Quantitative prediction of enantioselectivity of *Candida antarctica* lipase B by combining docking simulations and quantitative structure–activity relationship (QSAR) analysis

Jiali Gu, Ji Liu, Hongwei Yu*

Institute of Bioengineering, Department of Chemical and Biological Engineering, Zhejiang University, Yu Quan, Zhejiang, Hangzhou 310027, PR China

ABSTRACT

Prediction of enzyme enantioselectivity *in silico* could be of major utility for avoiding the expensive and time-consuming experiments. Herein, we aimed to develop a new approach to construct a quantitative enantioselectivity prediction model with high accuracy for *Candida antarctica* lipase B (CALB). In the work, Autodock was used to generate substrate conformations for improving the calculation efficiency, followed by the quantitative structure–activity relationship (QSAR) analysis. The effects of acyl donors and 5 molecular interaction fields (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields) on model construction were investigated. The results indicated that the application of actual acyl donors was indispensible for model construction. Inclusion of the relevant molecular interaction fields could significantly improve the predictive accuracy which suggested that enantioselectivity was a consequence of multiple molecular interactions. The final model was derived based on four molecular interaction fields (steric, electrostatic, hydrophobic, hydrogen bond acceptor fields) with actual acyl donors owning higher predictive accuracy ($R^2_{pred} = 0.92$) than previous report ($R^2_{pred} = 0.79$). Furthermore, the contour map produced by the model facilitated us to better elucidate the molecular basis of enzyme enantioselectivtiy, and was potential for the application of rational design of the enzyme.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Microbial lipases have been proven useful biocatalysts to obtain chiral compounds and intermediates. A substantial amount of research has been devoted to elucidate the mechanism and further

Abbreviations: A, hydrogen bond acceptor; CALB, Candida antarctica lipase B; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity analysis; cSDEP, the calculated cross-validated standard error of prediction in progressive scrambling; CT, active carbonyl carbon; D, hydrogen bond donor; DMIFs, differential molecular interaction fields; $dq^{2\prime}/dr^2_{yy\prime}$, slope of q^2_{ps} with respect to the correlation of the original dependent variables versus the scrambled dependent variables in progressive scrambling; E, electrostatic; *E*, enantiomeric ratio; Fratio, F-test value; H, hydrophobic; HB, hydrogen bonds; LOOq², leave one out cross-validated squared correlation coefficient; MD, molecular dynamics; MIFs, molecular interaction fields; MM, molecular mechanics; NULL, no field involved; ONC, optimum number of components; PLS, partial least square; QSAR, quantitative structure–activity relationship analysis; QM, quantum mechanics; q^2, the perturbation prediction in progressive scrambling; R^2, non-cross-validated correlation coefficient; R^2_{BS}, average correlation coefficient for bootstrapping; R^2_{pred}, the predictive correlation coefficient of the test set; *S*, steric; SEE, standard error of estimation; SEEBS, average standard error of estimate for bootstrapping; TI, tetrahedral intermediate.

* Corresponding author. Tel.: +86 571 8795 1873; fax: +86 571 8795 1873.
E-mail address: yuhongwei@zju.edu.cn (H. Yu).

improve the enantioselectivity of enzyme [1–7]. With the current advances in computational science, prediction of enantioselectivity of enzyme towards substrate molecules becomes one of the most heavily investigated areas [8–12] to avoid tedious conventional high throughput screening.

Generally, the methods for enantioselectivity prediction can be divided into 2 categories: one is based on the free energy difference between the fast- and slow-reacting enantiomers, such as quantum mechanics (QM) or molecular mechanics (MM), and the other is based on the structural difference, like quantitative structure–activity relationship (QSAR) analysis. While QM is too time-consuming to be attractive as predicting tools, MM often produces unacceptable predictive error as a large number of assumptions and approximations are made to improve the computation efficiency [2,10,13,14]. In contrast, QSAR, which correlates the molecule structure properties with the bioactivity based on statistical analysis, has been widely used for prediction with a reasonable accuracy in drug design [15,16]. However, success of a QSAR model strongly depends on the alignment of substrates [17]. Therefore, MM, an ideal substrate conformation searching and alignment approach, has been combined with QSAR to give satisfying qualitative predictions of enzyme activity or enantioselectivity [9,18,12,19,20]. In a pioneering work, Tomić and Kojic-Prodic proposed a 3D QSAR COMBINE approach to predict the

enantioselectivity of *Burkholderia cepacia* based on the assumption that the free energy differences between the two stereoisomeric lipase-substrate transitional intermediates could be expressed by weighted amino acid residue based energy differences. Nevertheless, the selection of residues involved in the free energy calculation was a serious drawback, which would significantly influence the resulting free energy values [12]. Braiuca et al. calculated enantiomeric ratio as the ratio of specificity constants of penicillin G amidase towards two enantiomers [19]. As little data was available about the specificity constant of enzyme towards molecules, the application of the model was hindered.

In a recent study, Braiuca et al. developed another model by application of differential molecular interaction fields (DMIFs) [9]. The predictive accuracy of the model reached 0.79. Nevertheless, the problem of efficiency mainly arising from the substrate conformation calculation step accomplished by molecular dynamics simulation (MD) could not be ignored. As MD uses a sophisticated force field, it is time-consuming and costly. In contrast, Autodock, based on a genetic algorithm, explores the vast conformational space of ligands in a short time and supplies a large library of candidates at a reasonable cost. Nevertheless, due to poor flexibility of the protein and lack of a widely applicable scoring function, the final substrate conformation could be unacceptable. Therefore, 2 conformational criteria were used in the paper to screen the docking conformations for generation of a more reliable substrate library [21–23]. Besides, in the model proposed by Braiuca et al., only "water" and "dry" probes were used in the calculation of molecular interaction fields. As enzymatic catalysis is a complex process involving various molecular interactions (such as hydrogen bond, ionic and lipophilic interactions) between substrates and enzyme, these two probes are not efficient to describe all above interactions. Thus, inclusion of more enantioselectivity relevant molecular interaction fields for model construction may further improve the model predictive accuracy.

Previous studies have indicated that the length of acyl chain would affect the success of kinetic resolution of enantiomers [24–26]. Ottosson and Hult proposed that the deep narrow cleft of the active site in CALB restricted the access of hairpin-shaped ester depending on the length of acyl chain [25]. It was also reported that the acyl donor might exert its influence through electron-withdrawing effect and resonance effect [27]. Therefore, the contribution of the acyl donor to the enantioselectivity prediction model must be investigated.

In the present study, we aim to derive a comprehensive and accurate quantitative enantioselectivity prediction model for CALB by combination of QSAR and docking simulations. Contributions of acyl donor and 5 relevant molecular interaction fields (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields) to the model construction will be explored and the origin of the enzyme enantioselectivity will be investigated based on the model. Furthermore, the potential of the contour map produced by the model in the rational design of the enzyme will be discussed.

