260). This suggests greater than 99.9% likelihood that the relationship derived from the correct assignment of the biological activity did not arise by chance.

To derive a statistically significant model with the training set of 20 compounds and to detect the outliers, various computational experiments were performed. Table 6 summarizes the results of 20 different models in which each of the 20 molecules was assumed to be an outlier. In such an experiment one would expect, a priori, that the model with a "true outlier" removed would show a high cross-validated

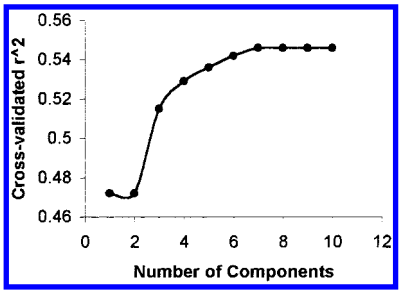


Figure 3. Cross-validated r^2 as a function of the number of components for model I (analysis 15).

Table 6. Effect of Elimination of Each Compound on the cv r^2 (Model I, Analysis 1)

compd excluded	cv r^2	compd excluded	cv r^2	compd excluded	cv r^2
1	0.150	**8**	0.188	**15**	0.202
2	0.182	**9**	0.157	**16**	0.465
3	0.249	**10**	0.157	**17**	0.249
4	0.196	**11**	0.190	**18**	0.190
5	0.224	**12**	0.162	**19**	0.143
6	0.227	**13**	0.206	**20**	0.192
7	0.235	**14**	0.129		

Table 7. CoMFA-PLS Analyses of 20 Inhibitors of Cytochrome P-450$_{14\alpha DM}$, Based on Apex-3D Alignment[a]

PLS analysis	30/30	5/30
cv r^2	0.207	0.385
SEP	1.123	1.026
R^2	0.997	0.998
SEE	0.070	0.052
no. of comp	5	6
fraction		
steric	0.578	0.533
electrostatic	0.422	0.467
cv r^2 [b]		
mean	−0.205	−0.269
std dev	0.217	0.329
max	0.186	0.308
min	−0.758	−1.137

[a] The cross-validated r^2 values and the explanatory R^2 (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with respect to the number of components listed in the table; the standard errors for both r^2 values (SEP, SEE) and the fractions of steric and electrostatic field contribution in the analysis are listed. [b] Results from scrambling experiments performed 50 times.

r^2 compared to others. Indeed, Table 6 shows such a result. When compound **16** was eliminated, the cv r^2 of the model jumped to 0.465. These experiments support our previous analyses.

We have also used different probe atoms to investigate the robustness of the CoMFA model. The leave-one-out cross-validation results using the entire training set showed some increase in the cross-validated r^2 from 0.207 to 0.220 with 5 optimum numbers of components when an sp^3 oxygen (O.3) probe with a charge of −1.0 was used instead of the sp^3 carbon (C.3) with a charge of +1.0. In most of the analyses, the models with a truncation value of the steric field set at 5 kcal/mol revealed better cv r^2. The results are shown in Table 7 from model I (20 compounds). These results obtained were not due to chance since the mean cv r^2 from the scrambling experiments were negative.

We wanted to see if we could improve the cross-validated r^2 results obtained from CoMFA analysis 15. In addition to

Table 8. Cross-Validated Runs Using Various Combinations of Columns, Model I (Analysis 1)

regressor	r^2	R^2	SEE	F value	cont.
a	0.207	0.997(5)	0.070	912.491	57.7/42.3
b	0.166	0.611(1)	0.693	28.313	100
c	0.300	0.691(1)	0.618	40.219	100
d	−0.144	0.119(1)	1.044	2.421	100
e	−0.315	0.004(1)	1.110	0.067	100
f	−0.202	0.027(1)	1.097	0.493	100
g	0.089	0.262(1)	0.955	6.404	100
a, d	0.129	0.996(5)	0.076	763.500	55.7/39.9/4.4
a, e	0.137	0.759(2)	0.562	26.757	55.5/35.1/9.4
a, f	0.232	0.997(5)	0.069	920.182	53.3/38.4/8.3
a, g	0.218	0.744(2)	0.579	24.663	53.0/30.5/16.5
a, d, e	0.237	0.624(2)	0.702	14.078	24.7/16.6/29.8/29
a, d, f	0.331	0.654(2)	0.674	16.033	22.1/14.6/30.9/32.4
a, d, g	0.196	0.454(1)	0.822	14.944	14.6/9.3/30.6/45.5
a, e, f	0.235	0.763(2)	0.557	27.404	45/28.2/7.4/19.5
a, e, g	0.138	0.808(3)	0.516	22.480	49.8/29.5/9.3/11.4
a, f, g	0.227	0.823(3)	0.496	24.769	47.9/26.4/14.3/11.4
a, e, d, g, f	0.409	0.919(4)	0.346	42.824	30.2/20.5/8.7/9.4/ 10/21.2

