Pergamon

SCIENCE DIRECT®

Tetrahedron: *Asymmetry* 14 (2003) 1807–1817

TETRAHEDRON: *ASYMMETRY*

Towards a novel explanation of *Pseudomonas cepacia* lipase enantioselectivity via molecular modelling of the enantiomer trajectory into the active site

David Guieysse,[a] Christophe Salagnad,[b] Pierre Monsan,[a] Magali Remaud-Simeon[a,*] and Vinh Tran[c]

[a]*Centre de Bioingénierie Gilbert Durand, Département de Génie Biochimique et Alimentaire, UMR CNRS 5504, UMR INRA 792, INSA, 135 Avenue de Rangueil, F-31077 Toulouse Cedex 4, France*

[b]*Aventis Pharma, Process Development Biotechnology, 9, quai Jules Guesde, F-94400 Vitry sur Seine, France*

[c]*Centre de Nantes, Unité de Physicochimie des Macromolécules, INRA, BP 71627, F-44316 Nantes Cedex 3, France*

Received 11 March 2003; accepted 7 May 2003

Abstract—In the transesterification reaction between (*RS*)-2-bromophenyl acetic acid ethyl ester and 1-octanol in *n*-octane, *Pseudomonas cepacia* lipase enantioselectivity towards the (*R*)-isomer is 57. Two strategies are described to investigate the structural basis involved in this enzyme enantioselectivity. Molecular modelling of the tetrahedral intermediate mimicking the transition state enables the identification of two potentially productive substrate-binding modes for each enantiomer. However, the conformations obtained with the faster and slower-reacting enantiomers have equivalent potential energies and most of them possess the hydrogen bonds essential for catalysis. On this basis, it is not possible to distinguish the diastereomeric complexes. The second approach is original and consists in a simple but robust protocol of pseudomolecular dynamics simulations under constraints to map the probable trajectory of the enantiomers in the active site. Enzyme/substrate interaction energy is always found to be lower for the faster-reacting enantiomer, which satisfactorily corroborates the experimental results. Energy differences are attributed to specific interactions of these substrates with a network of hydrophobic residues lining the access path. Furthermore, mechanistic details suggest that the pivoting side chains of the hydrophobic residues act in a concerted step–tooth gear motion whose basic role is to select and guide the substrates towards the active site. With this type of lipase, such dynamic features could be the key explanation of this as yet unexplored enantiorecognition. For the slower-reacting enantiomer, it appears that the concerted motion of the side chains is perturbed when the substrate passes through a bottleneck formed by Val266 and Leu17. The enantioselectivity of mutant Val266Leu with a more bulky side chain at this position supports our assumption: by narrowing the bottleneck, the enantioselectivity was considerably enhanced as much as up to 200.

1. Introduction

To reduce the cost of development of enantiopure molecules such as pharmaceuticals, many efforts are focused on the understanding of the molecular features that control enzyme enantioselectivity. The aim is to develop reliable predictive methods and to guide substrate and/or enzyme modifications to enhance kinetic resolutions. Lipases are serine hydrolases (EC 3.1.1.3) frequently employed in organic synthesis to catalyse kinetic resolutions of a wide range of substrates yielding enantiomerically pure compounds. These enzymes have been used to catalyse esterification transesterification, and aminolysis processes.[1–6] Among them, *Pseudomonas cepacia* (recently renamed *Burkholderia cepacia*) lipase (PCL) has turned out to be very useful for the resolution of racemic mixtures of primary and secondary alcohols [7–10] and carboxylic acids.[11–19]

The methods used to investigate the molecular factors of enantioselectivity generally combine kinetic results, structural data analyses and molecular modelling.[20,21] Many serine hydrolase structures are available [22–27] and the mechanistic details of the catalytic reaction are well known.[28–31] Proton transfer between the side-chain oxygen Oγ of the catalytic serine and His–Nε of the catalytic histidine enables nucleophilic attack by the oxygen of the catalytic serine on the substrate carbonyl function. This reaction yields a tetrahedral intermediate stabilized by a network of hydrogen bonds followed by

* Corresponding author. Tel.: +33-56-155-9446; fax: +33-56-155-9400; e-mail: remaud@insa-tlse.fr

0957-4166/03/$ - see front matter
doi:10.1016/S0957-4166(03)00374-4

the release of the alcohol and the formation of the acyl-enzyme. Then, the hydrolysis (or alcoholysis) of the acyl-enzyme proceeds via the same mechanism through the formation of a second tetrahedral intermediate followed by product release.

