Available online at www.sciencedirect.com
SCIENCE DIRECT®
Tetrahedron: Asymmetry
Tetrahedron: *Asymmetry* 15 (2004) 1163–1172

Stereoselectivity of *Burkholderia cepacia* lipase towards secondary alcohols: molecular modelling and 3D QSAR approach

Sanja Tomić,[a,*] Branimir Bertoša,[a] Biserka Kojić-Prodić[a] and Istvan Kolosvary[b]

[a]*Ruđer Bošković Institute, PO Box 180, HR-10002 Zagreb, Croatia*
[b]*Budapest University of Technology, Division of Chemical Information Technology, Gellért tér 4., Budapest H-1111, Hungary*

Received 21 January 2004; accepted 17 February 2004

Abstract—Lipase from *Burkholderia cepacia* (BCL, previously *Pseudomonas cepacia*) has proven to be a useful biocatalyst for obtaining enantiomerically pure compounds, in particular primary and secondary alcohols and their esters. Previously we derived 3D quantitative structure–activity relationships (QSARs) for predicting the enantioselectivity of BCL towards 3-(aryloxy)-1,2-propanediol derivatives. Herein we used these models to predict the stereoselectivity for a new set of diverse secondary alcohols. Despite significant differences in the data sets and the molecular modelling procedure utilised, the stereoselectivity was predicted in agreement with the results of kinetic resolution.We have combined the lipase–substrate complexes that we analysed earlier with those presented herein and derived a new model. Since this model is derived for the chemically diverse set of secondary alcohols, it is more general than those obtained previously for the 12 BCL-3-(aryloxy)-1,2-propanediols complexes and should be able to predict the stereoselectivity of BCL towards a wide range of secondary alcohols. This is demonstrated for two α-methylene-β-hydroxy esters. Additionally, we modelled a mutant BCL and, using the COMparative BINding Energy (COMBINE) analysis, predicted its stereoselectivity towards a few randomly selected secondary alcohols. The Monte Carlo based conformational search, used herein, turned out to be faster and more convenient for investigating the binding modes of secondary alcohols in the BCL active site than the systematic search used in our previous work.

1. Introduction

Syntheses and resolutions catalysed by microorganism enzymes are among the best methods for the preparation of enantiomerically pure compounds.[1–4] The most versatile biocatalysts are microbial lipases (EC 3.1.1.3, glycerol ester hydrolases) that carry out a wide range of bioconversions such as hydrolysis, interesterification, esterification, alcoholysis, acidolysis and aminolysis.[5–7] Therefore, their applications in the syntheses of enantiomerically pure products for the pharmaceutical and agrochemical industry are numerous. However, despite the large amount of experimental data available for key enzyme properties utilised in biocatalysis, enantioselectivity is still not completely understood. There are ongoing efforts to optimise the output of lipase-catalysed reactions[6,7] and better understanding of lipase enantioselectivity would be valuable. Therefore, we focused our research on theoretical approaches that should help in the understanding of the mechanism of enzyme stereoselectivity.

Lipase from *Burkholderia cepacia* (formerly *Pseudomonas cepacia*) has proven to be useful as a biocatalyst for obtaining enantiomerically pure compounds, in particular primary and secondary alcohols and their esters.[1,2,8,9] The recently determined crystal structure of the complex of *Burkholderia cepacia* lipase (BCL) with the secondary alcohol like inhibitor[10] enabled more reliable molecular modelling studies that have aided in understanding of lipase-catalysed reactions.

Herein is a continuation of our experimental and computational studies of the BCL complexes and the lipase catalysed acylation of primary[11] and secondary alcohols.[10,12] In our previous study, we derived several 3D QSAR COMBINE models for predicting the enantioselectivity of BCL towards 3-(aryloxy)-1,2-propanediols.[12] For this purpose, we approximated the free energy of the tetrahedral intermediate formation with a sum of weighted residue based energy terms (Δu_i):

* Corresponding author. Tel.: +385-1-4561025; fax: +385-1-4680245; e-mail: sanja.tomic@irb.hr

0957-4166/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetasy.2004.02.016

$$\Delta G = \sum_{i=1}^{n} w_i \Delta u_i^{\mathrm{sel}} + C \quad (1)$$

The free energy difference between the two stereoisomeric lipase–substrate tetrahedral intermediate complexes, formed with the two substrate enantiomers, is proportional to the ability of a lipase to stereoselectively catalyse esterification and hydrolysis of secondary alcohols and their esters, respectively, described by enantiomeric ratio, E,

$$\Delta\Delta G = -RT\{\ln E\} \quad (2)$$

where

$$E = (k_{\mathrm{cat}}/K_{\mathrm{M}})_{\text{fast reacting}}/(k_{\mathrm{cat}}/K_{\mathrm{M}})_{\text{slow reacting}} \quad (3)$$

We used the chemometric analysis to correlate the calculated energy term differences ($\Delta\Delta u_i$) with the results of kinetic measurements, expressed as an enantiomeric ratio E, and in this way calculate the contribution of each energy term to the binding free energy, w_i. The weights w_i determined in this way are models for predicting enantioselectivity of BCL towards 3-(aryloxy)-1,2-propanediols.

Herein, we extended our investigations to the more chemically diverse set of secondary alcohols: racemic compounds and two diastereomeric mixtures. We used a few QSAR models, with the best external predictive performances that we derived previously to predict selectivity of BCL to catalyse esterification of these compounds. Furthermore, using an extended set of the chemically diverse secondary alcohols, we derived a more robust model for predicting the stereoselectivity of BCL towards secondary alcohols.

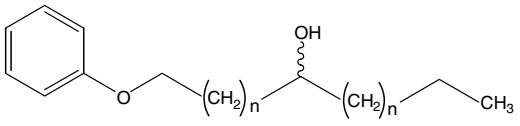
In our previous work, we utilised a systematic conformational search combined with molecular dynamic simulations to reveal the possible binding modes of 3-(aryloxy)-1,2-propanediols into the BCL active site. Herein, the conformational search was accomplished by Monte Carlo based methods. The purpose of the present analysis is to compare the outputs of the two different molecular modelling approaches utilised to investigate possible modes of binding of diverse secondary alcohols into the active site of the acylated BCL, and to check the ability of the earlier derived 3D QSAR COMBINE models to predict enantioselectivity of BCL towards the newly analysed set of secondary alcohols.

