J Mol Model (2002) 8:8–23
DOI 10.1007/s00894-001-0060-y

ORIGINAL PAPER

Wen-Hsin Huang · Richard E. Wilcox
Patrick J. Davis

Comparative molecular field analysis (CoMFA) for sulfoxidation reactions in *Mortierella isabellina* ATCC 42613 and *Helminthosporium* sp. NRRL 4671

Received: 18 June 2001 / Accepted: 23 October 2001 / Published online: 22 January 2002

Abstract Previous models for mechanisms of enzymatic sulfoxidation have been somewhat limited by a lack of knowledge of the essential features of substrate–enzyme versus product–enzyme relationships. Computerized methods for modeling ligand–protein (substrate–enzyme) interactions can overcome some of these limitations. Specifically, CoMFA (comparative molecular field analysis) provided a useful general approach in which to evaluate substrate–enzyme and product–enzyme relationships. The present investigation examined the relationship between substrate and product structure in predicting enantioselective sulfoxidation reactions using CoMFA for two species of microorganisms that have been used as models for mammalian metabolism, *Mortierella isabellina* and *Helminthosporium* sp. The overall enantioselectivity observed was based on the composite stereoselectivity of sulfoxide formation, sulfone formation (from the sulfoxide), and sulfoxide reduction back to the achiral substrate (sulfide).

Keywords Microbial synthesis · Drug metabolism · Sulfoxidation · Computational chemistry · CoMFA

Introduction

It is advantageous to assess the synthetic utility of enzymes and whole cell systems by predicting the kinetic parameters and the regiochemical and stereochemical outcome of proposed biotransformations. Various mechanism-based models [1] or "cubic-space" models [2] have been developed with qualitatively confirmed predictive abilities, but these models have limitations. Cubic-space descriptor models mapping the active site binding pockets for depicting bio-oxidation of various thioethers by microbial systems have been successfully proposed. The definition of a main hydrophobic pocket containing a strategic distal anchoring site (a polar hydrophobic binding site) and at least two smaller adjacent hydrophobic pockets defines the substrate accommodation primarily on steric/hydrophobic grounds. [3] Clearly there is a drawback in not considering the electronic contributions in the sulfide substrates used, since electronic and stereochemical effects contribute to the outcome of the majority of dynamic substrate–enzyme or ligand–receptor reactions. The "cubic-space" modeling approach is valuable in explaining and (to some extent) predicting stereochemical outcomes, but suffers from the limitation of being incapable of making predictions about either the expected yields of products or relevant kinetic parameters such as k*, Km and Vmax. Such kinetics resulting from substrates that can be oxidized at multiple sites are presumably determined by several factors. These include the intrinsic reactivity of the competing sites or competing multiple enzymes, their steric environment, hydrophobic binding, and the orientation of the substrate in the active site determined by protein–substrate interactions. [4]

Computerized techniques based on theoretical chemistry methods and experimental data can be used either to analyze molecules and molecular systems or to predict molecular and biological properties. Such approaches have grown rapidly since the 1980s. CoMFA (for abbreviations see Table 1), developed at Tripos, [5] is one of the most promising three-dimensional quantitative structure–activity relationship (3D QSAR) techniques. CoMFA has been widely used in the field of computer-aided drug design (CADD) since 1988 [6] to explore both statistical analyses and interactive graphics for correlating shapes and properties of molecules with their biological properties. The procedure addresses shape-dependent intermolecular interactions and accurately and easily de-

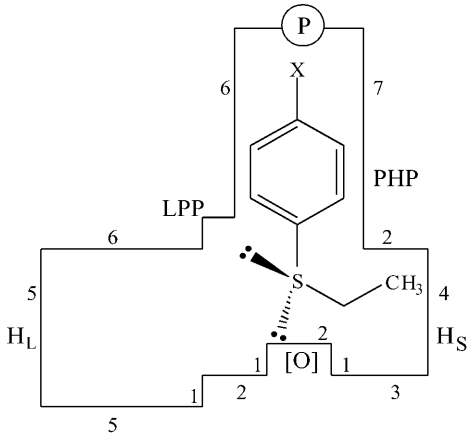
W.-H. Huang · P.J. Davis
Division of Medicinal Chemistry, College of Pharmacy, University of Texas at Austin, Austin, TX 78712, USA

R.E. Wilcox (✉)
Division of Pharmacology and Toxicology, College of Pharmacy, Wichita St., PHR 5.224D, University of Texas at Austin, Austin, TX 78712-1074, USA
e-mail: wilcoxrich@mail.utexas.edu
Tel.: +1-512-471-5199, Fax: +1-512-475-6088
URL: http://www.utexas.edu/pharmacy/divisions/pharmtox/faculty/wilcox.html

Table 1 Abbreviations

sp.		species
CoMFA		*Co*mparative *M*olecular *F*ield *A*nalysis
Solvents		
DMF	*N*,*N*-Dimethylformamide	
THF	Tetrahydrofuran	
Substrates		
TA	Thioanisole	$C_6H_5SCH_3$
EPS	Ethyl phenyl sulfide	$C_6H_5SC_2H_5$
n-PrPS	Propyl phenyl sulfide	C_6H_5S-*n*-C_3H_5
i-PrPS	Isopropyl phenyl sulfide	C_6H_5S-*i*-C_3H_5
c-PrPS	Cyclopropyl phenyl sulfide	C_6H_5S-*cyclo*-C_3H_5
4-FTA	*p*-Flurothioanisole	*p*-$FC_6H_4SCH_3$
4-ClTA	*p*-Chlorothioanisole	*p*-$ClC_6H_4SCH_3$
4-BrTA	*p*-Bromothioanisole	*p*-$BrC_6H_4SCH_3$
3-ClTA	*m*-Chlorothioanisole	*m*-$ClC_6H_4SCH_3$
3-BrTA	*m*-Bromothioanisole	*m*-$BrC_6H_4SCH_3$
4-CH_3TA	Methyl *p*-tolyl sulfide	*p*-$CH_3C_6H_4SCH_3$
4-CH_3OTA	1-Methoxy-4- (methylthio) benzene	*p*-$OCH_3C_6H_4SCH_3$
4-CH_3COTA	4'-(Methylthio) acetophenone	*p*-$COCH_3C_6H_4SCH_3$
4-NO_2TA	Methyl *p*-nitrophenyl sulfide	*p*-$NO_2C_6H_4SCH_3$
4-NH_2TA	4-(Methylthio) aniline	*p*-$NH_2C_6H_4SCH_3$
4-CNTA	4-(Methylthio) benzonitrile	*p*-$CNC_6H_4SCH_3$
BzMeS	Benzyl methyl sulfide	$C_6H_5CH_2SCH_3$
TC	Thiochroman	$C_6H_4S(CH_2)_3$

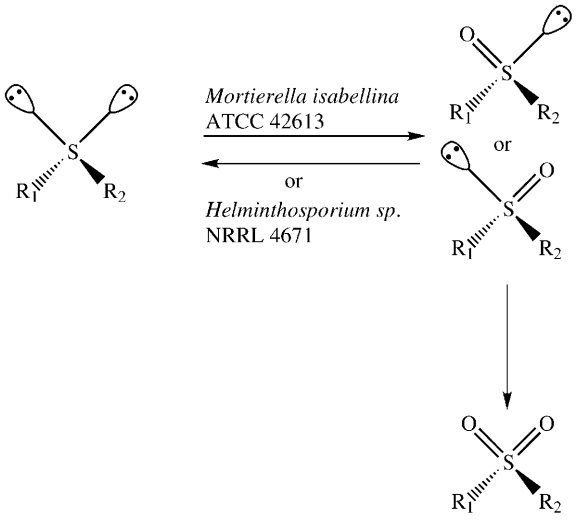
scribes the steric and electrostatic features of molecules in three-dimensional space. Few investigations have utilized the CoMFA approach to predict microbial biotransformations. One example is that of Faber et al. who used CoMFA to derive a quantitative substrate model for the enzymatic resolution of norbornanol esters by lipase from *Candida rugosa*. [7] Another example is that of Botta et al. who used the CoMFA approach for predicting the hydrolysis rates of 2-arylpropionic esters catalyzed by *Candida rugosa* lipase. [8] The present investigation examined the relationship between substrate and product structure and affinity for sulfoxidation reactions in two species of microorganisms that have been used as models for mammalian metabolism, *Mortierella isabellina* and *Helminthosporium* sp. using CoMFA. Previously, Holland and coworkers developed an active site model for the sulfoxidase of *Helminthosporium* species that contained a combination of two or more spatially defined hydrophobic binding regions (H_S and H_L, with S=small and L=large). A polar binding site and a polar site, P, located within the hydrophobic pocket (labeled PHP) were represented in the binding of *para*-substituted aryl substrates that contained a polar group capable of interacting with this site. Measurement of parameters from energy-minimized models suggested an optimum distance from site P to the oxidizing center of 8–10 Å (Fig. 1). [9]



Model for oxidation by Helminthosporium sp.
H_S: small hydrophobic pocket
H_L: large hydrophobic pocket
PHP: polar hydrophobic pocket
P: polar binding site
LPP: lone pair pocket
[O]: oxidation center
Dimensions in Angstrom

Fig. 1 Top view of active site model for sulfoxidation by *Helminthosporium* species NRRL 4671 – [Redrawn from Holland LL, et al (1997) Tetrahedron: Asymmetry 8:683–697]

Cytochrome P-450's heme prothetic group is responsible for the binding of oxygen, its activation, and delivery to the substrate of the oxidizing species. The cytochrome P-450 apoenzyme bearing hydrophobic pockets and/or polar binding sites is responsible for binding the substrate. [10] The apoenzyme decisively controls the substrate specificity and regio- and stereo-specificities of the bio-oxidation. Presumably, versatility in terms of induced fit is, to some extent, responsible for the range of substrate specificities and observed products. Therefore, the nature of the interaction between the apoenzyme and the substrate is the least understood aspect of oxygenase reactions. A moderate resolution X-ray structure is available for the P-450_{CAM} enzyme–substrate complex, which shows the substrate fitting tightly into a hydrophobic cavity of the enzyme and oriented toward the heme by



