Pharmacophoric Fingerprint Method (TOPP) for 3D-QSAR Modeling: Application to CYP2D6 Metabolic Stability

Simone Sciabola,[†] Iñaki Morao,[‡] and Marcel J. de Groot*[,‡]

Laboratorio di Chemiometria, Universita di Perugia, Via Elce di Sotto, 10, 1-06123 Perugia, Italy, and Pfizer Global Research and Development, Sandwich Laboratories, Sandwich, Kent CT13 9NJ, U.K.

Received April 20, 2006

The application of a new 3-point pharmacophore-fingerprinting package (TOPP, Triplets Of Pharmacophoric Points) to develop QSAR models is discussed. In the CYP2D6 metabolic stability case, these 3D pharmacophoric fingerprints have shown to be as valid as other 3D descriptors and 2D features. Interestingly, it was found in the 3D models that the use of more realistic substrate conformations, by an additional docking step, did not improve the statistical results significantly. A detailed analysis of the generated pharmacophoric hypotheses is consistent with the previously proposed dual interaction mode of substrates within the active site of CYP2D6.

INTRODUCTION

Cytochromes P450 (P450s) are part of a large superfamily of heme-containing drug metabolism enzymes, which are divided into families depending on the pairwise amino acid sequence identity.[1,2] P450s can metabolize (either oxidize or reduce) a large number of structurally different endo- and exogenous compounds due to broad substrate specificity and broad regio- and stereoselectivity. Seven of the 57 known human isoforms of P450s (CYP1A2, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, and CYP3A4)[3] are responsible for more than 90% of the metabolism of all pharmaceuticals in current clinical use.[4] Several of these isoforms display polymorphisms (e.g. CYP2D6,[5,6] CYP2C9[7−11]) which can result in a poor metabolism of drugs. It is therefore vitally important for the pharmaceutical industry to be able to predict at an early stage whether a drug candidate will interact with the P450s, which isoform the drug candidate will preferentially interact with, and consequently whether it is worthwhile spending time and money taking that candidate through to the next development steps.

A variety of computational chemistry methods have been applied to cytochromes P450, either concentrating on the small molecules that are metabolized by or inhibit the P450s, or on the P450 protein itself. These models of P450s are capable of predicting the possible involvement of P450s in the metabolism or inhibition of a specific compound. Therefore, they are important tools in drug discovery and development, primarily because of the clinical importance of P450s in the metabolism of xenobiotics and endogenous compounds and identification of potential drug−drug interactions.[12] The simplest (and fastest) computational approach to predict activity, binding, or inhibition of a compound is via quantitative structure−activity relationship (QSAR) methods. These methods try to find some kind of relationship between the chemical structures and the corresponding experimentally observed activities in a quantitative manner. In contrast, pharmacophore modeling overlays structures of ligands or features of these ligands in 3D-space, attempting to describe the physical, spatial, and chemical properties of the active/binding site. Homology protein modeling (or comparative modeling) predicts the 3D-structure of a protein based on the knowledge of its amino acid sequence and the 3D-structures of proteins with similar sequences (assuming that proteins with similar amino acid sequences adopt similar 3D-structures).[13]

The current work describes the successful application of a novel 3D pharmacophore fingerprinting approach (TOPP, Triplets Of Pharmacophore Points) to predict the inhibition of one specific P450: the polymorphic CYP2D6. The performance of this approach was compared to other 3D and 2D methodologies, commonly used in academia and industry. The influence of the conformational geometry employed was also explored.

COMPUTATIONAL DETAILS

A. Data Set. The experimental data used during this work consists of 150 compounds from the Bioprint database,[14] for which the 2D6 metabolic stability was reported as percentage. The QSAR models were built using 100 compounds in the training set. The remaining 50 compounds were used in the test set (25 stable, 25 unstable). A statistical subset selection was made using GOLPE[15] software to automatically select a balanced and chemically diverse test set. "Largest Minimum Distance" and "Most Descriptive Compound" variants were equally combined and then run over the PCA scores extracted from the TOPP descriptors. All QSAR models presented along this paper are categorical (qualitative: yes/no, positive/negative) and not numerical (quantitative). Therefore, the compounds are classified in two classes: unstable (U) (% metabolic stability < 60) and stable (S) (% metabolic stability ≥ 60).

B. Geometry Optimization. All data set compounds were directly converted to 3D structures using Corina.[16] The 3D

* Corresponding author phone: +44-1304-648746; fax: +44-1304-651817; e-mail: marcel.degroot@pfizer.com.
[†] Universita di Perugia.
[‡] Pfizer Global Research and Development.

10.1021/ci060143q CCC: $37.00
Published on Web 11/10/2006

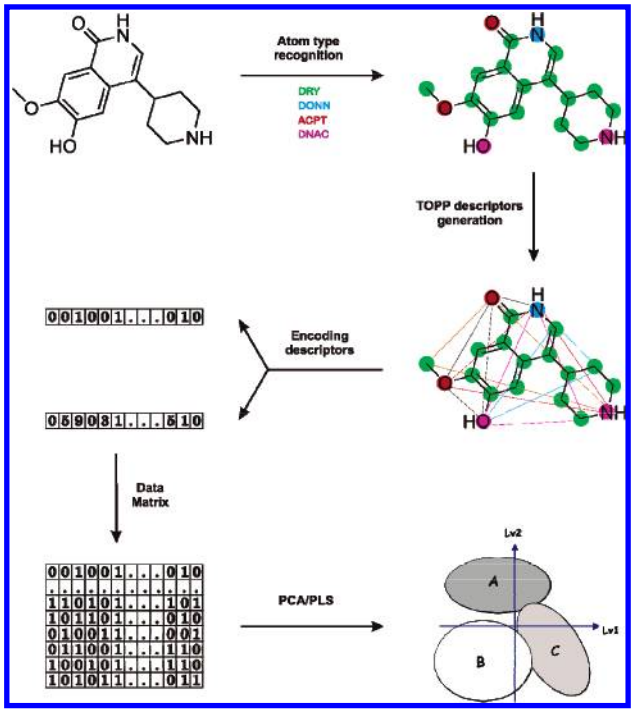

Figure 1. Working protocol in TOPP methodology. See Computational Section for details.

