# 4D-QSAR Analysis of a Series of Antifungal P450 Inhibitors and 3D-Pharmacophore Comparisons as a Function of Alignment

Jianzhong Liu, Dahua Pan, Yufeng Tseng, and Anton J. Hopfinger\*
Laboratory of Molecular Modeling and Design (M/C 781), College of Pharmacy,
The University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231

Received July 14, 2003

A training set of 55 antifungal P450 analogue inhibitors was used to construct receptor-independent fourdimensional quantitative structure-activity relationship (RI 4D-QSAR) models. Ten different alignments were used to build the models, and one alignment yields a significantly better model than the other alignments. Two different methodologies were used to measure the similarity of the best 4D-QSAR models of each alignment. One method compares the residual of fit between pairs of models using the cross-correlation coefficient of their residuals of fit as a similarity measure. The other method compares the spatial distributions of the IPE types (3D-pharmacophores) of pairs of 4D-QSAR models from different alignments. Optimum models from several different alignments have nearly the same correlation coefficients,  $r^2$ , and cross-validation correlation coefficients,  $xv-r^2$ , yet the 3D-pharmacophores of these models are very different from one another. The highest 3D-pharmacophore similarity correlation coefficient between any pair of 4D-QSAR models from the 10 alignments considered is only 0.216. However, the best 4D-QSAR models of each alignment do contain some proximate common pharmacorphore sites. A test set of 10 compounds was used to validate the predictivity of the best 4D-QSAR models of each alignment. The "best" model from the 10 alignments has the highest predictivity. The inferred active sites mapped out by the 4D-QSAR models suggest that hydrogen bond interactions are not prevalent when this class of P450 analogue inhibitors binds to the receptor active site. This feature of the 4D-OSAR models is in agreement with the crystal structure results that indicate no ligand-receptor hydrogen bonds are formed.

## INTRODUCTION

Azole antifungal agents are widely used for the treatment of topical or internal mycoses and, in particular, AIDs related mycotic pathologies.<sup>1–4</sup> These agents block the transformation of lanosterol to ergosterol by inhibiting 14-α-lanosterol demethylase (P450<sub>14DM</sub>, CYP51), which is a member of the enzyme cytochrome P450 dependent superfamily. The depletion of ergosterol and the accumulation of lanosterol and some other 14-methylsterols alter membrane properties, increase membrane permeability, and inhibit cell growth and replication.<sup>5–9</sup> The first generation of antifungal inhibitors of CYP51, fluconazole and itraconazole, have revolutionized treatment of some serious fungal infections.<sup>10</sup> However, the treatment of other types of infections is still far from satisfactory, and there is a need for both new broad- and narrow-spectrum antifungal agents.

In recent years some molecular modeling studies focusing on structural and physicochemical properties of active and inactive azole antifungal agents, using the CoMFA<sup>11</sup> and Apex-3D<sup>12</sup> methods, have been performed. These studies have suggested structure—activity relationships to design new, more specific, drug candidates and for understanding enzyme active site topography. The Catalyst method<sup>13</sup> and the SYBYL6.2 software package<sup>14</sup> have also been used to perform a pseudo 3D-QSAR study on a set of azole antifungal agents, which included the specific consideration of stereoisomers, and five pharmacophore features were

identified. The application part of the study reported in this paper employs the same data set as used in the Catalyst-based investigation but employs the 4D-QSAR analysis methodology.<sup>15</sup>

4D-QSAR analysis is a molecular modeling method that generates quantitative 3D-pharmacophore models as a function of alignment for a set of analogues both independent of (RI) and dependent (RD) upon receptor geometry. However, 4D-QSAR analysis lacks a tool to quantitatively and visually compare 4D-QSAR models, and their corresponding 3D-pharmacophores, which are generated from different alignment rules. This paper reports the development and implementation of such a tool for model and pharmacophore comparisons. The application of 4D-QSAR analysis, and the corresponding capacity to compare the similarity of 3D-pharmacophores of a series of azole P450 inhibitor antifungal agents, provides new insight and design criterion relative to previous studies.

# MATERIALS AND METHOD

Training Set of P450 Antifungal Analogues. Tafi et al.<sup>13</sup> have reported a structure—activity data set of 65 antifungal agents. The chemical structures of the selected training set of 55 antifungal analogues from the 65 compounds data set are given in Figure 1. A number of substituents on the phenyl ring, ranging in physicochemical properties, were chosen to understand the interaction of these compounds with cytochrome P450. There are three basic scaffold structures in Figure 1. The training set was chosen to span the three sets

<sup>\*</sup> Corresponding author phone: (312)996-4816; fax: (312)996-7107; e-mail: hopfingr@uic.edu.

Figure 1. The chemical structures of the 55 azole antifungal analogues used in the 4D-QSAR analysis training set.

of chemical structures and the entire range in  $-(MIC_{90})$ potency. The test set also spans the entire  $-\log(\text{MIC}_{90})$  range but is composed of only one of the these scaffold structure.

Receptor-Independent (RI) 4D-QSAR Analysis Applied to the Training Set. The operational formulation of the methodology for RI 4D-QSAR analysis, available in the 4D-QSAR software product, version 3.0,16 consists of 10 operational steps that have been described in detail in previous papers. This investigation employed a modification over the 10 steps in order to compare the similarity of the 3D-pharmacophores of the 4D-QSAR models as a function of alignment. This operational modification is summarized here only in terms of modeling the antifungal agents, and the 10 steps of 4D-QSAR analysis are as follows:

Step 1. An initial 3D structure of each of the 65 antifungal agents was constructed in the neutral form using the HyperChem 6.03 software.<sup>17</sup> Partial atomic charges were computed using the AM1 semiempirical method as implemented in the HyperChem program. Each structure was energy minimized also using AM1 without any geometric constraint. The energy-minimized structures are used as the initial structures in conformational ensemble sampling, step 3.

Step 2. The atoms of each molecule were classified into seven types of interaction pharmacophore elements (IPEs), which are defined in Table 1. A similarity matrix of the IPEs was constructed and is shown in Table 3 and discussed below. The IPE classification scheme is used to define the types of interactions involved with each pharmacophore site of a 4D-QSAR model. The IPE similarity matrix is used to facilitate the comparison of pharmacophores generated by 4D-OSAR analysis.