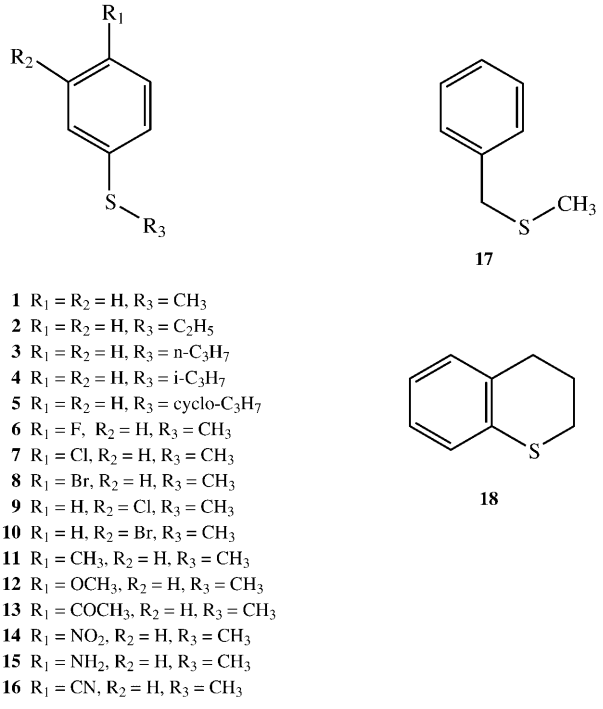
Scheme 1 Sulfide/sulfoxide/sulfone interconversions

H-bonding between the enzyme and substrate carbonyl, such that hydroxylation becomes regio- and stereo-selective. [11] The enantiomeric sulfoxides generated by fungal biotransformations may also undergo further oxidation to sulfone. A previous time course study also revealed the possibility that both *Mortierella isabellina* and *Helminthosporium* sp. reduce some sulfoxides back to the corresponding sulfides (Scheme 1), especially with substrates showing slow oxidation rates, such as 4-NO_2TA, 3-BrTA, 3-ClTA and 4-FTA (Huang and Davis, in preparation). Thus, kinetic models must take into account these competing processes for predicting kinetic parameters and stereochemical outcome.

Methods

Generation of 3D structures and conformational analyses of training set molecules

Starting 3D structures of the molecules were generated using the CONCORD program developed by Pearlman [12] and local minima were calculated using the MAXIMIN2 procedure within SYBYL. [5] Ligands probably do not bind to the protein in their global minimum energy conformations. This is because some degree of torsional change or rotatable bond flexion is required to adapt the ligand and protein to electrostatic and hydrogen-bonding distances to yield a ligand–protein complex of lower energy. The "minimum" energy conformation resulting from a MAXIMIN2 procedure is, therefore, only a useful starting point for possible candidate conformations for the compound. However, it is important to restrict the possible conformations of the ligand to those that can reasonably be obtained upon binding. Typically a 10 kcal mol^{-1} cutoff (difference between the local minimum and conformational energy) is considered reasonable. [13] As previously described, the template was selected based on the criteria of conformational rigidity and high biological activity. [13, 14, 15, 16]



1 $R_1 = R_2 = H, R_3 = CH_3$
2 $R_1 = R_2 = H, R_3 = C_2H_5$
3 $R_1 = R_2 = H, R_3 = n\text{-}C_3H_7$
4 $R_1 = R_2 = H, R_3 = i\text{-}C_3H_7$
5 $R_1 = R_2 = H, R_3 = cyclo\text{-}C_3H_7$
6 $R_1 = F, R_2 = H, R_3 = CH_3$
7 $R_1 = Cl, R_2 = H, R_3 = CH_3$
8 $R_1 = Br, R_2 = H, R_3 = CH_3$
9 $R_1 = H, R_2 = Cl, R_3 = CH_3$
10 $R_1 = H, R_2 = Br, R_3 = CH_3$
11 $R_1 = CH_3, R_2 = H, R_3 = CH_3$
12 $R_1 = OCH_3, R_2 = H, R_3 = CH_3$
13 $R_1 = COCH_3, R_2 = H, R_3 = CH_3$
14 $R_1 = NO_2, R_2 = H, R_3 = CH_3$
15 $R_1 = NH_2, R_2 = H, R_3 = CH_3$
16 $R_1 = CN, R_2 = H, R_3 = CH_3$

Scheme 2 Training set compounds

Alignment rule for training set compounds

Initially, a probable binding conformation of the template compound was selected. The template was selected based on substrates with high metabolic turnover followed by systematic search and multifit procedures to fit the seven heavy atoms of the phenyl-S group. The superimposition and alignment of the bioactive conformations of the training set were then done using flexible field fit procedures. [13, 14, 15, 16] CoMFA relates intermolecular interaction fields to biological activity (metabolism in this case). Thus, the field-fit method for improving alignment and correlation has been used successfully in a number of applications. There are two types of field-fit methods: the rigid fit method (the template-forcing approach based on atom overlapping) and the flexible field-fit method. The objective of the rigid "field-fit" approach is to minimize the root mean square (RMS) distances and/or the differences between field values of training set molecules and those of the template. The flexible field-fit allows changes in internal geometries to minimize the field differences between a compound and the template. Here, the rigid fit was performed first and then the flexible field fit was used for refining alignments. After the field-fit was completed, the CoMFA was rerun to optimize the correlation between molecules and biological activities. This is an "iterative" process in which those training set compounds having the highest residual value (difference between measured and predicted affinity) are re-fit using flexible field fit methods and the CoMFA and PLS redone (see below) until the model cannot be further improved. Scheme 2 shows the compounds of the training set.

Table 2 Logarithm of the apparent rate of substrate sulfoxidation by *Mortierella isabellina* ATCC 42613 (log k^*_T of total products; k^*_S: the apparent rate of (*S*)-sulfoxide formation; k^*_R: the apparent rate of (*R*)-sulfoxide formation; k^*_T: the apparent rate of total sulfoxide formation)

Entry	Substrate	log k^*_S	log k^*_R	log k^*_T	Substrate abbrev.
1	$C_6H_5SCH_3$	–1.050	–0.620	–0.500	TA
2	$C_6H_5SC_2H_5$	–0.731	–0.402	–0.295	EPS
3	C_6H_5S-*n*-C_3H_5	–0.847	–0.335	–0.325	n-PrPS
4	C_6H_5S-*i*-C_3H_5	–1.676	–0.812	–0.763	i-PrPS
5	C_6H_5S-*cyclo*-C_3H_5	–1.287	–0.699	–0.638	c-PrPS
6	*p*-$FC_6H_4SCH_3$	–0.921	–0.771	–0.614	4-FTA
7	*p*-$ClC_6H_4SCH_3$	–0.831	–0.416	–0.424	4-CITA
8	*p*-$BrC_6H_4SCH_3$	–1.010	–0.426	–0.436	4-BrTA
9	*m*-$ClC_6H_4SCH_3$	–1.866	–0.682	–0.668	3-CITA
10	*m*-$BrC_6H_4SCH_3$	–1.723	–0.733	–0.701	3-BrTA
11	*p*-$CH_3C_6H_4SCH_3$	–0.590	–0.372	–0.268	4-CH_3TA
12	*p*-$OCH_3C_6H_4SCH_3$	–0.443	–0.324	–0.083	4-CH_3OTA
13	*p*-$COCH_3C_6H_4SCH_3$	–1.146	–1.171	0.604[a]	4-CH_3COTA
14	*p*-$NO_2C_6H_4SCH_3$	–1.152	–1.073	–0.809	4-NO_2TA
15	*p*-$NH_2C_6H_4SCH_3$	–1.541	–1.023	–0.921	4-NH_2TA
16	*p*-$CNC_6H_4SCH_3$	–1.216	–0.719	–0.599	4-CNTA
17	$C_6H_5CH_2SCH_3$	–0.815	–0.795	–0.415	BzMeS
18	$C_6H_4S(CH_2)_3$	–1.755	–1.048	–0.970	TC

Comparative molecular field analysis (CoMFA)

The aligned active conformations of the compounds were placed in a grid box created by the software to calculate interaction energies at each lattice point surrounding the training set. Interaction energies include the steric and electrostatic interactions between a probe atom and each point of the grid box. Variants of the parameters examined in the present report include: fields (electrostatic, steric, or both fields); energy cutoffs (15–30 kcal mol^{-1}); dielectric function (constant or distance); column filtering, 1.0–4.0 kcal mol^{-1}); probe atom type (Csp^{3+}, O_3^-, H^+) (Csp^{3+} denotes a tetrahedral carbon with a +1 charge; O_3^- denotes an oxygen with a –1 charge; H^+ denotes a hydrogen with a +1 charge); grid step size (1 Å, 2 Å, 3 Å); and placement of the set of aligned compounds within the grid box (+0.5 Å, –0.5 Å, +1.0 Å, –1.0 Å, for each x, y, z direction).

Statistical analysis

Statistical analyses were performed to determine the relationship between the fields generated by the set of aligned structures and their biological activities. Partial least-squares (PLS) analyses using cross-validation [17] have been used successfully in many CoMFA studies and were used in the present investigation. PLS is an iterative procedure that produces its solutions, and summarizes the hundreds or thousands of descriptor variables such as CoMFA data using a few orthogonal new variables called latent variables (principal components of a factor analysis). PLS iteratively maximizes the degree of commonality between the set of descriptor variables and the biological data.

Contour plots

The processes described above (field-fit, CoMFA) are followed to iteratively adjust the alignment of the most poorly predicted compound(s) (higher residual values for measure–prediction) and the process is repeated until a mutual alignment is found which produces a 3D-QSAR having high predictive value. [13, 18]

Once the 3D QSAR has been optimized, the resultant coefficients for the columns in the table are usually plotted as contour diagrams. Typically there are two contour levels for each type of CoMFA energy field: the positive and negative contours. The contours are normally colored in green (magenta in the present figures) and yellow (blue in the present figures) for positive and negative steric effects, respectively. Positive steric contours show the regions where substituents increase the biological activity if occupied, and the negative steric contours show the area where substituents decrease the activity. Typically, blue (yellow in the present figures) and red (aqua in the present figures) are used for positive and negative electrostatic effects, respectively. The positive electrostatic contours indicate the regions where positive charges increase the activity, whereas the negative electrostatic contours display the regions where positive charges decrease the activity. Also, the coefficient plots depict the shape (active site) of the macromolecules (receptors or enzymes) in terms of intermolecular interaction fields.

