

A quantitative model for predicting enzyme enantioselectivity: application to *Burkholderia cepacia* lipase and 3-(aryloxy)-1,2-propanediol derivatives

Sanja Tomić*, Biserka Kojić-Prodić

Ruđer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia

Received 21 January 2002; received in revised form 21 May 2002; accepted 19 June 2002

Abstract

We describe a new approach for predicting the enantioselectivity of enzymes towards racemic compounds. It is based on comparative binding energy (COMBINE) analysis.

The approach is used to rationalise the enantioselectivity of *Burkholderia cepacia* lipase (BCL) towards thirteen racemic 3-(aryloxy)-1,2-propanediols in the process of acylation. According to our molecular modelling study the two 3-(aryloxy)-1,2-propanediols enantiomers bind in the BCL active site in different orientations. To derive a quantitative structure–activity relationship (QSAR), the difference in the interaction energy between two enantiomers with each amino acid residue was computed. These residue-based energy differences were then subjected to chemometric analysis and 3D QSAR models were derived. The models were able to unambiguously predict the fast-reacting enantiomer and the approximate magnitude of the enantioselectivity.

The study enabled identification of interactions between the substrate and the lipase amino acid residues that play key roles in secondary alcohol enantiodifferentiation. From the results, it was possible to propose modifications of both, substrate and protein, which would directionally modify enantioselectivity of BCL towards secondary aryl-alcohols.

© 2002 Elsevier Science Inc. All rights reserved.

Keywords: *Burkholderia cepacia* lipase; Kinetic resolution; Secondary alcohols; Enantioselectivity; Molecular modelling; COMBINE approach

1. Introduction

Microbial lipases have proven to be useful biocatalysts for obtaining chiral, non-racemic compounds [1]. Lipase from *Burkholderia cepacia* (EC 3.1.1.3, formerly known as *Pseudomonas cepacia*) very efficiently catalyses hydrolysis, transesterification, esterification, alcoholysis, acidolysis, and aminolysis [2–5]. In order to predict enantioselectivity of lipases towards alcohols, Kazlauskas and co-workers derived empirical rules [6,7]. However, these rules do not clarify the detailed mechanism of the enantioselectivity. In order to improve the usefulness of lipases as biocatalysts, an understanding of the reaction at submolecular level is needed. The crystal structures of *Burkholderia cepacia* lipase (BCL) [8,9] and its complexes [10,11] encouraged a number of

molecular modelling studies that accelerated understanding of lipase catalysed reactions.

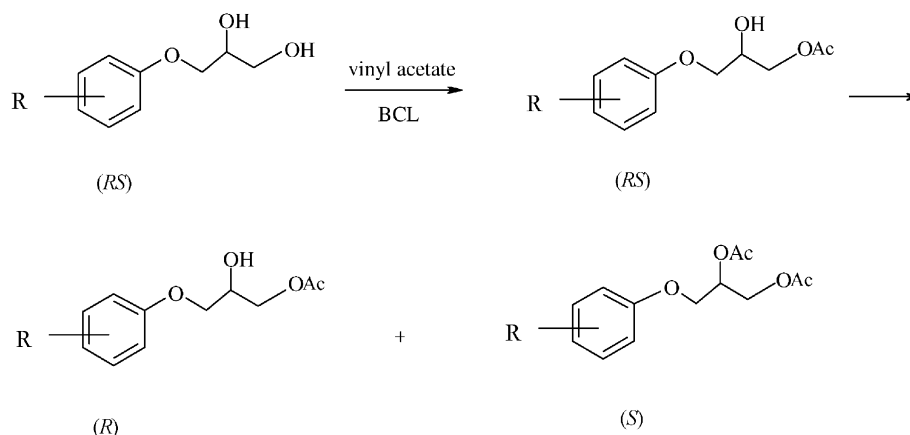
While a number of authors have used the potential energy of the tetrahedral lipase–substrate complexes to discriminate the fast-reacting from the slow-reacting enantiomer of secondary and primary alcohols [7,12,13] some have concentrated on the structural features solely [14,15]. Lately, energy unminimised complexes were used for study of enantioselectivity of lipase towards alcohols and their esters [15,16]. Schultz et al. [15] were able to distinguish substrates for which BCL shows high enantioselectivity from those for which enantiomers are poorly distinguished on the basis of the $H_{Ne}-O_{alcohol}$ distance for the slow-reacting enantiomer. Raza et al. studied enantioselectivity of *Candida antarctica* lipase B [16]. Considering the mean, function based, conformational energy of an unoptimised enzyme–substrate system, they obtained a reasonable agreement with the experimentally determined free energy of activation.

On the basis of molecular modelling, site-directed mutagenesis at the lipase active site can be planned [17,18]. Magnusson et al. [17] obtained mutants with significantly increased enantioselectivity towards ethyl-2-hydroxypropanoate.

Abbreviations: BCL, *Burkholderia cepacia* lipase; QSAR, quantitative structure–activity relationship; TI, tetrahedral intermediate; COMBINE, comparative binding energy; MD, molecular dynamics

* Corresponding author. Tel.: +385-1-4561025; fax: +385-1-4680245.

E-mail address: sanja.tomic@irb.hr (S. Tomić).



Scheme 1. Lipase-catalysed enantioselective transesterification of 3-(aryloxy)-1,2-propanediol derivatives.

Importance of enantiomerically pure 3-(aryloxy)-1,2-propanediols is related to their wide use in the pharmaceutical industry. They are used either as intermediates in the synthesis of some drugs (e.g. β -receptor blockers) or as building blocks for synthesis of crown ethers [19,20].

A few years ago, Theil et al. [20] accomplished kinetic measurements for a series of racemic monoacetates of 3-(aryloxy)-1,2-propanediols. The study of lipase-catalysed sequential transesterification of 1,2-diols with lipase from *Burkholderia cepacia* revealed that in the first step they are acylated regioselectively at the primary hydroxyl group without enantioselection, and the subsequent acylation at the secondary hydroxyl group of the primary monoacetate formed is responsible for high enantioselection (see Scheme 1).

They found that a substituent in the *para*-position of the aryl ring increases and in the *ortho*-position decreases the enantioselectivity. Compounds with a bulky substituent in the *ortho*-position were not catalysed at all. Such results could not be explained with the empirical rule [6], which relates the relative size of the substituents at the stereogenic centre with enantioselectivity, only. The aim of the present study is to give a rational explanation for these experimental results and to derive physical and structural parameters important for the enantioselectivity of BCL towards monoacetates of 3-(aryloxy)-1,2-propanediols.

2. COMBINE analysis

The recently determined X-ray structure of the complex of BC lipase with a transition state analogue of a secondary alcohol [11] enabled reliable modelling of the tetrahedral complexes of BCL with a secondary alcohol and critical evaluation of the existing models. Based on this structure, we carried out molecular modelling and, in order to derive the 3D QSAR model for predicting enantioselectivity of BCL to-

wards 3-(aryloxy)-1,2-propanediols performed COMBINE analysis.

The difference of the free energies of the lipase–substrate tetrahedral intermediate complex formed by the two enantiomers is directly proportional to the logarithm of the measured enantioselectivity expressed as the enantiomeric ratio E :

$$\Delta\Delta G = -RT\{\ln E\} \quad (1)$$

Here, $\Delta\Delta G = \Delta G_F - \Delta G_S$, where ΔG_F and ΔG_S denote the free energies of the tetrahedral intermediate complex formed with the fast (F)- and slow (S)-reacting enantiomer, respectively.

A free energy calculation by either perturbation or thermodynamic integration methods is computationally demanding and complex [21,22]. In order to maximise the ratio between the amount of information and the computational time required, in this work we considered an approximation of the binding free energy difference based on COMBINE analysis [23–25]:

$$\Delta\Delta G = \sum_{i=1}^n w_i \Delta\Delta u_i^{\text{sel}} + C \quad (2)$$

where $\Delta\Delta u_i^{\text{sel}}$ are residue-based energy difference terms and w_i are their weights as determined by chemometric analysis, C is a constant.