Step 3. Molecular dynamic simulation, MDS, was used to sample the conformational states available to each analogue and to generate the corresponding conformational

**Table 1.** Set of Interaction Pharmacophore Elements (IPEs)

IPE description	symbol	number code
any type of atom	A	0
nonpolar atom	NP	1
polar atom of positive partial charge	$\mathbf{P}^{+}$	2
polar atom of negative partial charge	$\mathbf{P}^{-}$	3
hydrogen bond acceptor	HA	4
hydrogen bond donor	HD	5
aromatic atoms of molecule	Ar	6

Table 2. Set of Trial Alignments Used in Constructing the 4D-QSAR Models

$$\begin{array}{c|c}
a & b \\
e & A \\
c & d \\
d & d \\
f & B \\
f &$$

align. no.	first atom	second atom	third atom
1	a	c	d
2	a	m	g
3	a	m	j
4	a	0	g
5	b	m	j
6	a	m	i
7	a	q	j
8	b	q	j
9	e	m	j
10	c	m	j

ensemble profile (CEP). The MDSs were done using the MOLSIM package<sup>18</sup> with an extended MM2 force field.<sup>19,20</sup> The simulation temperature for the MDS was set at 310K with a simulation sampling time of 40 ps with intervals of

**Table 3.** Interaction Pharmacophore Element (IPE) Similarity Matrix, S[IPE(A), IPE(B)]

	A(0)	NP(1)	P <sup>+</sup> (2)	P <sup>-</sup> (3)	HA(4)	HD(5)	Ar(6)
A(0)	1	1	1	1	1	1	1
NP(1)	1	1	0.1	0.1	0.1	0.1	1
$P^{+}(2)$	1	0.1	1	0.1	0.8	0.1	0.1
$P^{-}(3)$	1	0.1	0.1	1	0.1	0.8	0.1
HA(4)	1	0.1	0.8	0.1	1	0.1	0.1
HD(5)	1	0.1	0.1	0.8	0.1	1	0.1
Ar(6)	1	1	0.1	0.1	0.1	0.1	1

0.01 ps for a total sampling of 4000 conformations of each antifungal analogue. The atomic coordinates of each conformation and its intramolecular energy sampled during the MDS are recorded every 0.02 ps for a total of 2000 "frames", or steps, in the CEP of each analogue.

Step 4. The "standard" (RI) 4D-QSAR methodology uses three-ordered atom alignments to compare the molecules of a training set. <sup>21,22</sup> Ten alignments that systematically span the entire antifungal core structure were selected and are defined in Table 2. For some alignments, if the compound does not have the atom shown in Table 2, an atom at a similar position was chosen. In the case of alignment 4, for example, ring B of compound 39 is a six-member ring and has no nitrogen atom. The atom at the 3'-position of ring B is chosen for the nitrogen, because it is also two bonds lengths from the 1'-position of ring B, just as the nitrogen atom, see Table 2.

Step 5. Each conformation of an analogue from its CEP is aligned in the grid cell lattice using the invariant coordinates of the three-ordered alignment atoms. In this study, the size of the cubic grid cells of the lattice are 1 Å<sup>3</sup>, and the overall grid cell lattice size was chosen to fully enclose each compound of the training set. The normalized occupancy of each grid cell by each IPE atom type over the CEP for a given alignment forms a unique set of QSAR descriptors referred to as grid cell occupancy descriptors, GCODs. The GCODs are computed and used as the basis set of trial 4D-QSAR descriptors in 4D-QSAR analysis. Non-GCOD descriptors can also be added into the trial basis set by the user. No non-GCOD descriptors were considered in this study.

Step 6. A 4D-QSAR analysis generates an enormous number of trial QSAR descriptors, GCODs, because of the large number of grid cells and the seven IPEs. Partial least squares (PLS) regression analysis<sup>23</sup> is used to perform a data reduction fit between the observed dependent variable measures and the corresponding set of GCOD values.

Step 7. The most highly weighted PLS GCOD descriptors (currently the top 200), generated in step 6, are used to form the trial basis set for genetic algorithm (GA) model optimization. The specific GA currently used in the 4D-QSAR software<sup>16</sup> is modification of the genetic function approximation (GFA).<sup>24,25</sup> The GFA optimization is initiated using N (currently 300) randomly generated 4D-QSAR models. Mutation probability over the crossover optimization cycle is set at 10%. The smoothing factor, a GFA operations variable, controls the number of independent variables in the QSAR models and is varied in order to determine the optimal number of descriptors for the 4D-QSAR models. The diagnostic measures used to analyze the resultant 4D-OSAR models generated by the GFA include (i) descriptor usage as a function of crossover operation, (ii) linear crosscorrelation among descriptors and/or dependent variables (biological activity measures), (iii) number of significant and independent 4D-QSAR models, and (iv) indices of model significance including the correlation coefficient,  $r^2$ , leave-one-out cross-validation correlation coefficient,  $xv-r^2$ , and Friedman's lack of fit (LOF).<sup>26</sup> In this particular 4D-QSAR application, the alignment similarity comparisons were limited to models having same number GCODs.

Step 8. Steps 4–7 are repeated until all trial alignments are included in the 4D-QSAR analyses.

Step 9. Corresponds to the inspection and evaluation of the population of models obtained from the set of trial alignments. The goal of this step is to identify the best and distinct set of 4D-QSAR models which is referred to as the manifold model of the analysis.

Step 10. Hypothesize the "active" conformation of each compound in the training set. This is achieved by identifying all conformer states sampled for each compound that are within  $\Delta E$  of the global minimum energy conformation of the CEP. Currently,  $\Delta E$  is set at 2 kcal/mol. Each member of the resultant set of energy-filtered conformations is then individually evaluated in the best 4D-QSAR. The conformation within 2 kcal/mol of the apparent global minimum that predicts the highest activity in the best 4D-QSAR model is defined as the active conformation.

To fully explore and compare 4D-QSAR models, it is necessary to be able to compare, and establish, the similarity of models from different alignments. Such a comparison can be broken into two parts; comparison of activity endpoint predictions from models and, second, comparison of the 3Dpharmacophores of the models. Numerical activity endpoint predictions from pairs (A, B) of 4D-QSAR models from different alignments can be compared using the cross-correlation coefficient of their residuals of fit. This measure, the crosscorrelation coefficient of the residuals of fit to the common training set, has been used to identify equivalent and distinct models generated from a common alignment. Equivalent 4D-QSAR models will have highly correlated residuals of fit since they predict in the same way. Conversely, distinct models will have low cross-correlation coefficients owing to one model predicting differently from the other. This behavior of the cross-correlation coefficient of the residuals of fit holds for models from the same or from different alignments. Thus, the similarity of models from different alignments with respect to numerical activity endpoint prediction can be characterized by the cross-correlation coefficient of their residuals of fit to the common training set.