Biological data

As previously described (Huang and Davis, in preparation), we established the biological data based on metabolism (chemical yield, apparent rate k*) using HPLC to analyze substrates and products. Methods employed for biotransformation, sampling and analysis involved our routine two-stage fermentation procedure, [19] in which substrate is added after 24 h of growth in stage-2 culture, with aliquot sampling at periodic intervals. (Hourly intervals are utilized during the early phase and daily intervals are used in the later phase of the biotransformation.) Samples were kept frozen until pro-

Table 3 Logarithm of the apparent rate of substrate sulfoxidation by *Helminthosporium* sp. NRRL 4671 (log k^*_T of total products; k^*_S: the apparent rate of (*S*)-sulfoxide formation; k^*_R: the apparent rate of (*R*)-sulfoxide formation; k^*_T: the apparent rate of total sulfoxide formation)

Entry	Substrate	log k^*_S	log k^*_R	log k^*_T	Substrate abbrev.
1	$C_6H_5SCH_3$	–1.353	–1.119	–1.119	TA
2	$C_6H_5SC_2H_5$	–1.110	–1.331	–0.905	EPS
3	C_6H_5S-*n*-C_3H_5	–1.438	–1.570	–1.198	*n*-PrPS
4	C_6H_5S-*i*-C_3H_5	–1.699	–1.507	–1.301	*i*-PrPS
5	C_6H_5S-*cyclo*-C_3H_5	–1.772	–1.631	–1.395	*c*-PrPS
6	*p*-$FC_6H_4SCH_3$	–1.248	–1.453	–1.038	4-FTA
7	*p*-$ClC_6H_4SCH_3$	–1.169	–1.604	–1.033	4-CITA
8	*p*-$BrC_6H_4SCH_3$	–0.936	–1.510	–0.833	4-BrTA
9	*m*-$ClC_6H_4SCH_3$	–1.186	–1.747	–1.080	3-CITA
10	*m*-$BrC_6H_4SCH_3$	–1.521	–2.032	–1.405	3-BrTA
11	*p*-$CH_3C_6H_4SCH_3$	–1.439	–1.527	–1.180	4-CH_3TA
12	*p*-$OCH_3C_6H_4SCH_3$	–0.923	–1.514	–0.824	4-CH_3OTA
13	*p*-$COCH_3C_6H_4SCH_3$	–1.860	–2.310	–0.412[a]	4-CH_3COTA
14	*p*-$NO_2C_6H_4SCH_3$	–1.434	–2.347	–1.384	4-NO_2TA
15	*p*-$NH_2C_6H_4SCH_3$	–1.569	–2.222	–0.922[a]	4-NH_2TA
16	*p*-$CNC_6H_4SCH_3$	–1.536	–2.444	–1.485	4-CNTA
17	$C_6H_5CH_2SCH_3$	–1.260	–1.376	–1.113	BzMeS
18	$C_6H_4S(CH_2)_3$	–1.304	–1.409	–1.052	TC

Table 4 Inhibition studies of activity of *Mortierella isabellina* and *Helminthosporium* sp. with or without phenobarbital pretreatment

Pretreatment	Substrate control (ethyl phenyl sulfide)	Substrate plus methimazole	Substrate plus *n*-decylamine
Mortierella isabellina			
No pretreatment	100% k_R/k_S=3.26	16.1% (total) 38.8% (*S*)-sulfoxide 9.1% (*R*)-sulfoxide k_R/k_S=0.76	0% (complete inhibition)
Phenobarbital pretreated	75.7% (total sulfoxides) 89.2% (*S*)-sulfoxide 71.5% (*R*)-sulfoxide k_R/k_S=2.61	61.25% (total) 84.6% (*S*)-sulfoxide 53.9% (*R*)-sulfoxide k_R/k_S=2.08	5.8% (total) 7.2% (*S*)-sulfoxide 5.4% (*R*)-sulfoxide k_R/k_S=2.63
Interaction ratio	0.80	2.74	2.63
Helminthosporium sp.			
No pretreatment	100% k_S/k_R=1.38	100% (no inhibition)	18.0% (total) 18.9% (*S*)-sulfoxide 16.8% (*R*)-sulfoxide k_S/k_R=1.56

cessing by liquid–liquid extraction and analysis using chiral HPLC. Graphical analysis and curve fitting were utilized to determine kinetic parameters for the biotransformation of each substrate with each of the two organisms (*H.* species and *M. isabellina*) (Huang and Davis, in preparation).

The dependent variable, expressed as logarithmic or percentile values, was examined in the CoMFA. Tables 2 and 3 showed the log k* of fungal sulfoxidation of a series of alkyl aryl sulfides examined by *Mortierella isabellina* ATCC 42613 and *Helminthosporium* sp. NRRL 4671, respectively. The apparent rate of (*R*)-sulfoxide formation is consistently higher than that for (*S*)-sulfoxide formation for *Mortierella isabellina* (Table 2). In contrast, the apparent rate of (*S*)-sulfoxide formation was generally higher than that for (*R*)-sulfoxide formation in *Helminthosporium* sp. (Table 3).

Results

Results with *Mortierella isabellina*

Inhibition studies of activity for *Mortierella isabellina* and *Helminthosporium* sp. were performed and data measured by the formation of chiral sulfoxides are presented in Table 4. Under control conditions (without inhibitors of biotransformation) using *Mortierella isabellina*, the apparent rate of the (*R*)-sulfoxide formation is 3.3-fold over that for (*S*)-sulfoxide formation. The inhibition by methimazole of *M. isabellina* activity decreased the ratio of the rate of enantiomeric sulfoxide formation to k_R/k_S=0.76 (from 3.26). Addition of *n*-decylamine completely blocked *M. isabellina* sulfoxidation.

Phenobarbital pretreatment was used to produce induction of the CY-P450 enzyme complex. However, the treatment caused an almost 25% reduction in overall metabolism. Thus, following phenobarbital treatment, the

ratio of the apparent rate of enantiomeric sulfoxide formation is 2.61 for the *M. isabellina* a 20% inhibition (2.61/3.26). We also determined the ratio of the rate of enantiomeric sulfoxide formation following phenobarbital pretreatment in response to the inhibitor methimazole. The ratio of preferential activity of enzymes was slightly decreased from 2.61 in control assays to 2.08 by methimazole/phenobarbital. That is, methimazole produced less inhibition of metabolism following phenobarbital pretreatment leading to a 2.74-fold greater selectivity of metabolism (2.74=2.08/0.76). Similarly, less induction of metabolism of total sulfoxide formation in response to n-decylamine occurred after phenobarbital treatment of *M. isabellina*, 5.8% total sulfoxides versus 0%.

Results with *Helminthosporium* sp

Under control conditions (without inhibitors of biotransformation) using *Helminthosporium* sp., the apparent rate of (*S*)-sulfoxide formation was 1.38-fold over that for (*R*)-sulfoxide formation. Methimazole did not inhibit the activity of *Helminthosporium* sp. enzymes. The presence of *n*-decylamine induced an 82% inhibition in the activity of *Helminthosporium* sp. enzymes (18% activity remaining) yielding k_S/k_R=1.56. Thus, overall, these prototype metabolic inhibitors produced differential effects on sulfoxidation with a standard substrate in different species of fungi.

Compound alignment

In the present study, the initial conformations for alignment of the sulfides (Scheme 2) and each corresponding sulfoxide were created from CONCORD-derived structures [12] and minimized using the MAXIMIN2 procedure with the Gasteiger–Marsili charges within SYBYL. [5] Ligand–protein interactions usually occur at intermolecular interaction energies relatively close to the minimum energy (within 10 kcal mol^{-1}). Therefore, the "minimum" energy conformation derived from a SYBYL MAXIMIN2 procedure is a useful starting point for the initial bioactive conformations of compounds. Furthermore, a conformational search was employed for each compound to explore the probable low energy conformers. A systematic search was done for all flexible sulfide and sulfoxide compounds. This search explored the conformational space generated by searching all rotatable bonds in 30° increments. Thiochroman (compound **18**), the only rigid compound in this study, and its corresponding sulfoxides were subjected to random conformational searches to determine possible low energy conformers.

Generally, the high-biological activity compounds of the training set (compounds **8** and **12**) are quite flexible, which increases the problem of template selection. We used the lowest energy conformations (close to global minimum energy conformations) as candidate conformers, followed by the multifit procedure of SYBYL to fit

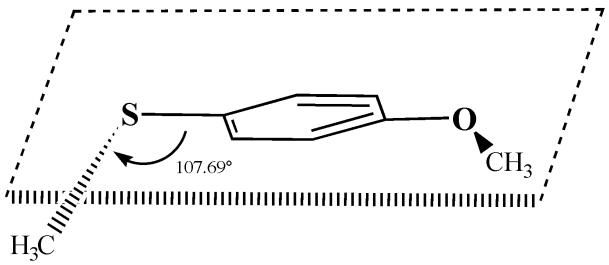


Fig. 2 The lowest energy conformation of CH_3OTA after systematic search

the seven heavy atoms of the phenyl-S group. The flexible field-fit procedure was then used to refine the alignment with CH_3OTA (compound **12**) as a template, but this procedure could not further improve the models in this study. We did not know whether CoMFA would correlate the biological activity with the structure of substrates or the corresponding products prior to performing the studies. Therefore, we did both substrate alignment and dominant product alignment for each *Mortierella isabellina* and *Helminthosporium* sp. We aligned the 17 compounds (excluding cyclopropyl phenyl sulfide, compound **5**, and its metabolites or cyclopropyl phenyl sulfoxides) as a training set for each model of *Mortierella isabellina* and *Helminthosporium* sp.

Alignment rule

We determined the lowest energy conformation after a systematic search. The results yielded a conformation in which the alkyl group was pendant in the phenyl-S plane shown in Fig. 2. The final aligned structures of the training set of substrates and products for *Mortierella isabellina* and *Helminthosporium* sp. are shown in Figs. 3, 4, 5, and 6, respectively. In the final alignment, CH_3COTA (compound **13**) and NH_2TA (compound **15**) were in opposite orientations compared with the rest of the sulfides or sulfoxides. NH_2TA was included in the alignment set based on the substrate-guided alignment for *Mortierella isabellina*. NH_3TA (protonated) was aligned on the basis of the substrate-guided alignment for *Helminthosporium* sp. [There were several noteworthy features of the alignments as follows. Using a product-guided alignment, the (*R*)-sulfoxide of CH_3COTA (compound **13**) was aligned with other (*R*)-sulfoxides of the training set for *Mortierella isabellina*. Similarly, the (*S*)-alcohol of CH3COTA was aligned with the (*S*)-sulfoxides of the training set for *Helminthosporium* sp. The N-acetylated compound **15** was aligned for both *Mortierella isabellina* and *Helminthosporium* sp. for the product-guided alignment (Figs. 5 and 6). Both the corresponding (*R*)-sulfoxides of *i*-PrPS and *cyclo*-PrPS (compounds **4** and **5**) were aligned for the training set and for predictions using product-guided alignment, respectively. *m*-Bromothioanisole and *m*-chlorothioanisole (compounds **8** and **9**) have opposite orientations in their alignments for Mortierella isabellina and *Helminthosporium* sp.]