The first stage of the present study, molecular modelling, led to identification of the lowest energy binding geometry for the two enantiomeric series of the substrate tetrahedral intermediates. Subsequent 3D QSAR analysis enabled derivation of the model for the enantioselectivity of BCL towards 3-(aryloxy)-1,2-propanediol derivatives of predictive value.

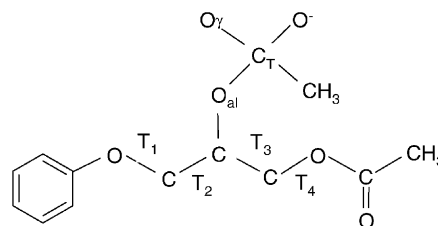
The same procedure can be used to determine enantioselectivity of lipase, as well the other enzymes, towards the other racemic compounds.

3. Methods

3.1. Modelling of the complexes

Starting structures were from the Protein Data Bank (PDB) [26] accession codes 5LIP [10] and 1HQD [11].

The acyl-enzyme structure was built using the topology of the BCL-inhibitor complex, 1HQD. According to the literature data [7] and the references therein, a phosphonate mimics transition state analogue in the process of ester hydrolysis. Based on this assumption, the $\text{P}(\text{CH}_3)\text{O}^-$ group covalently bound to Ser87 in 1HQD was replaced by the $\text{C}(\text{CH}_3)\text{O}^-$ group [27]. Hydrogens were added to correspond to pH 7.0 using the Biopolymer module in the program INSIGHTII [28]. Histidines were uncharged (singly protonated), aspartic and glutamic acids were negatively, and arginines and lysines positively charged. The protonation of the histidine imidazole ring (either N δ 1 or N ϵ 2) was adjusted manually depending on the possibilities of hydrogen bond formation. The catalytic His286 was in the doubly protonated form, according to the assumed reaction path. Parameterisation was performed in the AMBER force field [29] and charges of the tetrahedral carbon (C_T) bound to Ser87, O(Ser87), oxyanion (O^-) and O_alcohol (Scheme 2) were determined using MOPAC MNDO-ESP [30] calculations (0.79, −0.57, −0.81 and −0.57, respectively) and used as the forcefield parameters [29]. In order to remove internal strain, the initial structure was energy minimised without constraints by a conjugate gradient algorithm using 0.5 scaling of the 1–4 interaction. The bulky water screening



Scheme 2. Atomic and torsional numbering of the lipase substrate **1** reaction intermediate model (O_al = O_alcohol , C_T = $\text{C}_\text{tetrahedral}$, O^- = oxyanion).

of the electrostatic interactions was simulated by a distance dependent dielectric constant ($\epsilon = 4r$).

There are a number of buried water molecules found in the crystal structure of the native BCL and its complexes. In order to investigate the influence of water molecules on COMBINE models, molecular modelling was performed with and without the crystallographically determined water molecules.

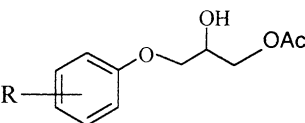
3.2. Determination of the possible conformations of a bound substrate

All substrates (Tables 1 and 2) were built and covalently bound to the acyl-enzyme. Each acyl-enzyme–substrate complex was energy minimised with the particular enantiomer of a substrate bound to the acylated Ser87.

The most detailed conformational search was performed for the lipase complex with compound **1** (Table 1).

Table 1

Kinetic resolution of diols (*R*, *S*): experimental data [20] and calculated potential energy difference of the tetrahedral complexes with fast- and slow-reacting enantiomer



Substrate ^a	R	Time (h)	Monoacetate (<i>R</i>)		Diacetate (<i>S</i>)		<i>c</i>	<i>E</i>	ΔE^b (kJ/mol)	ΔE^c (kJ/mol)
			Yield (%)	e.e. (%)	Yield (%)	e.e. (%)				
1	H	53	51	92.9	49	97.6	0.49	281	−0.4	−7.5
2	4-Me	54	58	65.0	42	99.0	0.40	391	−6.7	−17.6
3	4- <i>t</i> -Bu	100	47	99.1	49	98.4	0.50	671	−6.3	−41.4
4	4-CO-Ph	96	61	62.6	39	98.6	0.39	270	−2.9	−55.6
5	2,4-Me ₂	264	61	68.0	39	95.0	0.42	80	−0.4	−21.7
6	2-Me	240	65	39.1	35	89.9	0.30	27	−1.6	−13.0
7	4-Cl	72	62	57.2	38	98.8	0.37	296	−10.0	−46.8
8	2,4-Cl ₂	96	28	40.2	72	90.4	0.31	29	−8.4	−39.7
9	4-OPh	144	55	61.8	45	86.4	0.42	26	−36.4	−69.8
10	4- <i>t</i> -C ₈ H ₁₇	94	47	93.2	53	98.8	0.48	574	−3.3	−24.7
11	2-Cl	72	74	28.0	26	90.6	0.24	27	−3.3	−7.5
12	2-Ph	96	No substrate	—	—	—	—	—	−7.2 ^d (<i>S</i> - <i>R</i>)	−5.4 (<i>S</i> - <i>R</i>)
13	4-CH ₂ Ph	96	65	61.5	35	77.4	0.44	15	−35.1	−65.2

^a The large (L) substituent is R1 substituted phenyl $-\text{O}-\text{CH}_2$ group, and the medium substituent (M) is CH_2-OAc .

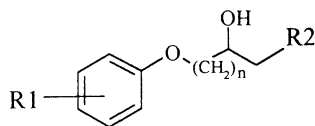
^b No water molecules considered.

^c The water molecules found in the crystal structure of the BCL–secondary alcohol-like inhibitor [11] were considered during minimisation.

^d Potential energy of the BCL–**12** (*R* and *S*) complex is about 3000 kJ/mol above those of the other complexes.

Table 2

Kinetic resolution of secondary alcohols (*R*, *S*): experimental data [20,36] and calculated potential energy difference of the tetrahedral complexes with fast- and slow-reacting enantiomer



Substrate	R1 (<i>n</i>), R2 ^a	Time (h)	Alcohol (X) ^b		Acetate (Y) ^b		<i>c</i>	<i>E</i>	ΔE^c (kJ/mol)	ΔE^c (kJ/mol)
			Yield (%)	e.e. (%)	Yield (%)	e.e. (%)				
14^d	H (1), OCO _n -C ₅ H ₁₁	73	55	67.1	43	98.2	0.41	223	9.2	−9.2
15^d	H (1), OCO- <i>i</i> -Pr	100	58	51.8	42	95.8	0.35	78	9.6	−1.3
16^d	H (1), N ₃	172	58	67.6	42	78.6	0.46	17	−4.5	−13.4
17^e	H (1), CH ₃	24	50	99.9	50	99.9	0.50	>200	−4.2	−7.9
18^e	H (3), C ₃ H ₇	300	19	12.4	81	54.1	0.19	4	5.0	−2.9

^a The large substituent (L) of **14–17** is Ph-O-CH₂ group, and of **18** is Ph-O-(CH₂)₃ group, the medium substituent (M) is equal to CH₂-R2 group.

^b X = *R*, Y = *S* for the substrates **14** and **15**, X = *S*, Y = *R* for the substrates **16–18**.

^c The same as notes (b, c) in Table 1.

^d From [20].

^e From [36].