Comparing the similarity of the 3D-pharmacohores embedded in a pair of 4D-QSAR models is dependent upon both the similarity between the IPEs of paired GCODs and the spatial similarity of the locations of the GCODs of the models. A set of paired similarity measures for the IPEs defined in Table 1 is given in the IPE similarity matrix in Table 3. These similarity measures were arrived at by noting how often one IPE type was found to substitute for another IPE type in past 4D-QSAR studies.

The algorithm for establishing the spatial superposition, and corresponding measure of similarity of the 3D-pharmacophore of a pair (A, B) of 4D-QSAR models from two different alignments, is as follows:

1. The GCODs of each model are ranked according to the absolute values of their regression coefficients from largest to smallest.

**Table 4.** Cross-Validation Correlation Coefficient,  $xv-r^2$ , and Correlation Coefficient,  $r^2$ , of the Best 4D-QSAR Model for Each Trial

alignment	1	2	3	4	5	6	7	8	9	10
$r^2$	0.751	0.746	0.808	0.796	0.817	0.801	0.824	0.829	0.816	0.823
$xv-r^2$	0.693	0.719	0.760	0.742	0.762	0.734	0.768	0.784	0.753	0.771

<sup>a</sup>Alignments 5–10 are derivatives of alignment 3[bold]. Alignment 8 [bold] gives the best model judged by  $xv-r^2$ .

- 2. GCODs of equivalent rank from each model A and B are paired to one another. That is, the GCOD with the largest (absolute value) regression coefficient of model A is paired to the GCOD of B having the largest (absolute value) regression coefficient, the second ranked GCOD of A is paired to the second ranked GCOD of B, and the pairing continued until the last GCOD of the smaller model of A and B is paired. Each unpaired GCOD of the larger model is paired to the closest GCOD in the smaller model after step 3 below is performed. In this application, each 4D-QSAR model is of the same size so this last operation is not necessary.
- 3. A rigid rotation and translation of one 3D-pharmacophore relative to the other 3D-pharmacophore is performed to maximize the initial 3D-pharmacophore similarity scoring function

$$S'(A, B) = \sum_{i=1}^{N_s} \frac{\sqrt{\bar{W}_A(i)\bar{W}_B(i)}}{1 + D_{AB}(i)} S_i[IPE(A), IPE(B)] \quad (1)$$

In eq 1  $N_s$  is the number of GCODs in the smaller 4D-QSAR model between A and B, i is the ith ranked corresponding pair of GCODs from A and B,  $D_{AB}(i)$  the distance between the ith pair of GCODs from A and B, and  $S_i[IPE(A), IPE(B)]$  is the IPE similarity value of the ith GCOD pair as given in Table 3.  $\overline{W}_{X}(i)$  is the weight of the ith GCOD of model X and given by

$$\bar{W}_{X}(i) = \frac{C_{X}(i)}{\sum_{i=1}^{N_{s}} C_{X}(j)}$$

$$(2)$$

where  $C_X(j)$  is the jth regression coefficient of the jth GCOD of 4D-QSAR model X.

The final 3D-pharmacophore similarity measure is obtained by "tacking on" to S'(A, B) the additional terms arising from GCODs of the larger model according to the rule in step 2, so that for model A being larger than model B

$$S(A, B) = S'(A, B) + \sum_{i=1}^{N_L} \frac{\sqrt{\bar{W}_A(i)\bar{W}_B(k^*)}}{1 + D_{AB}(i,k^*)}$$

$$S_{i,k^*}[IPE(A), IPE(B)] (3)$$

where  $N_{\rm L}$  is the number of additional GCODs in model A relative to model B, and k\* indicates which GCOD of model B is closest to the ith GCOD of A. The validation of this algorithm is discussed below.

### **RESULTS**

4D-QSAR models were constructed for each of the 10 trial alignments listed in Table 2. Only alignments 1-4 were

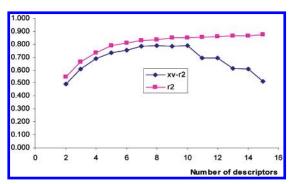


Figure 2. Plot of the number of 4D-OSAR model descriptors versus  $r^2$  and xv- $r^2$  for alignment 8.

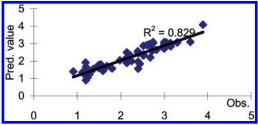


Figure 3. Predicted and observed activities [ $-\log(\text{MIC}_{90})$ ] of each analogue for alignment 8 using eq 4.

initially employed in the 4D-QSAR study. In alignment 1, only atoms of ring A define the alignment. Alignments 2-4 distribute the three-ordered atoms in systematic fashion across rings A, B, and C, respectively. Of these first four alignments, alignment 3 is obviously better than other three alignments based both  $r^2$  and  $xv-r^2$  from the best corresponding 4D-QSAR models which are given in Table 4. The subsequent six other alignments in Table 2 were chosen to explore alignment 3. The only differences among these six alignments as well as alignment 3 is that one, or two, of the three-ordered atoms in each of these alignments are one, or two, bond(s) distances from the atoms of alignment 3.

Figure 2 is a plot of  $r^2$  and  $xv-r^2$  as a function of number of model descriptors for alignment 8. From Figure 2 it is clear that maximum model significance is achieved for a model having seven descriptor terms, and there is no meaningful model improvement by the addition of more descriptor terms. Thus, to make comparisons, a seven-term 4D-QSAR model was built for each alignment. To avoid overfitting for any alignment by adopting a seven-term model, each alignment was checked by comparing the  $xv-r^2$ values of the seven-term model to that of the eight-term model. The results of these comparisons show that a seventerm model does not constitute overfitting for any alignment. 3D-pharmacophore similarity comparisons are based on seven-term models (seven pharmacophore sites). The corresponding best seven-term 4D-OSAR models, based on the cross-validated correlation coefficient,  $xv-r^2$ , are given in Table 4. Overall, alignment 8 yields the best overall model with respect to  $xv-r^2$  which is

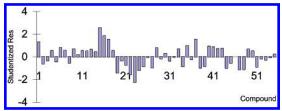


Figure 4. A plot of the studentized residuals, using eq 4, as function of compound number in Figure 1.