2. Materials and methods

2.1. General strategy

The general strategy (Fig. 1) to derive the comprehensive QSAR model with high precision included the following four steps: (1) exploration of acyl donor contribution to the model predictive ability, (2) generation and analysis of CoMFA model based on step 1, (3) exploration of different molecular interaction field contribution to the model predictive ability, and (4) construction and analysis of CoMSIA model based on the best field combination. In step 1, 2 different CoMFA models were constructed, one with vinyl acetate
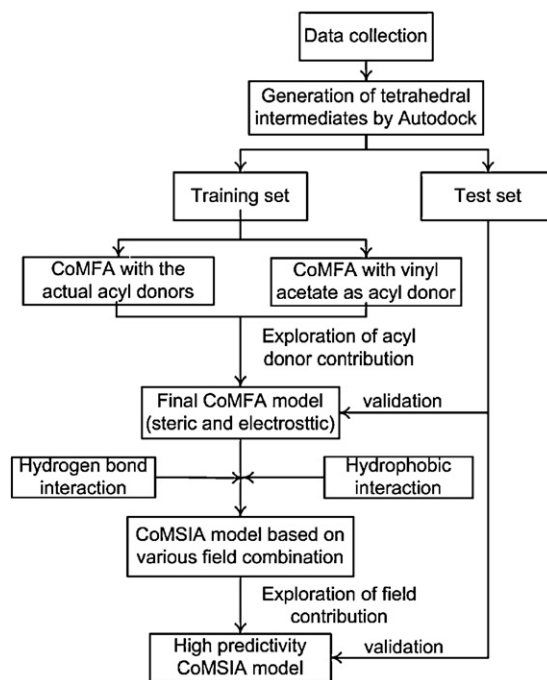

Fig. 1. General strategy for construction of the quantitative prediction model.

as acyl donor (Model 1) and the other with the actual acyl donors (Model 2). In step 3, CoMSIA models based on various molecular interaction fields combinations (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields) were constructed to figure out those enantioselectivity relevant fields. In step 2 and step 4, the CoMFA/CoMSIA models were constructed in standard implementation.

2.2. Data set

28 CALB catalyzed acylation reactions were selected from literature [5,28–36]. Only resolutions of alcohols were considered in order to generate a reliable model for predicting the enantioselectivity of CALB towards alcohol. It was different from the substrate set used by Braiuca et al., where amines were also included. Tetrahedral intermediate (TI) formed during acylation was used as substrate (Table 1). Enantiomeric ratio (E) of the reaction was converted into $pE(-\log E)$ and used as dependent variable. A total set of 28 substrates was divided into a training set (17 substrates) for model generation and a test set (11 substrates) for model validation. Both the training set and the test set were divided according to a representative range of pE values and acyl donor variation.
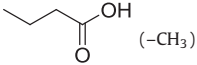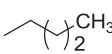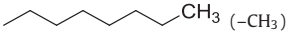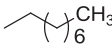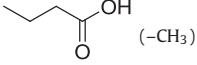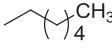
2.3. Generation of active substrate conformation

All tetrahedral intermediates (TIs) used for the construction of the data set were generated by Autodock 4.0 [21] on a 1.8 GHz Pentium Dual E2160 PC according to our previous report [8]. Chemoffice 2008 program (Cambridge Soft Corporation) was used to generate the initial TI structure, Chimera software [37] was used to minimize the energy and calculate the partial charges. 100 possible docking conformations of each substrate were obtained, all of them satisfied the following criteria: (1) the distance between serine OG atom and substrate CT atom was less than 1.7 Å, and (2) all catalytic hydrogen bonds (HBs) were formed [38], including HBs between alcohol oxygen of TI and HE atom of the protonated His 224, and ones between the carbonyl oxygen of TI and the

Table 1

Structures of tetrahedral intermediates

$$R_2 \overset{R_3}{\underset{}{\diagdown}} O \overset{O}{\underset{}{\diagup}} R_1$$

Sub.	R_1	R_2	R_3
Training set			
1	$-CH_3$ ($-CH_3$)	(phenyl)	$-CH_3$
2	$-CH_3$ ($-CH_3$)	(phenyl)	(alkyne)
3	$-CH_3$ ($-CH_3$)	(cyclohexyl)	$-CH_3$
4	$-CH_3$ ($-CH_3$)	(4-F-phenyl)	$-CH_3$
5	$-CH_2CH_2CH_3$ ($-CH_3$)	$-CH_2-Cl$	$-CH_3$
6	$-CH_2CH_2CH_3$ ($-CH_3$)	$-CH_2-Br$	$-CH_3$
7	(butanoic acid –OH, O) ($-CH_3$)	$(CH_2)_3CH_3$	$-CH_3$
8	(butanoic acid –OH, O) ($-CH_3$)	(branched alkene with two CH_3)	$-CH_3$
9	(octyl CH_3) ($-CH_3$)	$(CH_2)_4CH_3$	$-CH_3$
10	(octyl CH_3) ($-CH_3$)	$(CH_2)_4CH_3$	(alkene)
11	(butanoic acid –OH, O) ($-CH_3$)	$-CH_2CH_2CH_3$	$-CH_3$
12	$-CH_3$ ($-CH_3$)	$(CH_2)_3CH_3$	(alkene)
13	$-CH_3$ ($-CH_3$)	$(CH_2)_3CH_3$	(alkyne)
14	$-CH_3$ ($-CH_3$)	$(CH_2)_4CH_3$	$-CH_2CH_2CH_3$
15	$-CH_3$ ($-CH_3$)	$-CH_2CH_3$	(alkyne)
16	(heptyl CH_3) ($-CH_3$)	$(CH_2)_4CH_3$	$-CH(CH_3)_2$ (isopropyl)
17	(heptyl CH_3) ($-CH_3$)	$(CH_2)_4CH_3$	$-C(CH_3)_3$ (tert-butyl)
Test set			
18	$-CH_3$ ($-CH_3$)	(4-methylphenyl)	$-CH_3$
19	$-CH_3$ ($-CH_3$)	(methoxy-naphthyl)	$-CH_3$
20	$-CH_3$ ($-CH_3$)	(ethyl-phenyl)	$-CH_3$
21	$-CH_3$ ($-CH_3$)	(phenyl)	(allyl/alkene)
22	$-CH_3$ ($-CH_3$)	(phenyl-alkyne/alkene)	$-CH_3$
23	$-CH_3$ ($-CH_3$)	$(CH_2)_9CH_3$	$-CH_3$
24	$-CH_3$ ($-CH_3$)	(dienyl chain CH_3)	$-CH_3$

Table 1 (*Continued*)

Sub.	R_1	R_2	R_3
25	(butanoic acid / ester structure) (–CH$_3$)	(branched alkyl, CH$_3$ with subscript 2)	–CH$_3$
26	(heptyl chain) CH$_3$ (–CH$_3$)	(branched alkyl, CH$_3$ with subscript 6)	(alkyne ≡)
27	(butanoic acid / ester structure) (–CH$_3$)	–CH$_2$CH$_2$CH$_3$	–CH$_2$CH$_3$
28	–CH$_2$CH$_2$CH$_3$ (–CH$_3$)	(branched alkyl, CH$_3$ with subscript 4)	–CH$_2$–Cl

Sub.: substrate number; R_1 group (methyl) in the bracket is used for construction of CoMFA model with vinyl acetate as acyl donor (Model 1).

oxyanion hole residues (Gln106 and Thr40), which were clearly depicted by Ottosson et al. [39]. The final TI conformation chosen for construction of the data set was the one characterized by the lowest potential energy out of the above 100 docking conformations.