To tackle the problem of enantioselectivity, X-ray structures of lipase/enantiomer inhibitor complexes and comparison of the complexes obtained with the (*R*)- and (*S*)-inhibitors are very informative.[32–34] Lang et al.[33] crystallised *P. cepacia* lipase with an enantiopure analogue of triacylglycerol and showed that the model obtained with the slower-reacting enantiomer revealed unfavourable interactions between the carbonyl oxygen (O7) of the sn-2 chain and Cδ1 L287 and Cδ1 I290 atoms. The recently reported crystal structure of *P. cepacia* lipase complexed with a transition state analogue of a secondary alcohol also showed that slower-reacting enantiomer binding could not be productive because of its orientation.[32]

The methods combining kinetic results and structural analyses also provide some empirical rules predicting the enantiopreference of several lipases toward secondary alcohols[35] or chiral carboxylic acids.[36] These rules establish a correlation between the size of the substituents at the stereocentre and enantioselectivity. Although useful, they do not precisely identify the molecular determinants of enantiorecognition. The best described approach consists in modelling the tetrahedral intermediates formed with each enantiomer assuming that they are a good mimic of the transition state.[28] A lower potential energy is generally expected for the faster-reacting enantiomer, but there is not always a good correlation with experimental enantioselectivity.[37–39] Schulz et al.[40,41] developed a quantitative model to predict the enantioselectivity of *Pseudomonas cepacia* lipase[41] and of *Candida rugosa* lipase[40] towards a broad range of secondary alcohols. They established a correlation between the $His_{N\varepsilon}$–Oalc distance of the slower-reacting enantiomer and the experimental enantioselectivity values.

Few studies address the problem of the enantioselectivity of lipases toward 2-substituted carboxylic acids. Most computer simulations of lipase-catalysed resolutions have revealed steric interactions between the slower-reacting enantiomer and the amino acids of the active site[38,42–44] but none of them concerned *P. cepacia* lipase.

In the present study, our aim was to determine the structural basis of the *P. cepacia* lipase enantioselectivity in transesterification of an important building block used in drug synthesis: (*RS*)-2-bromophenyl acetic acid ethyl ester. We modelled the first tetrahedral intermediate obtained with each enantiomer assuming, from experimental information, that this step is responsible for the enantioselectivity of the overall reaction. Unfortunately, this approach did not provide a satisfactory explanation of the experimental data with the faster and slower-reacting enantiomers. Thus, we built a second protocol to tackle the probable trajectory of each enantiomer through the access path.

From a methodological point of view, this protocol can be assimilated to dynamics simulations under constraints. Analyses of the enzyme–substrate interactions obtained along the path towards the active site for each enantiomer are discussed with regards to enantioselectivity and to the final goal, which is to design more enantioselective mutants.

2. Results and discussion

2.1. *P. cepacia* lipase enantioselectivity towards (*RS*)-2-bromophenyl acetic acid ethyl ester

The PCL-catalysed transesterification between (*RS*)-2-bromophenyl acetic acid ethyl ester (50 mM) and 1-octanol at various concentrations (50–500 mM) was first performed in *n*-octane, with the water activity being controlled at 0.04. The transesterification rate was maximal at 100 mM and decreased slightly above 300 mM, and no significant variation of enantioselectivity was observed between 50 and 500 mM (Figures 1a and b). The transesterification reaction was then carried out with six aliphatic alcohols of different chain lengths. As reported in Table 1, all measurements revealed the same preference for the (*R*)-enantiomer. In addition, the enantioselectivity decreased slightly with the length of the alcohol chain but this effect was not very pronounced. From these results, it was assumed