$$-\log(\text{MIC}_{90}) = 1.83 - 10.88\text{GC1}(-4,0,7,\text{Ar}) + \\ 7.44\text{GC2}(-3,3,2,\text{P}^-) + 6.90\text{GC3}(-1,-3,9,\text{A}) + \\ 4.92\text{GC4}(0,1,6,\text{Ar}) + 2.52\text{GC5}(1,4,5,\text{A}) + \\ 1.32\text{GC6}(0,-1,1,\text{NP}) - 1.30\text{GC7}(1,-1,5,\text{NP}) \\ n = 55, r^2 = 0.829, \text{xv-}r^2 = 0.784$$
 (4)

In eq 4, GCi(x, y, z, X) is the *i*th significant GCOD descriptor located at (x, y, z) and having the X type IPE as defined in Table 1. The observed and predicted, using eq 4,  $-\log(\text{MIC}_{90})$  values are plotted in Figure 3. There are no outliers using eq 4 which can be gleaned from both Figure 3 and the residuals plot in Figure 4.

A stereographic representation of the 3D-pharmacophore embedded in the 4D-QSAR model given by eq 4, is shown in Figure 5. The reference analogues used in Figure 5 are the most active compound, 39, and the most inactive compound, 22, each in its predicted active conformation. Two GCODs of the 4D-QSAR model specify pharmacophore sites that decrease potency when occupied since they have negative regression coefficients. The benzene ring of compound 39 fits "between" these two "bad" GCODs in somewhat of a "sandwich" fashion. One of the carbons of the benzene ring of compound 22 occupies GC7, one of the "bad" GCODs. The other five GCODs have positive regres-

sion coefficients and occupancy of these GCODs by the correct IPE will increase compound potency. Compound 22 does not have any substitutent in GC3, which is a major source of loss in potency for this compound. Moreover, both compounds 22 and 39 do not have polar negative IPE at GC2, which will increase activity. This GCOD is significant for compound 7, whose stereographic representation is shown in Figure 5C. But compound 7 also occupies GC7 and does not occupy GC3. Thus, the overall  $-\log(\text{MIC}_{90})$  of compound 7 is very low.

In general, eq 4 provides a self-consistent pharmacophore explanation of inhibition potency as function of the three classes of chemical structures of the training set. The *o*-position polar negative substituents of ring C are predicted to greatly increase compound activity. Also, the presence of GC7 indicates that a two bond spacing between rings A and B is better than a three bond spacing, since the three bond spacing leads to occupancy of this GCOD and a corresponding significant decrease in inhibition potency.

Table 5 lists the seven-terms of each best 4D-QSAR model of each alignment. The similarity measures between each pair of 3D-pharmacophores of the best models of each alignment, using the GCODs of Table 5, have been computed using the algorithm given in the **Materials and Method** section. The similarity of the best 4D-QSAR models from each of the alignments has also been computed with respect to the cross-correlation coefficients of the residuals of fit.

The residuals of fit analysis shows that the most similar 4D-QSAR models are those of alignments 1 and 9 with a correlation cross-coefficient of 0.339. However, this is a low value for the highest cross-correlation of the residuals of fit and, overall, is indicative that all the best 4D-QSAR models from all the alignments are largely independent of one another. That is, each best 4D-QSAR model "explains" the training set data distinctly different from the best 4D-QSAR

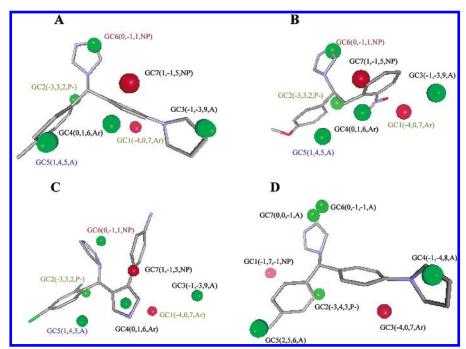


Figure 5. Stereoviews of the 3D pharmacophores relative to the most active compound 39 (A), the most inactive compound 22 (B), and compound 7 (C), in their respective active conformations as predicted using the best 4D-QSAR model [eq 4]. (D) is compound 39 in its active conformation as predicted using the best 4D-QSAR model of alignment 7. The GCODs are shown as spheres, although the actual grid cells are cubes in space.  $-\log(\text{MIC}_{90})$  enhancing grid cells are shown as green spheres, and grid cells that diminish  $-\log(\text{MIC}_{90})$  are shown as red spheres.

Table 5. Seven GCODs of Each Best 4D-QSAR Model of Each Alignment<sup>a</sup>

align.	GC1	GC2	GC3	GC4	GC5	GC6	GC7
1	-1530	-2640	0056	-3-460	-2-540	-3-3 8 1	1-3 5 1
	(-12.08)	(8.81)	(7.84)	(-7.65)	(-4.24)	(2.58)	(2.200)
2	-1010	-4-560	0076	4-3 5 1	0860	-3253	2-130
	(-9.68)	(5.02)	(-4.88)	(3.31)	(-1.85)	(-1.01)	(0.573)
3	-3-460	-2-231	2 4 4 0	1-570	0566	2220	-2-381
	(-8.72)	(-8.19)	(7.83)	(7.37)	(5.75)	(3.94)	(3.18)
4	2-251	-2161	0743	-1211	0-256	0670	0 - 146
	(-4.06)	(2.99)	(-2.92)	(1.94)	(-1.93)	(1.60)	(-1.19)
5	-4-273	0 6 1 3	0 8 5 1	2-265	-1-660	0-250	-1056
	(-13.63)	(-13.38)	(-8.86)	(-6.71)	(4.78)	(-1.48)	(0.88)
6	4256	0 - 480	-2-186	-1-550	-2261	-1-254	1-241
	(-15.44)	(11.14)	(-9.87)	(-9.57)	(-7.55)	(2.32)	(-1.88)
7	-17-11	-3433	-4076	-1-480	2560	$0 - 1 - 1 \ 0$	0 0-1 0
	(-10.37)	(8.15)	(-7.71)	(6.42)	(2.16)	(1.24)	(0.71)
8	-4076	-3323	-1-390	0166	1 4 5 0	$0-1\ 1\ 1$	$1-1\ 5\ 1$
	(-10.88)	(7.44)	(6.90)	(4.92)	(2.52)	(1.32)	(-1.30)
9	-4-151	-1-166	-1-530	-3-440	$0\ 1-1\ 0$	1 3 4 1	-1-1-20
	(-12.64)	(8.43)	(-6.71)	(-4.42)	(-2.60)	(2.12)	(1.59)
10	1831	-2356	-4266	3-230	-2-541	-2250	$1-1\ 5\ 1$
	(-18.77)	(14.00)	(-7.87)	(7.07)	(-6.58)	(-4.93)	(1.77)

<sup>a</sup> The first three numbers are the lattice coordinates of the GCOD, and the fourth number is the IPE type. The GCODs are listed across the table in descending order of the absolute value of their regression coefficients. The number below each GCOD is its corresponding regression coefficient.