2.4. Comparative molecular field analysis (CoMFA) and comparative molecular similarity analysis (CoMSIA) interaction energies

In comparative molecular field analysis (CoMFA), the steric (Lennard–Jones 6-12 potential) and electrostatic (Coulombic potential) molecular interaction fields were calculated by placing the aligned molecules into a 3D cubic lattice with a 2 Å grid spacing and 4 Å unit extending in all directions. A sp^3 carbon probe atom with charge of +1.0 was used. The steric and electrostatic energy values were truncated at 30 kcal mol^{-1} to avoid infinity of energy values inside the molecule. All parameters were used with the default SYBYL 8.0 (Tripos, St. Louis, MO) settings.

In comparative molecular similarity analysis (CoMSIA), 5 molecular interaction fields (steric, electrostatic, hydrophobic, hydrogen bond (HB) donor and acceptor fields) were calculated with the same lattice box used in CoMFA and a sp^3 carbon probe atom with charge of +1, hydrophobicity of +1, HB donor of +1 and HB acceptor of +1 properties. All parameters were used with the default SYBYL 8.0 (Tripos, St. Louis, MO) settings.

2.5. Calculation of differential molecular interaction fields (DMIFs)

As enantioselectivity is a property of 2 enantiomers, the structural information contained in the molecular interaction fields (MIFs) of each enantiomer had to be merged into a single entity. Differential molecular interaction fields (DMIFs) proposed by Braiuca et al. [9] was applied and used as independent variables instead of MIFs. DMIF was generated by subtracting each variable in MIF of the slow-reacting enantiomer from the corresponding variable in MIF of the fast-reacting enantiomer. Consequently, in DMIFs, null values or small absolute values indicated that the 2 enantiomers established identical or similar molecular interactions with the enzyme, whereas high absolute values indicated that the enantiomers established different interactions with the enzyme.

2.6. Partial least square (PLS) analysis and models validation

Partial least squares analysis [40,41], an extension of multiple regression analysis, was used to derive the QSAR model. The model was first evaluated by leave-one-out cross-validation method which gave the LOOq^2 value as a statistical index of predictive power [42]. LOOq^2 is defined as:

$$q^2 = 1 - \frac{\sum y - y_{\text{pred}}}{\sum (y - \widetilde{y})^2}$$

where y and y_{pred} are experimental and predicted pE values of the training set compounds, respectively; $\widetilde{y}$ is the mean value. To minimize the tendency of overfit, optimum number of components (ONC) was chose according to both values of LOOq^2 and standard error of estimation (SEE). The final model was derived using non-cross-validated PLS analysis with ONC.

The model was further evaluated by the following methods.

1) The predictivity of model was evaluated by the test set (details in supplemental material). The predictive correlation coefficient (R^2_{pred}) was calculated using the same formula as LOOq^2, except that y and y_{pred} are changed to the corresponding experimental and predicted pE values of the test set.
2) The statistical confidence of the model was assessed by bootstrapping analysis [43] (100 runs) with ONC. Bootstrapping involves the random generation of new datasets from the original one, during which, one sample may be extracted more than once in one run, or it may never be extracted in whole runs. The statistical calculation is performed for each of the bootstrap samplings and gives the statistical parameter of average correlation coefficient for bootstrapping (R^2_{BS}) and average standard error of estimate for bootstrapping (SEE$_{\text{BS}}$).
3) The sensitivity of the model to chance correlations was evaluated by progressive scrambling [44]. It produces three statistics, namely the perturbation prediction (q^2), the calculated cross-validated standard error of prediction (cSDEP), and slope of (q^2_{ps}) with respect to the correlation of the original dependent variables versus the scrambled dependent variables ($dq^{2'}/dr^2_{yy'}$). The susceptibility of the model to chance correlation is gauged by $dq^{2'}/dr^2_{yy'}$. An ideal model has the slope equal to 1, in a model where the correlation of activity is not by chance, the slope should be from the range of 0.8–1.2.

3. Results and discussion

3.1. Contribution of acyl donor to the model construction

In Braiuca et al. work, only three different acyl donors were considered in model construction, as they aimed to quantify the effect of the nucleophile structure on enantioselectivity. Since a considerable amount of work had indicated that acyl donor significantly influenced the enantioselectivity of CALB [26,45,46], Herein, we explored the contribution of acyl donor to the model predictive performance. Comparative molecular field analysis (CoMFA), one of the most popular tools in QSAR analysis since its publication by Cramer et al. [47,48], was applied. Two CoMFA models were constructed: Model 1 with vinyl acetate as acyl donor for all substrates, and Model 2 with the actual acyl donors, including vinyl acetate, S-ethyl thiooctanoate, succinic anhydride and vinyl butyrate. The result clearly showed that Model 2 performed much better ($LOOq^2 = 0.612$) than Model 1 ($LOOq^2 = -0.007$), indicating the importance of acyl donor on model construction.

3.2. Generation and analysis of CoMFA model

Based on the above results, the CoMFA model with actual acyl donors was generated. It employed non-cross-validated PLS analysis with ONC. To further assess the statistical confidence limits of the derived models, bootstrapping analysis was carried out with 100 runs. A R^2_{ps} of 0.993 ± 0.007 and a SEE_{BS} of 0.085 ± 0.085 suggested a good internal consistency of the model. Besides, the progressive scrambling was performed to evaluate the sensitivity of the CoMFA model to chance correlations. In case of ONC, the slope of q^2 ($dq^{2'}/dr^2_{yy'} = 1.07$) was close to 1, indicating the model was not derived by chance correlation. All statistical parameters supporting the CoMFA model are listed in Table 2 and the weight values of dependent variables and intercept values are listed in supplementary material.

The predictive ability of CoMFA model was further validated by the test set. The predicted pE values were in agreement with the experimentally obtained values, except for substrates 25 and 27, as shown in Fig. 2 (raw data in Table S1). However, substrates 7, 8, 11 in the training set, which have similar structures, were predicted accurately. McCabe and Taylor [49] correlated the lack of activity of CALB towards the 2-oxo-adipic acid with the formation of HB between the 2-oxo-carbonyl of this substrate and the Thr40 side chain hydroxyl, illustrating the importance of this HB to stabilize the oxyanion in the tetrahedral intermediate. Thus, we suspected whether there were additional HBs formed between these substrates and CALB except the catalytic HBs [38] since all these substrates possess succinic anhydride as acyl donor. Analysis

Table 2
Summary of the results obtained from the CoMFA and CoMSIA analyses.