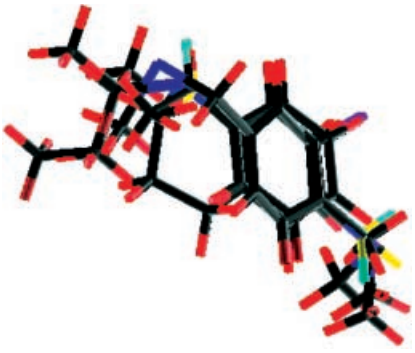


Fig. 3 Aligned substrates for *Mortierella isabellina*

Fig. 5 Aligned products for *Mortierella isabellina*

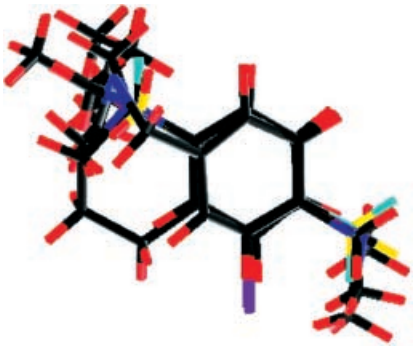


Fig. 4 Aligned substrates for *Helminthosporium* sp.

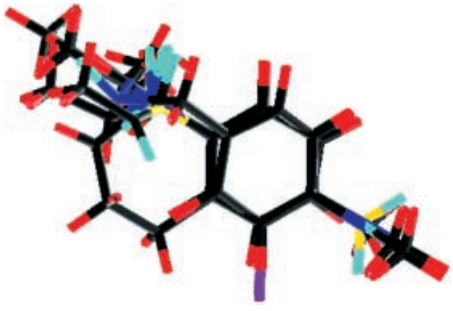


Fig. 6 Aligned products for *Helminthosporium* sp.

Table 5 CoMFA results by various options at aligned sulfides (substrates) for *Mortierella isabellina*

Fields	E cutoff (kcal mol–1)	Diele. func.	CoMFA Region	Min. σ	Grid step size (Å)	Probe atom type	No. of columns	q^2	No. of comp.	scv
Both	30/30	1/r	default	2.0	2.0	Csp^{3+}	85/1152	0.713	3	0.219
Steric	30	1/r	default	2.0	2.0	Csp^{3+}	70/648	–0.009	3	0.411
Elec.	30	1/r	default	2.0	2.0	Csp^{3+}	51/648	0.020	4	0.422
Both	50/50	1/r	default	2.0	2.0	Csp^{3+}	89/1152	0.716	3	0.218
Both	35/35	1/r	default	2.0	2.0	Csp^{3+}	85/1152	0.720	3	0.217
Both	25/25	1/r	default	2.0	2.0	Csp^{3+}	84/1152	0.717	3	0.218
Both	30/30	1.0	default	2.0	2.0	Csp^{3+}	137/1152	0.579	3	0.266
Both	35/35	1.0	default	2.0	2.0	Csp^{3+}	136/1152	0.591	4	0.272
Both	35/35	1/r	default	1.0	2.0	Csp^{3+}	136/1152	0.657	3	0.240
Both	35/35	1/r	default	3.0	2.0	Csp^{3+}	76/1152	0.719	3	0.217
Both	35/35	1/r	default	2.0	2.0	O_3^-	138/1152	0.493	3	0.291
Both	35/35	1/r	default	2.0	2.0	H^+	148/1152	0.482	3	0.294
Both	35/35	1/r	+0.2(x,y,z)	2.0	2.0	Csp^{3+}	147/1152	0.474	3	0.325
Both	35/35	1/r	–0.2(x,y,z)	2.0	2.0	Csp^{3+}	133/1152	0.509	3	0.287
Both+C*logP	35/35	1/r	default	2.0	2.0	Csp^{3+}	86/1153	0.668	4	0.245

CoMFA

After the aligned bioactive conformations were obtained, initial CoMFAs were run using the Sybyl default settings (Tables 5, 6, 7, and 8). The target property (log k*) was regressed against predictors (independent variables) calculated as steric and electrostatic features of the intermolecular interaction fields using PLS regression. We systematically investigated the various CoMFA options to determine the final optimal CoMFA model as described in "Methods" and shown in Tables 5, 6, 7, and 8. The "leave one out" method of cross-validation in PLS was used to find the optimum options and number of components based on the standard error of prediction. After determining the optimum CoMFA model, we derived the final model without the cross-validation to display the contour plots. Tables 9, 10, 11, and 12 show the actual, predicted and residual values for the training set compounds for the four models.

A column of computed log P values for each of the compounds calculated from the program CLOGP [20] was used in addition to the CoMFA column as a hydro-

Table 6 CoMFA results by various options at aligned sulfides (substrates) for *Helminthosporium* sp.

Fields	E cutoff (kcal mol^{-1})	Diele. func.	CoMFA Region	Min. s	Grid step size (Å)	Probe atom type	No. of columns	q^2	No. of comp.	s_{cv}
Both	30/30	1/r	default	2.0	2.0	Csp^{3+}	177/1152	0.485	2	0.257
Steric	30	1/r	default	2.0	2.0	Csp^{3+}	57/576	–0.233	2	0.398
Elec.	30	1/r	default	2.0	2.0	Csp^{3+}	157/576	0.333	1	0.283
Both	50/50	1/r	default	2.0	2.0	Csp^{3+}	185/1152	0.470	2	0.261
Both	35/35	1/r	default	2.0	2.0	Csp^{3+}	179/1152	0.491	2	0.256
Both	25/25	1/r	default	2.0	2.0	Csp^{3+}	176/1152	0.476	2	0.260
Both	30/30	1.0	default	2.0	2.0	Csp^{3+}	605/1152	0.218	2	0.317
Both	30/30	1/r	default	2.0	2.0	O_3^-	178/1152	0.473	2	0.260
Both	30/30	1/r	default	2.0	2.0	H^+	178/1152	0.390	2	0.280
Both	30/30	1/r	+0.2x, y, z	2.0	2.0	Csp^{3+}	180/1152	0.533	2	0.245
Both	30/40	1/r	+0.2x, y, z	2.0	2.0	Csp^{3+}	180/1152	0.546	2	0.242
Both	30/40	1/r	+0.2x, y, z	2.5	2.0	Csp^{3+}	135/1152	0.553	2	0.240
Both	30/40	1/r	+0.2x, y, z	3.9	2.0	Csp^{3+}	76/1152	0.578	2	0.233
Both+C*logP	30/40	1/r	+0.2x, y, z	3.9	2.0	Csp^{3+}	77/1153	0.470	3	0.0.271

Table 7 CoMFA results by various options at aligned *R*-sulfoxides (products) for *Mortierella isabellina*

Fields	E cutoff (kcal mol–1)	Diele. func.	CoMFA Region	Min. σ	Grid step size (Å)	Probe atom type	No. of columns	q^2	No. of comp.	s_{cv}
Both	30/30	1/r	default	2.0	2.0	Csp^{3+}	83/1280	0.706	6	0.183
Steric	30	1/r	default	2.0	2.0	Csp^{3+}	58/640	–0.014	3	0.297
Elec.	30	1/r	default	2.0	2.0	Csp^{3+}	61/640	0.460	5	0.236
Both	50/50	1/r	default	2.0	2.0	Csp^{3+}	87/1280	0.663	6	0.195
Both	35/35	1/r	default	2.0	2.0	Csp^{3+}	83/1280	0.660	6	0.196
Both	20/20	1/r	default	2.0	2.0	Csp^{3+}	81/1280	0.738	6	0.172
Both	20/20	1.0	default	2.0	2.0	Csp^{3+}	171/1280	0.603	6	0.212
Both	30/30	1.0	default	2.0	2.0	Csp^{3+}	173/1280	0.564	6	0.222
Both	20/20	1/r	default	1.0	2.0	Csp^{3+}	150/1280	0.664	6	0.195
Both	20/20	1/r	default	2.5	2.0	Csp^{3+}	81/1280	0.776	6	0.159
Both	20/20	1/r	default	2.0	2.0	O_3^-	170/1280	0.543	6	0.227
Both	20/20	1/r	default	2.0	2.0	H^+	171/1280	0.566	6	0.222
Both	20/20	1/r	–0.2(x,y,z)	2.0	2.0	Csp^{3+}	166/1280	0.455	5	0.237
Both	20/20	1/r	+0.2(x,y,z)	2.0	2.0	Csp^{3+}	173/1280	0.524	6	0.232
Both+C*logP	20/20	1/r	default	2.5	2.0	Csp^{3+}	71/1281	0.736	6	0.173

Table 8 CoMFA results by various options at aligned *S*-sulfoxides (products) for *Helminthosporium* sp.

Fields	E cutoff (kcal mol–1)	Diele. func.	CoMFA Region	Min. σ	Grid step size (Å)	Probe atom type	no. of columns	q^2	No. of comp.	s_{cv}
Both	30/30	1/r	default	2.0	2.0	Csp^{3+}	84/1296	0.310	3	0.270
Steric	30	1/r	default	2.0	2.0	Csp^{3+}	54/648	0.228	1	0.265
Elec.	30	1/r	default	2.0	2.0	Csp^{3+}	67/648	0.487	3	0.232
Both	50/50	1/r	default	2.0	2.0	Csp^{3+}	88/1296	0.293	3	0.273
Both	35/35	1/r	default	2.0	2.0	Csp^{3+}	86/1296	0.307	3	0.270
Both	S32.5/E30	1/r	default	2.0	2.0	Csp^{3+}	85/1296	0.316	3	0.268
Both	S32.5/E30	1.0	default	2.0	2.0	Csp^{3+}	161/1296	0.233	1	0.266
Both	S32.5/E30	I/r	default	2.0	2.0	O_3^-	86/1296	0.280	3	0.275
Both	S32.5/E30	1/r	default	2.0	2.0	H^+	83/1296	0.283	2	0.265
Both	S32.5/E30	1/r	+0.1x,y,–0.3z	2.0	2.0	Csp^{3+}	85/1296	0.375	3	0.257
Both	S32.5/E30	1/r	+0.1x,y,–0.3z	3.0	2.0	Csp^{3+}	64/1296	0.310	3	0.270
Both	S32.5/E30	1/r	+0.1x,y,–0.3z	4.0	2.0	Csp^{3+}	50/1296	0.396	3	0.252
Both	S32.5/E30	1/r	+0.1x,y,-0.3z	3.9	2.0	Csp^{3+}	50/1296	0.397	3	0.252
Both+C*logP	S32.5/E30	1/r	+0.1x,y,–0.3z	3.7	2.0	Csp^{3+}	53/1297	0.477	3	0.235

phobic term in the final model to estimate the entropic effect in this study. We tested the predictive ability of our final model using the compound cyclopropyl phenyl sulfide that was excluded in the training set for each model. Cyclopropyl phenyl sulfide and the corresponding sulfoxides (*R*- and *S*-) were optimized and aligned in the same manner as described above.