Corina geometries were also used as the starting coordinates for the docking studies.

C. Docking Studies Using GOLD. To analyze the conformational effects in QSAR models docking calculations were carried out. As no crystal structure of CYP2D6 isoform was available at the time of the study,[17] the consensus homology protein model (derived from CYP101, CYP102, and CYP108 templates as previously published by de Groot et al.[18]) was used instead. GOLD (Genetic Optimization for Ligand Docking[19]) was chosen for the docking procedure. For each of the 30 independent genetic algorithm (GA) runs, a maximum number of 50 000 GA operations were performed on a single population of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. To allow poor nonbonded contacts at the start of each GA run, the maximum distance between hydrogen donors and fitting points was set to 3 Å, and nonbonded van der Waals energies were cut off at a value equal to 6. The best-scored solution by GOLD has been taken for each compound as starting point geometry for following 3D-QSAR studies.

D. TOPP (Triplets Of Pharmacophoric Points). TOPP[20] is a QSAR approach using 3-point pharmacophores as 3D descriptors and uses GOLPE[15] to perform the multivariate statistical analysis. The working protocol carried out by TOPP is represented in Figure 1. First, the atoms of each molecule are classified by the GRID force field parametrization. In this way, atoms are described according to their charge and hydrogen bonding properties: DRY (hydrophobic), DONN (HBD, hydrogen bond donor), ACPT (HBA, hydrogen bond acceptor), and DNAC (both HBD and HBA). After the classification of the atoms an iterative procedure generates all possible combinations of three points with the four different atom types (DRY, DONN, ACPT, and DNAC), being encoded in two possible ways. One approach is to store

Chart 1

Step1→ Read compound 1→ X-Matrix:

Cmpd ID	Var A	Var C	Var E	Var F	Var Z
Mol001	1	1	1	1	1

Step2→ Read compound 2→ Compound 2's bit string:

Cmpd ID	Var A	Var B	Var C	Var E	Var F
Mol001	1	1	1	1	1

New X-Matrix:

Cmpd ID	Var A	Var B	Var C	Var E	Var F	Var Z
Mol001	1	0	1	1	1	1
Mol002	1	1	1	1	1	0

only the presence or the absence of a 3-point pharmacophore combination. The other approach is to count the number of times that a combination is present in the molecule. Although TOPP can also handle the conformational space of molecules by using a conformer generator, no conformational analysis was carried out in this work. TOPP also allows the user to define a common anchor point (an atom or functional group in the molecule) from which every triangle will be built (FIX atom option). This option is particularly advantageous with data sets where some well-known substituents are crucial for the activity. The calculations described above are performed for all the compounds, and the results, i.e., the combinations of 3-point pharmacophores, are stored in an X-matrix of descriptors. As this final matrix of descriptors is built up in a dynamic way no length of the bit string is initially fixed. When the first compound is processed, the first row of the X-matrix consists exclusively of those variables present in this compound. Therefore, all the bits so far are set to "1". When the program processes the second compound, the X-matrix will change according to the triangles present in this second compound. Then, the program will fill with "0"s these new positions in the first row (compound 1) due to the processing of the second compound and also the positions of the second row due to the first compound (see Chart 1).

The final X-matrix is then submitted for statistical analysis, i.e., a linear regression model using Principal Component Analysis (PCA) and Partial Least Square (PLS).

The first criterion used to encode triplets of pharmacophoric points into the TOPP bitstring is based on atom type recognition (i.e. DRY, HB donor, HB acceptor, and HB donor/acceptor). For ordering purposes, type DRY prevails over type DONN (HB donor), which prevails over type ACPT (HB acceptor), which in turn, prevails over type DNAC (HB donor/acceptor). All the possible 3-point combinations using four different atom types are encoded in the TOPP bitstring as follows: AAA, AAB, AAC, AAD, ABB, ABC, ABD, ACC, ACD, ADD, BBB, BBC, BBD, BCC, BCD, BDD, CCC, CCD, CDD, and DDD.

The second criterion used to encode triplets is based on interatomic distances. We used a maximum distance cutoff of 20 Å and a distance binning customizable by the user (1 Å by default). For each class of interaction (i.e. AAA, ..., DDD), distances range according to the scheme reported below:

AAA010101, AAA010102, ..., AAA010120,
AAA010201, AAA010202, ..., AAA010220, ...,
AAA012020, AAA020101, AAA020102, ...,
AAA020120, AAA020201, AAA020202, ...,
AAA020220, ..., AAA200101,
AAA200102, ..., AAA200120,
AAA200201, AAA200202, ..., AAA200220, ...,
AAA202001, AAA202002, ...,
AAA202020, ..., DDD010101,..., DDD202020

Using this scheme, each bin is associated with a specific triplet of interaction defined by the type of the vertices comprising the triplet and the relative distances between each atom pairs. Due to the class definition, AAB, ABA, or BAA belong to the same triplets of interaction AAB. Every time a triplet is made of 2 dry atoms and 1 donor atom, the first criterion will be used to associate it with the class AAB, and the second criterion (based on atom pair distances) will be used to correctly locate the triplet inside the AAB class within the TOPP bitstring. The same rule is applied to all 3-point pharmacophores (AAA, ABC, ..., DDD) combinations.

For instance, let us take these four combinations of PP (Pharmacophoric Points) depicted in Figure 2. Case (**a)** (Figure 2a) shows a PP triplet made only of hydrophobic atoms. Atom pair distances range from 4 to 7 Å, so the question is which bit this triplet will set to 1. All vertices belong to the same class; therefore, we use the distance criterion in order to assign the triplet to the correct position within the TOPP bitstring.