The goal of the COMBINE analysis was the identification of the amino acid residues that determine the stereoselectivity of BCL towards secondary alcohols and their esters. These findings should enable rational modification of the lipase in order to optimise output of the catalysed reactions. The set of the amino acid residues that had the highest weights in models derived for 3-(aryloxy)-1,2-propanediols is compared with the one that the present study revealed. Reappearance of amino acid residues in different models confirmed their importance for BCL resolution of secondary alcohols. For comparison with the recent investigations of Guieysse et al.[13] we modelled BCL with three point mutations at positions 221, 266 and 287 and used in the COMBINE analysis to predict the stereoselectivity of the mutant towards four randomly selected secondary alcohols.

2. Materials and methods

2.1. Compounds considered in the study

For the racemic compounds **N1**–**N4** (Scheme 1) and the diastereomeric mixtures **N5** (C5-*S*, C5-*R*) and **N6** (C7-*S*, C7-*R*) (Scheme 2), stereoselectivity of acylation catalysed by BCL was measured by Ljubović and Šunjić.[14,15]

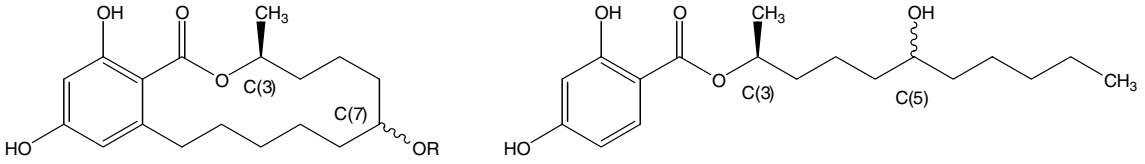


Scheme 1. Racemic secondary alcohols **N1** ($n = 0$, $E > 200$), **N2** ($n = 1$, $E = 1.5$), **N3** ($n = 2$, $E = 3.8$), **N4** ($n = 3$, $E = 2.4$).[14]

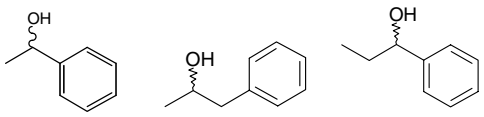
They found that BCL predominantly acetylated the (*R*)-enantiomer of the racemic (1:1) mixtures of **N1**–**N4**. When applied to diastereomeric mixtures, **N5** and **N6**, BCL predominantly acetylated acyclic diastereomer (3*S*,5*R*)-**N5**, and cyclic diastereomer (3*S*,7*S*)-**N6**.

For compounds **N7**–**N9** (Scheme 3), measurements were performed by Kazalauskas et al.[16] They found that BCL predominantly catalysed the hydrolysis of the (*R*)-enantiomer of the racemic (1:1) mixture of esters of the secondary alcohols **N7**–**N9**. This reaction was reversible to the esterification of these alcohols. It is justifiable to use the E values determined for the BCL catalysed esterification to study the BCL catalysed hydrolysis of the formed esters and vice versa.

Recently, Nascimento et al.[17] measured the resolution of two α-methylene-β-hydroxy esters by *Pseudomonas* sp. lipases, closely related to BCL. We built their tetrahedral complexes with BCL and used the COMBINE analysis to predict the enantiomeric ratios.



Scheme 2. Diastereomeric mixture of the cyclic secondary alcohols **N6** (3*S*) ($D = 60$) and their acyclic precursors **N5** (3*S*) ($D = 2.4$).[15]



Scheme 3. Racemic secondary alcohols **N7** ($E > 50$), **N8** ($E > 50$) and **N9** ($E > 50$).[16]

2.2. Modelling of the substrates and their complexes with the Ser87-acetylated BCL

The starting structure for the modelling of the lipase–substrate tetrahedral intermediate complexes was the crystal structure of the enzyme-inhibitor complex taken from the Protein Data Bank (PDB)[18] accession code 1HQD.[10] The bound substrates were built using the inhibitor in 1HQD as a template in which the $P(CH_3)O^-$ group covalently bound to Ser87, was replaced by the $C(CH_3)O^-$ group. Hydrogens were added, to correspond to pH 7.0, using the Biopolymer module in the program INSIGHTII.[19] Histidines were uncharged (mono-protonated), aspartic and glutamic acids were negatively charged, and arginines and lysines were positively charged. The catalytic His286 was in the doubly protonated form, according to the assumed reaction path. Parameterisation was performed in the all atom AMBER force field[20] (for details of the parameterisation see our previous work[12]).

All the water molecules found in the BCL-inhibitor complex 1HQD, (altogether 291 molecules), were included in the molecular modelling procedure, while the bulk water influence was modelled by a distance dependent dielectric constant.

2.3. Determination of the possible conformations of a bound substrate

Geometry optimisation and the conformational search were performed by the program MACROMODEL.[21] For each enzyme–substrate **N1**–**N9** [both (*R*)- and (*S*)-enantiomers; and in the case of diastereomeric mixtures **N5** and **N6** two compounds, which have opposite configuration at the C5 and the C7 atoms, respectively] complex, a conformational search was performed in the following way: all amino acid residues within a sphere of approximately 15 Å around a substrate molecule with the water molecules found in the 1HQD were free to move during the conformational search, while the rest of the protein atoms were fixed using frozen atoms mode. The conformational search was performed by two algorithms: The Monte Carlo Multiple Minimum (MCMM) and Low Mode Conformational Search (LMCS).[22] Both algorithms proved to be equally efficient that is, in each case 2000 MC steps were enough for the complete investigation of the conformational space of the bound substrate.

All conformers within 6 kcal/mol from the global minima were visually analysed and those containing H-bonds essential for the tetrahedral intermediate (TI): Oδ2(Asp264)···Hδ1(His286), Hε(His286)···Oγ(Ser87), Hε(His286)···O(alcohol) and hydrogen bonds between oxyanion (O^-) and protonated amide nitrogens of Leu17 and Gln88 were considered in further analysis.

The most stable lipase–substrate TI complexes determined for (*S*)- and (*R*)-enantiomers of each substrate **N1**–**N4**, **N7**–**N9** and the two mixtures of diastereomers **N5** and **N6** were subjected to the COMBINE analysis.