Dihedrals T₂ to T₄ (Scheme 2) were varied from 0 to 360° with a step size of 60°. Each rotation was followed by a limited minimisation procedure (combined steepest descent and conjugate gradient algorithms). The protein back bone was constrained during the conformational search, energy minimisation and the subsequent molecular dynamics (MD) simulations. The complexes containing H-bonds characteristic for the tetrahedral intermediate (TI) (Fig. 1): Oδ2(Asp264)···Hδ1(His286), Hε(His286)···Oγ(Ser87), Hε(His286)···O(alcohol) and hydrogen bonds between oxyanion (O[−]) and protonated amide nitrogen of Leu17 and Gln88 were subjected to short (10 ps) MD simulations at room temperature. In the case of unstable conformations, transition occurred after only a few picoseconds of the simulation. In this way, productive binding modes (all hydrogen bonds necessary for the reaction to occur are formed) of both enantiomers of compound **1** were determined. They were further used as templates for modelling TIs of the other 1,2-diols. However, the substrates differ in the size and nature of their substituents and consequently they interact differently with the amino acid residues in the binding pocket. In order to allow readjustment of a substrate and the amino acid residues side chains each newly constructed lipase–substrate TI complex was subjected to short MD (a few 10 ps) simulations at increasing temperature (300–350 K) followed by energy optimisation. In this way, the most probable conformation of a TI was determined.

The most stable lipase–substrate TI complexes determined for (*S*)- and (*R*)-enantiomers of each substrate were subjected to COMBINE analysis.

3.3. Calculation of the residue-based terms to be used in COMBINE analysis

In this study, interaction energy u_{ij} (van der Waals u_{ij}^{vdW} and electrostatic u_{ij}^{ele}) and solvent accessible surface area

SA_j (polar SA_j^{p} and non-polar SA_j^{np}) terms were selected to approximate the binding free energy:

$$\Delta\Delta G = \sum_{i=1}^{n\text{SUB}} \sum_{j=1}^{n\text{ENZ}} w_{ij} \Delta\Delta u_{ij}^{\text{vdW}} + \sum_{i=1}^{n\text{SUB}} \sum_{j=1}^{n\text{ENZ}} w_{ij} \Delta\Delta u_{ij}^{\text{ele}} - \sum_{j=1}^{n\text{SUB}+n\text{ENZ}} (w_j \Delta\Delta SA_j^{\text{p}} + w_j \Delta\Delta SA_j^{\text{np}}) \quad (3)$$

The ground-state free energy is the same for both unbound enantiomers and terms $\Delta\Delta u_{ij}$ and $\Delta\Delta SA_j$ in Eq. (3) reduce to the differences between the u_{ij} and SA_j terms of tetrahedral complexes of the faster and the slower secondary alcohol enantiomers, Δu_{ij} and ΔSA_j terms. The u_{ij} term is the non-bonded interaction energy (either electrostatic or van der Waals) between the group *i* of the covalently bound secondary alcohol ester and amino acid residue *j* of the lipase; SA_j is solvent accessible surface area of either the amino acid residue or of the covalently bound acyl-secondary alcohol group *j*. The latter was represented by five groups: C_T-O[−] (where C_T denotes the asymmetric carbon bound to the catalytic Ser87), methyl group (attached to C_T), O_{alcohol}-C_{chiral}-H, large substituent (L) at alcohol chiral centre C_{chiral} and smaller, so called [1] medium substituent (M) at C_{chiral}. L and M of each substrate are identified in Tables 1 and 2. Summation is performed over substrate groups, *n*SUB, and over enzyme amino acid residues, *n*ENZ.

The solvent accessible surface area (SA) was evaluated, using the program NACCESS [31]. The difference between the SA of the complex with the slower and the faster reacting enantiomer was subjected to the chemometric analysis.

The weight, *w*, of each term in Eq. (3) was determined by PLS analysis [32].

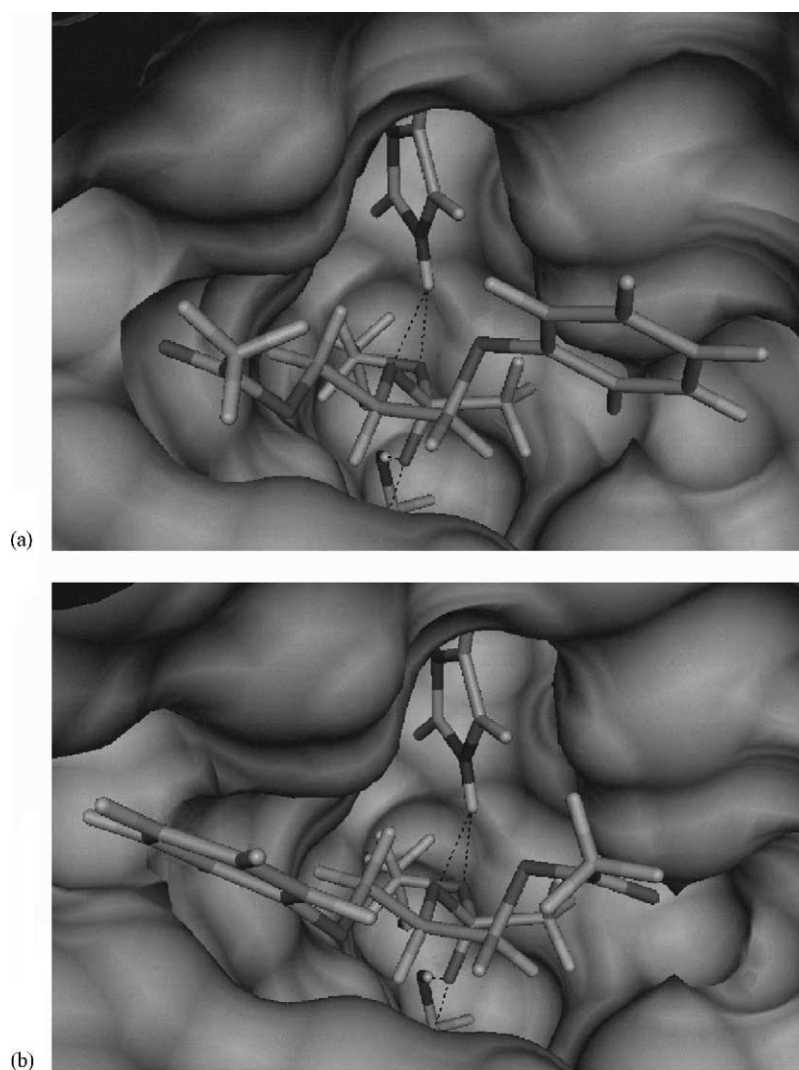


Fig. 1. Lipase–substrate **1** tetrahedral intermediate complex: (a) lipase–substrate (*S*)-**1** tetrahedral intermediate complex, (b) lipase–substrate (*R*)-**1** tetrahedral intermediate complex.

3.4. Preparation of data for chemometric analysis

Before being applied to statistical analysis, the X-variables were pretreated using different procedures: zeroing, minimum standard deviation cut-off and block unscaled weights (BUW). In the present study, X-variables were divided into four blocks: van der Waals block, electrostatic block, the conformational energy difference of the diastereomeric complexes and solvent accessible surface area block. BUW pretreatment was used to give the same importance to each block of variables.

3.5. Chemometric analysis

The PLS analysis [32] was performed to derive QSAR models and determine the variables that are the most important for specificity of binding.

Regarding the type of X-variables used in chemometric analysis, two basic types of models were derived: (a) only

enzyme–substrate tetrahedral intermediate (TI) interaction energy terms (I0 model), and (b) interaction energy and the surface area terms (ISA model).

For the complexes optimised with water molecules, besides the model I0, including only the substrate–protein interaction energy terms, two additional models (I0+ and I0++) were derived. The I0+ model, which also includes the substrate–water interaction energy terms, and the I0++ model, which besides the terms included in I0+ includes non-bonded substrate–substrate interaction energy terms as well.

As the 1D *Y* matrix in the PLS analysis, we used logarithmic (base 10) values of the measured enantiomeric ratio (*E*) [20].