To do this the vertices of the triangle must be ordered in a way that $d(A_1A_2) > d(A_1A_3) > d(A_2A_3)$, obtaining

$$A_1\ A_2\ A_3\ 07\ 06\ 04$$

where

$$07 \rightarrow \text{distance between vertices } A_1 - A_2$$

$$06 \rightarrow \text{distance between vertices } A_1 - A_3$$

$$04 \rightarrow \text{distance between vertices } A_2 - A_3$$

Case (**b)** (Figure 2b) shows a PP triplet made of 2 hydrophobic and 1 HB donor features. As mentioned before, feature A prevails over feature B, so this triplet will be associated with the following class:

$$A\ A\ B\ d_1\ d_2\ d_3$$

We still have to assign the three distance values between the edges of the triangle. We first set d_1 to A_1A_2 distance, that in our case is 4 Å:

$$A_1\ A_2\ B\ 04\ d_2\ d_3$$

Now, to assign d_2 and d_3, the following rule is followed:

if: $d(A_1B) > d(A_2B)$ then:

the bit set to 1 will be $A_1\ A_2\ B\ 04\ d(A_1B)\ d(A_2B)$

else if: $d(A_2B) > d(A_1B)$ then:

the bit set to 1 will be $A_1\ A_2\ B\ 04\ d(A_2B)\ d(A_1B)$

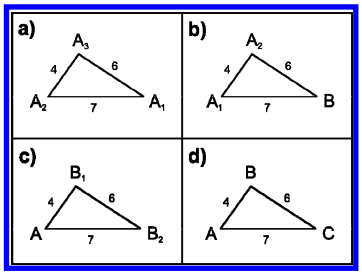

Figure 2. Determination of pharmacophoric points.

In our example the bit in question will be

$$A_1\ A_2\ B\ 04\ d(A_1B)\ d(A_2B) \rightarrow A\ A\ B\ 04\ 07\ 06$$

The same scheme is also used in case (**c)** (Figure 2c), where the PP triplet is made of 1 hydrophobic and 2 HB donor features; therefore, this triplet will be associated with the following class:

$$A\ B\ B\ d_1\ d_2\ d_3$$

Now we have two vertices belonging to the same class B with lower priority compared to class A. In this kind of cases, we first assign d_3 to B_1B_2 distance; in our case it will be

$$A\ B_1\ B_2\ d_1\ d_2\ 06$$

Then, to assign d_1 and d_2, the following rule is applied:

if: $d(AB_1) > d(AB_2)$ then:

the bit set to 1 will be $A\ B_1\ B_2\ d(AB_1)\ d(AB_2)\ 06$

else if: $d(AB_2) > d(AB_1)$ then:

the bit set to 1 will b: $A\ B_1\ B_2\ d(AB_2)\ d(AB_1)\ 06$

In our example the bit set to 1 will be

$$\mathbf{A\ B_1\ B_2}\ d(AB_2)\ d(AB_1)\ 06 \rightarrow A\ B\ B\ 07\ 04\ 06$$

In the case (**d**) (Figure 2d) all the vertices of the triangle belong to different classes. The rule to follow is based on the concept of "class prevalence" (A > B > C > D). In terms of distance values we consider $d(AB) > d(AC)$ and $d(AC) > d(BC)$. The bit associated to triple in case (**d**) is therefore

$$A\ B\ C\ 04\ 07\ 06$$

E. GRIND Descriptors. GRid-INdependent Descriptors (GRIND[21]) were generated and analyzed using the software ALMOND.[22] The GRIND working procedure involves three steps: (a) computing a set of molecular interaction fields (MIFs), (b) filtering the MIFs to extract the most relevant regions, and (c) computing the product of the interaction energy for each pair of filtered points (nodes). In the present work the standard probes were used (DRY, N1, O, and TIP).

F. DRAGON Descriptors. DRAGON is an application for the calculation of molecular descriptors originally developed by the Milano Chemometrics and QSAR Research Group.[23] For comparison purposes with TOPP software, only the 3D descriptors were selected (i.e., charge descriptors, aromaticity indexes, Randic molecular profiles, and geometrical, RDF, 3D-MoRSE, WHIM, and GETAWAY descriptors[24]).

Table 1. Statistical Results Obtained Using PLS as Diagnostic Model[a]

	descriptors	GEOM[b]	PC[c]	r^2	q^2	S^d	U^e	SDEP	TP	TN	conc	kappa
1a	TOPP	C	3	0.87	0.66	2	1	0.60	92	96	94	0.88
1b	TOPP	G	4	0.93	0.79	2	4	0.63	92	84	88	0.76
2a	ALMOND	C	4	0.73	0.40	7	4	0.84	72	84	78	0.56
2b	ALMOND	G	4	0.75	0.50	5	1	0.79	80	96	88	0.76
3a	DRAGON	C	3	0.53	0.26	6	6	0.80	76	76	76	0.52
3b	DRAGON	G	3	0.58	0.34	6	4	0.89	80	92	86	0.72
4a	QUASAR	C	2	0.50	0.30	4	4	0.67	84	84	84	0.68
4b	QUASAR	G	4	0.50	0.25	4	4	0.75	88	92	90	0.80
5	ECFP6	smi	3	0.81	0.23	1	4	0.71	96	80	88	0.76
6	MDL keys	smi	2	0.61	0.38	1	2	0.57	96	96	96	0.92
7	BCI 4096	smi	3	0.78	0.59	5	5	0.85	80	80	80	0.60

[a] The descriptors, statistical values from the confusion matrix, and PLS related values are explained in the Computational Section. [b] GEOM: origin of the 2-3D structure used as starting geometry. For the 3D descriptors: (C) Corina convertor and (G) after docking using GOLD program. For 2D descriptors: (smi): smile definition. [c] PC: the best principal component according to the results of cross-validation. [d] *S*: the number of wrong predictions in the stable test set. [e] *U*: the number of wrong predictions in the unstable test set.

G. QuaSAR Descriptors. Standard QuaSAR descriptors were calculated using Molecular Operating Environment (MOE) package from Chemical Computing Group Inc.[25] In this work all 2D and 3D descriptors generated by default were utilized to build the model.