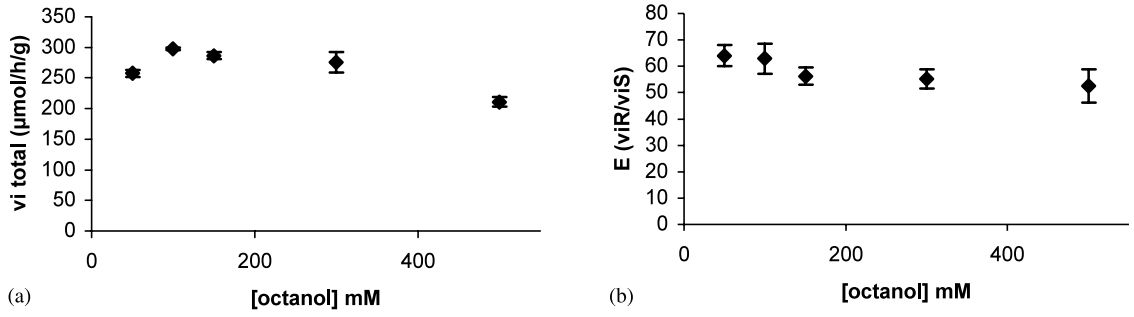


Figure 1. Effect of 1-octanol concentration (a) on initial rate and (b) on *P. cepacia* lipase enantioselectivity (*E*) for transesterification between α-bromophenyl acetic acid ethyl ester (50 mM) and 1-octanol (50–500 mM) in *n*-octane at a water activity of 0.04.

Table 1. PCL-catalysed transesterification of (*RS*)-2-bromophenyl acetic acid ethyl ester in *n*-octane. Effect of aliphatic alcohol chain length on *P. cepacia* lipase enantioselectivity

Alcohols	*E* (vi*R*/vi*S*)	% Conversion	e.e.$_s$ (%)	e.e.$_p$ (%)
Butanol-1	70 (±3)	43 (96 h)	72	87
Pentanol-1	63 (±6)	44 (96 h)	76	86
Hexanol-1	64 (±2)	46 (96 h)	77	86
Octanol-1	57 (±7)	47 (84 h)	77	84
Decanol-1	53 (±7)	49 (96 h)	76	77

E: enantioselectivity factor. e.e.$_s$ and e.e.$_p$: enantiomeric excess of substrate and product, respectively.

that the enzyme enantioselectivity is almost independent on the deacylation step of the reaction. Consequently, our modelling protocols were focused on the primary stage of the reaction up to the formation of the first tetrahedral intermediate.

2.2. Molecular modelling

2.2.1. Static molecular modelling of the tetrahedral intermediates

2.2.1.1. Active site environment. The *P. cepacia* lipase catalytic cavity (the active site plus the access path) has three pockets[33] as shown in Figures 2a and b. A first inner hydrophobic pocket (HA), anchoring the acyl part of the substrate, another inner hydrophobic/hydrophilic pocket (HH), anchoring the alkyl part of the substrate and finally, along the main axis of the enzyme cavity, a third outer pocket (HB) can be seen beyond Leu17 (Fig. 2a, above the plane of the paper). The active site is deep inside the enzyme with a contiguous access path like a funnel but with a pronounced bottleneck in the middle, formed by Leu17 and Val266.

2.2.1.2. Description of binding modes. Among the four tetrahedral intermediates that can be constructed (named *RR*, *SR*, *RS* and *SS*), only two solutions (*RR* and *SR*) were found to be productive on the basis of the hydrogen bonding network necessary for catalysis.

Two binding modes could exist (named I and II) in both tetrahedral intermediates *RR* and *SR* as shown in Figures 3a, b, c and d. Figures 3a and c (left side) show the first binding mode for the *RR* and *SR* intermediate types, respectively. The bromine atom points down towards pocket HA, whereas the aromatic ring points up towards pocket HB. Figures 3b and d (right side) show the second binding mode with reverse orientations for the bromine atom and the aromatic ring. For each binding mode, the alkyl part of the substrate (-OCH_2-CH_3) lies in pocket HH (bottom part of Fig. 3).