Statistical parameters	CoMFA model	CoMSIA model
$LOOq^2$	0.612	0.755
ONC	4	6
R^2	0.989	0.997
SEE	0.113	0.063
F_{ratio}	279.259	599.108
R^2_{pred}	0.81[a]	0.92
R^2_{BS}	0.993 ± 0.007	0.998 ± 0.002
SEE_{BS}	0.085 ± 0.085	0.045 ± 0.049
q^2	0.465	0.581
cSDEP	0.803	0.784
$dq^{2'}/dr^2_{yy'}$	1.037	1.071
Contribution		
Steric	0.567	0.289
Electrostatic	0.433	0.309
Hydrophobic	–	0.143
Hydrogen bond acceptor	–	0.259

Abbreviation: $LOOq^2$: correlation coefficient of the leave-one-out cross-validation, ONC: optimum number of components, R^2: correlation coefficient of non-cross-validation, SEE: standard error of estimate, F_{ratio}: F-test value, R^2_{pred}: predictive correlation coefficient for the test set, R^2_{BS}: average correlation coefficient for boot-strapping, SEE_{BS}: average standard error of estimate for bootstrapping, q^2: the perturbation prediction in progressive scrambling, cSDEP: the calculated cross-validated standard error of prediction in progressive scrambling, $dq^{2'}/dr^2_{yy'}$: slope of q^2_{ps} with respect to the correlation of the original dependent variables versus the scrambled dependent variables in progressive scrambling.

[a] Predictive correlation coefficient for the test set excluding substrates 25 and 27.

of the Autodock results revealed that a HB was formed between the terminal carboxyl hydrogen of sub.25 and Ile 189 which might account for the underestimation of sub.25. However, no HBs were formed between the other four substrates and CALB. As there were only steric and electrostatic fields calculated in the CoMFA model, we speculated that the model was inefficient to extract enough useful information, thus we were encouraged to include more molecular interaction fields for model construction and explore whether it could improve the model predictive accuracy.

3.3. Contribution of various molecular interaction fields to model construction

Herein, CoMSIA, an extension of CoMFA method, was applied to investigate the contribution of molecular interaction fields to model construction because it involves more fields [50,51]. The model with addition of the explored molecular interaction field was named as new model, respected to the original model without the corresponding explored field. Each model was generated using the number of component giving the highest $LOOq^2$ value. The predictive ability of the model was determined by the $LOOq^2$ value. The explored field was believed necessary for model construction
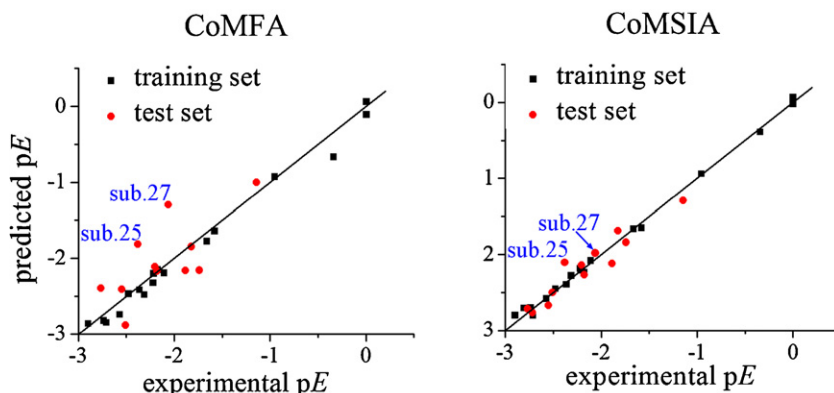

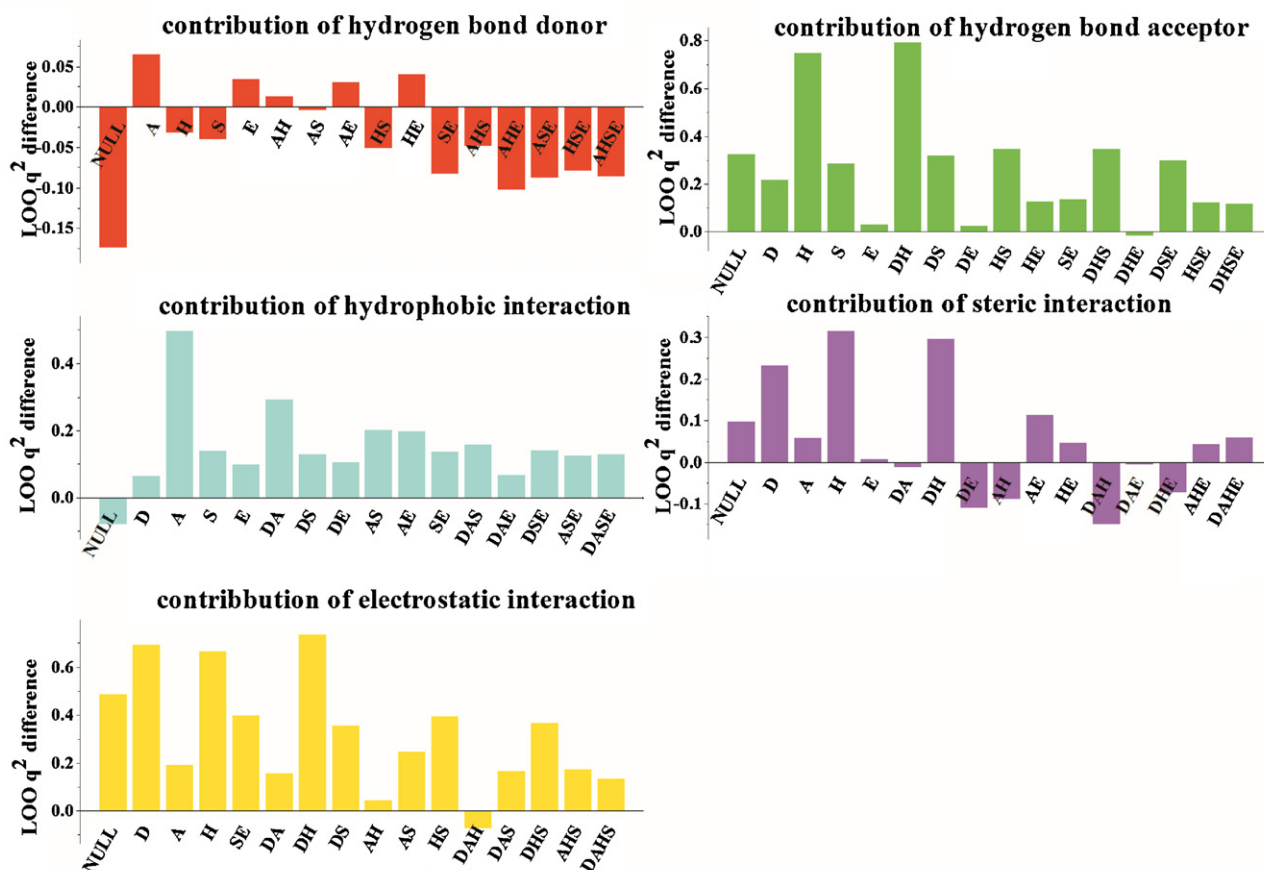**Fig. 2.** Plots of experimental pE versus predicted pE based on CoMFA and CoMSIA models.