H. ECFP Fingerprints. Extended Connectivity Fingerprints (ECFPs) are a new class of fingerprints developed by Sci-Tegic.[26] These are based on a variant of the Morgan algorithm (originally used in isomorphism issues[27]). The generation of these fingerprints begins with the assignment of an initial atom code to every heavy atom present in the molecule based on the following features: the number of connections, the element type, the charge, and the atomic mass. This initial fingerprint is called "ECFP_0" as the maximum diameter explored was only around each atom. An iterative process is used to generate larger structural neighborhoods until a desired size is reached. In this case, the standard ECFP_6. The advantages of these circular fingerprints are that the calculation is very fast and the features (size up to 4 billion) are not predefined in a limited fragment dictionary. These descriptors have been well tested in conjunction with the Laplacian-modified Bayesian analysis and have proved to be efficient in similarity-based virtual screening.[28,29]

I. MDL Public Keys. Structure-based 2D descriptors based on two common MDL keysets: one containing 960 keybits and the other containing a subset of 166 keybits.[30,31]

J. Barnard Fingerprints. BCI structural fingerprints are 4096-long bit strings based on the presence or absence of 2D structural features of a molecule, listed in a predefined fragment dictionary, which contains six different families of fragments: augmented atoms, atom/bond sequences, atom pairs, ring composition fragments, ring fusion fragment, and ring ortho fragments.[32]

K. Statistical Analysis. Principal Component Analysis (PCA) and Partial Least Squares (PLS) analysis were performed within GOLPE[15] software. Variables Selection using Fractional Factorial Design (FFD) was used as implemented in GOLPE to filter out all the noisy variables present in the training set.

The predictivity of each model was evaluated both using (1) cross-validation (CV) as "internal validation" technique and (2) an external test set as "external validation".

(1) Cross-validation works building reduced models (models for which some of the objects were removed) and using them to predict the Y-variables of the objects held out. In our case the objects in the training set were split at random into 5 groups, each one containing an equal number of objects. Models were then built keeping one of these groups out of the analysis until all the objects were kept once. The formation of the groups and the validation were repeated 20 times.[15] At the end, predicted Y is compared to experimental Y, for each dimensionality using the computed parameters q^2 and SDEP (Standard Deviation of Errors of Prediction)

$$q^2 = 1 - \left[\frac{\sum(y - y')^2}{\sum(y - \bar{y})^2}\right] \quad \text{(i)}$$

$$\text{SDEP} = \sqrt{\frac{\sum(y - y')^2}{N}} \quad \text{(ii)}$$

where y is the experimental value, y' is the predicted value, and $\bar{y}$ is the average value.

The optimal dimensionality of each PLS model was then chosen according to the results of cross-validation.

(2) Another way to evaluate the efficiency of the model is to use an external prediction set. In this approach the objects in the original data set are split up into two groups from the very beginning of the analysis. The first one, the learning set, will be used to build the PLS model. The other, the prediction set, will be used to compare their experimental Y-values with the predictions made by the PLS model. In this analysis the same eq ii was used to compute the uncertainty of predictions for the external test set compounds, and the corresponding data were reported in Table 1.

L. Confusion Matrix. A confusion matrix contains information about actual and predicted classifications done by a classification system. The following table shows the confusion matrix for a two-class classifier

		actual	
		positive	negative
predicted	positive	a	b
	negative	c	d

where *a* is the number of correct predictions that an instance is positive (True Positive, TP), *b* is the number of incorrect predictions that an instance is negative (False Negative, FN),

c is the number of incorrect of predictions that an instance is positive (False Positive, FP), and *d* is the number of correct predictions that an instance is negative (True Negative, TN). The statistical terms concordance and kappa have been used to compare the methods. The concordance gives the number of correct predictions (for both stable and unstable compounds), while kappa is a weighted statistic. When kappa equals 0, the model is equivalent to that expected by chance. When it equals 1, there is a perfect agreement between actual and predicted values. Concordance and kappa are defined as

$$\text{conc} = \frac{a+d}{a+b+c+d} \quad \text{and} \quad \text{kappa} = \frac{\text{conc}/(100-\delta)}{1-\delta}$$

$$\text{with} \quad \delta = \frac{(a+c)(a+b)+(b+d)(c+d)}{(a+b+c+d)^2}$$

RESULTS AND DISCUSSION

Comparative Study of Different Descriptors/Fingerprints. Three-dimensional (3D) pharmacophore fingerprints have been originally applied in high throughput virtual screening and molecular similarity searching of large databases because of their extremely fast generation and their ability to handle different classes of ligands.[33] However, during the last years pharmacophore fingerprinting has become a common alternative definition of molecules for QSAR and focused library design applications.[34] The question that arises is how good these fingerprint-based QSAR models really are. Moreover, are QSAR models better with 3D rather than 2D descriptors? Therefore, seven QSAR models using seven well-known molecular definition packages have been constructed. The descriptors utilized in this work have been widely used and have proven to be very successful according to literature, i.e., Almond (GRIND) 3D descriptors, DRAGON 3D descriptors, QuaSAR 2D/3D descriptors, circular ECFP_6 fingerprints, MDL structural keys, and Barnard 4096 fingerprints (see Computational Section for details). These results have been compared with the new 3D pharmacophore fingerprinting software based on triplets of pharmacophoric points (TOPP). One of the typical problems to compare different methods is the miscellaneous selection of the diagnostic methodology. For instance, ECFP fingerprint are normally combined with naïve Bayesian classifiers, while GRIND descriptors make use of linear regression techniques. In this work, the partial least squares (PLS) method using GOLPE has been used for all comparisons. This will enable us to give a fairly objective answer about the relative performance of the different descriptors. Another issue to address is the relative importance of molecular geometry used by employing two completely different approaches in all 3D methods studied. The first one consists of the well-known Corina conversion, which is mainly based on crystallographic data. The second approach is the use of the most likely conformation of the molecule within the active site. To retrieve such geometry an additional computational docking step is required together with molecular knowledge of the target under study. The biological example for all these comparisons was the stability/instability of P450 2D6 substrates.