The first criterion, which characterizes the productive binding modes of the substrate in the active site is the potential energy. The relative energies are 0, 4, 4 and 5 kcal/mol for $SR_{\text{mode I}}$, $SR_{\text{mode II}}$, $RR_{\text{mode I}}$ and $RR_{\text{mode II}}$, respectively. Thus, $SR_{\text{mode I}}$ should be the preferred solution because all others are about 5 kcal/mol higher. But these results are unable to explain the *R* enantiopreference of *P. cepacia* lipase. The enantioselectivity could also be analysed according to the network of five hydrogen bonds required for the catalysis as already described in the literature.[30,31,45] Table 2 reports all hydrogen bonds shown in Figure 3. Within a maximum distance of 3.1 Å, all intermediate conformations have the five expected linkages except for $SR_{\text{mode I}}$ with the missing His286[Nε2H···Oγ]Ser87 bond necessary for the activation of the catalytic serine Oγ. In the latter case, the absence of an essential H-bond seems to be inconsistent with the lowest potential energy.

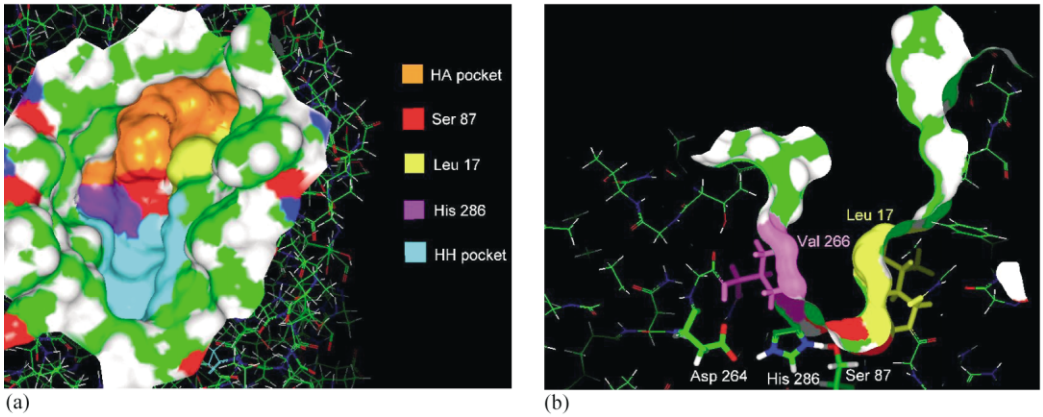


Figure 2. Representation of the catalytic cavity with Connolly surfaces. (a) Hydrophobic pocket (HA, in orange), anchoring the acyl part of the substrate, is lined by Leu17, Pro113, Phe119, Leu164, Leu167, Val266 and Val267. Hydrophobic/hydrophilic pocket (HH, in blue), anchoring the alkyl part of the substrate, is lined by Gln16, Thr18, Tyr23, His86, Leu287, Ile290, Gln292 and Leu293. Residues Leu17, Ser87 and His286 (yellow, red, purple, respectively) separate pockets HA and HH. (b) Perpendicular cross section of PCL cavity (roughly at level Ser87) showing the funnel shape of the access path (Ser87 at the bottom and His286, Asp264 at the left). The bottleneck formed by Val266 and Leu17 delimits the inner sites (HA and HH) and the outer one (HB, at the top). Ser87, His286 and Leu17 are in the plane of this figure and pockets HA and HH are in the background and in the foreground, respectively. The bottleneck formed by Val266 and Leu17 delimits the inner sites (HA and HH) and the outer one (HB, at the top).