The quality of the model was evaluated by checking its predictive abilities. For this purpose, internal and external validation was performed. For internal validation: “leave one out” (LOO) and “random groups cross-validation” using five random groups and 20 randomisations were performed. The

predictive ability of a model is described with standard deviation of error of prediction (SDEP) [33] and with the correlation between the experimental and predicted enantioselectivity, Q^2 :

$$Q^2 = \frac{\sum(Y - Y')^2}{\sum(Y - \langle Y \rangle)^2}$$

where Y is the experimental value, Y' the predicted value and $\langle Y \rangle$ is the average experimental value of $\log(E)$.

External validation of the models derived for the 12-object training set (Table 1) was performed on three- and five-object test sets (Table 2).

In order to extract the most predictive X-variables and to optimise model dimensionality, selection was performed by a fractional factorial design (FFD) strategy [34]. For each model one to three cycles of FFD variable selection, depending of the improvement of predictive ability (internal and external), were performed. In the first cycle we used 20%, and in subsequent cycles 33% of dummy variables. The combination to variables ratio was two in the first cycle and three in the subsequent cycles. Uncertain variables were retained in each cycle.

The FFD procedure might results with model over-fitting. The consequence would be lower predictive performances for the compounds in the external, test sets. To avoid this, the external prediction of the model was verified after each cycle of FFD (in the case an increase in the SDEP was noticed, the latest cycle was abandoned). Application of FFD resulted with decrease of the internal SDEP for about 30–40%. The external SDEP, for the three-object test set, decreased up to 10%. Prediction of the enantiomeric ratio for the remaining two, the most dissimilar compounds in the test set, did not improved.

4. Results

Molecular modelling and 3D QSAR analysis were used to explain differences in enantioselectivity in acetylation of a series of racemic monoacetates of 3-(aryloxy)-1,2-propanediols (Table 1).

4.1. Modelling substrate in the lipase active site

Modelling of the tetrahedral intermediates of the racemic monoacetate esters revealed that the two enantiomers bind into the BCL binding site differently (Fig. 1). The conformational search revealed a few different binding modes for both enantiomers. In the lowest energy productive binding mode (all hydrogen bonds necessary for the reaction to occur are formed), the fast-reacting enantiomer accommodates its large substituent into the hydrophobic, so called HA pocket [10] similarly to how it was found in the crystal structure of the complex of BCL with the secondary alcohol-like inhibitor [11] (Fig. 2). On the other hand, the slow-reacting

enantiomer in the productive binding mode inserts its large substituent into the partly hydrophilic, so called HH pocket [10] (Fig. 1). Similar binding of (*S*)- and (*R*)-enantiomers of esters of secondary alcohols, with opposite orientations of large and medium substituents, and the same orientation of the hydrogen atom at the stereogenic centre into the active site, was found by some other authors as well: Zuegg et al. in their study on BCL and *Candida rugosa* lipase [12], as well as by Haeffner et al. in their study on *Candida antarctica* lipase B [13]. Recently performed investigations of Gascoyne et al. [35] on *Chromobacterium viscosum* lipase, closely related to BCL, and chiral bisphenols indicated that two enantiomers bind into the active site with different orientation of the substituents at the alcohol chiral centre. They concluded that this might likely to be the reason for the enantioselectivity.

Depending on the size and the shape of the substrate, the large substituent (L) of the fast-reacting enantiomer occupies and protrudes out of a large hydrophobic pocket with the closest amino acid residues being Phe119, Val123, Ala247, Leu248, Thr251 and Val266. Its medium size (M) substituent fits into the HH pocket defined with residues: Tyr23, Tyr29, Leu287, Ile290 and Leu293 (Fig. 2). Such binding is in accord with the results of kinetic measurements, i.e. even the compounds with bulky L substituents (such as, for example, compounds **3** and **10** in Table 1) were catalysed by BCL (Table 1). Exceptions are compounds with bulky substituents such as *tert*-butyl or phenyl substituted at the position 2 of the phenyl ring. Such a big substituent displaces Leu287 from its original position, reduces the rotational entropy of the His286 side chain (Fig. 3) and interacts with the terminal CH₃ group of the M substituent. The potential energy of (*S*, *R*)-**12**-acyl-BCL is about 20% higher than the mean energy calculated for the other complexes (Table 1) without a bulky substituent at the *ortho*-position. The potential energy of the members of this set deviates less than 1% of the mean value. Apparently, a bulky substituent at the *ortho*-position destabilises the complex.

Conformational analysis revealed a bound substrate with T₂ and T₃ (Fig. 1) predominantly in *synclinal* ($\pm(60 \pm 30^\circ)$), and T₁ and T₄ in *antiperiplanar* ($\pm(180 \pm 30^\circ)$) conformations. Orientations of the side chains of the amino acid residues in the binding pocket, in most cases, slightly deviate from their position in the free enzyme. The largest deviation was noted for the His286 imidazole ring. In some complexes, it is oriented about $\pm 30^\circ$ from the position determined in the crystal structures of the free enzyme (67° in LIP2) [8,9]. These findings are in accord with the high similarity between the crystal structures of the native enzyme and its complexes.

4.2. Models for predicting enantioselectivity

The potential energy difference between the diastereomeric (acyl-BCL–alcohol) complexes in all cases correctly predicted the fast-reacting enantiomer (Table 1). For the

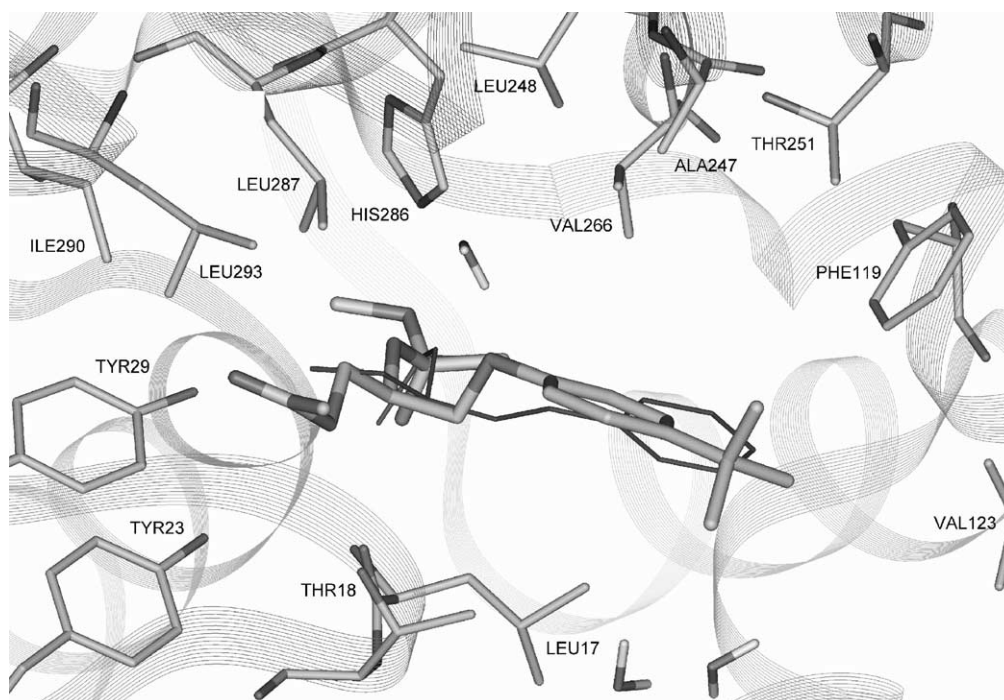


Fig. 2. BCL–inhibitor and BCL–(*S*)-**3** complexes. Superposition of (*S*)-**3** in the productive binding mode and the crystal structure of the complex of BCL and the secondary alcohol-like inhibitor (thin line) [11]. Amino acid residues in the binding pocket are displayed. The water molecules that turned out to be important for enantiodifferentiation are displayed.