The main statistical results of the evaluation of these descriptors (see Computational Section) are shown in Table 1. The additional docking step does not improve the statistical results significantly (compare b vs a entries). TOPP pharmacophore fingerprints are suitable 3D descriptors for QSAR models and only GRIND (Almond) descriptors of the 3D models can give us statistically values close to the TOPP performance. DRAGON and QuaSAR descriptors do not achieve such a level of accuracy (r^2 and cross-validated q^2) found for the two previously mentioned approaches. It is also noteworthy that 2D fingerprints afford worse statistical results than 3D descriptors except for the BCI 4096 case. Nevertheless, the validity of any model is finally confirmed by external validation. As expected, the number of compounds incorrectly predicted is consistent with the statistical q^2 values. The only exceptions found in 3D models are the unexpectedly elevated four faulty molecules with TOPP descriptors using GOLD docked conformation and the relatively small number of erroneous compounds in the QuaSAR prediction. The model generated by MDL public keys only predicts incorrectly a total of 3 molecules, while the best statistically 2D methodology (BCI) fails in 10 cases. The main reason of these discrepancies is due to the small number of test molecules analyzed (25 stable and 25 unstable). The Standard Deviation of Errors of Prediction (SDEP) was also calculated for the test set (see Table 1).

The analysis of the molecules incorrectly predicted is summarized in Table 2. The chemical structures of compounds 1−16s and 1−13u are pictured in Figures S1 and S2 of the Supporting Information. The ratio between incorrectly predicted stable and unstable molecules is moderately well balanced for 3D models. Indeed, 19 (15) stable versus 15 (9) unstable molecules were found using Corina (GOLD) geometries, respectively. TOPP descriptors do not share the same incorrectly predicted molecules using Corina and GOLD docked geometries (3u is the only exception). The other 3D methods show more similar behavior between both geometries. A possible explanation is that changing the definition of the 3D geometries employed by the molecules during the development of the QSAR model, resulting in the selection of a different group of variables during the PLS procedure. Although the number of failures is still small, these variables seemed more appropriate to explain the training set compounds but are slightly less effective to recognize an external validation set.

This misbehavior can also be explained by the fact that no evidence exists telling us that the best scored solution of a docking software, no matter how good it is, corresponds always to the 3D geometry of the compound cocrystallized with the enzyme. In fact, from previous literature[35] the best case scenario for a docking study is the misprediction of about 50−55% of the entire decoy data set as the top scored solution using a 2.00 Å RMSD cutoff with GOLD, GLIDE, SURFLEX, and FLEXX.

Interestingly, 1s and 6s are incorrectly predicted using three different 3D methods (with both Corina and GOLD geometries), while 16s is three times erroneously predicted but exclusively with Corina geometries. In addition, this compound is quite rigid, and the superimposition of Corina and GOLD geometries was almost identical. So, the molecules of the training set must be quite flexible in order to show structural differences between Corina and GOLD geometries and, in turn, to generate different models. A significant characteristic in 2D models is that there is no

Table 2. Table of the Erroneous Predictions in the Different QSAR Models[a−c]

S	d1	d2	d3	d4	d5	d6	d7	U	d1	d2	d3	d4	d5	d6	d7
1s	■	● ■		● ■				1u	■		■			▲	
2s		●	● ■				▲	2u					▲		
3s					▲			3u	● ■			●			▲
4s							▲	4u	■		●	● ■			
5s			■	● ■			▲	5u			●				
6s	●		● ■	● ■				6u			●		▲		▲
7s				●				7u		● ■	● ■	●			▲
8s		●	● ■					8u		●	●				
9s		■						9u		●					
10s		■					▲	10u							▲
11s						▲		11u	■	●			▲		▲
12s	■		●					12u			●	● ■			
13s		●						13u					▲		
14s		● ■													
15s		● ■	● ■				▲								
16s	●	●	●												

[a] S and U stand for stable and unstable compounds. Their chemical structures are pictured in Figure S1-2 of the Supporting Information. [b] D1−7 correspond to the different descriptors used in this comparative study: TOPP (d1), Almond (d2), DRAGON (d3), QuaSAR (d4), PLP (d5), MDL (d6), and BCI (d7), respectively. [c] Nature of the geometries in the QSAR models: Corina (●), GOLD docking (■), smiles notation (▲).

overlap of incorrectly predicted molecules (2 out of 15). Therefore, consensus models can be straightforwardly deduced by the reader to improve the performance of individual predictions.[36,37] Stable substrates 4s and 10s contain the same bicyclic scaffold and are expected to behave similarly in each model. Therefore, it was not surprising that the BCI model failed with both substrates. A very high degree of independence between 2D and 3D methods is observed (21 out of 29). It is difficult to answer the question which model is the best as these classes of methods explore different parts of chemical space. Focusing on the unstable compounds, the TOPP method is much better using Corina geometries. Conversely, Almond and DRAGON models estimate more efficiently using GOLD conformations. Interestingly, compound 7u is incorrectly predicted in four different methods. In fact, 7−9u substrates hold the same tetradibenzoquinoline framework. Analyzing the trained structures a single molecule with this core was found. Thus, due to the poor training set in terms of this particular core structure, poor performance in predictions is not unexpected. Finally, compound 1u is (twice) erroneously predicted when GOLD geometries are used. This molecule is very flexible (around the saturated carbon chain), and significant structural differences after alignment were found between Corina and docked structures. Instead, rigid 8u is incorrectly predicted (twice) when Corina geometries are used. This result is unexpected and opposite to the rigid counterpart 7u which is incorrectly predicted in both Corina and GOLD geometries.

TOPP Analysis of Key Features. As described above, TOPP fingerprints can be used in efficiently describing molecules for QSAR purposes. Another important feature of any QSAR procedure is to be able to identify the variables that play an important role in such a model. TOPP also allows the analysis of the key features of the model and to inspect the molecules, enabling the detection of those regions around the molecules that are most relevant to explain the stability/instability of the different compounds.