In order to compare the most favourable conformation of a substrate with the conformation of its tetrahedral intermediate bound into the BCL active site, a conformational search was performed for unbound substrates as well, in vacuum and chloroform. For this purpose, we used the MM3* force field,[23] distance dependent dielectric constant and generalised Born-solvent accessible surface area (GB/SA) continuum solvent model[24] as implemented in the MACROMODEL program. Chloroform was used to mimic different organic solutions used in the experiments.[14–16]

The tetrahedral complexes with **N10** and **N11** were built using the molecular modelling procedure consisting of the systematic search and molecular dynamic simulation as described in our previous article.[12]

The tetrahedral complexes of the mutated BCL (Val266 → Leu, Leu287 → Ile, Phe221 → Leu) and the compounds **N1**, **N2**, **N3**, **N7** and **N11** were built from the complexes with the native protein. The mutations were performed in the following way: The side chains of Val266, Leu287 and Phe221 were replaced by the new ones, Leu, Ile and Leu, respectively, and the mutated complexes were reoptimised until a convergence of 0.01 kcal/mol was reached.

2.4. COMBINE analysis

2.4.1. Energy terms considered in the COMBINE analysis. For the purpose of being able to apply the 3D QSAR COMBINE models that we derived in our previous work[12] to predict BCL-enantioselectivity for the new set of secondary alcohols, we calculated the same energy terms (see Eq. 1) as before. Therefore, the binding free energy was (as in the previous work)[12] approximated by the residue based interaction energy terms calculated between substrate subunits and protein, substrate and water and different substrate subunits. Each substrate was, for this purpose, divided into five subunits, groups, of which the most important for resolution were large (L) and medium (M) substituents at the alcohol stereogenic centre (see earlier work[12] for details).

$$\Delta\Delta G = \sum_{i=1}^{nSUB}\sum_{j=1}^{nENZ} w_{ij}^{vdW}\Delta\Delta u_{ij}^{vdW} + \sum_{i=1}^{nSUB}\sum_{j=1}^{nENZ} w_{ij}^{ele}\Delta\Delta u_{ij}^{ele} - \sum_{k=1}^{nSUB+nENZ}\left(w_k^{p}\Delta\Delta SA_k^{p} + w_k^{np}\Delta\Delta SA_k^{np}\right) \quad (4)$$

where u_{ij} is the energy of interaction (van der Waals u_{ij}^{vdW} and electrostatic u_{ij}^{ele}) between the amino acid residue (or water molecule) j and the substrate subgroup i and SA_k (polar, SA_k^p and nonpolar, SA_k^{np}) is the solvent accessible surface area for either an amino acid residue ($k = j$) or a substrate group ($k = i$). The solvent accessible surface area per residue and substrate group was calculated using the NACCESS 2.1.1 program.[25] Decomposition of the interaction energy to the residue-based terms was performed by the DISCOVER program.[26] A separate program was written to extract the calculated energy terms from the DISCOVER and NACCESS outputs and to generate the energy matrix of nine rows and, in the case when the interactions among the water molecules and substrate groups as well as those among substrate groups were not considered, $5\times320\times2+325\times2=3860$ terms in a row. In the case when the energy terms describing the interaction between the water molecules and substrate groups and different substrate groups, in addition, each row has 6120 terms and 6140 terms, respectively. In the model where the interaction with the water molecules is included, the surface area terms are excluded. As an additional term, the conformational energy difference between the two tetrahedral complexes was also included in some models.

2.4.2. Chemometric analysis. The chemometric analysis was performed by the GOLPE program.[27] The set of substrates considered herein presents a test set for the earlier derived models.[12] Further tests on the new models were derived for the 24 compounds, those considered before and during this work. Prior to statistical analysis, the X-variables were pretreated using different procedures: Zeroing (X-variables with the absolute value below 0.01 were excluded from the analysis), minimum standard deviation cut-off (X-variables with the standard deviation below 0.01 were excluded from the analysis) and block unscaled weights (BUW). In the present study, as in the previous one[12] the X-variables were divided into blocks of variables: The van der Waals block, the electrostatic block, and the solvent accessible surface area block. BUW pretreatment was used to give the same relevance to each block of variables. Such treatment of X-variables was based on our experience gained in the COMBINE based study of protein–DNA interactions.[28,29]

PLS analysis[30] was performed to derive 3D QSAR models. The quality of the models is determined by the correlation between the calculated and experimental $\log(E)$ values, R^2, and by its predictive abilities. For internal validation, 'leave one out' (LOO) cross-validation was performed. The predictive ability of a model is established by the correlation between the experimental and predicted enantioselectivity, Q^2, and with the standard deviation of error of prediction (SDEP):

$$\mathrm{SDEP} = \left[\sum_i \frac{(Y_i - Y_i')^2}{N}\right]^{1/2} \quad (5)$$

Table 1. The calculated lipase–substrate interaction energies in the tetrahedral complexes with the fast and slow reacting secondary alcohol enantiomer and the enantiomeric ratio determined from the experiments

Compound	Energy (kcal/mol)		$E(D)$
	Slow	Fast	
N1	−29.7	−25.7	>200[a]
N2	−31.9	−28.2	1.5[a]
N3	−28.3	−33.5	3.7[a]
N4	−35.7	−36.0	2.4[a]
N5	−32.7	−36.0	2.3[b]
N6	−34.0	−35.6	64[b]
N7	−26.8	−23.4	>50[c]
N8	−26.3	−26.8	>50[c]
N9	−28.0	−25.5	>50[c]

[a] Experimentally determined E values.[14]
[b] D values determined for the diastereomeric mixtures **N5** (C5-*S*, C5-*R*) and **N6** (C7-*S*, C7-*R*).[15]
[c] Experimentally determined E values.[16]

Y_i is experimental and Y_i' is predicted lipase stereoselectivity value, either $\log(E)$ or $\log(D)$ (Table 1).

In order to improve predictive abilities of the models and to determine the most important X-variables, selection of variables was performed by fractional factorial design (FFD).[31]

Regarding the type of X-variables used in chemometric analysis, two basic types of models were derived: (a) I-models (only interaction energy terms are included), and (b) ISA model (both interaction energy and the solvent accessible surface area terms are included). Regarding the interaction terms considered, three different I-models were derived: I0 model (only enzyme–substrate tetrahedral intermediate (TI) interaction energy terms included), I0+ model, in addition to I0 terms the substrate–water interaction energy terms are included, and the I0++ model, including the I0+ terms and nonbonded substrate–substrate interaction energy terms.

3. Results and discussion

3.1. Binding modes identification

The main predisposition for successful 3D QSAR analysis is a reliable structure of the enzyme–substrate complex. This problem is also present in the determination of the binding mode of a secondary alcohol tetrahedral intermediate (TI) into the lipase active site. Conformational searches obtained by MCMM and LMCS algorithms enabled us to locate the global minimum and other important, energetically favourable structures for stereo isomeric complexes between BCL and a substrate.