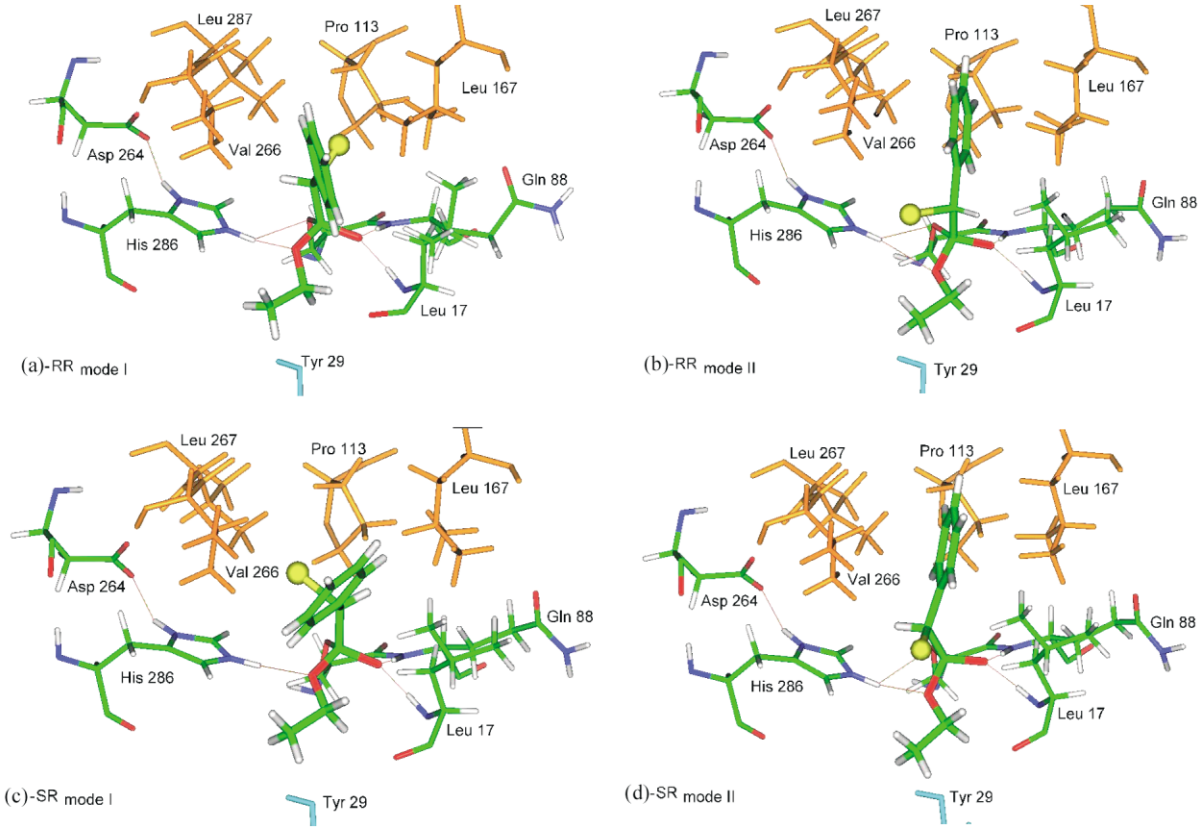


Figure 3. For the sake of clarity, the enzyme orientation is identical to that of Figure 2a. Pockets HA and HB are at the top of the schema (below and above plane of the figure, respectively) and pocket HH is at the bottom. (a) $RR_{\text{mode I}}$, (b) $RR_{\text{mode II}}$, (c) $SR_{\text{mode I}}$, (d) $SR_{\text{mode II}}$. Residues in orange (Leu267, Pro113 and Leu167) represent the HA pocket (at the top and in the background). The HB pocket is in above the plane of the paper while Tyr29 denotes the HH pocket (below). In binding mode I, the bromine atom points down toward the HA pocket and the aromatic ring points up toward the HB pocket. Binding mode II has the opposite orientation.