The cross-validated R^2 (q^2) values resulting from the various CoMFA options against log k* for *Mortierella isabellina* and *Helminthosporium* sp. are shown in

Table 9 Actual, predicted and residual values, and relative contributions from the final model of aligned sulfides for *Mortierella isabellina*

Substrate	Actual (log k*)	Predicted (log k*)	Residual (Actual–Predicted)	Relative contributions (fraction)	
$C_6H_5SCH_3$	–0.50	–0.56	0.06	Steric (Norm. Coeff.)	Electrostatic (Norm. Coeff.)
$C_6H_5SC_2H_5$	–0.30	–0.31	0.01	0.381 (1.444)	0.619 (2.343)
C_6H_5S-*n*-C_3H_5	–0.32	–0.36	0.04		
C_6H_5S-*i*-C_3H_5	–0.76	–0.82	0.06		
p-$FC_6H_4SCH_3$	–0.61	–0.58	–0.03		
p-$ClC_6H_4SCH_3$	–0.42	–0.42	–0.0011		
p-$BrC_6H_4SCH_3$	–0.44	–0.40	–0.04		
m-$ClC_6H_4SCH_3$	–0.67	–0.69	0.02		
m-$BrC_6H_4SCH_3$	–0.70	–0.70	–0.0037		
p-$CH_3C_6H_4SCH_3$	–0.27	–0.29	0.02		
p-$OCH_3C_6H_4SCH_3$	–0.08	–0.05	–0.03		
p-$COCH_3C_6H_4SCH_3$	0.60	0.60	0.0044		
p-$NO_2C_6H_4SCH_3$	–0.81	–0.73	–0.08		
p-$NH_2C_6H_4SCH_3$	–0.92	–0.88	–0.04		
p-$CNC_6H_4SCH_3$	–0.60	–0.62	0.02		
$C_6H_5CH_2SCH_3$	–0.42	–0.41	–0.01		
$C_6H_4S(CH_2)_3$	–0.97	–0.98	0.01		

Table 10 Actual, predicted and residual values, and relative contributions from the final model of aligned sulfides for *Helminthosporium* sp.

Substrate	Actual (log k*)	Predicted (log k*)	Residual (Actual–Predicted)	Relative contributions (fraction)	
$C_6H_5SCH_3$	–1.12	–0.99	–0.13	Steric (Norm. Coeff.)	Electrostatic (Norm. Coeff.)
$C_6H_5SC_2H_5$	–0.91	–0.88	–0.03	0.431 (2.431)	0.569 (3.211)
C_6H_5S-*n*-C_3H_5	–1.20	–1.18	–0.02		
C_6H_5S-*i*-C_3H_5	–1.30	–1.32	0.02		
p-$FC_6H_4SCH_3$	–1.04	–1.11	0.07		
p-$ClC_6H_4SCH_3$	–1.03	–1.09	0.06		
p-$BrC_6H_4SCH_3$	–0.83	–0.85	0.02		
m-$ClC_6H_4SCH_3$	–1.08	–1.21	0.13		
m-$BrC_6H_4SCH_3$	–1.40	–1.35	–0.05		
p-$CH_3C_6H_4SCH_3$	–1.18	–1.15	–0.03		
p-$OCH_3C_6H_4SCH_3$	–0.82	–0.83	0.01		
p-$COCH_3C_6H_4SCH_3$	–0.41	–0.41	0.0008		
p-$NO_2C_6H_4SCH_3$	–1.38	–1.40	–0.01		
p-$NH_2C_6H_4SCH_3$	–1.98	–1.98	–0.0018		
p-$CNC_6H_4SCH_3$	–1.49	–1.41	–0.07		
$C_6H_5CH_2SCH_3$	–1.11	–1.15	0.04		
$C_6H_4S(CH_2)_3$	–1.05	–1.02	–0.03		

Tables 5, 6, 7, and 8. The choice of CoMFA options described below was based on maximizing the q^2 value and minimizing the standard error of prediction.

CoMFA for the apparent sulfoxidation rate in Mortierella isabellina *based on substrate-guided alignment (Tables 5 and 9)*

At the default settings including both steric and electrostatic fields, we observed a q^2 of 0.713 and a standard error of prediction of 0.219 with three principal components for *Mortierella isabellina* (Table 5). The final model was generated without cross-validation at three principal components. It utilized the following options: both steric and electrostatic fields with 35 kcal mol^{-1} cutoffs, 1/r for the electrostatic function, a 2.0 Å step size, 2.0 kcal mol^{-1} of column filtering, a Csp^{3+} probe atom, and a grid box set at SYBYL's default position. This model had an observed q^2 of 0.720 with three principal components, a standard error of prediction of 0.219, an R^2 value of 0.990, a standard error of estimate of 0.047 and an F (6, 10) value of 162.812. The final cross-validated model utilized 85 of 1152 columns for the analysis. Figure 7 shows the relationship between the calculated and measured log k* values for the non-cross-validated analysis of the *Mortierella isabellina* based on the substrate-guided alignment. The measured and predicted parameters, residuals and relative contributions of steric and electrostatic coefficients are shown in Table 9. Incorporating a C*log P term in the CoMFA analysis, yielded a lower q^2 value of 0.668 with four principal

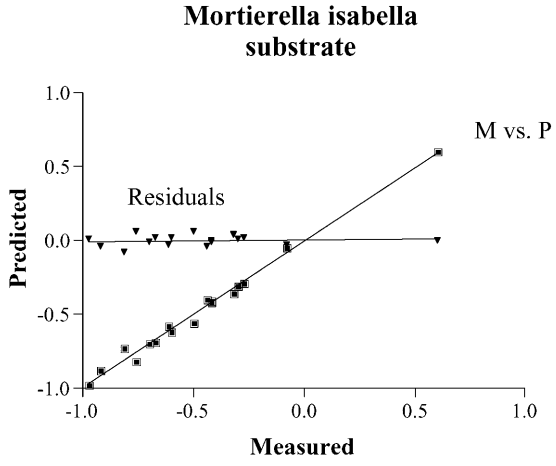


Fig. 7 Relationship between calculated and measured log k* values for the non-cross-validated analysis of the *Mortierella isabellina* based on the substrate-guided alignment

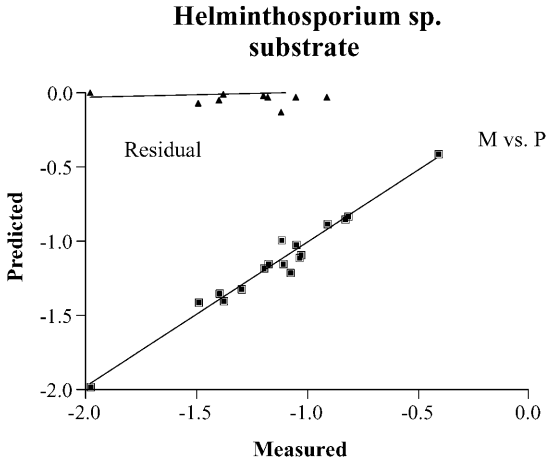


Fig. 8 Relationship between calculated and measured log k* values for the non-cross-validated analysis of the *Helminthosporium* sp. based on the substrate-guided alignment

Table 11 Actual, predicted and residual values, and relative contributions from the final model of aligned products for *Mortierella isabellina*

Products of	Actual (log k^*_R)	Predicted (log k^*_R)	Residual (Actual–Predicted)	Relative contributions (fraction)	
$C_6H_5SCH_3$	–0.62	–0.58	–0.04	Steric (Norm. Coeff.)	Electrostatic (Norm. Coeff.)
$C_6H_5SC_2H_5$	–0.40	–0.39	–0.01	0.554 (2.011)	0.446 (1.622)
C_6H_5S-*n*-C_3H_5	–0.34	–0.34	0.007		
C_6H_5S-*i*-C_3H_5	–0.81	–0.86	0.05		
p-$FC_6H_4SCH_3$	–0.77	–0.75	–0.02		
p-$ClC_6H_4SCH_3$	–0.42	–0.49	0.07		
p-$BrC_6H_4SCH_3$	–0.43	–0.43	–0.0002		
m-$ClC_6H_4SCH_3$	–0.68	–0.70	–0.02		
m-$BrC_6H_4SCH_3$	–0.73	–0.70	–0.03		
p-$CH_3C_6H_4SCH_3$	–0.37	–0.39	0.02		
p-$OCH_3C_6H_4SCH_3$	–0.32	–0.32	0.001		
p-$COCH_3C_6H_4SCH_3$[a]	–1.17	–1.22	0.05		
p-$NO_2C_6H_4SCH_3$	–1.07	–1.02	–0.05		
p-$NH_2C_6H_4SCH_3$[a]	–0.62	–0.61	–0.01		
p-$CNC_6H_4SCH_3$	–0.72	–0.67	–0.05		
$C_6H_5CH_2SCH_3$	–0.80	–0.80	–0.0035		
$C_6H_4S(CH_2)_3$	–1.05	–1.05	–0.0005		

[a] Dominant (*R*)-sulfoxide of 4-CH_3COTA; acetylation of NH_2TA

components and with a standard error of prediction of 0.245 for the substrate-guided alignment (Table 5). For the CoMFA model based on substrate-guided alignment in *Mortierella isabellina* electrostatic features of the model predominated over steric features (Table 9) in accounting for variance in the data. A residual plot indicated that there was no systematic basis for variation in the predicted log k* values.