Visual examination of the structures indicated several possible orientations of the substrate molecule in the enzyme active site while these orientations differed from

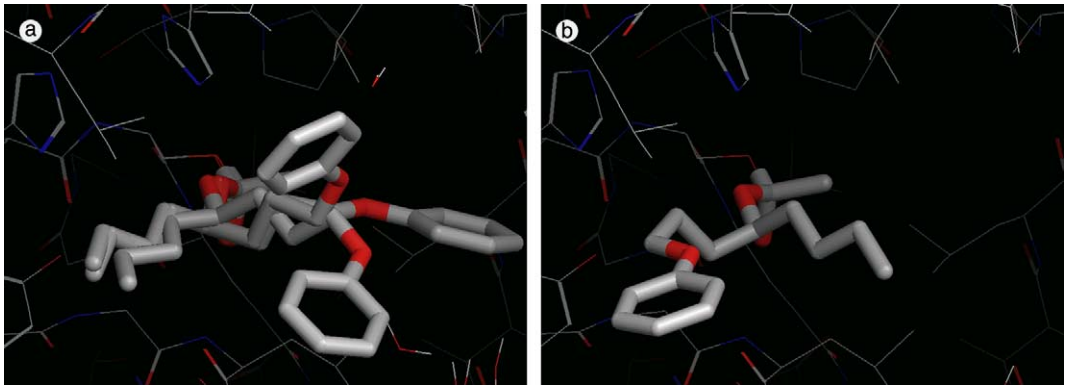


Figure 1. Complex between BCL (Ser87 acetylated) and the substrate **N3** tetrahedral intermediate: (a) for (*R*)-**N3** three low energy tetrahedral intermediate binding modes were found, (b) BCL–(*S*)-**N3** tetrahedral intermediate complex, only one low-energy binding mode was determined.

substrate to substrate. However, orientations of the two stereoisomeric tetrahedral intermediates in the lipase active site significantly changed in all cases. Similarly as determined by the systematic search,[12] the Monte Carlo based methods also revealed binding of two enantiomers into the enzyme active site with the different orientation of their substituents at the alcohol stereogenic centre. While in the case of the slow reacting stereoisomer the large substituent (L) is mostly buried into the partly hydrophilic, the so called HH pocket;[32] in the case of the fast reacting stereoisomer, it either accommodates into the large hydrophobic HA pocket, or points out of the lipase binding site (Fig. 1).

There is also a difference in the number of the low energy (threshold of 6 kcal/mol was used) TIs. In most cases there is only one energetically favourable tetrahedral intermediate orientation of the slow reacting stereoisomer, with the exception of **N5** (C5-*S*) and the relatively small secondary alcohols **N7**–**N9**. On the other hand, for the fast reacting stereoisomer at least two energetically similarly, or in the case of **N5** (C5-*R*) even six, TI orientations were found (see e.g., Fig. 1).

3.2. Conformations of the free substrates

Free substrates were modelled in all the conformations found in the lipase and were subjected to molecular mechanic calculations. In most cases, the conformation that corresponded to the global minimum of the complex enzyme–substrate, was the most stable one.

3.3. Energies of the substrate tetrahedral intermediates–lipase complexes

Lipase enantioselectivity is exponentially proportional to the binding free energy difference of the two stereoisomeric substrate TIs–lipase complexes ($\Delta\Delta G$, see Introduction). The enthalpic part of the binding free energy difference ($\Delta G = \Delta H - T\Delta S$) might be approximated by the potential energy difference of these complexes, and the entropic part by the solvation free energy and the rotation entropy differences. However, evaluation of these terms is an extremely demanding task, mostly because of the complexity of the experimental conditions. The activity of the lipases significantly depends on the organic solvent and additives, but what precisely is their distribution around the active site is hard to guess.

As observed before,[12,13,33–36] the prediction of the stereoselectivity of the lipase towards secondary alcohols is not possible on the basis of the conformational energies of the tetrahedral complexes. Hult and co-workers[33,34] used the truncated potential energy of the enzyme–substrate complexes to distinguish the fast reacting enantiomer from the slow one. However they were not able to quantitatively predict the enantiomeric ratio. The potential energy of big systems, such as lipase–substrate–water complexes, is very sensitive value since a small (for practical purposes insignificant) change in conformation could lead to a large change (even more than 10 kcal/mol) in potential energy. On the other hand, substrate protein interactions directly reflect the orientation of the substrate in the binding pocket. Although the interaction energy measures the strength of a substrate protein binding, it does not necessarily reflect the conversion ratio. In contrast, we noticed that the interaction energy was lower in complexes with a slow reacting enantiomer than in those with a fast reacting one (see Table 1 for the energies determined for the global minima of the substrate TI–lipase complexes). This may be indicative of the different speeds of the reaction for two stereoisomers, and is in accord with the assumption that after the cleaving of the ester-catalytic serine bond the ester is held in close proximity to the active site greatly favouring the reformation of the cleaved bond.[36]

The rotational entropy of the system could be roughly correlated with the number of substrate orientations in the active site. For fast reacting stereoisomers, the larger number of the orientations, in comparison to slow reacting ones, could be one of the reasons for their faster catalytic reaction. Also, the number of productive binding modes versus the number of low energy non-productive binding modes for the slow reacting enantiomer is lower than those for the fast reacting one;

apparently, the slow reacting enantiomer has a larger probability to be trapped in the 'wrong' orientation.

3.4. Quantitative predicting of enantioselectivity

The 3D QSAR COMBINE approach is based upon the assumption that the free energy difference between the two stereoisomeric lipase–substrate TI complexes can be approximated by weighted amino acid residue based energy differences[28,29,37,38] (see Eq. 1). These sums are then correlated with the enantioselectivity of lipase, expressed as an enantiomeric ratio E (Eq. 2).

In our previous study we derived a few 3D QSAR models for predicting the enantioselectivity of BCL towards 3-(aryloxy)-1,2-propanediols.[12] In this work we used some of these models, to predict differences in the stereoselectivity of BCL towards the eight racemic phenyl secondary alcohols with a variable alkyl chain length: **N1**–**N4**, **N7**–**N9** and two diastereomeric mixtures **N6** (C7-*S*, C7-*R*,*S*) and **N5** (C5-*S*, C5-*R*,*S*) as well as two alkyl secondary alcohols, **N10** and **N11**. Since in the previous work[12] we showed the importance of including water molecules in the optimisation and in the 3D QSAR model construction procedure, the chemometric analysis presented herein is performed only on the complexes optimised with the water molecules. Three types of COMBINE models, differing in the type of interactions included (see Methods), were considered for this purpose: I0, I0+ and I0++ models. The fit and dimensionality of the model, for which the best prediction was achieved, are summarised in Table 2.