Table 2. Network of hydrogen bonds in tetrahedral intermediate geometries

Donor	Acceptor	Distances between heavy atoms (Å)[a]			
		$RR_{\text{mode I}}$	$RR_{\text{mode II}}$	$SR_{\text{mode I}}$	$SR_{\text{mode II}}$
His 286 Nδ1H	Oδ2 Asp 264	2.7	2.7	2.7	2.7
His 286 Nε2H	Oγ Ser 87	3.0	3.0	3.5	3.1
His 286 Nε2H	O-Et substrate	3.0	3.1	3.0	3.0
Gln 88 NH	O^- substrate	2.8	2.8	2.9	2.9
Leu 17 NH	O^- substrate	2.7	2.7	2.7	2.7

[a] The hydrogen bond, can be detected with geometrical criteria: short distance between the donor and the acceptor atom and likely a linear arrangement between the donor atom, hydrogen and the acceptor atom.[45] These criteria are implemented in the InsightII package to automatically visualize H-bonds.

Once more, this fact enhances the discrepancy between this modelling protocol and the experimental results of enantioselectivity. This means that despite our efforts to detect, as exhaustively as possible, all the probable intermediate geometries, those found were unable to satisfactorily explain the PCL preference for the (*R*)-enantiomer. Nevertheless, it must be mentioned that modelling of this tetrahedral intermediate had to be attempted first because, in several studies, this approach was found appropriate to predict the lipase enantioselectivity.[38,40,41,45–48]

2.2.2. Molecular modelling of the substrate trajectories. As a consequence of the failure of the above model, we assumed that the enantiomer discrimination possibly occurred before the formation of the covalent bond between the catalytic serine and the substrate during the transport of the enantiomers to the active site located at the bottom of a funnel. Thus, the preliminary trajectory of each enantiomer in the active site and for each mode was analysed. Indeed, if the access to the active site is facilitated for one enantiomer, that will affect the dissociation constant and the kinetic resolu-

tion. As already discussed, the funnel-like shape of the access path technically yielded non-chronological dynamics simulations and it was assumed that all trajectories can be reversibly interpreted. Furthermore, since there was no way to discriminate one of the two binding modes, at this stage all initial $SR_{\text{mode I}}$, $SR_{\text{mode II}}$, $RR_{\text{mode I}}$ and $RR_{\text{mode II}}$ conformations were subjected to this dynamic protocol.

2.2.2.1. Interaction energies along the trajectory. The curves of the potential energy of the enzyme/substrate interactions versus the $D_{(B1–B2)}$ values (described in Section 4) are reported in Figure 4. The minimum $D_{(B1–B2)}$ value (4.55 Å) corresponds solely to the virtual disruption of the O–C bond between the catalytic serine and the substrate without any movement away of the substrate. Due to the inherent incapacity for molecular mechanics to properly model covalent bond disruption (only a quantum scheme would be able to do that), an initial abnormally high interaction energy is observed for all starting geometries (the O–C bond being now considered as too short a non-bonded distance between the two entities). Therefore, the first part of these curves must be discarded from the analysis.

As concerns the rest of the curves, two phenomena are observed. The first general tendency is the slow but continuous increase of the energy whatever the enantiomer type and the binding mode. In the reverse direction, this must be interpreted as a global regular gradient of substrate attraction through the access path. More interestingly, approximately from 5.6 to 9.0 Å, the (*R*)-enantiomer trajectories always yield lower interaction energies compared to the (*S*)-enantiomer whatever the binding mode. It seems reasonable to consider that this part of the curve is much more meaningful than the other one (from 9.0 to 12.5 Å) and this preliminary result is consistent with the experimental preference for the (*R*)-enantiomer.