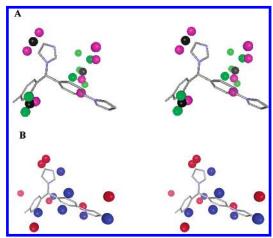


Figure 6. Stereoviews of the 3D pharmacophores overlaid with respect to the most active compound 39, in its minimum energy conformation. The GCODs are shown as spheres, although the actual grid cells are cubes in space. The pink GCODs define the 3D-pharmacophores of alignment 1, green alignment 9, red alignment 7, and blue alignment 8. The black spheres in part A are the cluster centers of three different clusters formed by the GCODs of alignments 1 and 9. All cluster analyses performed in this study used the K-mean iterate and classify method.<sup>27</sup>

models of the other 9 alignments! The 3D-pharmacophores of the best models of alignments 1 and 9 are optimally overlaid in Figure 6A based upon eqs 1-3. An inspection of this figure shows that these two 3D-pharmacophores have distinctly different pharmacophores sites, which is consistent with the low cross-correlation coefficient of the residuals fit for their corresponding 4D-QSAR models.

The 3D-pharmacophore similarity matrix, based on eqs 1-3, is given in Table 6. The most similar pair of 3D-pharmacophores comes from alignments 7 and 8, which are optimally overlaid in Figure 6B. However, the similarity measures is only 0.216 suggesting that the 3D-pharmacophores, like the corresponding 4D-QSAR models, of the 10 alignments are all different from one another. The 3Dpharmacophores of the models from alignments 7 and 8 are optimally superimposed in Figure 6B. Some similarities

among GCODs (pharmacophores sites) can be inferred. But overall, these two 3D-pharmacophores are significantly different from one another. The correlation coefficient between the corresponding elements of the cross-correlation of the residuals of fit matrix and the 3D-pharmacophore similarity matrix (Table 6) is only 0.126. This finding suggests that the relative similarity of the 4D-QSAR equations is not transferred to the corresponding spatial representations of the equations, namely the 3D-pharmcophores.

Both the  $r^2$  and  $xv-r^2$  of the best models from each of the alignments are all within 13% of those of the overall best 4D-QSAR model, eq 4, from alignment 8. One would also expect the average errors in estimation of  $-\log(\text{MIC}_{90})$  to be in the 15% region. Thus, it is difficult to exclude any best 4D-QSAR model of any alignment based on lack of model significance. It would seem that each best 4D-OSAR model of each alignment is not only unique but also as significant as any other best model.

A test set of 10 compounds was used to evaluate the predictivity of the best 4D-QSAR model of each alignment. However, the test set also served to definitively identify the 4D-QSAR model of alignment 8 as the best of the best models. Compounds 56-65, whose structures and observed -log(MIC<sub>90</sub>) values are given in Table 7, are the test set compounds. The compounds of the test set span the entire training set range in  $-\log(\text{MIC}_{90})$  value. The predicted  $-\log$ - $(MIC_{90})$  values and the  $r^2$  between the observed and predicted -log(MIC<sub>90</sub>) values for the best model of each alignment for the test set are given in Table 8. Table 8 shows that alignment 8 has highest predictivity for the test set. By comparing the  $r^2$  values between the observed and predicted  $-\log(\text{MIC}_{90})$  values of Table 8 with the training set  $r^2$  values in Table 4, it is seen that only alignment 8 has a test set  $r^2$ value anywhere near the training set  $r^2$  values.

Still, a test set  $r^2 = 0.631$  for alignment 8 is substantially less than the training set  $r^2 = 0.829$ . Thus, an exploration of the relationship between predicted test set  $-\log(\text{MIC}_{90})$ values and GCOD occupancy values was undertaken for each best 4D-QSAR model of each alignment. This exploration

Table 6. 3D-Pharmacophore Similarity Matrix from the Best 4D-QSAR Models of the Ten Trial Alignments Given in Table 2

	align1	align2	align3	align4	align5	align6	align7	align8	align9	align10
align1	1									
align2	0.128	1								
align3	0.109	0.164	1							
align4	0.154	0.125	0.129	1						
align5	0.113	0.091	0.141	0.036	1					
align6	0.162	0.173	0.157	0.127	0.091	1				
align7	0.148	0.155	0.156	0.121	0.101	0.182	1			
align8	0.127	0.139	0.143	0.139	0.098	0.14	0.216	1		
align9	0.138	0.148	0.175	0.159	0.056	0.174	0.116	0.139	1	
align10	0.061	0.067	0.075	0.068	0.053	0.062	0.063	0.086	0.068	1

**Table 7.** Ten Azole Antifungal Analogues Which Make Up the Test Set and Their Observed  $-\log(MIC_{90})$  Values

compd	R	$R_1$	$R_2$	$[-\log(\mathrm{MIC}_{90})]$
56	Cl	Н	Ph	3.34
57	Cl	Me	Ph	3.05
58	F	H	Ph	3.00
59	OMe	2,4-Cl2-Bn	Ph	1.39
60	H	2,4-Cl2-Bn	Ph	1.66
61	Cl	2,4-Cl2-Bn	Ph	1.40
62	Me	Me	Ph	2.42
63	H	Me	Ph	2.40
64	Me	2,4-Cl2-Bn	Ph	1.38
65	F	Me	Ph	2.40

was limited to identifying the two largest outliers for each 4D-QSAR model that are indicated in bold print in Table 8. The GCOD occupancy values for the two largest outliers of each best 4D-QSAR model of each alignment have been analyzed to identify the GCODs most responsible for each of these poorly predicted  $-\log(\text{MIC}_{90})$  values. The results are given in Table 9. For alignment 1, compounds 56 and 57 are two outliers. The GC4 values for the 4D-QSAR model given in Table 5 for these two compounds are 0.376 and 0.438, respectively, and both values are larger than the occupancy values of other test compounds. But, in addition, the regression coefficient of this GCOD is -7.65, resulting in these two compounds having low predicted  $-\log(\text{MIC}_{90})$  values. For alignment 2, the GC3 occupancy value of