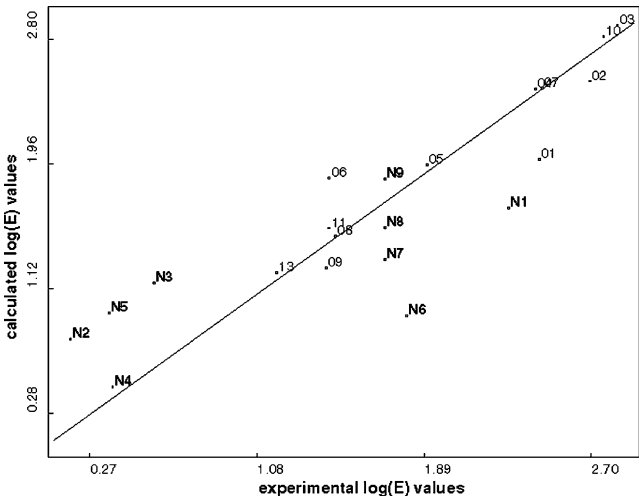


Figure 2. Predicted versus measured enantioselectivity, log(E) values, for 9-phenyl-substituted secondary alcohols **N1**–**N9** determined using the I0+ model derived earlier for 12 diols.[12]

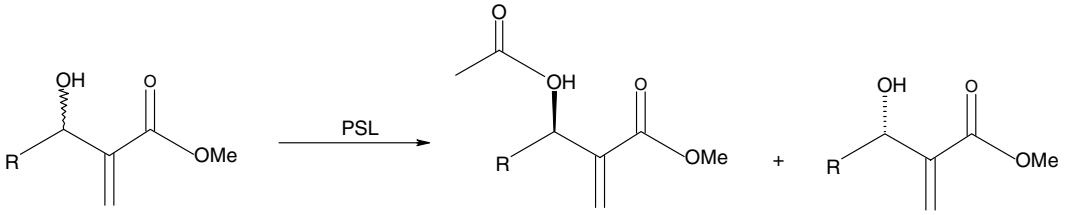
The enantiomeric ratios (E) for compounds **N2**–**N5** were out of range of the E values used to derive the QSAR models. However, for most of them, enantioselectivity was quite well predicted, especially with the I0+ model (Fig. 2).

In the case of the prediction obtained using the other models, **N2** and **N5** are outliers that is, their log(E) is by the 3D QSAR models significantly overestimated (see e.g., Fig. 3). The log(D) value of the diastereomeric mixture of the macrocyclic secondary alcohol **N6** (C7-*S* and C7-*R*) is in most models significantly underesti-

Table 2. Predictive performances of the 3D QSAR COMBINE models derived earlier for the BCL-acyl-3-aryloxy-1,2-propanediol complexes,[12] and applied for the racemic secondary alcohols **N1**–**N4**, **N7**–**N11** and diastereomeric mixtures **N5** (C5-*S*, C5-*R*) and **N6** (C7-*S*, C7-*R*)

Model[a]	R^2 (# components)	SDEC	Q^2	SDEP	SDEPext[b]	SDEPext[c]
I0	0.89 (3)	0.19	0.67	0.34	0.73	0.16
I0[d]	0.92 (4)	0.17	0.72	0.32	0.68	0.42
I0+	0.89 (4)	0.20	0.67	0.34	0.63	0.31
I0+[d]	0.89 (5)	0.20	0.64	0.36	0.52	0.10
I0++[e]	0.92 (5)	0.17	0.59	0.38	0.61	0.50

[a] Identifies type of the model: I0 includes only the substrate–protein interaction energy terms, I0+ besides these includes the substrate–water interaction energy terms, the I0++ model in addition includes nonbonded substrate–substrate interaction energy terms as well.
[b] SDEP calculated for the external data sets consisting of 9-phenyl-substituted secondary alcohols (Schemes 1–3).
[c] SDEP calculated for two α-methylene-β-hydroxy esters (Scheme 4).
[d] The model I0+ was recalculated including conformational energy difference.
[e] The I0 and I0++ models were recalculated using slightly different FFD variable selection procedure (larger number of LV was retained, and number of dummy variables and the combination to variable ratio were 33% and 3.5%, respectively).



Scheme 4. Schematic representation of the PSL catalysed esterification of alkyl substituted secondary alcohols: **N10**, R=CH_3 ($E > 50$) and **N11**, R=$CH_3CH_2CH_2$ ($E > 50$).[17]

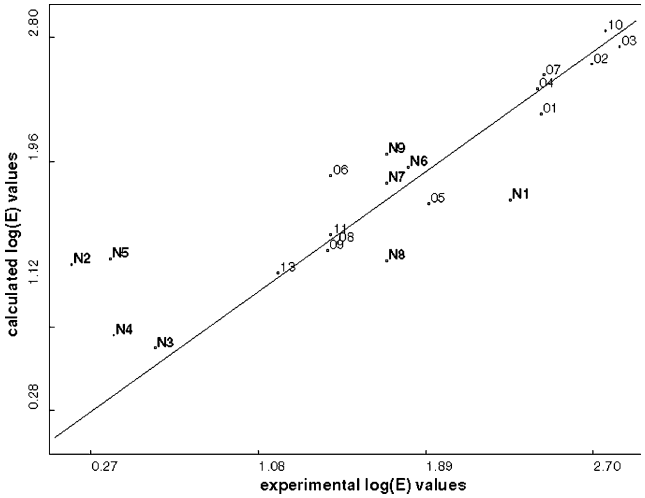


Figure 3. Predicted versus measured log(E) for 9 secondary alcohols considered in this work **N1**–**N9** determined using the I0 model derived earlier for 12 diols.[12]