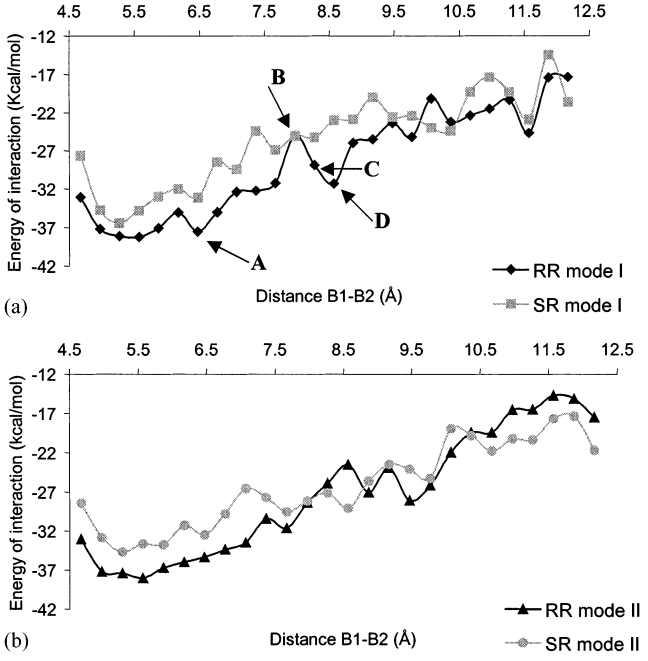


Figure 4. Curves of non-corrected interaction energy between the enzyme and the substrate versus $D_{(B1–B2)}$.

All along these trajectories, the potential energies of both molecular entities were also checked. No interesting information could be derived from the truncated enzyme energy. On the contrary, those of the substrate (Fig. 5) gave relevant information to discriminate the two binding modes. The substrate potential energies are always higher in binding mode II regardless of the enantiomer type. This feature is particularly true for low $D_{(B1–B2)}$ values (near the formation of the covalent bond), which means that to produce binding mode II, the substrate must be progressively distorted to adopt the final geometry. These artificial distortions are correlated with the [Φx,Φy (Scheme 1)] values corresponding to high energy domains on the maps drawn at the first stage of the former modelling protocol (maps and data not shown). In contrast, for binding mode I, the (Φx,Φy) values correspond to low minima regions. Furthermore, as expected for enantiomeric forms, the (*R*)- and (*S*)-substrate energies are similar within the small fluctuations necessary to adapt their geometries to the access path context. Thus, binding mode I must be the only process leading to the tetrahedral intermediate.

To sum up, at this stage in the dynamics protocol, these calculations clearly show the (*R*)-enantiopreference, which supports the more realistic binding mode I with the bromine atom located in pocket HA and the aromatic ring in pocket HB.

2.2.2.2. Correction and analysis of the trajectories. The *R* and *S* energy curves have several significant variations. It appears that some other oscillating phenomena are superimposed on the background continu-

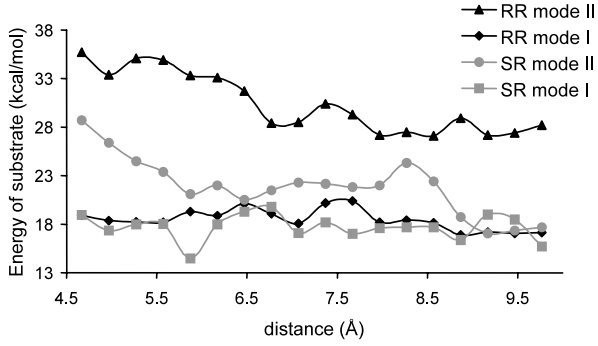
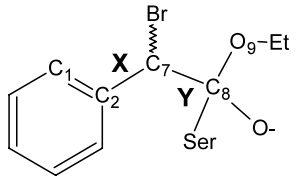


Figure 5. Curves of potential energy of (*RS*)-2-bromophenyl acetic acid ethyl ester versus $D_{(B1–B2)}$.