CoMFA for the apparent sulfoxidation rate in Helminthosporium *sp. based on substrate-guided alignment (Tables 6 and 10)*

Using the default settings, which included both steric and electrostatic fields, we observed a q^2 of 0.485 and a standard error of prediction of 0.257 with two principal components for *Helminthosporium* sp (Table 6). The final non-cross validated model was obtained using two principal components. The model was derived using the following options. Steric fields with 30 kcal mol^{-1} cutoffs and electrostatic fields with 40 kcal mol^{-1} cutoffs were used. The electrostatic function was 1/r. A 2.0 Å step size, 3.9 kcal mol^{-1} of column filtering, a Csp^{3+} probe atom, and a grid box set at (+0.2x, y, z) relative to SYBYL's default position completed the parameterization. This model had a q^2 of 0.578 with two principal components, a standard error of prediction of 0.233, an R^2 value of 0.968, a standard error of estimate of 0.075 and an F (6, 10) value of 50.982. The final cross-validated model utilized 76 of 1152 columns for the analysis. Figure 8 showed the relationship between the calculated

Table 12 Actual, predicted and residual values, and relative contributions from the final model of aligned sulfides for *Helminthosporium* sp.

Products of	Actual ($\log k^*_S$)	Predicted ($\log k^*_S$)	Residual (Actual–Predicted)	Relative contributions (fraction)	
$C_6H_5SCH_3$	–1.35	–1.43	0.08	Steric (norm. coeff.)	Electrostatic (norm. coeff.)
$C_6H_5SC_2H_5$	–1.11	–1.14	0.03	0.617 (2.195)	0.279 (0.991)
C_6H_5S-*n*-C_3H_5	–1.44	–1.44	0.0036	C*log P (norm. coeff.)[a]	
C_6H_5S-*i*-C_3H_5[b]	–1.51	–1.51	–0.0016	0.141 (0.371)	
p-$FC_6H_4SCH_3$	–1.25	–1.20	–0.05		
p-$ClC_6H_4SCH_3$	–1.17	–1.15	–0.02		
p-$BrC_6H_4SCH_3$	–0.94	–1.13	0.19		
m-$ClC_6H_4SCH_3$	–1.19	–1.31	0.12		
m-$BrC_6H_4SCH_3$	–1.52	–1.28	–0.24		
p-$CH_3C_6H_4SCH_3$	–1.44	–1.40	–0.04		
p-$OCH_3C_6H_4SCH_3$	–0.92	–0.87	–0.05		
p-$COCH_3C_6H_4SCH_3$[b]	–0.41	–0.41	0.0018		
p-$NO_2C_6H_4SCH_3$	–1.43	–1.48	0.05		
p-$NH_2C_6H_4SCH_3$[b]	–0.92	–0.95	0.03		
p-$CNC_6H_4SCH_3$	–1.54	–1.49	–0.05		
$C_6H_5CH_2SCH_3$	–1.26	–1.29	0.03		
$C_6H_4S(CH_2)_3$	–1.30	–1.22	–0.08		

[a] Log k^*_S=–1.689+(0.221)*C log P
[b] Dominant (*R*)-sulfoxide of *i*-PrPS; dominant (*S*)-alcohol of 4-CH_3COTA; acetylation of NH_2TA

and measured log k* values for the non-cross-validated analysis of the *Helminthosporium* sp. based on the substrate-guided alignment. For this CoMFA model steric features accounted for 43% while electrostatic features accounted for 57% of the variance in the data. The measured and predicted parameters, residuals and relative contributions of steric and electrostatic coefficients are shown in Table 10. Incorporating a C*log P term with the final model in the CoMFA analysis, yielded a lower q^2 value of 0.470 with three principal components and with a standard error of prediction of 0.271.

CoMFA for the apparent sulfoxidation rate at Mortierella isabellina *based on product-guided alignment (Tables 7 and 11)*

At the default settings including both steric and electrostatic fields, we observed a q^2 of 0.706 and a standard error of prediction of 0.183 with six principal components for *Mortierella isabellina* (Table 7). The final non-cross validated model used six principal components. The model was generated using the following options: both steric and electrostatic fields with 20 kcal mol^{-1} cutoffs, 1/r for the electrostatic function, a 2.0 Å step size, 2.5 kcal mol^{-1} of column filtering, a Csp^{3+} probe atom, and a grid box set at SYBYL's default position. This model had a q^2 of 0.776 with six principal components, a standard error of prediction of 0.159, an R^2 value of 0.984, a standard error of estimate of 0.043 and an F (6,10) value of 101.464. The final cross-validated model utilized 81 of 1280 columns for the analysis. In this model steric features accounted for 55% while electrostatic features accounted for 45% of the variance. Figure 9 shows the relationship between calculated and

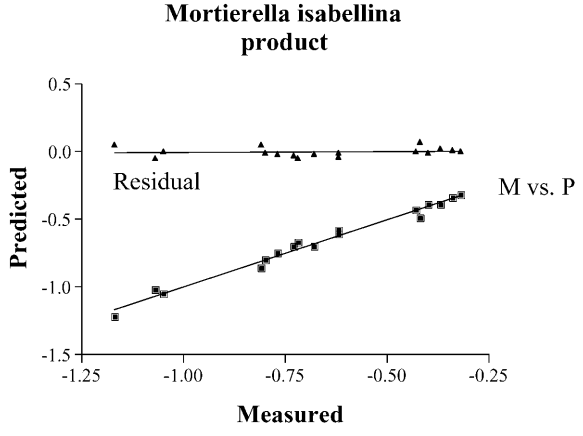


Fig. 9 Relationship between calculated and measured log k* values for the non-cross-validated analysis of the *Mortierella isabellina* based on the product-guided alignment

measured log k* values for the non-cross-validated analysis of the *Mortierella isabellina* based on the product-guided alignment. The measured data, predicted data, residuals and relative contributions of steric and electrostatic coefficients are shown in Table 11. Incorporating a C*log P term in the CoMFA analysis, yielded a lower q^2 value of 0.736 with six principal components and with a standard error of prediction of 0.173.

CoMFA for apparent sulfoxidation rate at Helminthosporium *sp. based on product-guided alignment (Tables 8 and 12)*

Using the default settings, we observed a q^2 of 0.310 and a standard error of prediction of 0.270 with three princi-

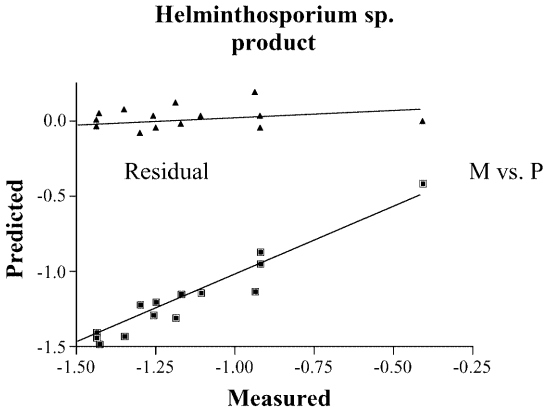


Fig. 10 Relationship between calculated and measured log k* values for the non-cross-validated analysis of the *Helminthosporium* sp. based on the product-guided alignment

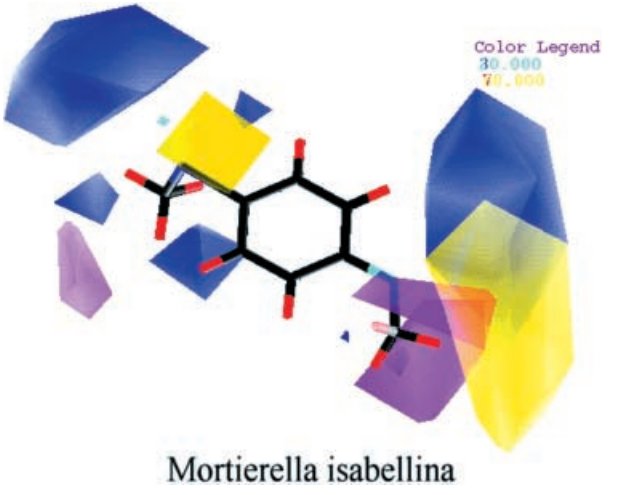


Fig. 11 Contour map for *Mortierella isabellina* based on substrate guided alignment *Yellow* (blue in a non-inverted figure)=positive charge favors high reaction rate; *aqua* (red in a non-inverted figure)=positive charge does not favor high reaction rate; *magenta* (green in a non-inverted figure)=steric bulk favors high reaction rate; *blue* (yellow in a non-inverted figure)=steric bulk does not favor high reaction rate

pal components for *Helminthosporium* sp. (Table 8). The final model incorporated a C*log P term in addition to the CoMFA columns. That model (no cross-validation at three principal components) had the following parameters: steric, 32.5 kcal mol^{-1} electrostatic, 30 kcal mol^{-1}, 1/r, a 2.0 Å step size, 3.7 kcal mol^{-1} of column filtering, a Csp^{3+} probe atom, and a grid box set at (+0.1x, y, –0.3z) of SYBYL's default position. This model had an observed q^2 of 0.477 with three principal components and with a standard error of prediction of 0.235, an R^2 value of 0.898 with a standard error of estimate of 0.118 and with an F (6,10) value of 14.718. The final cross-validated model utilized 53 of 1297 columns for the analysis. In this model steric features accounted for 62% while electrostatic features accounted for 28% of the variance. Figure 10 shows the relationship between calculated and measured log k* values for the non-cross-validated analysis of the *Helminthosporium* sp. based on the product-guided alignment. The measured and predicted parameters, residuals and relative contributions of steric coefficient, electrostatic coefficient, and hydrophobic coefficient are shown in Table 12.