The detection of the most important variables is carried out analyzing the PLS coefficients profile plot according to the optimal dimensionality of the model (Figure 3). In this plot all the variables (TOPP descriptors) are reported sequentially along the *X*-axis. Each division corresponds to

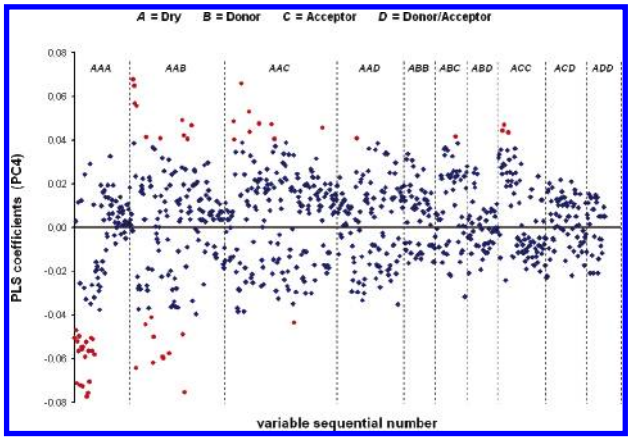

Figure 3. PLS coefficient plot using TOPP fingerprints.

a particular class of 3-point interactions. The red circles (the highest coefficients) represent those variables that contribute most to the biological activity of the compounds in the training set. The cutoff value was arbitrary set to 0.04, which seems to include all those variables that discriminate stable/unstable compounds. As can be seen from the highest PLS coefficients only a few groups of variables are able to discriminate stable from unstable compounds, i.e., the triplets of interaction AAA, AAB, AAC, and ACC. Triplets from the combination AAA (DRY-DRY-DRY, three hydrophobic regions) are present only on compounds classified as unstable, so these features can be linked directly to the substrate behavior. Table 3 also confirms such a statement, showing the presence of these triplets in unstable compounds and the absence in stable compounds.[38] Moreover, compounds that present the AAB interaction (DRY-DRY-DONN, two hydrophobic and one hydrogen-bond donor regions) can be either stable or unstable depending on the particular distances analyzed (e.g. AAB-040604 and AAB-060501 in Table 3). On the other hand, the presence of triplets AAC (DRY-DRY-ACPT) is mainly displayed by stable compounds. Other groups of variables (AAD, ABB, ABC, ABD, ACC, and ADD) do not affect at all or participate only marginally to the linear combination of the final QSAR model. It is worth noticing that all these variables come from the statistical analysis of the TOPP descriptors generated for

Table 3. Some of the Relevant Triplets Analyzed in Compounds **S1-2** and **U1-2**[a]

	(+)-coeff					(−)-coeff				
compd	AAC 02 03 01	AAC 03 04 03	AAC 04 06 03	ACC 04 03 06	ACC 05 04 05	AAA 05 06 06	AAA 07 06 06	AAA 07 07 06	AAB 04 06 04	AAB 06 05 01
S1	1	1	1	1	0	0	0	0	1	0
S2	1	1	1	1	1	0	0	0	0	0
U1	0	0	0	0	0	1	1	1	0	1
U2	0	0	0	0	0	1	1	1	1	1

[a] For the chemical structures of compounds **S1-2** and **U1-2** see Figure 4. "1"s and "0" s denotes that the 3-point interaction is present or absent, respectively.

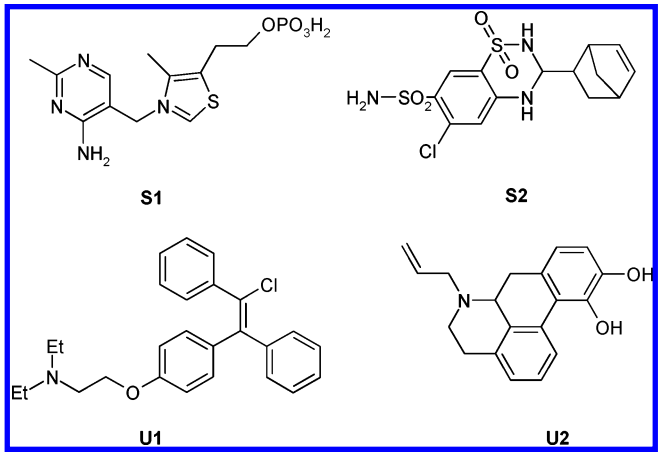

Figure 4. Four commercially available compounds (**U1-2** and **S1-2**) used in the training set. U1-2 are the examples for the interpretation of TOPP fingerprints.

the training set molecules, and, therefore, no information related to test set compounds has been taken into account. Although the structures of these trained compounds cannot be displayed for proprietary reasons, we show the chemical structure (Figure 4) and the most relevant variables of four commercial compounds (**S1-2** and **U1-2**) that were used in the external test set.

Once the statistically relevant triplets involved in the activity are identified, a retro searching around the molecular space of the unstable compounds can be performed. This step will identify the arrangements of chemical features required by a substrate to interact with the active site of P450 2D6 homology model.[39] Grouping the most relevant (highest) negative coefficients of the unstable compounds, the combination of triplets AAA and AAB essentially, a 4-point pharmacophore is generated (Figure 5). This common feature consists of the combination of three hydrophobic and one hydrogen bond donor (HBD) regions, as the basic nitrogen is positively charged. While the hydrophobic subregions overlays quite well, the HBD areas are completely dissimilar. Therefore, these two different arrangements define two possible ways of interaction within the binding site. This hypothesis is confirmed from a detailed analysis of the binding site of P450 2D6 (according to the homology protein model)[18] with the 3D coordinates of these substrates U1-2 docked inside the cavity (Figure 6). The hydrophobic pattern of interaction showed by both substrates is located in a flat and large lipophilic region (green region, MIF generated with probe C1=) [40] involving Leu121, Leu213, Ala305, Val370, and Thr309. Instead, the basic nitrogen atoms are independent each other (see blue regions, MIFs generated with probe N1+).[40] This is in agreement with the previously published