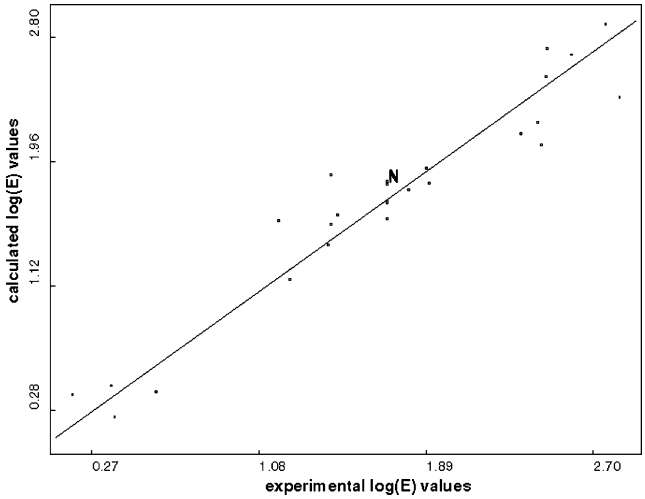


Figure 4. Predicted (by I0++ model for 24 secondary alcohols) versus experimentally determined logarithm of enantiomeric ratio for two α-methylene-β-hydroxy esters **N**. The correlation between the calculated and the experimental log(E) and log(D) values for 24 diverse secondary alcohols: 15 aryloxy-1,2-propanediol derivatives analysed in our previous publication,[12] and 9 newly analysed secondary alcohols **N1**–**N9** are presented with smaller squares.

mated (about 1 in logarithmic scale), with the exception of the I0 model (Fig. 3). In all cases the fast reacting enantiomer was correctly predicted.

3.5. New 3D QSAR COMBINE model

Furthermore, a set of different COMBINE models (Table 3) was derived for the diverse set of twenty-four lipase–substrate diastereomeric complexes, including 15 compounds considered earlier (12 BCL-3-(aryloxy)-1,2-propanediols for which the models were derived and three additional compounds from the same series of measurements[40] that served as a test set in our previous work) and compounds considered in this work for which the Monte Carlo based conformational search was performed.

Since derived for the diverse set of compounds, these models are expected to be more general than those obtained previously for the 12 BCL-3-(aryloxy)-1,2-propanediol complexes and should be able to predict enantioselectivity of BCL towards the wide range of secondary alcohols.

The correlation between the experimentally determined E and D (in the case of the diastereomeric mixtures **N5** and **N6**) values and those calculated with the I0+ and I0++ models is high while the resolution of the two α-methylene-β-hydroxy esters by the lipase is correctly predicted as shown in Figure 4. Enantiomeric ratios for these two compounds are by different 24-object COMBINE models, predicted more consistently than by the models derived for 12-object. In the models I0+ and I0++, similarly to those derived earlier for the smaller data set,[12] the electrostatic interactions between substrate and a few crystallographic water molecules located at the entrance of the binding pocket are the significant X variables. The other high weighted X-terms are similar to those in the I0 model (Fig. 4).

Table 3. 3D QSAR COMBINE models derived for the 24 BCL-secondary alcohol complexes

Model[a]	R^2 (components)	SDEC	Q^2	SDEP	SDEPext[b]
ISA	0.68 (1)	0.43	0.56	0.51	0.40
I0	0.94 (6)	0.20	0.67	0.43	0.23
I0+	0.93 (5)	0.21	0.74	0.39	0.07
I0++	0.93 (5)	0.21	0.75	0.38	0.11

[a] Identifies type of the model, see Table 2 for the reference, ISA means that the interaction energy terms and the solvent accessible surface area terms are included.

[b] External SDEP calculated for two α-methylene-β-hydroxy esters (Scheme 4).

3.6. The lipase–substrate interactions important for stereoselectivity

The main purpose of the COMBINE analysis was to identify the amino acid residues that are the most responsible for differences in the BCL catalysed reaction of esterification of secondary alcohols. The residues Leu17, Thr18, Tyr29, Gln88, Val123, Leu167, Val266 and Ile290 appear, beside those of the catalytic triad (Ser87, Asp264 and His286), in the all COMBINE models with high weights (see e.g., Fig. 5). Apparently they are most important for the stereoselectively catalysed esterification of the secondary alcohols by BCL. Furthermore, it is obvious that the proximity of the substrate to His286 (interactions: L-His286, M-His286 and O_{al}-His286, see Fig. 5) improves its catalysis.

There are significant differences in the individual terms of the different COMBINE models and the set of compounds considered; however, all of them revealed the same amino acid residues from the BCL active site as important. The conclusions drawn in our earlier work

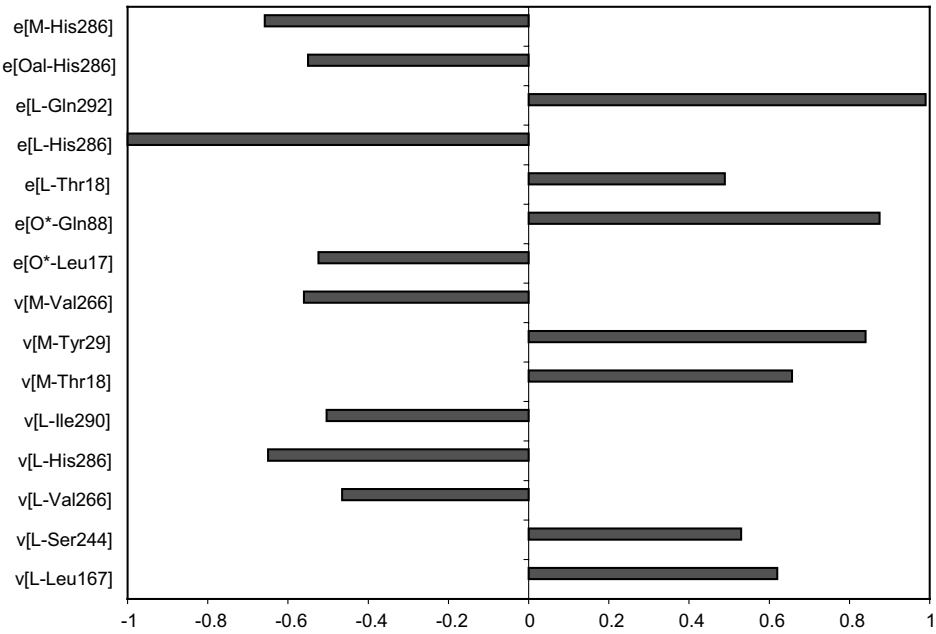


Figure 5. The highest weighted interactions in the I0 model obtained for the 24 complexes. Interactions are given on the *y*-axis and their normalised weights on the *x*-axis. v[...]=van der Waals, e[...]=electrostatic interactions. L and M = large and medium substituents at the alcohol stereogenic centre.

remain valid: (a) the attractive interactions between the fast reacting enantiomer and the amino acid residue Thr18 have an opposite effect on enantioselectivity, (b) desolvation of the fast reacting enantiomer, expressed as a tightness of its fit into the HA pocket, is proportional to the rate of catalysis.