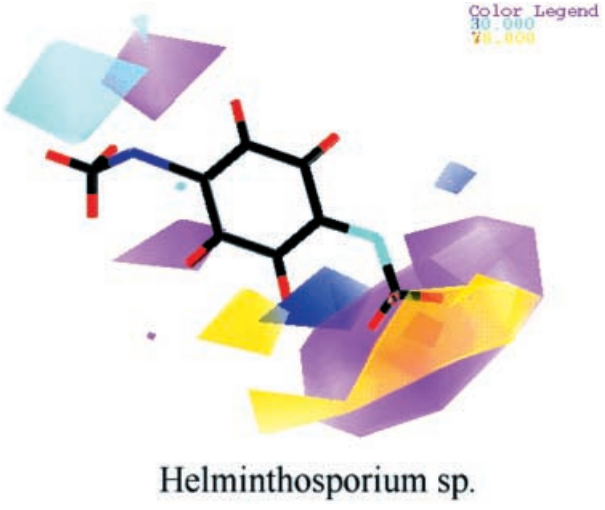


Fig. 12 Contour map for *Helminthosporium* sp. based on substrate-guided alignment

Contour maps

The *STDEV*COEFF* field type was used for all models to plot the contour maps. Figures 11, 12, 13, and 14 show the contour maps of the final CoMFA models for apparent rate of *Mortierella isabellina* and *Helminthosporium* sp. based on both substrate- and product-guided alignments. Both steric and electrostatic fields are utilized for those final models to facilitate comparisons among them.

The contour map for substrate-guided alignment in *Mortierella isabellina* demonstrates prominent electrostatic features near those portions of the training set molecules most capable of hydrogen bonding to the enzyme. In particular two regions in which positive charge favors high substrate affinity for the enzyme are depicted at the two ends of the template molecule. This is entirely consistent with electrostatic features of this model accounting for 62% of the variance of the data. Steric features of the model are more complex, in part because these features can account for a smaller proportion of the experimental data than do the electrostatic features. However, as with the electrostatic features, the steric features of the model are concentrated at the two ends of the template and appear balanced between regions of the active site in which steric bulk favors and is unfavorable for high substrate affinity.

The contour map for substrate-guided alignment in *Helminthosporium* sp shows important similarities and differences with respect to that above for *Mortierella*

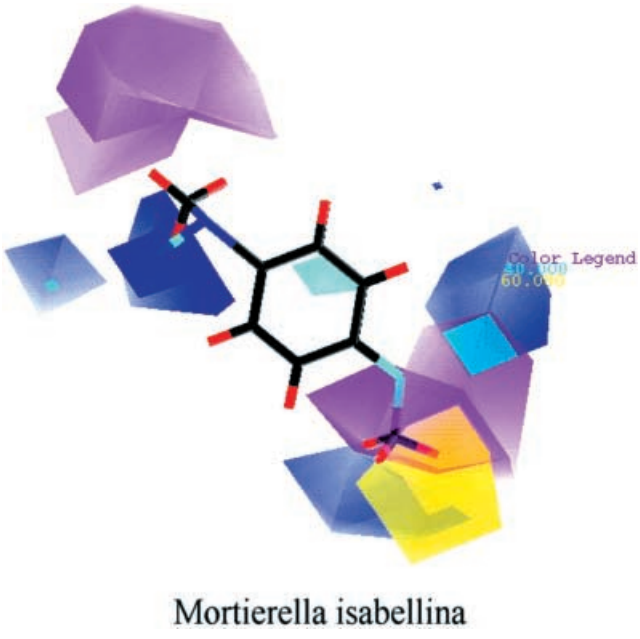


Fig. 13 Contour map for *Mortierella isabellina* based on product-guided alignment

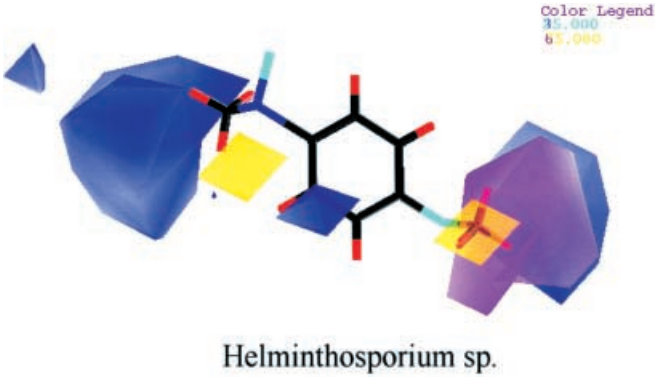


Fig. 14 Contour map for *Helminthosporium* sp. based on product-guided alignment

isabellina. Similarities include the fact that electrostatic features predominate and are localized within the active site near those regions of the template molecule most likely to show hydrogen bonding (the two ends) and away from the aromatic ring. A difference is that there is a marked asymmetry in that partial charge is generally favored at the aryl end of the molecule whereas positive charge is unfavorable at the alkyl end. Steric features are similar but not identical to those of the previous model also. For example, while steric aspects of substrate binding at the two ends of the template predominate there is also a region present in the *H.* sp. model not present in the *M. isabellina* model, namely a region near the aromatic ring in which steric bulk favors high substrate affinity.

The contour map for product-guided alignment in *Mortierella isabellina* demonstrates the predominance of steric features shown by the greater (55%) relative contribution of these aspects of the model to product formation by the enzyme. Generally, the ends of the template molecule remain of greater importance than the center ring in the product model. Steric bulk is favored near the oxygen-containing portion of the template molecule but disfavored near the sulfur-containing portion for products of the sulfoxidation. The importance of positive charge for product affinity remains near the oxygen-containing portion of the template but is lost near the sulfur-containing portion of the molecule.

The contour map for product-guided alignment in *Helminthosporium* sp. shows major differences from the substrate-guided model. Steric features predominate (and account for 62% of the variance). Steric bulk is highly favored near the oxygen-containing portion of the template but quite unfavored near the sulfur-containing portion of the molecule. While there are clear regions of the active site in which positive charge is favored for high affinity of the product, their locations are shifted somewhat from their positions in the substrates.

Table 13 Prediction of cyclopropyl phenyl sulfide and its sulfoxide which were not used in deriving CoMFA models

Apparent rate[a]		*Mortierella isabellina*		*Helminthosporium* sp.	
		S-guided	SO-guided	S[b]-guided	SO[c]-guided[d]
Log k*	Measured	–0.638	–0.699	–1.395	–1.631
	Predicted	–0.630	–0.706	–1.209	–1.635
	Residual	–0.008	0.007	–0.186	0.004

[a] unit: μM/h; [b] S: substrate; [c] SO: product; [d] Measured and predicted (*R*)-sulfoxide

Predictive abilities of the models

We tested the predictive ability of the final models by examining cyclopropyl phenyl sulfide and its corresponding sulfoxides, which were not in the training set. From the sets of conformers after conformational searches, representative ones were selected to fit *i*-PrPS using the FIT ATOMS routine within SYBYL. The predictive results shown in Table 13 were quite good overall. Thus, the residual values for log k* were generally less than 0.009 except for the substrate-guided alignment in *Helminthosporium* sp.

CoMFA model evaluation

Evaluation of the four CoMFA models was done by cross-validation analysis using groups of six compounds. For each model cross-validation was performed 20 times (Tables 14, 15, 16, 17). The mean q^2 for cross-validated models for *Mortierella isabellina* was 0.62 (SEM=0.02; Table 14), while that for product-guided alignment was 0.69 (SEM=0.02; Table 16). The mean q^2 for substrate-guided alignment for *Helminthosporium* sp. was 0.47 (SEM=0.02; Table 15), while that for product-guided

Table 14 Cross-validation of CoMFA model for *Mortierella isabellina* substrate-guided alignment

	q^2	No. of comp.	SEP
1	0.549	3	0.275
2	0.648	3	0.243
3	0.533	3	0.28
4	0.717	3	0.218
5	0.763	3	0.199
6	0.649	5	0.264
7	0.525	3	0.282
8	0.579	3	0.266
9	0.636	5	0.268
10	0.747	3	0.206
11	0.731	3	0.212
12	0.487	3	0.293
13	0.723	3	0.215
14	0.504	2	0.278
15	0.462	3	0.3
16	0.615	3	0.254
17	0.576	2	0.257
18	0.505	6	0.328
19	0.747	3	0.206
20	0.669	3	0.235
mean	0.62	3.25	0.25
SD	0.10	0.97	0.04
SEM	0.02	0.22	0.01

Table 15 Cross-validation of CoMFA model for *Helminthosporium* sp. substrate-guided alignment

	q^2	No. of comp.	SEP
1	0.342	2	0.291
2	0.546	2	0.242
3	0.466	1	0.253
4	0.571	2	0.235
5	0.499	2	0.254
6	0.506	2	0.252
7	0.39	2	0.28
8	0.409	2	0.276
9	0.439	2	0.269
10	0.553	2	0.24
11	0.49	2	0.256
12	0.525	2	0.247
13	0.593	2	0.229
14	0.492	2	0.256
15	0.502	2	0.253
16	0.554	2	0.239
17	0.122	1	0.325
18	0.573	2	0.234
19	0.513	2	0.25
20	0.353	2	0.288
mean	0.47	1.90	0.26
SD	0.11	0.31	0.02
SEM	0.02	0.07	0.01

Table 16 Cross-validation of CoMFA model for *Mortierella isabellina* product-guided alignment

	q^2	No. of comp.	SEP
1	0.8	5	0.143
2	0.633	6	0.204
3	0.685	4	0.172
4	0.772	6	0.161
5	0.764	6	0.163
6	0.694	5	0.177
7	0.321	3	0.243
8	0.726	6	0.176
9	0.742	6	0.171
10	0.685	6	0.189
11	0.785	6	0.156
12	0.613	6	0.209
13	0.712	5	0.172
14	0.752	6	0.167
15	0.739	5	0.164
16	0.721	5	0.169
17	0.592	6	0.215
18	0.618	4	0.19
19	0.778	6	0.158
20	0.737	6	0.172
mean	0.69	5.40	0.18
SD	0.11	0.88	0.02
SEM	0.02	0.20	0.01

Table 17 Cross-validation of CoMFA model for *Helminthosporium* sp. product-guided alignment

	q^2	No. of comp.	SEP
1	0.344	4	0.273
2	0.289	3	0.274
3	0.367	3	0.258
4	0.458	3	0.239
5	–0.553	1	0.377
6	0.174	2	0.284
7	0.444	3	0.242
8	–0.052	3	0.333
9	–0.117	2	0.331
10	0.336	3	0.264
11	0.417	3	0.248
12	0.365	3	0.259
13	0.322	3	0.267
14	0.371	3	0.257
15	0.125	2	0.292
16	0.341	3	0.263
17	0.33	3	0.26
18	0.417	3	0.248
19	0.317	3	0.268
20	0.094	2	0.298
mean	0.24[a]	2.75	0.28
SD	0.24	0.64	0.03
SEM	0.05	0.14	0.01

[a] The lower predictive q^2 showed that the sequential oxidation of sulfides to sulfone products by *H.* sp. is more difficult compared to these same processes in *M. isabellina*

alignment was 0.24 (SEM=0.05, Table 17). These results confirm that the CoMFA models described above are likely to be of some predictive value when extended to other related compounds. Furthermore, the higher predictive validity for the *Mortierella isabellina* models in comparison to those for *Helminthosporium* sp. suggests a more consistent relationship between ligand–protein associations within the former organism. The lower predictions for the product-guided alignment for *Helminthosporium* sp. suggests that somewhat different structure–activity relationships may hold for subsets of the training set compounds.