compound 56, one major outlier, is 0.335 and the regression coefficient of GC3 is -4.88. Thus, compound 56 is predicted to have a low  $-\log(\text{MIC}_{90})$  value. The major outlier for alignment 2 is compound 64 which has a GC3 occupancy value of zero, leading to relatively high predicted -log-(MIC<sub>90</sub>) value. For alignment 3, compounds 60 and 61 are the two largest outliers. The GC4 occupancy values for these two compounds are 0.215 and 0.261, respectively, and both these values are relatively larger than the occupancy values of the other test compounds. The positive value regression coefficient of GC4 leads to these two compounds having high predicted  $-\log(\text{MIC}_{90})$  values. Using alignment 4, the GC1 occupancy value for compound 61 is zero and leads to this compound being the second largest outlier in the test set. The largest outlier for alignment is compound 64. The GC4 occupancy value for this compound is 0.315, which is larger than that other compound GC4 values. The GC5 occupancy value of compound 59, an outlier for alignment 5, is 0.202, and its regression coefficient is 4.78, resulting in compound 59 to have a relatively high predicted -log-(MIC<sub>90</sub>) value. For the other outlier, compound 64, the GC3 occupancy value is 0.004 resulting in compound 64 having a high predicted  $-\log(\text{MIC}_{90})$  value. In alignment 6, the GC5 occupancy values for both major outliers, compounds 61 and 64, are zero and these two compounds have high predicted -log(MIC<sub>90</sub>) values. Both GC3 occupancy values for the alignment 7 major outliers, compound 61 and 64 are zero so that these two compounds have relatively high predicted -log(MIC<sub>90</sub>) values. For alignment 8, the GC1 occupancy value for compound 62, one of the two major outliers, is 0.030, and the negative regression coefficient of GC1, -10.88, leads to this compound having a low predicted -log-(MIC<sub>90</sub>) value. The GC5 occupancy value for the other major

**Table 8.** Predicted  $-\log(\text{MIC}_{90})$  Values of the Test Set in Table 7 Using the Best 4D-QSAR Model for Each Alignment and the Correlation Coefficients of the Predictions to the Observed  $-\log(\text{MIC}_{90})$  Values of the Test Set<sup>a</sup>

	-log(MIC <sub>90</sub> ) predicted value										
compd no.	align. 1	align. 2	align. 3	align. 4	align. 5	align. 6	align. 7	align. 8	align. 9	align. 10	
56	-0.642	1.518	3.600	2.705	2.838	3.255	3.148	3.413	3.184	2.953	
57	0.396	1.735	4.334	2.470	3.750	2.799	2.782	3.085	3.317	3.023	
58	2.602	2.959	2.535	2.49	3.719	2.421	2.188	2.425	2.753	2.763	
59	2.145	2.848	2.181	2.335	2.803	2.562	2.110	2.186	2.886	2.417	
60	2.801	2.842	3.052	2.701	2.729	2.725	2.559	1.625	2.971	2.742	
61	2.452	2.767	3.429	2.736	2.764	2.784	2.497	1.888	2.929	2.748	
62	1.593	1.281	1.406	2.412	2.789	1.803	2.392	1.816	1.124	2.801	
63	2.283	1.745	1.952	1.984	3.572	2.365	2.270	2.071	4.894	3.272	
64	2.381	2.845	2.656	2.825	2.755	2.857	2.523	1.804	2.977	2.700	
65	2.520	2.616	2.341	2.708	2.598	2.643	2.432	2.475	3.096	2.357	
$r^2$	0.439	0.364	0.061	0.037	0.308	0.004	0.204	0.631	0.002	0.203	

<sup>&</sup>lt;sup>a</sup> The two largest outliers of the test set for each alignment are indicated by bold print.

Table 9. Two Largest Outliers for Each Alignment and the GCOD(s) Value Most Responsible for the Poorly Predicted  $-\log(\text{MIC}_{90})$  Values

align.	outlier 1	GCOD	occupancy	reg coeff	outlier 2	GCOD	occupancy	reg coeff
1	56	GC4	0.376	-7.65	57	GC4	0.438	-7.65
2	56	GC3	0.335	-488	64	GC3	0	-4.88
3	61	GC4	0.261	7.37	60	GC4	0.215	7.37
4	61	GC1	0	-4.06	64	GC4	0.315	1.94
5	59	GC5	0.202	4.78	64	GC3	0.004	-8.86
6	61	GC5	0	-7.55	64	GC5	0	-7.55
7	61	GC3	0	-7.71	64	GC3	0	-7.71
8	62	GC1	0.030	-10.88	58	GC5	0.215	2.52
9	63	GC3	0	-6.71	64	GC3	0	-6.71
10	61	GC3	0	-7.87	64	GC3	0	-7.87

**Table 10.** Cross-Validated Correlation Coefficient,  $xv-r^2$ , of the Best 4D-QSAR Model for Each Trial Alignment Using Training Set(II) Described in the Discussion Section

align.	1	2	3	4	5	6	7	8	9	10
xv-r <sup>2</sup>	0.734	0.748	0.752	0.642	0.678	0.725	0.775	0.783	0.719	0.691

**Table 11.** Average Cross-validation Correlation Coefficient,  $\langle xv-r^2 \rangle$ , and Average Correlation Coefficient,  $\langle r^2 \rangle$ , of the Best 4D-QSAR Model for Each Trial Alignment from the Random Scrambling Experiments of the Observed  $-\log(\text{MIC}_{90})$  Values

align.	1	2	3	4	5	6	7	8	9	10
$\langle r^2 \rangle$	0.579	0.545	0.555	0.590	0.564	0.596	0.566	0.573	0.570	0.574
$\langle XV - r^2 \rangle$	0.486	0.359	0.411	0.533	0.447	0.440	0.467	0.455	0.479	0.457

outlier, compound 58, is 0.313, which is larger than that of other compounds. And the regression coefficient of GC5 is 2.52, which leads compound 58 to have a high predicted —log(MIC<sub>90</sub>) value. Using alignment 9 both GC3 occupancy values for the two largest outliers, compounds 63 and 64, are zero so that these two compounds are predicted to have too high of —log(MIC<sub>90</sub>) values. And alignment 10 has the same outlier features as alignment 9, both GC3 occupancy values for compound 61 and 64 are zero, resulting in these two outliers having high predicted —log(MIC<sub>90</sub>) values.