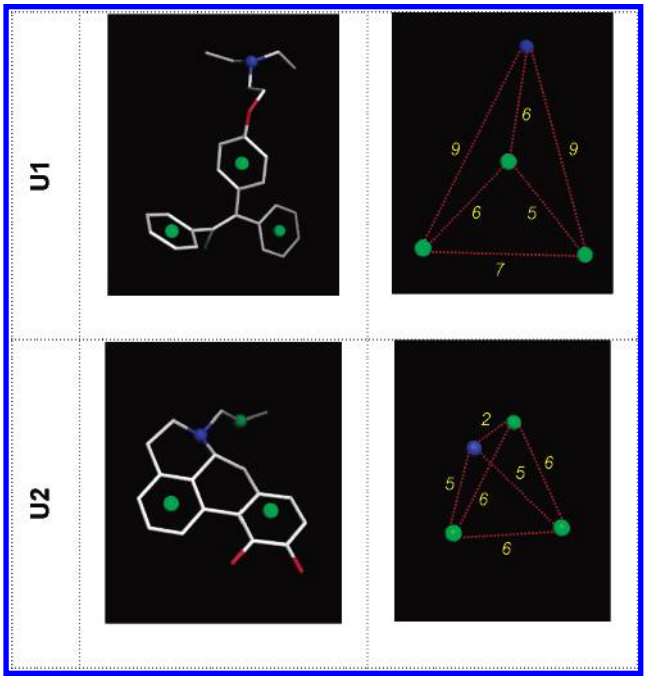

Figure 5. Four-point pharmacophore features of compounds **U1-2**.

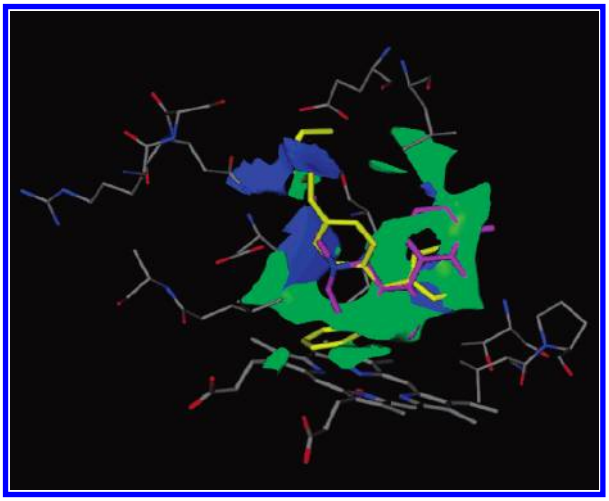

Figure 6. Hydrophobic (green) and hydrogen bonding (blue) regions in the active site of CYP2D6.

results,[18] where the CYP2D6 substrates were classified in two main groups, 5/7 Å substrates (interacting with Asp301) and 10 Å substrates (interacting with Glu216). In our QSAR docking model, compound U1 can be considered a 10 Å substrate, where the nitrogen atom forms hydrogen bonds with the residues Glu216 and Gln117. In contrast, compound U2 behaves as a 5 Å substrate that exclusively interacts with

Asp301 residue.[41] In fact, CoMFA contour plots have shown two different electrostatic regions to locate the ion-pair interaction depending on the distance between the protonated nitrogen atom and the aromatic ring under hydroxylation.[42] However, for more complex compounds, i.e., more functionalized than U1-2, an additional hydrogen bond acceptor (HBA) pharmacophoric feature is also needed.[43] Moreover, this hydrogen bonding competes or even becomes the primary binding interaction rather than the charge-pair interaction of the positively charged nitrogen.[44,45]

SUMMARY

TOPP pharmacophoric fingerprints can be successfully applied in QSAR approaches, as shown for the CYP2D6 stability example. The method displays a q^2 that is at least as good as other approaches available for QSAR analysis. The descriptors can be analyzed and displayed in 3D space and seem consistent with the current understanding of the CYP2D6 active site. The results from this study also show that the structures derived from docking in a protein (model) has no a significant influence on the q^2 of the QSAR approaches in this specific system. It is not certain if this will be the case for other protein structures. The TOPP approach is a 3D method with a speed resembling 2D fingerprint approaches, thereby outperforming commonly 3D-QSAR approaches evaluated during this study. Combining these features, TOPP can be used as a fast 3D-filtering methodology to identify molecules with potentially inappropriate (ADMET) properties in compound collections and during compound design.

ACKNOWLEDGMENT

Pfizer, Inc. is acknowledged for continuing support to this research. The authors gratefully thank Molecular Discovery for the evaluation version of the TOPP software. We also wish to acknowledge Prof. Gabriele Cruciani and Dr. Jonathan S. Mason for their helpful discussions during the work and preparation of this paper and Dr. Massimo Baroni for the development of TOPP.

Supporting Information Available: ISIS pictures of all the molecules incorrectly predicted in the different models studied. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Nelson, D. R.; Kamataki, T.; Waxman, D. J.; Guengerich, F. P.; Estabrook, R. W.; Feyereisen, R.; Gonzalez, F. J.; Coon, M. J.; Gunsalus, I. C.; Gotoh, O.; Okuda, K.; Nebert, D. W. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* **1993**, *12*, 1−51.

(2) Nelson, D. R.; Koymans, L.; Kamataki, T.; Stegeman, J. J.; Feyereisen, R.; Waxman, D. J.; Waterman, M. R.; Gotoh, O.; Coon, M. J.; Estabrook, R. W.; Gunsalus, I. C.; Nebert, D. W. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **1996**, *6*, 1−42.

(3) http://drnelson.utmem.edu/human.genecount.html.

(4) Smith, D. A.; Jones, B. C. Speculations on the substrate structure−activity relationship (SSAR) of cytochrome P450 enzymes. *Biochem. Pharmacol.* **1992**, *44*, 2089−2098.

(5) Mahgoub, A.; Idle, J. R.; Dring, L. G.; Lancaster, R.; Smith, R. L. Polymorphic oxidation of debrisoquine in man. *Lancet* **1977**, *311*, 584−586.

(6) Kroemer, H. K.; Eichelbaum, M. "It's the genes, stupid". Molecular bases and clinical consequences of genetic cytochrome P450 2D6 polymorphism. *Life Sci.* **1995**, *56*, 2285−2298.

(7) Sullivan-Klose, T. H.; Ghanayem, B. I.; Bell, D. A.; Zhang, Z. Y.; Kaminsky, L. S.; Shenfield, G. M.; Miners, J. O.; Birkett, D. J.; Goldstein, J. A. The role of the CYP2C9-Leu359 allelic variant in the tolbutamide polymorphism. *Pharmacogenetics* **1996**, *6*, 341−349.