Discussion

In the present study, we examined the requirements for S-oxygenation by *Mortierella isabellina* and *Helminthosporium* sp. and determined the stereochemical outcome of sulfoxide formation to derive CoMFA models. This was done to complement mechanism-based or "cubic-space" models for understanding substrate–enzyme interactions. We also comparatively evaluated both substrate- (sulfide) and product- (sulfoxide) guided alignments for distinct CoMFA models. Based on the flexibility of alkyl aryl sulfides or the corresponding sulfoxides, we chose the lowest energy conformers as a starting point in processing alignments.

The results demonstrated that CH_3COTA and NH_2TA are crucial in defining the model, even though they exhibit a slow rate of S-oxygenation and undergo other routes of metabolism. CH_3COTA and NH_2TA adopt an opposite orientation compared with the rest of the sulfides and sulfoxides in the training set. This was because CH_3COTA exhibited a greater degree of ketone reduction than heteroatom-oxygenation. Also, acetylation of NH_2TA is slightly faster than S-oxygenation. NH_2TA was aligned only for *Mortierella isabellina* and NH_3TA (protonated form) was aligned only for *Helminthosporium* sp. via trial and error. Thus, there are some exceptions (e.g. CH_3COTA and NH_2TA) to the uniform alignment rule generated by the training set as described in the "Results" section. After successful alignment of the compounds, we could not use flexible field-fit procedures to further optimize the training set alignment because no suitable template was identifiable and the training set was too flexible to refine further.

CoMFA for substrate-guided alignment

Our findings with the CoMFA models derived from substrate-guided alignment showed that the enzymes of *Mortierella isabellina* and *Helminthosporium* sp. have similar geometric features associated with the aryl moiety but show opposite geometric maps for the alkyl group Fig. 11 versus Fig. 12. Our results suggest that the enzyme or enzymes of *Mortierella isabellina* favor heteroatom-oxygenation from direction B of Oae [21] resulting in formation of the (*R*)-enantiomer as the dominant product. In contrast, the enzymes of *Helminthosporium* sp. prefer direction A for heteroatom-oxygenation resulting in (*S*)-sulfoxide formation. This visualization of contour results is consistent with the directional model originally proposed by Oae. [21] That, in turn, is similar to the model known as Prelog's rule. [22]

Enzymes within both *Mortierella isabellina* and *Helminthosporium* sp. show some level of steric hindrance associated with the *meta* position of the aryl ring, but on opposite sides of the aromatic ring (Fig. 11 versus Fig. 12). These figures show that positive charge is favored at the *para* position of the aryl ring for *Mortierella isabellina,* and at both *para* and *meta* positions of the aryl moiety for *Helminthosporium* sp. Positive charge is not favored around the alkyl group for *Helminthosporium* sp. These contour results imply that the S-oxygenation for both *Mortierella isabellina* and *Helminthosporium* sp. favors electron-donating (induction or resonance) groups (i.e. $-OCH_3$, –Br, –Cl, $-CH_3$ etc.) at the *para* position of the aryl ring, and does not favor electron-withdrawing groups (i.e. $-NO_2$, –CN, –F etc.) in the *para* position. These contour results confirm our rate data in the present study, and are consistent with the Hammett plots previously developed for enzyme–substrate interactions in these organisms (Huang and Davis, in preparation). The Hammett plot results were σ^+=–0.42 for *M. isabellina* and σ^+=–0.36 for *Helminthosporium* sp. Thus, the enzymatic reactions in these organisms probably involve a radical cation "sulfenium" intermediate. Also, the enzymatic reactions are enhanced by an electron-donating group and retarded by an electron-withdrawing group. Above the S atom, the electrostatic map shows the field favoring a positive charge that implies that S-oxygenation favors the attack by the activated "oxenoid" complex associated with the enzyme (Fig. 11).

CoMFA based on product-guided alignment

Figures 12 and 13 show contour maps based on the product-guided alignment for *Mortierella isabellina* and *Helminthosporium* sp., respectively. Compared to the substrate-guided maps, the product-guided maps show similar steric and electrostatic features associated with the aryl ring, but the positive charge is less of a factor. The major difference was in the steric field around the alkyl moiety and was opposite to what was observed versus the substrate-guided maps. The steric geometry is also opposite in *Mortierella isabellina* versus *Helminthosporium* sp. (Figs. 11, 12, 13, and 14). These results show that the formation of the (*R*)-enantiomer is dominant in *Mortierella isabellina*. The results also show that the (*S*)-enantiomer is dominant in *Helminthosporium* sp. Furthermore, *Mortierella isabellina* favors a second S-oxygenation from direction A to yield the achiral sulfones. However, *Helminthosporium* sp. allows the oxidation from direction B to yield the sulfone products. [21] Nevertheless, there were rare sulfones produced by *Helminthosporium* sp. in laboratory work.

In the substrate-guided models, the results (Tables 9 and 10) show that the electrostatic feature (a 62% contribution for *Mortierella isabellina* and 57% contribution for *Helminthosporium* sp.) is more important than the steric field. The product-guided results (Tables 11 and 12) reveal a contrasting effect. Here, the steric coefficient (55% contribution for *Mortierella isabellina* and 62% contribution for *Helminthosporium* sp.) is more dominant than the electrostatic coefficient. An additional hydrophobic contribution (C*log P; 14%) for *Helminthosporium* sp. may account for the difficulty in the further conversion of a sulfoxide to an achiral sulfone.

Both the substrate-guided and product-guided models are in agreement in explaining the stereochemical outcome and reaction requirements for each of the organisms. We suggest that geometric (steric) and/or electrostatic features of the enzymes are involved in the regio- and stereo-specificities of the reaction, while electrostatic features control the rate of reaction.

Our CoMFA results (Tables 5, 6, 7, and 11) show moderate q^2 values (0.720 and 0.776) for *Mortierella isabellina* and lower q^2 values (0.578 and 0.477) for *Helminthosporium* sp (in CoMFA, a q^2 value greater than 0.5 is usually considered acceptable). However, the overall results are significant enough to explain the stereochemical outcome, mechanism and reaction rate based on both steric and electrostatic contributions for *Mortierella isabellina* and *Helminthosporium* sp. The predictive abilities of the final CoMFA models (Table 13) demonstrate that the CoMFA is a reliable and powerful tool in the application of QSAR for the study of sulfoxidation reactions.

We made a decision to use rate data rather than chemical yield at a particular point in time as the dependent variable in the CoMFA models. This is because such point data may not fully represent the dynamics of metabolism in the fungal systems under study. For instance, the highest apparent rate for a substrate (CH_3OTA) among the 17 aligned compounds at microbial sulfoxidation does not yield the highest chemical yield at 48 h. This is because the product is further oxidized to the sulfone. This was confirmed by quite low q^2 values for CoMFA models based on chemical yield (data not shown).

Acknowledgements *Helminthosporium* sp. NRRL 4671 was kindly provided by the Northern Regional Research Laboratories, U.S. Department of Agriculture.

References

1. Faber K (1997) Biotransformations in organic chemistry: a textbook. Springer, Berlin Heidelberg New York, pp 11–85
2. Faber K (1997) Biotransformations in organic chemistry: a textbook. Springer, Berlin Heidelberg New York, p 87
3. Beecher J, Willetts A (1998) Tetrahedron: Asymmetry 9:1899–916
4. Mattos C, Ringe D (1993) Multiple binding modes. In: Kubinyi H (ed) 3D QSAR in drug design: theory, methods, and applications. ESCOM, Leiden, pp 226–254
5. SYBYL (1999) Molecular modeling system (version 6.6). TRIPOS Associates, St. Louis, Mo.
6. Cramer III RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
7. Faber K, Griengl H, Honig H, Zuegg J (1994) Biocatalysis 9:227–239
8. Botta M, Cernia E, Correlli F, Manetti F, Soro S (1996) Biochim Biophys Acta 1296:121–126
9. Holland HL, Brown FM, Lakshmariah G, Larsen BG, Patel M (1997) Tetrahedron: Asymmetry 8:683–697
10. Holland HL (1988) Chem Rev 88:473–485
11. Poulos TL, Finzel BC, Gunslus ILC, Wagner GC, Kraut J (1985) J Bio Chem 260:16122
12. CONCORD (1999) User's manual version 3.0.1. TRIPOS Associates, St. Louis, Mo.
13. Wilcox RE, Huang WH, Brusniak KMY, Wilcox DM, Pearlman RS, et al (2000) J Med Chem 43:3005–19
14. Brusniak MY, Pearlman RS, Neve KA, Wilcox RE (1996) J Med Chem 39:850–9
15. Wilcox RE, Tseng T, Brusniak MY, Ginsburg B, Pearlman RS, et al (1998) J Med Chem 41:4385–99
16. Wilcox RE, Ragan JE, Eglen RM, Bonhaus DW, Miller JD (2000) High-affinity agonist interactions at recombinant 5HT7 receptors – estimates using CoMFA (comparative molecular field analysis). Soc Neurosci Abstr 26:2157
17. Wold S, Johnsson E, Cocchi M (1993) PLS-Partial least-squares projections to latent structures. In: Kubinyi H (ed) 3D QSAR in drug design: theory, methods, and applications. ESCOM, Leiden, pp 523–550
18. Marshall GR, Cramer III RD (1988) Trends Pharmacol Sci 9:285–9
19. Yang W, Davis P (1991) Drug Metab Dispos 20:38–46
20. TRIPOS (1999) SYBYL user's manual, MedChem software 3.6. St. Louis, Mo.
21. Oae S, Mikami A, Matsura T, Ogawa-Asada K, Watanabe Y, et al. (1985) Biochem Biophys Res Commun 131:567
22. Takata T, Yamazaki M, Fujimori K, Kim YH, Oae S, Iyanagi T (1980) Chem Lett 1441–4