Fig. 3. Exploration of field contribution to model construction.

if more than half of the new models (>7) possessed higher LOOq^2 value than the original one.

As shown in Fig. 3 (raw data in Table S2), steric, electrostatic, hydrophobic and hydrogen bond acceptor molecular interaction fields all significantly improved the predictive ability of the original model, as most of the new models possessed higher LOOq^2 values than the original one (steric: 10/15; electrostatic: 14/15; hydrophobic: 15/15; hydrogen bond acceptor: 14/15). Among the models based on one molecular interaction field, the one generated by the electrostatic field owned the highest LOOq^2 value (LOOq^2 = 0.486), suggesting that the electrostatic interaction might be the most important factor. In contrast, the model based on the hydrophobic field alone possessed a negative LOOq^2 (LOOq^2 = -0.078), indicating that the hydrophobic interaction was an important but not determinant factor in discrimination of enantiomers by CALB. However, in amount of work, the different steric and electrostatic interactions between 2 enantiomers and the active site residues were explored and used for explaining the enantiopreference of CALB [5,52,53]. Thus, the result might raise another perspective to understand the origin of enantioselectivity, as hydrophobic interactions indeed affect the regioselective behavior of CALB, according to the results from Oliveira et al. [54]. The hydrogen bond donor field was excluded for the final model construction due to the worse prediction performances of most of the new models (9/15). It was probably because that the structural information about the hydrogen bond donor was too little to be used for statistical analysis since there were only three hydrogen bond donors (sub.7, 8, 11) in the model.

To systematically elucidate the role of molecular interaction field on model predictive ability, the performance of models based on 1, 2, 3, 4 fields were compared (Fig. 4), in which the highest LOOq^2 values are represented. The LOOq^2 values gradually augmented, the predictive accuracy of the model based on EHAS (LOOq^2 = 0.755) was much higher than that of the model based on E alone (LOOq^2 = 0.486). The results clearly showed that the model based on EHAS was more reliable for prediction, and reflected that the enzyme enantioselectivity was a consequence of multiple molecular interactions. Moreover, the results might also explain why our model gave better prediction than the model proposed by Braiuca et al. [9], excepting different substrate conformation searching tools and substrates set used for model construction.
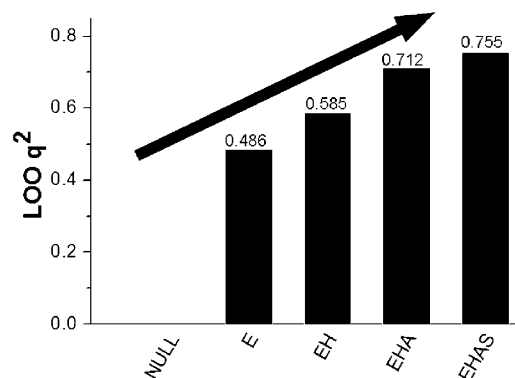

Fig. 4. LOOq^2 values of CoMSIA model based on E, EH (E + H), EHA (E + H + A), EHAS (E + H + A + S).
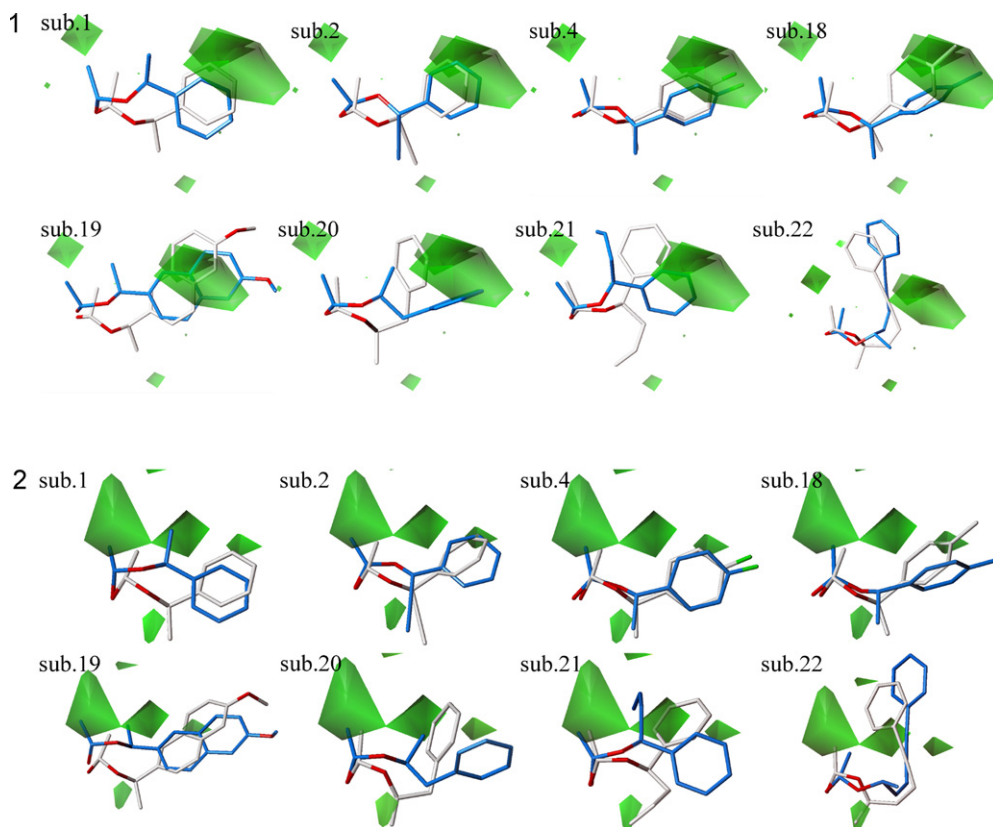
Fig. 5. Steric contour maps of CoMFA and CoMSIA models with substrates with benzene/naphthalene.
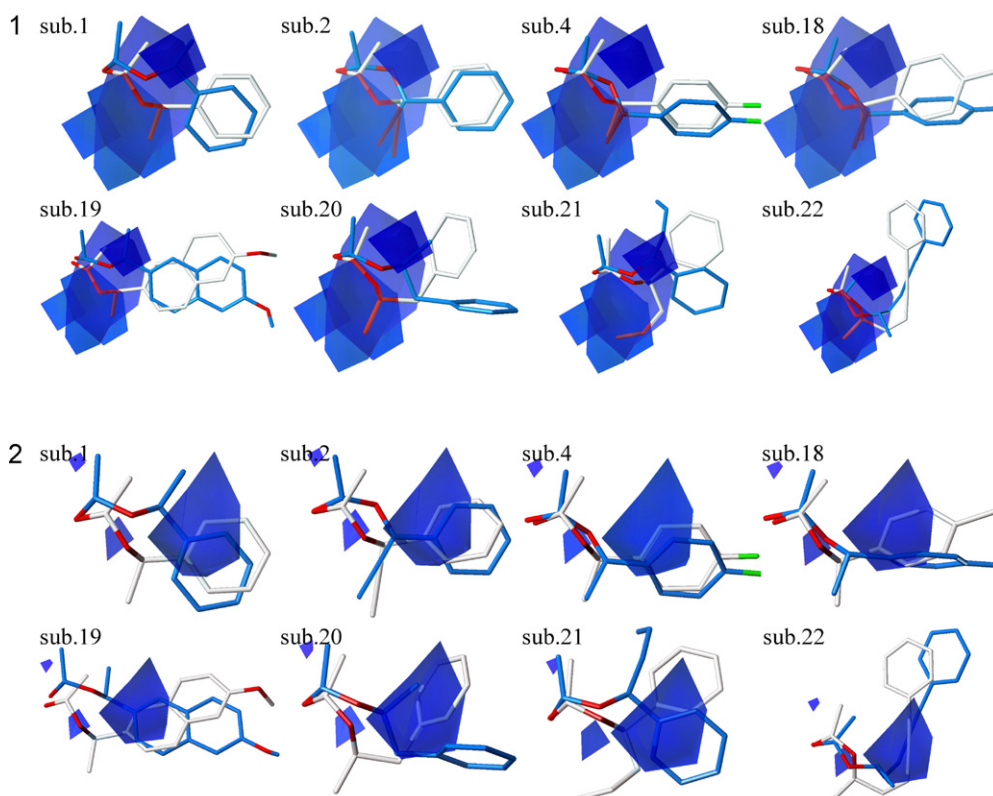