(8) Aithal, G. P.; Day, C. P.; Kesteven, P. J.; Daly, A. K. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* **1999**, *353*, 717−719.

(9) Kidd, R. S.; Straughn, A. B.; Meyer, M. C.; Blaisdell, J.; Goldstein, J. A.; Dalton, J. T. Pharmacokinetics of chlorpheniramine, phenytoin, glipizide and nifedipine in an individual homozygous for the CYP2C9*3 allele. *Pharmacogenetics* **1999**, *9*, 71−80.

(10) Kidd, R. S.; Curry, T. B.; Gallagher, S.; Edeki, T.; Blaisdell, J.; Goldstein, J. A. Identification of a null allele of CYP2C9 in an African-American exhibiting toxicity to phenytoin. *Pharmacogenetics* **2001**, *11*, 803−808.

(11) Takahashi, H.; Echizen, H. Pharmacogenetics of warfarin elimination and its clinical implications. *Clin. Pharmacokinet.* **2001**, *40*, 587−603.

(12) Wrighton, S. A.; Stevens, J. C. The human hepatic cytochromes P450 involved in drug metabolism. *Crit. Rev. Toxicol.* **1992**, *22*, 1−21.

(13) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **1986**, *5*, 823−826.

(14) http://www.cerep.fr/cerep/users/pages/Collaborations/Bioprint.asp.

(15) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.

(16) http://www.mol-net.com/software/category/gen3dcoord.html.

(17) After completion of the work described, a crystal structure of CYP2D6 became available: Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281*, 7614−7622.

(18) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C., Novel approach to predicting P450 mediated drug metabolism. The development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 1515−1524.

(19) CCDC. Software *Gold*, *version 2.2*; CCDC Software: 12 Union Rd, Cambridge CB2 1E2, U.K., 2004.

(20) A commercial version will be available soon from Molecular Discovery Ltd., 215 Marsh Road, 1st Floor HA5 5NE, Pinner, Middlesex, U.K.

(21) Fontaine, F.; Pastor, M.; Sanz, F. Incorporating molecular shape into the alignment-free GRid-INdependent Descriptors. *J. Med. Chem.* **2004**, *47*, 2805−2815.

(22) Molecular Discovery Ltd. *Almond*, *version 3.3.0*; Molecular Discovery Ltd.: 215 Marsh Road, 1st Floor HA5 5NE, Pinner, Middlesex, U.K., 2005.

(23) Todeschini, R.; Consonni, V.; Pavan, M. *Dragon*, *version 2.1*; Milano Chemometrics and QSAR Research Group, Department of Environmental Sciences: Milano, Italy, 2002.

(24) Todeschini, R.; Consonni, V. *Handbook of molecular descriptors;* Wiley-VCH: Weinheim, 2000; Vol. 11, p 667.

(25) http://www.chemcomp.com/journal/descr.htm.

(26) Scitegic Inc. *PipeLine Pilot 4.5.2*, *version 4.5.2*; Scitegic Inc.: 9665 Chesapeake Dr., Suite 401, San Diego, CA 92123, U.S.A., 2005.

(27) Morgan, H. L. Generation of a unique machine description for chemical structures − a technique developed at Chemical Abstracts Service. *J. Chem. Docum.* **1965**, *5*, 107−113.

(28) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.

(29) Rogers, D.; Brown, R. D.; Hahn, M. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen.* **2005**, *10*, 682−686.

(30) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.

(31) McGregor, M. J.; Pallai, P. V. Clustering of large databases of compounds: using the MDL "keys" as structural descriptors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 443−448.

(32) Barnard, J. M.; Downs, G. M. Chemical fragment generation and clustering software. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 141−142.

(33) Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity profiling and design using 3D pharmacophores: pharmacophore-derived queries (PDQ). *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1214−1223.

(34) McGregor, M. J.; Muskal, S. M. Pharmacophore fingerprinting. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569−574.
(35) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 225−242.
(36) Arimoto, R.; Prasad, M.-A.; Gifford, E. M. Development of CYP3A4 inhibition models: comparisons of machine-learning techniques and molecular descriptors. *J. Biomol. Screening* **2005**, *10*, 197−205.
(37) O'Brien, S. E.; de Groot, M. J. Greater than the sum of its's parts: combining models for useful ADMET prediction. *J. Med. Chem.* **2005**, *48*, 1287−1291.
(38) For clarity, Table 3 only shows a small subset of the original binary bit string focused on the selected relevant triplets in Figure 2. Unstable (stable) compounds participate with negative (positive) sign of coefficients in the linear combination of original variables, therefore inversely (directly) correlated with the 2D6 metabolic stability property, respectively.
(39) Using a homology model can introduce uncertainties in the identification of amino acids responsible for interactions. However, the CYP2D6 system is well investigated, and the results obtained in this work are in line with known experimental results, including the recently obtained CYP2D6 crystal structure.
(40) Goodford, P. J. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(41) Paine, M. J.; McLaughlin, L. A.; Flanagan, J. U.; Kemp, C. A.; Sutcliffe, M. J.; Roberts, G. C.; Wolf, C. R. Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J. Biol. Chem.* **2003**, *278*, 4021−4027.
(42) Haji-Momenian, S.; Rieger, J. M.; Macdonald, T. L.; Brown, M. L. Comparative molecular field analysis and QSAR on substrates binding to cytochrome P450 2D6. *Bioorg. Med. Chem.* **2003**, *11*, 5545−5554.
(43) Snyder, R.; Sanger, R.; Wang, J.; Ekins, S. Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *Quant. Struct.-Act. Relat.* **2002**, *21*, 357−368.
(44) Hutzler, J. M.; Walker, G. S.; Wienkers, L. C. Inhibiton of cytochrome P450 2D6: structure−activity studies using a series of quinidine and quinine analogues. *Chem. Res. Toxicol.* **2003**, *16*, 450−459.
(45) Vaz, R. J.; Nayeem, A.; Santone, K.; Chandrasena, G.; Gavai, A. V. A 3D-QSAR model for CYP2D6 inhibition in the aryloxypropanolamine series. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3815−3820.

CI060143Q