#### DISCUSSION

The new 4D-QSAR methodology feature developed and applied in this study is the numerical and graphical comparison of the 3D-pharmacophores embedded in 4D-QSAR models from different alignments. It is found that different alignments will produce significantly different models for the chemical classes of P450 inhibitors investigated. From Table 4 it is seen that the range in  $r^2$  as a function of alignment is from 0.746 to 0.853 for seven-term 4D-QSAR models, and  $xv-r^2$  varies from 0.693 to 0.784. Thus, selection of alignment is a key step in building 4D-QSAR models for P450 inhibitors of the training set. However, several similar alignments have similar  $r^2$  and  $xv-r^2$  values, and there is also

the possibility that both the way the training set and test set are selected influences alignment 8 to be the 4D-QSAR model with the highest  $xv-r^2$  value. To address these concerns and ambiguities, a new training set was selected to repeat a 4D-QSAR analysis employing the same procedure as described above. From the 65 compounds data set, twelve compounds, namely 2, 4, 12, 16, 28, 32, 33, 37, 41, 46, 53, and 57, were chosen as a second test set (II), and the remainder of the compounds defined a second training set (II). Both training set (II) and test set (II) span the three types of chemical scaffolds present in the parent compound set, and also span the complete range in  $-\log(\text{MIC}_{90})$  potency. Table 10 lists the  $xv-r^2$  for the best 4D-QSAR model of each alignment using training set (II). Among the first four alignments, the 4D-QSAR model of alignment 3 still has the highest  $xv-r^2$  value, and among all 10 alignments, alignment 8 again produces the 4D-QSAR model with the highest  $xv-r^2$  value.

There is also the possibility of chance correlations in model building using the training sets. Hence, the dependent variable,  $-\log[\text{MIC}_{90}]$ , was randomly scrambled for each alignment 10 times, and the average  $r^2$  and  $xv-r^2$  values of the corresponding best 4D-QSAR models for each alignment were computed and are listed in Table 11. The scrambled

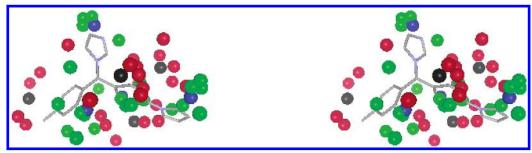


Figure 7. An overlay of the 3D pharmacophores of the best 4D-QSAR models for alignment 3, 5-10 and the cluster centers, shown relative to the most active compound 39, in its minimum energy conformation. The GCODs are shown as spheres, although the actual grid cells are cubes in space.  $-\log(\text{MIC}_{90})$ -enhancing grid cells are shown as green spheres, and grid cells that diminish  $-\log(\text{MIC}_{90})$  are shown as red spheres. The blue spheres are the cluster centers of the potency enhancing GCODs, and the black spheres are the cluster centers of potency diminishing GCODs.

average  $r^2$  and  $xv-r^2$  values for all alignments are significantly less than the  $r^2$  and  $xv-r^2$  values in Table 4. These scrambling studies indicate that the best models are the result of chance correlations is remote.

Since each best 4D-QSAR model of each alignment is unique and relatively significant, the question persists if the best 4D-QSAR models of the alignments contain any similar GCOD features. To address this question, cluster analysis of the spatial locations of the GCODs of the best models of alignments 1 and 9 was performed after applying eqs 1-3. Three clusters are formed with  $P \le 0.05$  among the GCODs of the best 4D-QSAR models of alignments 1 and 9. The cluster centers are defined as black spheres in Figure 6A. Cluster analysis was also applied to the GCODs from the best 4D-QSAR models for alignments 3, 5, 6, 7, 8, 9, and 10. Figure 7 shows the composite set of superimposed GCODs using eqs 1-3 from these seven alignments. The GCODs are color coded as activity enhancing (green) and activity diminishing (red). The most active inhibitor, compound 39 of Figure 1, in its minimum energy (not predicted active conformation) conformation is shown with the composite set of GCODs for reference. Most of the activity enhancing GCODs form local clusters in space. Cluster analysis indicates there are four significant cluster centers  $(P \le 0.05)$ , which define regions of space where occupancy by the appropriate inhibitor IPE type increases inhibitor activity. The same type of clustering is seen for the GCODs that diminish activity, but these clusters are not as compactly distributed in space as the enhancing GCOD clusters. Around ring B a "red region" interfaces very closely with a "green" region. This close interfacing of activity enhancing and diminishing regions suggests ligand molecular structure exhibits high specificity to inhibition potency at this location of the inhibitor. However, it is not readily possible to assign this specificity to flexibility or to inhibitor-receptor interactions. Thus, even though both the matrix of the crosscorrelation coefficients of the residuals of fit and the 3Dpharmacophore similarity matrix show that detailed model similarity is low across the 10 alignments, the best 4D-QSAR models of each alignment still contain common GCOD information.

The distribution of GCODs shown in Figure 7 is, in general, in agreement with the HYPO1 "pharmacophores model" from the Catalyst analysis. But in addition, the 4D-QSAR models have a smaller residual of fit to the training set than realized from the Catalyst study. The 4D-QSAR models seemingly provide a higher-resolution, more quantitative description of the binding pharmacophore than does the catalyst model. It is surprising that there are only two GCODs having hydrogen bond interaction IPE types among the seventy GCODs listed in Table 5. The two hydrogen bond GCODs are GC4 (hbd) in alignment 5 and GC6 (hba) in alignment 6. GC4 in alignment 5 has a negative regression coefficient, -6.71, indicating a hydrogen donor at that position is unfavorable. Although the regression coefficient of GC6 in alignment 6 is positive, its regression coefficient is very small relative to other GCODs of the model. Overall, these observations suggest that hydrogen bonding is not a significant interaction when these ligands bind to the receptor active site. This finding is in agreement with the crystal structure recently published by Podust.<sup>28</sup> This crystal structure reveals that the active-site domain in the case of fluconazole

is largely composed of phenylalanine residues (F78, M79, F83, and F255) above the porphyrin plane. Moreover, the crystal structure shows that F83 and F255 provide nonbonded contacts to the ligand, and no hydrogen bonds are formed. These observations indirectly support that our RI 4D-QSAR models meaningfully map the active site of the receptor.