[a] CoMFA (S,E). [b] Steric. [c] Electrostatic. [d] HOMO. [e] ClogP. [f] Total_Hydrophobicity. [g] ΔH_{conf}.

the CoMFA steric and electrostatic fields generated by the SYBYL QSAR program, HOMO, ClogP, Total_Hydrophobicity, and ΔH_{conf} value were included in the model. ClogP (calculated log *P*) values represent a measure of lipophilicity calculated using the Biobyte 1.0 program. Total_Hydrophobicity values were calculated from Apex-3D program. HOMO energies were obtained from MOPAC AM1 semiempirical method, and ΔH_{conf} values were obtained from the difference between the energy of the molecule in its aligned conformation and energy of the molecule at the nearest local minimum. Multiple cross-validated runs were performed using various combinations of the additional columns. The results are listed in Table 8. In all the analyses 20 cross-validation groups were used and a minimum σ was set to 2.0. Several models had cross-validation results better than CoMFA analysis I, which included the entire training set. Unfortunately no improvement over CoMFA analysis 15 cv r^2 was observed, even when the additional regressors HOMO, ClogP, Total_Hydrophobicity, and ΔH_{conf} were included.

At present no logical reasoning can be given for these outliers as the close analogues of these compounds are well predicted by the CoMFA model. Compounds **6** and **16** are true outliers. There could be factors other than those which we have considered in the present study that are involved in the activity of these outliers. All further analyses such as test set predictions and generation of CoMFA contour maps were preformed using the model derived from analysis 16.

As already mentioned, the four compounds **21**−**24** (test set) were used to evaluate the predictive power of the CoMFA model I (analysis16). The r^2_{pred} (0.778) shows the significance of the CoMFA model I. The actual and corresponding CoMFA predicted MIC values for the compounds of the test set are shown in Table 9 and plotted in Figure 2. The high bootstrapped (30 samplings) r^2 value of $0.999 \pm 0.001$and small standard deviation (0.047 ± 0.045) suggest a high degree of confidence in analysis 16. The results are shown in Table 10.

Shown in Figure 4a are the important steric contour maps from the CoMFA model and the most active analogue, compound **19** (red), and least active analogue, compound **9** (blue), as reference structures. The steric fields are colored green, where an increase in biological activity correlated with increased steric bulk, or colored yellow, where a decrease in biological activity correlated with increased steric bulk. The phenyl/thienyl/benzothiophene/naphthyl ring in the least active compounds **5**, **7**−**11**, **18**, and **20** occupy sterically unfavorable regions. The large green polyhedron located around the long lipophilic moieties benzyloxy/piperazinylphenyl/benzothiphene/aryloxy of the most active compounds such as **1**, **3**, **12**, **15**, and **19** indicate favorable steric interactions of these structural moieties with a large hydrophobic pocket in the active site.[35] The lack of a large favorable steric field at our pharmacophore aromatic ring centroid can easily be explained: this region does not vary. Every compound had an aromatic ring centroid, and every compound was aligned by this centroid. CoMFA only correlates structural regions that vary in biological activity.

Table 9. Observed and Predicted MIC Values of the Compounds **21**−**24** Forming the Test Set from Model I (Analysis 16)

compd	obsd	predicted	residual
21	−0.8696	−1.2000	0.3304
22	−0.8096	−1.1810	0.3714
23	0.9935	0.4820	0.5115
24	−0.3979	−0.6150	0.2171

Table 10. Statistics of the Calibration CoMFA Model Based on Model I (Analysis 16) Including Predictive r^2 for the Entire Test Set (Compounds **21**−**24**)

R^2	0.998	steric and electrostatic	56.1/43.9
SEE	0.054	contribution (%)	
F test	1492.58	r^2_{boot}[a]	0.999 ± 0.001
r^2_{pred}	0.778	SEE_{boot}[a]	0.047 ± 0.045

[a] Results of the bootstrapped analysis (30 samplings, minimum $\sigma = 2.0$).