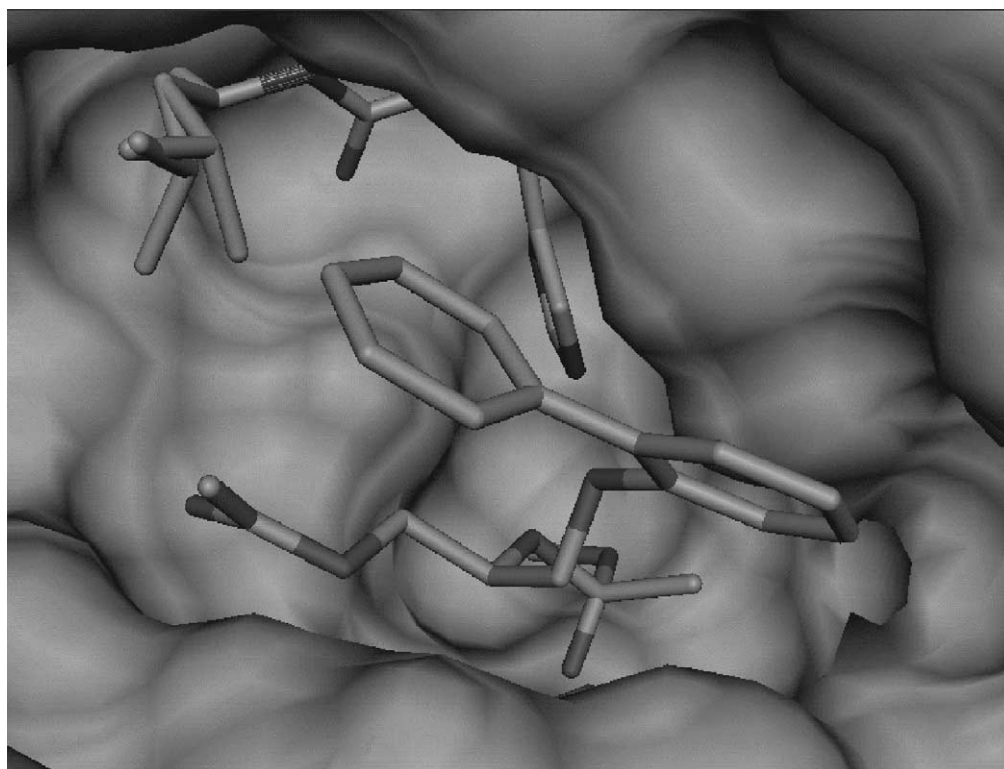


Fig. 3. Destabilisation effect of the large substituent at the *ortho*-position of the phenyl ring on the tetrahedral BCL–acyl-3-aryloxy-1,2-propanediol complexes. The phenyl ring substituted at the *ortho*-position displaces Leu287 from its original position (green). Position of the Leu287 side chain in the tetrahedral BCL complexes with the substrate without the big substituent at the *ortho*-position (green) and in the BCL–**12** tetrahedral complex.

Table 3
Predictive performances of the 3D QSAR COMBINE models

Objects	Model	Optimisation ^a	Substrate ^b	Water ^c	R^2 (components)	SDEC	Q^2	SDEP	SDEP ^d (external: 3/5)
12	I0	–	–	–	0.85 (2)	0.23	0.67	0.34	0.80/1.02
12	I0	+	–	–	0.89 (3)	0.19	0.67	0.34	0.32/0.62
12	ISA	–	–	–	0.97 (3)	0.10	0.80	0.26	0.30/0.90
12	ISA	+	–	–	0.99 (4)	0.05	0.75	0.27	0.27/0.81
12	I0+	+	–	+	0.89 (4)	0.20	0.67	0.34	0.33/0.47
12	I0++	+	+	+	0.92 (5)	0.17	0.56	0.40	0.32/0.42
17	I0	–	–	–	0.92 (3)	0.15	0.75	0.26	–
17	I0	+	–	–	0.88 (3)	0.18	0.65	0.31	–
17	ISA	–	–	–	0.96 (3)	0.13	0.84	0.23	–

^a Identifies type of complexes for which model is derived: complexes optimised without water molecules (–), complexes optimised with water molecules found in the crystal structure of the lipase–secondary alcohol-like inhibitor, 1HQD (+).

^b Identify interactions used to derive the model: (+) the non-bonded substrate–substrate interactions are included.

^c Identify interactions used to derive the model: (+) the intermolecular substrate–water interactions are included.

^d SDEP calculated for the external datasets consisting of three compounds [20] and the extended five compounds dataset [36].

complexes optimised by including the water molecules found in the 1HQD crystal structure, these energy differences are significant [11]. However, the quantitative evaluation of the enantioselectivity based on these differences is not possible, i.e. there is no direct correlation between the energy difference and the measured enantiomeric ratio. Obviously, entropic effects due to desolvation and different binding modes of two enantiomers should be included as

well. In order to derive a quantitative model for predicting enantioselectivity from structural parameters, the COMBINE approach was applied. Models are derived for the set of tetrahedral BCL–3-(aryloxy)-1,2-propanediol complexes (Table 1). Their predictive performances were tested on the two external test sets. The first includes three representatives of different derivatives of aryloxy-secondary alcohols (enantioselectivity determined by Theil et al. [20]) and the

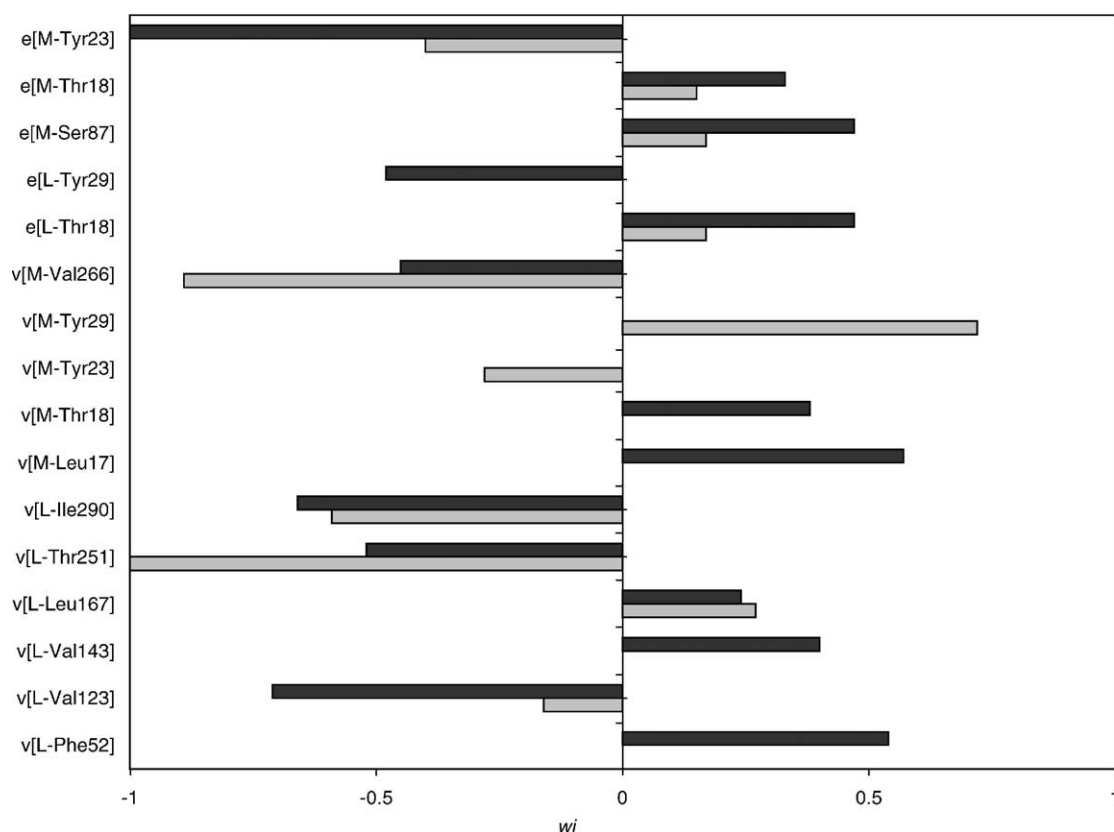


Fig. 4. X-variables in different I0 models. The highest weighted interactions in the I0 models: obtained for the complexes optimised without water molecules (light grey) and with water molecules (dark grey). v[···] = van der Waals, e[···] = electrostatic interactions.