Different from the results obtained for 3-(aryloxy)-1,2-propanediols, the inclusion of solvent accessible surface area terms did not improve predictive ability of the I0 model. Interestingly, the additional terms resulted in a decrease in model dimensionality (see Table 3). Apparently the ISA model derived in this work is more robust than those derived earlier.[12] Dominating terms in this model are those describing stabilisation of the fast reacting isomer.

3.7. Prediction of the mutant BCL stereoselectivity towards N1, N2, N7 and N11

Recently Guieysse et al.[13] also approached BCL enantioselectivity by molecular modelling. They used pseudomolecular dynamic simulations under constrains to map the probable trajectories of the two stereoisomers of 2-bromophenyl acetic acid ethyl ester, and found that the enzyme–substrate interaction energy was always higher for the slow reacting enantiomer. They argued that the valine at position 266 is the most responsible for this, and in order to prove this, tested the enantioselectivity of the mutated (Phe221 → Leu, Val266 → Leu, Leu287 → Ile) BC lipase. The enantioselectivity of the mutant enzyme towards (*RS*)-2-bromophenyl acetic acid ethyl ester was much higher than of the native one. On the other hand Hirose et al.[41] determined a decrease or even an inversion of stereoselectivity for 1,4 dihydropyridines using the mutated BC lipase.

In accordance with the results of pseudo molecular dynamics simulations, Val266 appears as important amino acid residue in the all COMBINED models derived in this and our previous work. The residue Leu287 was less important while Phe221 appeared to be of no significance at all.

We used the COMBINE analysis to examine the effect of the enzyme mutations on stereoselective catalysis. Enantioselectivity of the mutated enzyme towards a set of selected secondary alcohols **N1**, **N2**, **N7** and **N11** was predicted using the models derived for 24 compounds. Results revealed different influences of the mutations on different secondary alcohols. Stereoselectivity for the lipase towards the smallest secondary alcohols, **N7** and **N11**, remained the same whereas for **N1** and **N2**, it slightly increased.

The **N1** and **N2** mean log(*E*), calculated using the COMBINE models I0, I0+ and I0++, for the native enzyme are 1.8 and 1.7, respectively, while for the mutant are 1.9 and 1.8, respectively. Comparison of the interaction energies for **N2** in the native and the mutated enzyme complexes (Table 4) revealed the following changes. Strength of the L-Val(→Leu)266 interaction increased for **N2** fast reacting enantiomer in the complex with the mutant. However the strength of the L-Leu(→Ile)287 interaction decreased for the slow reacting enantiomer. The mutations also influenced the orientation of the side chains of the neighbouring amino acid residues. In this way they induced changes of the energy terms in which they were not explicitly involved. The Leu287 to Ile mutation influenced the positions of

Table 4. The interaction energies in the native and mutated **N1**- and **N2**-BCL complexes

Substrate	Enzyme[a]	vdW[L-266][b]	vdW[L-287][b]	vdW[L-Leu167][b]	ele[L-His286][b]
R-**N2**	**n**	−0.40	−0.19	−0.39	0.37
R-**N2**	**m**	−1.18	−0.20	−0.11	0.08
S-**N2**	**n**	−0.09	−2.03	−0.02	0.05
S-**N2**	**m**	−0.09	−1.77	−0.01	0.00

[a] Native enzyme (**n**), and the (Phe221 → Leu, Val266 → Leu, Leu287 → Ile) mutant (**m**).
[b] Van der Waals (vdW) and electrostatic (ele) interactions between the large substituent at the alcohol stereogenic centre and an amino acid residue at the position specified (kcal/mol).

His286 and Gln292 side chains. The Val266 to Leu mutation was reflected on the L-Leu167 interactions since the fast reacting enantiomer was in a 'sandwich' between these two residues. In the complex of the mutant with (*R*)-**N2**, the strength of the L-His286 interaction increased while that of the L-Leu167 interaction decreased. The interactions discussed above are practically identical in the native and mutant complexes of the opposite stereoisomer meaning the changes of the interactions discussed above could be directly correlated with the changes of the enantioselectivity.

The L-Val(→Leu)266, L-His286 and L-Leu167 interactions are highly weighted in the I0 COMBINE model (Fig. 5). The first two interactions are negatively correlated with the enantiomeric ratio, while the last one is positively. Apparently, the predicted increase in enantioselectivity of the mutant towards **N2** is correlated with a decrease of strength of the L-Leu167 interaction (energy difference becomes more positive), as well as with an increase of strength of the van der Waals L-Val(→Leu)266 and L-His286 interactions (energy difference becomes more negative).

4. Conclusion

The purpose of this work was to investigate further the reasons for the high enantioselectivity of BCL towards secondary alcohols, as well as to check the ability of the earlier derived 3D QSAR COMBINE models to predict the stereoselectivity of BCL towards the set of secondary alcohols different from those analysed earlier.

The results of the Monte Carlo based conformational search utilised herein are consistent with those of the systematic search combined with the molecular dynamic simulation used in our earlier investigation. Both approaches reveal different binding modes for two stereoisomers: while the fast reacting one, binds in a way that its large substituent predominantly occupies the large hydrophobic pocket or protrudes out of the active site, the slow reacting one binds into the smaller, partly hydrophilic pocket on the opposite side of the catalytic serine 87. In most cases, both methods revealed a larger number of favourable binding modes for the fast reacting stereoisomer. This finding points to entropy as an important factor for the lipase stereoselection, and is in agreement with the experimental results of Ema et al.[39] who determined an increase in enantioselectivity of BCL towards 1,1-diphenyl-2-propanol via an increase in temperature.

It was not possible to quantitatively predict the stereoselectivity of the lipase by only considering the conformational energy (and/or the substrate–protein interaction energy) difference between the stereoisomeric lipase tetrahedral complexes and the population of the binding modes. On the other hand, using the 3D QSAR COMBINE models derived earlier, a satisfactory prediction of the stereoselectivity of BCL towards the secondary alcohols was achieved despite differences in the molecular modelling approaches utilised and sets of secondary alcohols considered.

Herein, we have derived new COMBINE models for a broad selection of chemically diverse secondary alcohols. They are more general than those obtained previously and should be able to predict the enantioselectivity of BCL towards a wide range of secondary alcohols as has been successfully done for two α-methylene-β-hydroxy esters.