The important electrostatic contributions in the CoMFA model are shown in Figure 4b along with the most active analogue, compound **19** (green), and least active analogue, compound **9** (yellow), as reference structures. The electrostatic fields are colored blue, where an increase in activity was correlated with increased positive electrostatic charge, or colored red, where an increase in activity was correlated with increased negative charge. The presence of a blue polyhedron near the azole nucleus suggests that low electron density at this nucleus enhances the biological activity, which is indeed true for highly active compound **19**. A red polyhedron near the ortho position of the phenyl ring suggests that higher electron density at this position may increase the activity. Oxolane ring/etherial/alcoholic oxygen occupying a red electrostatic field indicates that high electron density at this position increases the activity of a compound as observed in highly active compounds **1**, **3**, **12**, **14**, **15**, and **19**. Blue and red regions near the benzyloxy part of compound **9** suggest the respective low and high electron density favorable for activity. Another red polyhedra at the methylene group of the benzyloxy moiety of least active compounds such as **2**, **5**, **7**, **9**−**11**, **18**, and **20** suggest a bad interaction within the active site. A small red region at the para position of the benzyloxy group suggests that high electron density at this position may increase the biological activity.

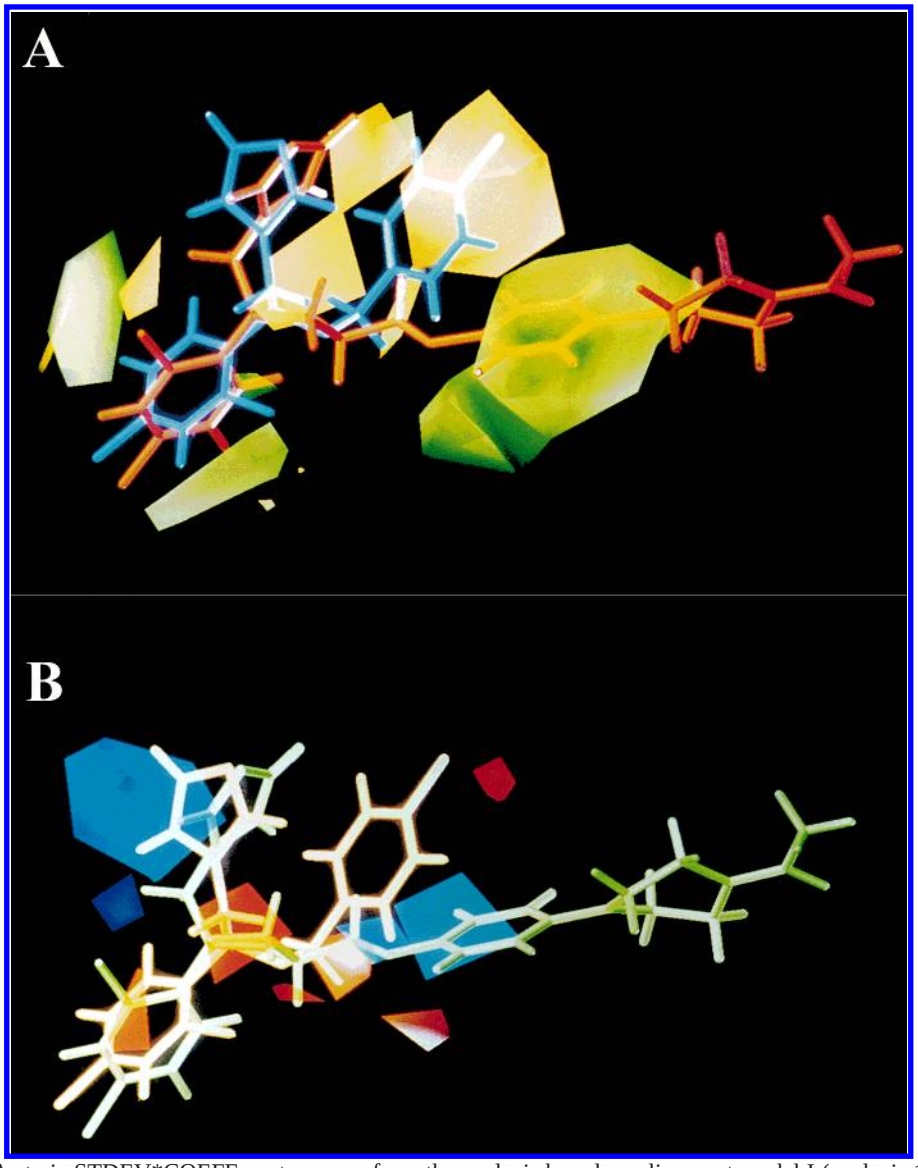


Figure 4. (a) CoMFA steric STDEV*COEFF contour map from the analysis based on alignment model I (analysis 16). Sterically favored areas (contribution level 80%) are represented by green polyhedra. Sterically unfavored areas (contribution 20%) are represented by yellow polyhedra. The most active compound **19** and the less active compound **9** are also represented as red and a blue structure, respectively. (b) CoMFA electrostatic STDEV*COEFF contour map from the analysis based on alignment model I (analysis 16). Negative charge unfavored areas (contribution level 80%) are represented by red polyhedra. Negative charge-favored areas (contribution level 20%) are represented by blue polyhedra. The more active compound **19** and the less active compound **9** are also represented as a green and a yellow structure, respectively.

A green region along with yellow polyhedra near phenyl rings of both compounds **9** and **19** indicates that activity decreases by structural moieties larger than a phenyl ring. A red region at oxolane ring with a yellow cap indicates that high electron density with less steric bulk is optimum for activity, as seen in compound **19**.

CONCLUSIONS