Fig. 6. Electrostatic contour maps of CoMFA and CoMSIA models with substrates with benzene/naphthalene.

3.4. Generation and analysis of CoMSIA models with the best field combination

The final CoMSIA model was derived based on steric, electrostatic, hydrophobic, hydrogen bond acceptor fields, and the weight values of dependent variables and intercept values were listed in the supplementary material. The contribution of each molecular interaction field to the model was 0.289, 0.309, 0.143, and 0.259, respectively. It clearly showed that the electrostatic interaction was the most important while the hydrophobic interaction was less important. The bootstrapping analysis suggested a good internal consistency of the model, while progressive scrambling demonstrated the model did not dependent on chance correlation (Table 2). The CoMSIA model was further validated by the same test set as used in CoMFA and the results demonstrated that the model was robust to predict both high and low enantioselectivity, as shown in Fig. 2 (raw data in Table S1). Besides, the CoMSIA model was more predictive ($R^2_{pred} = 0.92$) than the CoMFA model ($R^2_{pred} = 0.81$), especially for compounds with phenyl/naphthalene ring (sub.18–22). Previous report has illustrated that the steric and resonance effect induced by the phenyl ring of substrate

would dramatically change the catalytic property of enzyme [27]. The CoMFA and CoMSIA steric and electrostatic contour maps were compared (Figs. 5 and 6), which respectively displayed the regions where important steric and electrostatic interactions took place. The area covering benzene/naphthalene ring was shown in both CoMFA and CoMSIA steric contour maps. However, regarding the electrostatic contour map, only in the CoMSIA model, the electrostatic interaction between benzene/naphthalene and CALB was predicted to be important for enantioselectivity, which might account for the higher predictive accuracy in case of sub.18–22. Another huge improvement was observed for substrates 25 and 27 which were poorly predicted in CoMFA (SEE = −0.638, SEE = 0.839 for sub.25 and sub.27, respectively) for lack of the hydrogen bond interaction, though there was only hydrogen bond acceptor field involved in CoMSIA (SEE = −0.273, SEE = 0.078 for sub.25 and sub.27, respectively). In both CoMFA and CoMSIA models, the predictive errors of sub.20 were relative big (SEE = −0.2149, SEE = −0.3149 in CoMFA and CoMSIA respectively). It might due to the special molecular property of sub.20 itself, as a number of reports have demonstrated that the reaction selectivity (regioselectivity or enantioselectivity) was not only
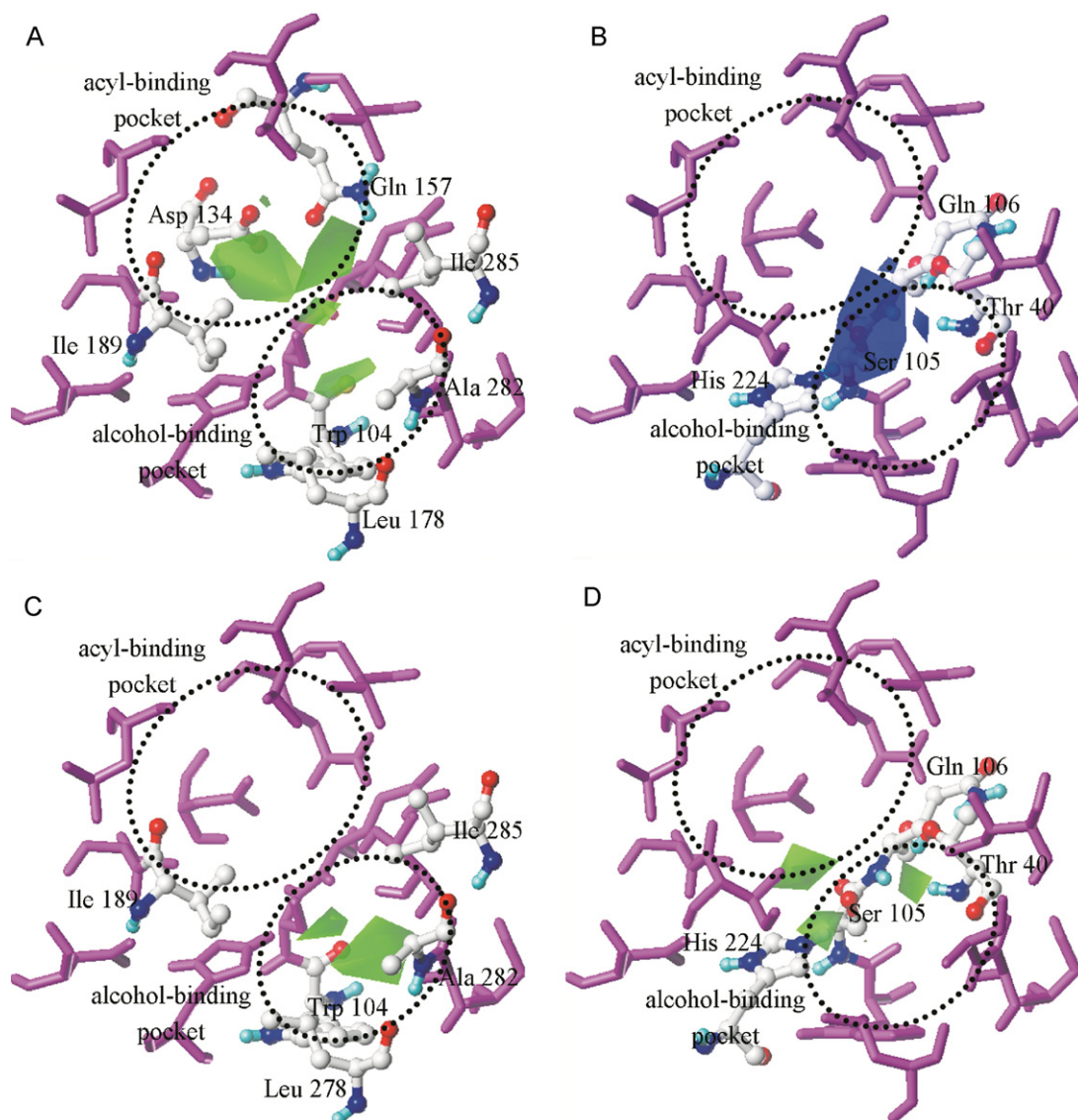

Fig. 7. COMSIA contour maps with the active sites of CALB.

dependent on the interactions between substrate and enzyme, but also on the chemical reactivity of the substrate molecule itself [55–57].