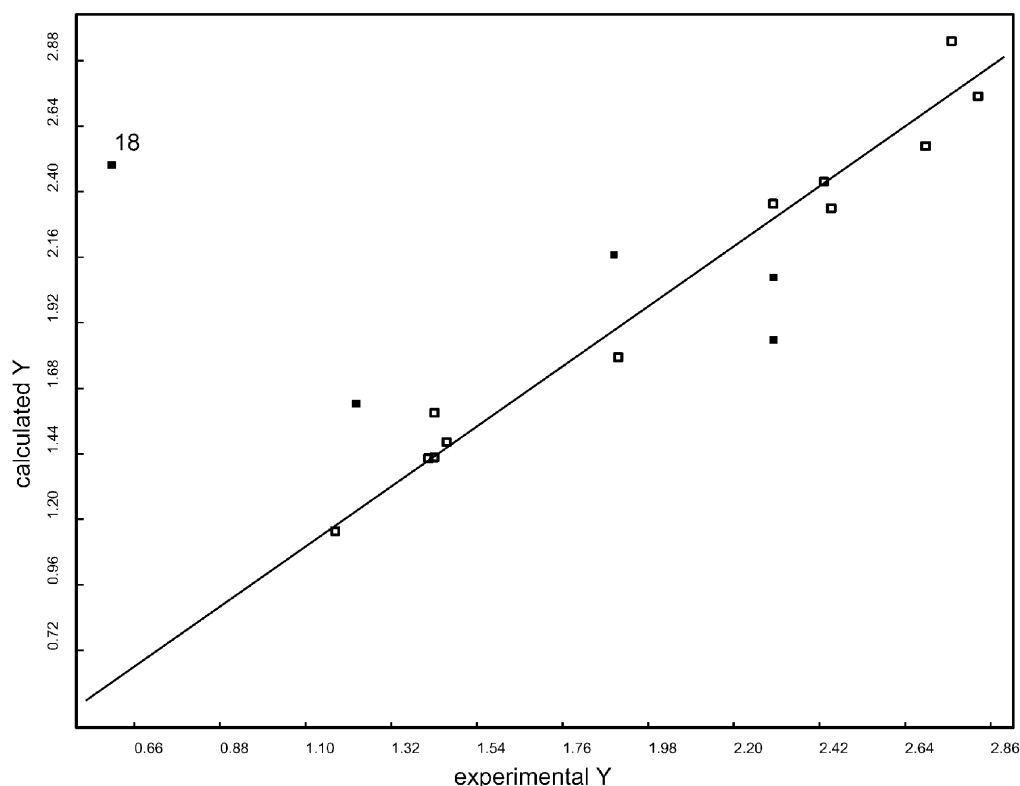


Fig. 5. Measured E vs. the one predicted with the ISA model.

second one extended by two additional aryloxy-secondary alcohols (enantioselectivity measured by Ljubović and Šunjić [36]) (Table 2).

4.3. I0 models (interaction energy terms only)

I0 models were derived for the complexes optimised with and without water molecules. Inclusion of water molecules during optimisation significantly increases predictive performances of I0 model (Table 3) for the compounds in two external test sets. In these two models sets of the highest weighted terms are similar (Fig. 4), namely those describing electrostatic and van der Waals interactions between substrate TI and the amino acid residues Thr18, Tyr23 and Ser87, and Val123, Leu167, Thr251, Val266 and Ile290, respectively. However, there are a few terms that appear to be much more significant in one model than in the other. For example, the electrostatic interaction between the large substituent (L) at the alcohol chiral centre and Tyr29 is significant only in the model derived for the complexes optimised with the water molecules as well as van der Waals interactions between the medium substituent (M) and residues Leu17, Thr18 and Phe52. On the contrary, the van der Waals interactions between the medium substituent and Tyr23 and Tyr29 are significant in the model obtained for the complexes optimised without water molecules, only (Fig. 4).

Inclusion of the substrate–water interaction terms into the model, I0+ and I0++, significantly increases its predictive performances on the external, larger test set (Table 3). In these models, the electrostatic interactions between substrate and a few crystallographic water molecules located at the entrance of the binding pocket (Fig. 2) have significant weights. The other important terms are as in the I0 model, electrostatic M-Tyr23 and L-Tyr29 and van der Waals interactions between the large substituent and the residues Phe52, Val123, Val143 and Ile290.

4.4. ISA model (based on interaction energy and solvent accessible surface area terms)

The ISA model is derived for complexes optimised with and without water molecules. In both cases, the interaction terms describing substrate–enzyme interactions are considered, only. As inclusion of water molecules as “COMBINE residues” together with the solvent accessible surface area terms (SA) is questionable. Predictive performances of ISA models are better than of the I0 models, but the external predictive performances are worse than of the I0+ model (Table 3). This model well predicted enantiomeric ratio (E) for all compounds in the external test sets except the compound **18**, having E significantly overestimated (Fig. 5). An explanation for this is complete desolvation of the amino

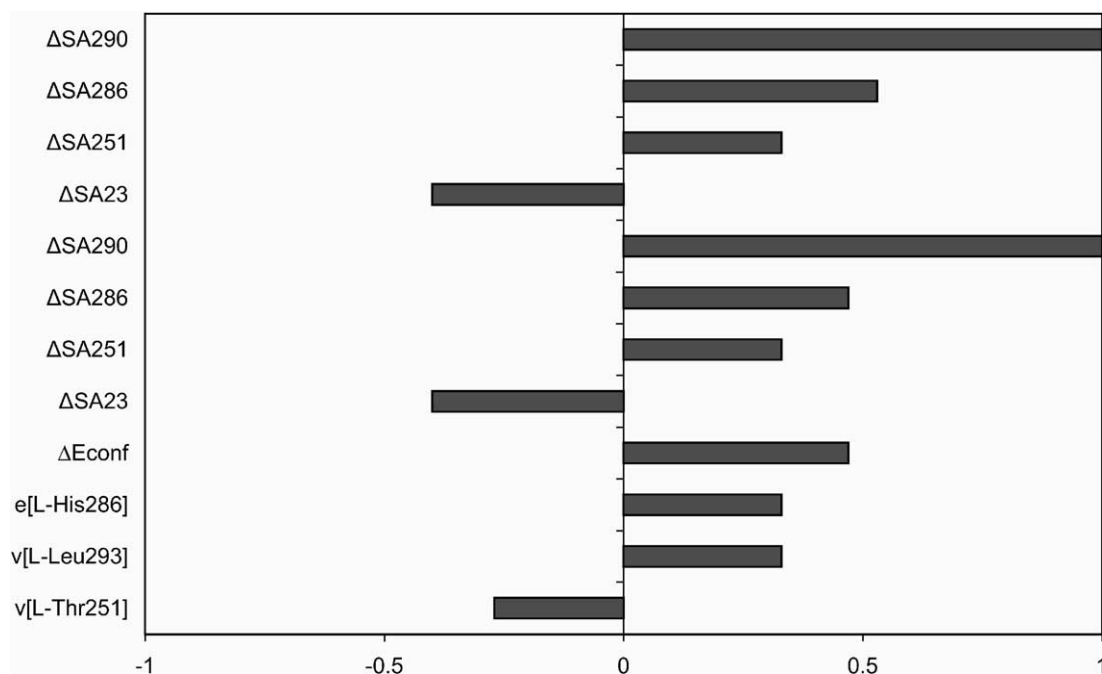


Fig. 6. The most significant X-variables in the ISA model.

acid residue Ile290 in both BCL–**18** complexes with the *S*- and *R*-enantiomer ($\Delta SA_{290} = 0$). The solvent accessible surface area difference of Ile290 is the highest weighted X-variable in the ISA model (Fig. 6), and for the other compounds with low enantioselectivity (**13**, **16**), it is negative, i.e. the solvent accessible surface of Ile290 (Fig. 6) is larger in the complexes with the faster enantiomer. Besides, there is a strong dispersive van der Waals interaction between the phenyl ring of the slower reacting enantiomer of **18** and Tyr23. Such stabilisation interactions between the slower enantiomer and enzyme decrease enantioselectivity, but in the ISA model the term mentioned above has very low weight.