Using Apex-3D alignment as a criterion for the superimposition of the molecules, the conformational flexibility and structural diversity of a diverse set of azole antifungals has been studied. On the basis of this alignment, we have proposed a three-component pharmacophore for these molecules. This model justifies the importance of the main pharmacophoric groups (N3 atom of azole ring, etherial oxygen, and aromatic ring centroid) as well as of their relative distances.

This Apex-3D derived pharmacophore model served as a suitable mode of alignment for subsequent CoMFA studies. Elimination of two compounds from the initial CoMFA model of choice resulted in a model with significant internal predictability that correlated the steric and electrostatic properties of 18 compounds with their MIC values. This CoMFA model also accurately predicted the biological activities of four compounds, which were not used during model generation. The qualitative and quantitative information from the model should assist in the design of additional inhibitors of the cytochrome P-$450_{14\alpha DM}$ receptor.

ACKNOWLEDGMENT

The authors thank the All India Council for Technical Education (AICTE), University Grants Commission (UGC), New Delhi, for the financial support under its special assistance and COSIST programs. The authors also thank Mr. Vijay Gokhale for useful discussion.

REFERENCES AND NOTES

(1) Tafi, A.; Anastassopoulou, J.; Theophanides, T.; Botta, M.; Corelli, F.; Massa, S.; Artico, M.; Costi, R.; Di Santo, R.; Ragno, R. Molecular Modeling of Azole Antifungal Agents Active Against *Candida albicans*. 1. A Comparative Molecular Field Analysis Study. *J. Med. Chem.* **1996**, *39*, 1227−1235.

(2) Hitchcock, C. A.; Dickinson, K.; Brown, S. B.; Evans, E. G. V.; Adams, D. J. Interaction of Azole Antifungal Antibiotics with Cytochrome P-450 Dependent 14 α-Sterol Demethylase Purified from *Candida albicans*. *Biochem. J.* **1990**, *266*, 475−480.

(3) Yoshida, Y.; Aoyama, Y. Interaction of Azole Antifungal Agents with Cytochrome P-$450_{14\alpha DM}$ Purified from *Saccharomyces cerevisiae* Microsomes. *Biochem. Pharmacol*. **1987**, *36*, 229−235.

(4) Goldman, R. C.; Klein, L. L. Problems and Progress in Opportunistic Infections. In *Annual Reports in Medicinal Chemistry*; Bristol, A. J., Ed.; Academic Press: San Diego, CA, 1994; Vol. 29, pp 155−164.

(5) Odds, F. C.; Schmid, J.; Soll, D. R. Epidemiology of Candida Infections in AIDS. In *Mycoses in AIDS Patients*; Vanden Bossche, H., Ed.; Plenum Press: New York, 1990; pp 67−74.

(6) Golender, V.; Vorpagel, E. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 137−149.

(7) Van de Waterbeemd, H. In *Advanced Computer-Assisted Techniques in Drug Discovery*; Van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1994; Vol. 3.

(8) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA) 1. Effect of Shape in Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(9) Fromtling, R. A. *Drugs Future* **1985**, *10*, 983−984.

(10) Sammes, P. G. In *Topics in Antibiotic Chemistry*; Sammes, P. G., Ed.; Ellis Horwood Ltd.: Chichester, U.K., 1982; Vol. 6, Chapter 1, pp 22−97.

(11) *Drugs Future* **1981**, *6*, 444−445.

(12) *Drugs Future* **1988**, *13*, 624−626.