In conclusion, COMBINE analysis enables insight into the importance of individual amino acid residue for the stereo differentiation, and based on this, the site-directed mutagenesis can be rationally planned. This procedure has been demonstrated in the example of the (Phe221 → Leu, Val266 → Leu, Leu287 → Ile) mutant.

Acknowledgements

The work was supported by the grant no 0098036 of Ministry of Science and Technology, Croatia.

S. Tomić acknowledges continuing financial support from the Alexander von Humboldt Foundation. This work is also part of Croatian–German bilateral project HRV 01/010. We thank Dr. G. Cruciani for enabling us to use GOLPE program and Dr. R. C. Wade for valuable comments and suggestions.

References and notes

1. Faber, K. *Biotransformations in Organic Chemistry*, 3rd ed.; Springer: Berlin, 1997.
2. Schmid, R. D.; Verger, R. *Angew. Chem., Int. Ed.* **1998**, *37*, 1608–1633.

3. Schmid, A.; Hollmann, F.; Byung, J.; Bühler, B. *Prot. Technol.Commercial Enz.* **2002**, *13*, 359–366.
4. Straathof, A.; Panke, S.; Schmid, A. *Curr. Opin. Biotechnol.* **2002**, *13*, 548–556.
5. Muralidhar, R. V.; Chirumamilla, R. R.; Marchant, R.; Ramachandran, V. N.; Ward, O. P.; Nigam, P. *World J. Microbiol. Biotechnol.* **2002**, *18*, 81–97.
6. Bornscheuer, U. T.; Bessler, C.; Srinivas, R.; Krishna, S. H. *Trends Biotechnol.* **2002**, *20*, 433–437.
7. Reetz, M. T. *Curr. Opin. Biotechnol.* **2002**, *6*, 145–150.
8. Laumen, K.; Schneider, M. P. *J. Chem. Soc., Chem. Commun.* **1988**, 598–600.
9. Kazlauskas, R. J.; Bornscheuer, U. T. Biotransformations with lipases. In *Biotransformations with Lipases in Biotechnology*; Rehm, H. J., Reed, G., Pühler, A., Stadler, P. J. W., Kelly, D. R., Eds.; VCH: Weinheim, Germany, 1998; Vol. 8.
10. Luić, M.; Tomić, S.; Lešćić, I.; Ljubović, E.; Šepac, D.; Šunjić, V.; Vitale, L. J.; Saenger, W.; Kojić-Prodić, B. *Eur. J. Biochem.* **2001**, *268*, 3964–3973.
11. Tomić, S.; Dobovičnik, V.; Šunjić, V.; Kojić-Prodić, B. *Croat. Chem. Acta* **2001**, *74*, 343–357.
12. Tomić, S.; Kojić-Prodić, B. *J. Mol. Graphics Modell.* **2002**, *21*, 241–252.
13. Guieysse, D.; Salagnad, C.; Monsan, P.; Remaud-Simeon, M.; Tran, V. *Tetrahedron: Asymmetry* **2003**, *14*, 1807–1817.
14. Ljubović, E.; Šunjić, V. *Croat. Chem. Acta* **1998**, *71*, 99–117.
15. Ljubović, E.; Šunjić, V. *Tetrahedron Lett.* **2000**, *41*, 9135–9138.
16. Kazalauskas, R. J.; Weissfloch, A. N. E.; Rappaport, A. T.; Cuccia, L. A. *J. Org. Chem.* **1991**, *56*, 2656–2665.
17. Nascimento, M. G.; Zanotto, S. P.; Melegari, S. P.; Fernandes, L.; Sá, M. M. *Tetrahedron: Asymmetry* **2003**, *14*, 3111–3115.
18. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
19. INSIGHTII, 1997, Molecular Simulations, San Diego.
20. Cornell, W. D.; Cieplak, P.; Payly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
21. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440–467.
22. Kolossváry, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671–1684.
23. Allinger, N. L.; Yuh, Z. H.; Lii, J.-H. The MM3 force field for hydrocarbons 1–3. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566, 8566–8575, 8576–8582.
24. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
25. NACCESS, Hubbard, S. J.; Thornton, J. M. Department of Biochemistry and Molecular Biology, University College London: London, 1993.
26. DISCOVER, 1997, Molecular Simulations, San Diego.
27. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. *Quant. Struct.–Act. Relat.* **1993**, *12*, 9–20.
28. Tomić, S.; Nilsson, L.; Wade, R. C. *J. Med. Chem.* **2000**, *43*, 1780–1792.
29. Tomić, S.; Wade, R. C. *Croat. Chem. Acta* **2001**, *74*, 295–314.
30. Wold, S.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjoestroem, M. In *Chemometrics–Mathematics and Statistics in Chemistry*; Kowalski, B. R., Ed.; Reidel: Dordrecht, 1984; pp 17–95.
31. Cruciani, G.; Clementi, S.; Baroni, M. Variable selection in PLS analysis in 3D QSAR. In *Drug Design Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993.
32. Lang, D. A.; Mannesse, M. L. M.; De Haas, G. H.; Verheij, H. M.; Dijkstra, B. W. *Eur. J. Biochem.* **1998**, *254*, 333–340F.
33. Haeffner, F.; Norin, T.; Hult, K. *Biophys. J.* **1998**, *74*, 1251–1262.
34. Raza, S.; Fransson, L.; Hult, K. *Protein Sci.* **2001**, *10*, 329–338.
35. Zuegg, J.; Hönig, H.; Schrag, J. D.; Cygler, M. *J. Mol. Catal. B: Enzym.* **1997**, *3*, 83–98.
36. Felluga, F.; Pitacco, G.; Valentin, E.; Coslanich, A.; Fermeglia, M.; Ferrone, M.; Pricl, S. *Tetrahedron: Asymmetry* **2003**, *14*, 3385–3399.
37. Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681–2691.
38. Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1997**, *40*, 1136–1148.
39. Ema, T.; Kageyama, M.; Korenaga, T.; Sakai, T. *Tetrahedron: Asymmetry* **2003**, *6*, 1323–1344.
40. Theil, F.; Lemke, K.; Ballschuh, S.; Kunath, A.; Schick, H. *Tetrahedron: Asymmetry* **1995**, *6*, 1323–1344.
41. Hirose, Y.; Kariya, K.; Nakanishi, Y.; Kurono, Y.; Achiwa, K. *Tetrahedron Lett.* **1995**, *36*, 1063–1066.