As aforementioned, more molecular interaction fields were considered in CoMSIA. Besides, a Gaussian-type function was used for molecular interaction calculation, which avoided the occurrence of singularities at the atomic position, overcame the problem of arbitrary determination of steric and electrostatic cut-off values in CoMFA [50]. All these factors might account for the higher prediction accuracy in CoMSIA model. However, due to the CoMFA/CoMSIA methods are based on the principle that a suitable sampling of molecular interactions surrounding a set of ligand molecules provides all the information necessary for the observed biological properties, so the rules of other factors, such as solvent, are ignored. Hult et al. had already revealed the influences of solvent molecules in the CALB catalysis, and proposed that solvent acted as a competitive inhibitor for CALB [39,58]. Tomić et al. also pointed out that inclusion of solvent molecules in model construction significantly improved the model predictive performance [18,12]. Therefore, the application of other approaches to complement the effect of solvent was necessary in the future work.

3.5. Analysis of CoMFA/CoMSIA contour map

As well as Tomić et al. work [18,12], the amino acid residues which were the most responsible for the enantiopreference of CALB towards substrates were indentified. They were displayed in the contour map and allowed us a better understanding of the molecular basis of CALB enantioselectivity. The CoMSIA steric contour (Fig. 7A) displayed three regions: one surrounded by the residues Ala282, Leu278, Ile285 and Ile189, the other one right above the residue Trp104 and the third one in the vicinity of the residues Asp134 and Gln157 in the acyl-binding pocket. All above residues were supposed to exhibit their influence through steric effect. Since residues Ala282, Leu278, Ile285 and Ile189 locate at the narrow tunnel leading to CALB active site, the steric hindrance between these residues and substrate dramatically affect the access of substrate to the catalytic residues, in order to form the tetrahedral intermediate. The residue Trp104, which makes up the bottom of stereospecificity pocket, strictly delimites the size of substrate group and unfavorable steric interactions will arise for the bulky substrate group. The electrostatic contour (Fig. 7B) was mainly around the residues Gln106, Thr40, His224 and Ser105, which are essential catalytic functional units. Residues Ser105 and His224 are directly involved in the reaction serving as nucleophilic attacking group and general acid–base catalytic elements, respectively. Residues Gln106 and Thr40 constitute the oxyanion hole to stabilize the high-energy intermediate through hydrogen bond. The hydrophobic contours displayed the areas in the proximity of the residues Ala282, Leu278, Ile285, Ile189 and Trp104 (Fig. 7C), which consist the hydrophobic entrance tunnel of CALB [52]. However, no region was shown in the acyl-binding pocket. Since the crystallographic structure of CALB also illustrated that the acyl-binding pocket was quite hydrophobic [59], thus it was not possible to affirm with certainty that the enantioselectivity was independent from the hydrophobic interaction between enzyme and acyl donor. In the hydrogen bond acceptor map (Fig. 7D), the donor groups of CALB were predicted to be located in the areas around the residues Gln106, Thr40 and His224, Ser 105. With the contour maps, we acquired a better understanding of enantioselectivity of CALB. More importantly, the contour map efficiently demonstrated the important residues, the role of which had been confirmed by other groups [52,60]. Though, the positive or negative influence of these residues on enantioselectivity, which were clearly revealed by Tomić et al., could not be identified in this

article, the contour map was still potential for rational design of CALB.

4. Conclusion

In this study, docking simulations and QSAR analysis were combined to derive a quantitative enantioselectivity prediction model. Owing to the application of Autodock, the computational time was significantly reduced. A whole QSAR model including a data set of 28 compounds could be developed within about 5 days using standard computational facilities. The results indicated that the enzyme enantioselectivity was a consequence of multiple molecular interactions with different contributions. Among the explored five molecular interaction fields, contribution of the electrostatic field to enantioselectivity was the most significant. The model was derived successfully based on actual acyl donors, which could be applied for screening of an ideal acyl donor for a given reaction. Besides, the acyl donor was possible to affect the enantiorecognition of enzyme through steric effect as shown in the steric contour map. However, it remained unclear whether the hydrophobic interaction between the acyl donor and the enzyme would influence the enantioselectivity. The contour map produced by the model could indicate the amino acid residues playing important roles in enzyme enantioselectivity to facilitate the understanding of the origin of enantioselectivity at molecular level, which was potential for the rational design of the enzyme. This modeling protocol must now be applied to other lipases and substrate, in order to confirm its efficiency as a predictive tool for the enantioselectivity in lipase-catalyzed reactions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molcatb.2011.06.011.

References

[1] M. Cygler, P. Grochulski, R.J. Kazlauskas, J.D. Schrag, F. Bouthillier, B. Rubin, A.N. Serreqi, A.K. Gupta, J. Am. Chem. Soc. 116 (1994) 3180–3186.
[2] T. Ema, Curr. Org. Chem. 8 (2004) 1009–1025.
[3] A.J.J. Straathof, J.A. Jongejan, Enzyme Microb. Technol. 21 (1997) 559–571.
[4] T. Xu, L.J. Zhang, E.Z. Su, D.B. Cui, X.D. Wang, D.Z. Wei, J. Mol. Catal. B: Enzym. 62 (2010) 288–296.
[5] D. Rotticci, J.C. Rotticci-Mulder, S. Denman, T. Norin, K. Hult, Chembiochem 2 (2001) 766–770.
[6] P. Berglund, Biomol. Eng. 18 (2001) 13–22.
[7] P.L.A. Overbeeke, S.C. Orrenius, J.A. Jongejan, J.A. Duine, Chem. Phys. Lipids 93 (1998) 81–93.
[8] J. Liu, X.L. Tang, H.W. Yu, Biotechnol. Bioeng. 105 (2010) 687–696.
[9] P. Braiuca, K. Lorena, V. Ferrario, C. Ebert, L. Gardossi, Adv. Synth. Catal. 351 (2009) 1293–1302.
[10] Y. Zhou, in, PhD thesis, Delft University of Technology, 2006.
[11] J.C. Hermann, E. Ghanem, Y.C. Li, F.M. Raushel, J.J. Irwin, B.K. Shoichet, J. Am. Chem. Soc. 128 (2006) 15882–15891.
[12] S. Tomic, B. Kojic-Prodic, J. Mol. Graph. Modell. 21 (2002) 241–252.
[13] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, J. Med. Chem. 49 (2006) 5912–5931.
[14] A.R. Leach, B.K. Shoichet, C.E. Peishoff, J. Med. Chem. 49 (2006) 5851–5855.
[15] M.A. Khanfar, D.T.A. Youssef, K.A. El Sayed, Eur. J. Med. Chem. 45 (2010) 3662–3668.
[16] A. Strasser, H.J. Wittmann, Mol. Inf. 29 (2010) 333–341.