(13) Benfield, P.; Clissold, S. P. Sulconazole: A Review of Its Antimicrobial Activity and Therapeutic Use in Superficial Dermatomycoses. *Drugs* **1988**, *35*, 143−153.

(14) Xavier, R.; Hopkins, S. J. Neticonazole: A New and Highly Active Imidazole Antifungal agent. *Drugs Today* **1994**, *30*, 497−505.

(15) Plempel, M.; Regel, E.; Buchel, K. H. Antimycotic Efficacy of Bifonazole in Vitro and in Vivo. *Arzneim.-Forsch.* **1983**, *33*, 517−524.

(16) Bartoli, J.; Alguero, M.; Boncomte, E.; Forn, J. Synthesis and Antifungal Activity of a Series of Difluorotritylimidazoles. *Arzneim.-Forsch.* **1992**, *42*, 832−835.

(17) Michel, G. *Drugs Today* **1993**, *29*, 307−310.

(18) Walker, K. A. M.; Braemer, A. C.; Hitt, S.; Jones, R. E.; Matthews, T. R. 1-[4-(4-Chlorophenyl)-2-(2,6-dichlorophenylthio)-*n*-butyl]-1H-imidazole Nitrate, a New Potent Antifungal Agent. *J. Med. Chem.* **1978**, *21*, 840−842.

(19) *Drugs Future* **1996**, *21*, 160−166.

(20) Veronese, M.; Salvaterra, M.; Barzaghi, D. Fenticonazole: A New Imidazole derivative with Antibacterial and Antifungal Activity. *Arzneim.-Forsch.* **1981**, *31*, 2133−2137.

(21) Rane, D. F.; Desai, J. A.; Pike, R. E. 2-Azolylmethyl-3-difluorobenzyloxy-2,3-dihydrofluorobenzo[*b*]thiophenes and compositions containing them. *Chem. Abstr.* **1986**, *105*, P152918e.

(22) Fromtling, R. A. *Drugs Future* **1985**, *10*, 293−294.

(23) Drouhnet, E.; Dupont, B. *Drugs Future* **1996**, *21*, 160−166.

(24) *INSIGHT II* (Version 97.0), October, 1995; Molecular Simulations Inc.: San Diego, CA, 1995.

(25) *SYBYL* (Version 6.22) *Molecular Modeling System*; Tripos Associates Inc.: St. Louis, MO, 1995.

(26) Talele, T. T.; Kulkarni, V. M. Three-Dimensional Quantitative Structure−Activity Relationship (QSAR) and Receptor Mapping of Cytochrome P-$450_{14\alpha DM}$ Inhibiting Azole Antifungal Agents. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 204−210.

(27) Hariprasad, V.; Kulkarni, V. M. A proposed common spatial pharmacophore and the corresponding active conformations of some peptide leukotriene receptor antagonists. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 284−292.

(28) Stahle, L.; Wold, S. Partial Least Squares Analysis with Cross-Validation for the Two-Class Problem: A Monte Carlo Study. *J. Chemom.* **1987**, *1*, 185−196.

(29) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18−25.

(30) Greco, G.; Novellino, E.; Silipo, C.; Vittoria, A. Comparative Molecular Field Analysis on a Set of Muscarinic Agonists. *Quant. Struct.−Act. Relat.* **1991**, *10*, 289−299.

(31) Kim, K. H.; Martin, Y. C. Direct Prediction of Dissociation Constants (pK_a's) of Clonidine-like Imidazolines, 2-Substituted Imidazoles, and 1-Methyl-3-substituted Imidazoles from 3D Structures Using a Comparative Molecular Field Analysis (CoMFA). *J. Med. Chem.* **1991**, *34*, 2056−2060.

(32) Kulkarni, S. S.; Kulkarni, V. M. Three-Dimensional Quantitative Structure−Activity Relationship of Interleukin 1-β Converting Enzyme Inhibitors: A Comparative Molecular Field Analysis Study. *J. Med. Chem.* **1999**, *42*, 373−380.

(33) Dewar, M. J. S. E.; Zoebisch, G.; Healy, E. F. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(34) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and limitations. In *3D QSAR in drug design*; Kubinyi, H., Ed.; Escom: Leiden, The Netherlands, 1993; pp 583−618.

(35) Talele, T. T.; Hariprasad, V.; Kulkarni, V. M. Docking Analysis of a Series of Cytochrome P-$450_{14\alpha DM}$ Inhibiting Azole Antifungals. *Drug Des. Discovery* **1998**, *15*, 181−190.

CI990020O