#### ACKNOWLEDGMENT

Partial funding for this study was provided by National Institutes of Health Grant P01-GM 62195. We also gratefully acknowledge support from the Laboratory of Molecular Modeling & Design at UIC and from The Chem21 Group, Incorporated.

#### REFERENCES AND NOTES

- Fromtling, R. A. Imidazoles as Medically Important Antifungal Agents: An Overview. *Drugs Today* 1986, 20, 235-349.
- (2) D'Arcy, P. F.; Scott, E. M. Antifungal Agents. Prog. Drug Res. 1987, 22, 94–147.
- (3) Kerridge, D. Antifungal Drugs. Drugs Today 1988, 24, 705-715.
- (4) Koltin, Y. Target for Antifungal Drug Discovery. Annu. Rep. Med. Chem. 1990, 25, 141–148.
- (5) Yeagle, P. L.; Martin, R. B.; Lala, A. K.; Lin, H.; Block, K. Differential Effects of Cholesterol and Lanosterol on Artificial Membranes. *Proc. Natl. Acad. Sci. U.S.A.* 1977, 74, 4924–4926.
- (6) Vanden Bossche, H.; Bellens, D.; Cools, W.; Gorrens, P. M.; Verhoeven, H.; Willemsens, G.; De Coster, R.; Beerens, D.; Haelterman, C.; Coene, M. C.; Lauwers, W.; Le Jeune, L. Cytochrome P450: Target for Itraconazole. *Drug Dev. Res.* 1986, 8, 287–298.
- (7) Hichcock, A.; Brown, B.; Evans, E. G. V.; Adams, D. J. Cytochrome P-450-Dependent 14α-Demethylation of Lanosterol in *Candida albi*cans. Biochem. J. 1989, 260, 549-556.
- (8) Adams, J. L.; Metcalf, B. W. Therapeutic Consequences of the Inhibition of Sterol Metabolism. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, England, 1990; Vol. 2, pp 333–364.
- (9) Asai, K.; Tsuchimori, N.; Okonogi, K.; Perfect, J. R.; Gotoh, O.; Yoshida, Y. Formation of Azole-Resistant *Candida albicans* by Mutation of Sterol 14-Demethylase P450. *Antimicrob. Agents Chemother.* 1999, 43, 1163–1169.
- (10) Buchel, K. H.; Draber, W.; Regel, E.; Plempel, M. Synthesen und Eigenschften von Clotrimazol und Weiteren Antimykotischen 1-Triphenylmethylimidazolen (Syntheses and Properties of Clotrimazole and Further Antimycotic 1-(Triphenylmethyl)imidazoles). *Arzneim.-Forsch.* 1972, 22, 1260–1272.
- (11) 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.
- (12) Talele, T. T.; Kulkarni, V. M. Three-dimensional Quantitative Structure—Activity Relationships (QSAR) and Receptor Mapping of Cytochrome P-450<sub>14αDM</sub> Inhibiting Azole Antifungal Agents. J. Chem. Inf. Comput. Sci. 1999, 39, 204–210.
- (13) Tafi, A.; Costi, R.; Botta, M.; Di Santo, R.; Corelli, F.; Massa, S.; Ciacci, A.; Manetti, F. Artico, M. Antifungal Agents. 10. New Derivatives of 1-[(Aryl)[4-aryl-1H-pyrrol-3-yl]methyl]-1H-imidazole, Synthesis, Anti-Candida Activity, and Quantitative Structure Analysis Relationship Studies. J. Med. Chem. 2002, 45, 2720-2732.
- (14) Ji H.; Zhang, W.; Zhou, Y.; Zhang, M.; Zhu, J.; Song, Y.; Lu, J.; Zhu, J. A Three-Dimensional Model of Lanosterol 14α-Demethylase of *Candida albicans* and Its Interaction with Azole Antifungals. J. Med. Chem. 2000, 43, 2493–2505.
- (15) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J. Am. Chem. Soc. 1997, 119, 10509-10524.
- (16) 4D-QSAR User's Manual, Version 3.0, The ChemBats21 Group, Inc., 1780 Wilson Dr., Lake Forest, IL 60045, 2003.
- (17) HyperChem Program Release 6.03 for Windows; Hypercube, Inc.; 1999.
- (18) Doherty, D. C. MOLSIM User's Guide, The ChemBats21 Group, Inc., 1780 Wilson Dr., Lake Forest, IL 60045, 1997.
- (19) Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J. Am. Chem. Soc. 1977, 99, 8127–8134.

- (20) Hopfinger, A. J.; Pearlstein, R. A. Molecular mechanics force-field parametrization precedures. J. Comput. Chem. 1984, 5, 486-492.
- (21) Duca, J. S.; Hopfinger, A. J. Estimation of Molecular Similarity Based on 4D-QSAR Analysis: Formalism and Validation J. Chem. Inf. Comput. Sci. 2001, 41, 1367–1387.
- (22) Krasowski, M. D.; Hong, X.; Hopfinger, A. J.; Harrison, N. L. 4D-QSAR Analysis of a Set of Propofol Analogues: Mapping Binding Sites for an Anesthetic Phenol on the GABAA Receptor. *J. Med. Chem.* **2002**, *45*, 3210–3221.
- (23) Glen, W. G.; Dunn, W. J., III.; Scott, D. R. Principal components analysis and partial least squares. *Tetrahedron Comput. Methods* **1989**, 2, 349–354.
- (24) Rogers, D. G/SPLINES; A hybrid of Friedman's multivariate adaptive regression splines (MARS) algorithm with Holland's genetic algorithm. The Proceedings of the Fourth International Conference on Genetic Algorithm; San Diego, 1991; pp 38–46.
- (25) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure activity relationships and quantitative structure property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854—866.
- (26) Friedman, J. Multivariate adaptive regression splines. Technical Report No. 102; Laboratory for Computational Statistics, Department of Statistics, Stanford University, Stanford, CA, November 1988 (revised August 1990).
- (27) Brian S. E.; Sabine L.; Morven L. Cluster analysis, 4th ed.; Oxford University Press: 2001.
- (28) Podust, L. M.; Poulos, T. L.; Waterman, M. R. Crystal Structure of Cytochrome P450 14α-sterol Demethylase (CYP51) from *Mycobacterium tuberculosis* in Complex with Azole Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 2001, 98, 3068–3073.

CI034142Z