[17] T. Tuccinardi, G. Ortore, M.A.L. Santos, S.R.M. Marques, E. Nuti, A. Rossello, A. Martinelli, J. Chem. Inf. Modell. 49 (2009) 1715–1724.
[18] S. Tomic, B. Bertosa, B. Kojic-Prodic, I. Kolosvary, Tetrahedron: Asymmetry 15 (2004) 1163–1172.
[19] P. Braiuca, L. Boscarol, C. Ebert, P. Linda, L. Gardossi, Adv. Synth. Catal. 348 (2006) 773–780.
[20] S.M. Chandrasekaran, S. Bhartiya, P.P. Wangikar, Biotechnol. Bioeng. 94 (2006) 554–564.
[21] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, J. Comput. Chem. 28 (2007) 1145–1152.
[22] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Proteins: Struct. Funct. Bioinf. 47 (2002) 409–443.
[23] H. Alonso, A.A. Bliznyuk, J.E. Gready, Med. Res. Rev. 26 (2006) 531–568.
[24] A. Sobolev, M.C.R. Franssen, B. Vigante, B. Cekavicus, R. Zhalubovskis, H. Kooijman, A.L. Spek, G. Duburs, A. de Groot, J. Org. Chem. 67 (2002) 401–410.
[25] J. Ottosson, K. Hult, J. Mol. Catal. B: Enzym. 11 (2001) 1025–1028.
[26] B.H. Hoff, H.W. Anthonsen, T. Anthonsen, Tetrahedron: Asymmetry 7 (1996) 3187–3192.
[27] Z.Y. Wang, M.H. Zong, Biotechnol. Prog.: Asymmetry 15 (2004) 3117–3122.
[28] C. Raminelli, J.V. Comasseto, L.H. Andrade, A.L.M. Porto, Tetrahedron: Asymmetry 15 (2004) 3117–3122.
[29] Y.H. Wang, R. Wang, Q.S. Li, Z.M. Zhang, Y. Feng, J. Mol. Catal. B: Enzym. 56 (2009) 142–145.
[30] C. Orrenius, N. Öhrner, D. Rotticci, A. Mattson, K. Hult, T. Norin, Tetrahedron: Asymmetry 6 (1995) 1217–1220.
[31] E. García-Urdiales, F. Rebolledo, V. Gotor, Tetrahedron: Asymmetry 12 (2001) 3047–3052.
[32] X.M. Wu, W. Sun, J.Y. Xin, C.G. Xia, World J. Microb. Biotechnol. 24 (2008) 2421–2424.
[33] T. Ohtani, H. Nakatsukasa, M. Kamezawa, H. Tachibana, Y. Naoshima, J. Mol. Catal. B: Enzym. 4 (1998) 53–60.
[34] E. Garcia-Urdiales, F. Rebolledo, V. Gotor, Tetrahedron: Asymmetry 12 (2001) 3047–3052.
[35] F. Rebolledo, R. Liz, J. Chem. Educ. 82 (2005) 930–933.
[36] R. Didier, H. Fredrik, O. Christian, N. Torbjörn, H. Karl, J. Mol. Catal. B: Enzym. 5 (1998) 267–272.
[37] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.
[38] J. Kraut, Annu. Rev. Biochem. 46 (1977) 331–358.
[39] J. Ottosson, L. Fransson, J.W. King, K. Hult, BBA-Protein Struct. 1594 (2002) 325–334.
[40] S. Wold, M. Sjostrom, L. Eriksson, Chemom. Intell. Lab. 58 (2001) 109–130.
[41] M. Clark, R.D. Cramer, Quant. Struct. Act. Relat. 12 (1993) 137–145.
[42] S. Wold, Technometrics 20 (1978) 397–405.
[43] R.D. Cramer, J.D. Bunce, D.E. Patterson, I.E. Frank, Quant. Struct. Act. Relat. 7 (1988) 18–25.
[44] R. Clark, P. Fox, J. Comput. Aided Mol. Des. 18 (2004) 563–576.
[45] R. Chenevert, N. Pelchat, P. Morin, Tetrahedron: Asymmetry 20 (2009) 1191–1196.
[46] K. Hirose, H. Naka, M. Yano, S. Ohashi, K. Naemura, Y. Tobe, Tetrahedron: Asymmetry 11 (2000) 1199–1210.
[47] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[48] M. Clark, R.D. Cramer Iii, D.M. Jones, D.E. Patterson, P.E. Simeroth, Tetrahedron Comput. Methodol. 3 (1990) 47–59.
[49] R.W. McCabe, A. Taylor, Enzyme Microb. Technol. 35 (2004) 393–398.
[50] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130–4146.
[51] G. Klebe, Perspect. Drug Discov. 12 (1998) 87–104.
[52] Z. Marton, V. Léonard-Nevers, P.O. Syrén, C. Bauer, S. Lamare, K. Hult, V. Tranc, M. Graber, J. Mol. Catal. B: Enzym. 65 (2010) 11–17.
[53] A.O. Magnusson, J.C. Rotticci-Mulder, A. Santagostino, K. Hult, Chembiochem 6 (2005) 1051–1056.
[54] E.B. De Oliveira, C. Humeau, L. Chebil, E.R. Maia, F. Dehez, B. Maigret, M. Ghoul, J.-M. Engasser, J. Mol. Catal. B: Enzym. 59 (2009) 96–105.
[55] E.B. De Oliveira, C. Humeau, E.R. Maia, L. Chebil, E. Ronat, G. Monard, M.F. Ruiz-Lopez, M. Ghoul, J.-M. Engasser, J. Mol. Catal. B: Enzym. 66 (2010) 325–331.
[56] S. Fiorucci, J. Golebiowski, D. Cabrol-Bass, S. Antonczak, J. Agric. Food Chem. 55 (2007) 903–911.
[57] S. Antonczak, J. Mol. Struct. 856 (2008) 38–45.
[58] M. Graber, R. Irague, E. Rosenfeld, S. Lamare, L. Franson, K. Hult, BBA-Proteins Proteomics 1774 (2007) 1052–1057.
[59] V. Gotor-Fernández, E. Busto, V. Gotor, Adv. Synth. Catal. 348 (2006) 797–812.
[60] P.B. Juhl, K. Doderer, F. Hollmann, O. Thum, J. Pleiss, J. Biotechnol. 150 (2010) 474–480.