Only in ISA models, the conformational energy difference (ΔE_{conf}) has significant weight.

Measured versus predicted enantioselectivity values (*E*) determined using the ISA model for 12-object training dataset (empty square) and for five-object test dataset (filled square).

All models described above correctly predicted the fast-reacting enantiomer not only for 3-(aryloxy)-1,2-propanediols but also for the other secondary alcohols considered (Tables 2 and 3).

4.5. Models for the 17-object dataset

Further on, I0 and ISA models were derived for the all 17 lipase–substrate diastereomeric complexes (Tables 1 and 2). The later one has better fit and better prediction performances (Table 3). The most significant terms in these models differ from that derived for the smaller dataset in the following: the terms describing interactions with the residues in the

HH pocket (Tyr23, Ile290) have lower, and those describing interactions with the residues in the HA pocket (Ala120, Leu164, Leu167), higher weights. The residues important in both models are Phe119 and Thr251. It seems that goodness of fit of the larger substituent into the HA pocket is one of the regulators of BCL enantioselectivity towards secondary alcohols. The electrostatic interaction between the larger substituent at the alcohol chiral centre and Thr18 appears to be significant in the I0 model.

5. Discussion

The present study points to the importance of including the effects of solvent molecules to predict both substrate binding and degree of enantiodifference in lipase catalysed esterification of secondary alcohols.

Inclusion of solvent accessible surface area increased internal predictive performances. Consideration of the crystallographically determined water molecules into the binding pocket enabled more reliable molecular modelling of a substrate binding. The model that explicitly includes interactions between tetrahedral intermediate and the neighbouring water molecules (I0+, I0++) correctly predict enantiomeric ratio for all compounds in the test sets (Table 3). Apparently, the water molecules close to the active site play an important role in enantiodifferentiation of secondary alcohols by BCL. From the models derived we can make some educated guess about the role of individual amino acid residues in prediction of enantioselectivity. It seems that amino acid residues Thr18, Tyr23, Phe119, Val123, Leu164, Leu167, Thr251 and Ile290 play important roles in BCL catalysed

esterification of secondary alcohols with an aromatic ring.

The attractive interaction between Tyr23 and the medium substituent of the faster enantiomer positively influences enantioselectivity, while that between the large substituent of the slower enantiomer and Tyr29 negatively influences enantioselectivity.

The fit of the fast-reacting enantiomer into the binding pockets is defined by dispersive van der Waals interactions between: (a) the large substituent and Val123 and Thr251; and (b) the medium substituent and Tyr23 and Ile290. The magnitude of these interactions is positively correlated with enantioselectivity. Favourable van der Waals and electrostatic interactions between the fast-reacting enantiomer and the amino acid residue Thr18 have an opposite effect on enantioselectivity. These interactions increase substrate binding (K_M), but probably hinder the suitable geometrical rearrangement necessary for catalysis and are negatively correlated with $E = RT \ln[(k_{\text{cat}}/K_M)_{\text{fast-reacting}}/(k_{\text{cat}}/K_M)_{\text{slow-reacting}}]$.

Positive solvent accessible surface area term in the ISA model refers to larger solvent accessible surface area in the complex with the slower reacting enantiomer than the faster one. The majority of these terms have positive weights (an exception is the term considering solvent accessible surface area difference of Tyr23). Apparently, the rate of catalysis is proportional to desolvation of the faster reacting enantiomer.

6. Conclusion

The present study describes the application of a modified COMBINE approach to the prediction of enantioselectivity of BCL towards aryloxy-1,2-diols. As a result 3D QSAR models of predictive value were derived and the enzyme residues important for the catalysis (besides the catalytic triad) were identified.

Since 3D structures of the complexes were not available, molecular modelling study was prerequisite for the analysis. Molecular modelling revealed different productive binding for the two secondary alcohol enantiomers and recognised possible substrate–protein interactions. It offered a rational explanation to the fact that compounds with a bulky substituent at the *ortho*-position of the phenyl ring are not catalysed by BCL and provided qualitative explanation for the measured enantioselectivity.

Modified COMBINE analysis enabled quantification of molecular modelling results and recognition of key amino acid residues. Derived 3D QSAR models correctly predicted the fast-reacting enantiomer not only for the 1,2-diols but also for the other secondary alcohols in the external test set and gave rough estimation of their enantiomeric ratio.

Based on this study, the substrate modifications and lipase mutations are suggested aiming to improve the enzyme enantioselectivity towards acyclic-1,2-propanediol derivatives. We expect that increasing the size of the large aromatic substituent (e.g. replacement of phenyl with naphthal-

ene or diphenyl) and adequate modifications of the medium substrate (which would improve its fit into the HH pocket) should improve enantioselectivity. Similar effect on enantioselectivity could be achieved by lipase modifications, namely replacement of Thr251 by Phe.

The mutation of Leu287 to Ala would enlarge the HH pocket and probably facilitate binding of 1,2-propanediols derivatives with a bulky substituent at the *ortho*-position. However, this as well as Tyr23 and Tyr29 to Phe, Val266 to Ile and Val123 to Ala mutations would decrease difference in size and nature of the HA and HH pockets and very probably decrease or even invert the lipase enantioselectivity.

We plan to perform the site-directed mutagenesis proposed by the model and to test the modified BCL for changes in the predicted direction.

The approach described in this work should be applicable to study enzyme catalysed resolution of chiral compounds in general. However, the COMBINE models are rather specific. They are sensitive to fine tuning of the substrate–enzyme interactions but not robust enough. Prediction of enantioselectivity on the external test set is better for compounds that are more similar to those in the test set. Larger and more diverse test set would increase model robustness. Therefore, we have initiated molecular modelling on the other set of BCL secondary alcohol complexes for which experimental data are available.

Acknowledgements

S. Tomić gratefully acknowledges continuing financial support from Alexander von Humboldt Foundation. This work was supported by grants 00980608 from the Ministry of Science and Technology of Croatia and Volkswagen-Stiftung, I/72566, Germany. We acknowledge Dr. R.C. Wade for valuable comments. We thank Gabriele Cruciani for the hospitality and for providing us Golpe programme in his laboratory.

References

- [1] R.J. Kazlauskas, U.T. Bornscheuer, Biotransformations with lipases, in: H.J. Rehm, G. Reed, A. Pühler, P.J.W. Stadler, D.R. Kelly (Eds.), Biotransformations with Lipases in Biotechnology, vol. 8, VCH Publishers, Weinheim, Germany, 1998, pp. 37–191.
- [2] K. Faber, Biotransformations in Organic Chemistry, 3rd ed., Springer, Berlin, 1997.
- [3] R.D. Schmid, R. Verger, Lipases: interfacial enzymes with attractive applications, Angew. Chem. Int. Ed. Engl. 37 (1998) 1608–1633.
- [4] S.M. Roberts, Preparative biotransformations, J. Chem. Soc. Perkin Trans. 1 (2000) 611–633.
- [5] K. Laumen, M.P. Schneider, A highly selective ester hydrolase from *Pseudomonas* sp. for the enzymatic preparation of enantiomerically pure secondary alcohols chiral auxiliaries in organic synthesis, J. Chem. Soc. Chem. Commun. (1988) 598–600.
- [6] R.J. Kazlauskas, A.N.E. Weissfloch, A.T. Rappaport, L.A. Cuccia, A rule to predict which enantiomer of a secondary alcohol reacts faster in reactions catalysed by cholesterol esterase, lipase from

- Pseudomonas cepacia*, and lipase from *Candida rugosa*, J. Org. Chem. 56 (1991) 2656–2665.
- [7] W.V. Tuomi, R.J. Kazlauskas, Molecular basis for enantioselectivity of lipase from *Pseudomonas cepacia* toward primary alcohols. Modelling, kinetics, and chemical modification of Tyr29 to increase or decrease enantioselectivity, J. Org. Chem. 64 (1999) 2638–2647.
 - [8] J.D. Schrag, Y. Li, M. Cygler, D. Lang, T. Burgdorf, H.J. Hecht, R. Schmid, D. Schomburg, T. Rydel, J.D. Oliver, L.C. Strickland, C.M. Dunaway, S.B. Larson, J. Day, A. McPherson, The open conformation of a *Pseudomonas* lipase, Structure 5 (1997) 187–202.
 - [9] K.K. Kim, H.K. Song, D.H. Shin, K.Y. Hwang, S.W. Suh, The crystal structure of a triacylglycerol lipase from *Pseudomonas cepacia* reveals a highly open conformation in the absence of a bound inhibitor, Structure 5 (1997) 173–185.
 - [10] D.A. Lang, M.L.M. Mannesse, G.H. De Haas, H.M. Verheij, B.W. Dijkstra, Structural basis of the chiral selectivity of *Pseudomonas cepacia* lipase, Eur. J. Biochem. 254 (1998) 333–340.
 - [11] M. Luić, S. Tomić, I. Lešćić, E. Ljubović, D. Šepac, V. Šunjić, L.J. Vitale, W. Saenger, B. Kojić-Prodić, Complex of *Pseudomonas cepacia* lipase (PCL) with transition state analogue of 1-phenoxy-2-acetoxybutane: biocatalytic, structural and modelling study, Eur. J. Biochem. 268 (2001) 3964–3973.
 - [12] J. Zuegg, H. Hönig, J.D. Schrag, M. Cygler, Selectivity of lipases: conformational analysis of suggested intermediates in ester hydrolysis of chiral primary and secondary alcohols, J. Mol. Catal. B: Enzymatic 3 (1997) 83–98.
 - [13] F. Haefner, T. Norin, K. Hult, Molecular modelling of the enantioselectivity in lipase-catalysed transesterification reactions, Biophys. J. 74 (1998) 1251–1262.
 - [14] A. Tafi, A. van Almsick, F. Corelli, M. Crusco, K.E. Laumen, M.P. Schneider, M. Botta, Computer simulations of enantioselective ester hydrolysis catalysed by *Pseudomonas cepacia* lipase, J. Org. Chem. 65 (2000) 3659–3665.
 - [15] T. Schultz, J. Pleiss, R.D. Schmid, Stereoselectivity of *Pseudomonas cepacia* lipase toward secondary alcohols: a quantitative model, Protein Sci. 9 (2000) 1053–1062.
 - [16] S. Raza, L. Fransson, K. Hult, Enantioselectivity in *Candida antarctica* lipase B: a molecular dynamic study, Protein Sci. 10 (2001) 329–338.
 - [17] A. Magnusson, K. Hult, M. Holmquist, Creation of an enantioselective hydrolase by engineered substrate-assisted catalysis, J. Am. Chem. Soc. 123 (2001) 4354–4355.
 - [18] C. Heiss, M. Laivenieks, G.J. Zeikus, R.S. Phillips, The stereospecificity of secondary alcohol dehydrogenase from thermoanaerobacter ethanolicus is partially determined by active site water, J. Am. Chem. Soc. 123 (2001) 345–346.
 - [19] F. Theil, J. Weidner, S. Ballschuh, A. Kunath, H. Schick, Kinetic resolution of acyclic-1,2-diols using a sequential lipase-catalysed transesterification in organic solvents, J. Org. Chem. 59 (1994) 388–393.
 - [20] F. Theil, K. Lemke, S. Ballschuh, A. Kunath, H. Schick, Lipase-catalysed resolution of 3-(aryloxy)-1,2-propanediol derivatives—towards an improved active site model of *Pseudomonas cepacia* lipase (Amano PS), Tetrahedron: Asymm. 6 (1995) 1323–1344.
 - [21] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford University Press, Oxford, 1987.
 - [22] P.A. Bash, U.C. Singh, R. Langridge, P.A. Kollman, Free energy calculations by computer simulation, Science 236 (1997) 564–568.
 - [23] A.R. Ortiz, M.T. Pisabarro, F. Gago, R.C. Wade, Prediction of drug binding affinities by comparative binding energy analysis, J. Med. Chem. 38 (1995) 2681–2691.
 - [24] S. Tomić, L. Nilsson, R.C. Wade, Nuclear receptor–DNA binding specificity: a combine- and free-Wilson QSAR analysis, J. Med. Chem. 43 (2000) 1780–1792.
 - [25] R.C. Wade, A.R. Ortiz, F. Gago, Comparative binding energy analysis, Perspect. Drug Discovery Design 9 (1998) 19–34.
 - [26] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
 - [27] S. Tomić, V. Dobovičnik, V. Šunjić, B. Kojić-Prodić, Enantioselectivity of the *Pseudomonas cepacia* lipase towards 2-methyl-3-(or 4)-arylalkanols: an approach based on stereoelectronic theory and the molecular modelling, Croat. Chem. Acta 74 (2001) 343–357.
 - [28] INSIGHTII, Molecular Simulations Inc., San Diego, 1997.
 - [29] W.D. Cornell, P. Cieplak, C.I. Payly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
 - [30] J.J.P. Stewart, MOPAC Manual, 2000.
 - [31] S.J. Hubbard, J.M. Thornton, NACCESS, Department of Biochemistry and Molecular Biology, University College London, London, 1993.
 - [32] S. Wold, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström, in: B.R. Kowalski (Ed.), Chemometrics-Mathematics and Statistics in Chemistry, Reidel, Dordrecht, 1984, pp. 17–95.
 - [33] S. Tomić, R.C. Wade, COMBINE analysis of nuclear receptor–DNA binding specificity: comparison of two datasets, Croat. Chem. Acta 74 (2001) 295–314.
 - [34] G. Cruciani, S. Clementi, M. Baroni, Variable selection in PLS analysis in 3D QSAR, in: H. Kubinyi (Ed.), Drug Design: Theory, Methods and Applications, ESCOM Publishers, Leiden, 1993.
 - [35] D.G. Gascoyne, H.L. Finkbeiner, K.P. Chan, J.L. Gordon, K.R. Stewart, R.J. Kazlauskas, Molecular basis for enantioselectivity of lipase from *Chromobacterium viscosum* towards the diesters of 2,3-dihydro-3-(4'-hydroxyphenyl)-1,1,3-trimethyl-1*H*-inden-5-ol, J. Org. Chem. 66 (2001) 3041–3048.
 - [36] E. Ljubović, V. Šunjić, Effect of distant groups on the enantioselectivity in kinetic resolution of secondary alcohols catalysed by microbial lipases, Croat. Chem. Acta 71 (1998) 99–117.