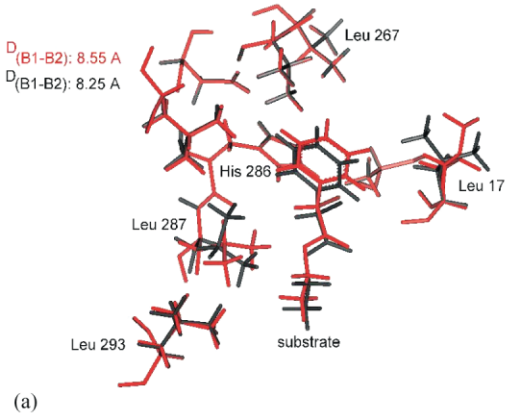
Scheme 1. Serine/2-bromophenyl acetic acid ethyl ester. Dihedrals angles Φ_X(C1–C2–C7–C8) and Φ_Y(C2–C7–C8–O9).

ous variation of the energy. For example, points A, B, C and D were added to the $RR_{\text{mode I}}$ curve of Figure 4a to illustrate a complete cycle of oscillations comprised between two minima (points A and D) with a local peak (point B) and a transitory state (point C). These intriguing fluctuations can neither be correlated to the potential energy of the enzyme nor to that of the substrate and must be attributed to local unrelieved steric conflicts between the two entities. The best way to identify the molecular origin of these conflicts, was to superimpose two subsequent snapshots (at 0.3 Å interval) of the whole system geometry with significantly different energies. As an example, Figure 6a shows such a superposition of the C and D points of the $RR_{\text{mode I}}$. Except for the imposed displacement of the substrate, the only differences come from the side chain orientations of Leu287 and Leu17. In the latter case, the enzyme residue is very close to the substrate.

From a mechanical point of view, on the one hand, the backbone of Leu17 is basically fixed and the only degree of freedom is its side chain rotation. On the other hand, due to the very close contacts, any tiny displacement imposed on the substrate would require a larger Leu17 side chain rotation to avoid steric clashes. In other words, this figure suggests that the turnstile-like motion of Leu17 could be especially designed to monitor the substrate displacement along the access path. Now, with this schematic mechanism in mind, one must go back to the protocol background consisting in successive small displacements for the substrate. It happens that, despite our efforts to fully relax the system after each distance constraint (combination of pseudo-dynamics simulations and long minimisation process), it was technically impossible to relieve all local steric conflicts as in any optimisation algorithm applied to a large system with a huge number of internal parameters. Once more, this is the clear manifestation of the local false minima problem. To understand this feature better, a whole cycle of energy fluctuations will be described in relation with the local side chain rotation. The starting point A is a relaxed energy position (Leu17 and the substrate are not yet in conflict). At point B, with a new constraint imposed on one entity (elementary displacement of the substrate), the algorithm was not able to properly diffuse this constraint toward the other entity and the frustrated rotation of the Leu17 side chain, not exactly in phase with the imposed substrate location at the end of the minimisation process, yielded a momentary energy increase. This frustration can last for other subsequent increases of $D_{(\text{B1–B2})}$ (point C) until the local conflict becomes too important inducing a flip of the Leu17 side chain. This finally occurred at point D where the new Leu17 side chain position returned back in phase with the substrate location. It is important to mention that this unrealistic fluctuation of energy certainly occurs if the reverse trajectory is created because this artefact is basically attached to the minimisation protocol.

To corroborate these technical and physical explanations, we tried to circumvent this behaviour by locally exploring this trajectory in the reverse direction. In so doing, it was expected that, starting from the next relaxed state (point D), it could be possible to relieve uphill frustrations generated during the former propagation direction and therefore to decrease the interaction energy (especially for point B). Finally this local correction was successfully carried out and new lower energies replaced former ones to better represent the substrate trajectory. These non-automatic corrections were systematically performed on both $RR_{\text{mode I}}$ and $SR_{\text{mode I}}$ trajectories from each local minima detected on the previous energy curves (Fig. 4a). These manual corrections were very efficient and yielded much better smoothed curves assumed to represent the realistic substrate trajectories (Fig. 7).

These curves confirm the enzyme enantiopreference whose mode I is the only acceptable binding mode because the energy for the faster-reacting enantiomer (R) is always lower than that of the slower one (S) and above all, at any step of the trajectory. The energy difference fluctuates between approximately 3 and 5 kcal/mol. On this basis, PCL appears to be designed to attract the (R)-enantiomer to form the first tetrahedral intermediate.



(a)

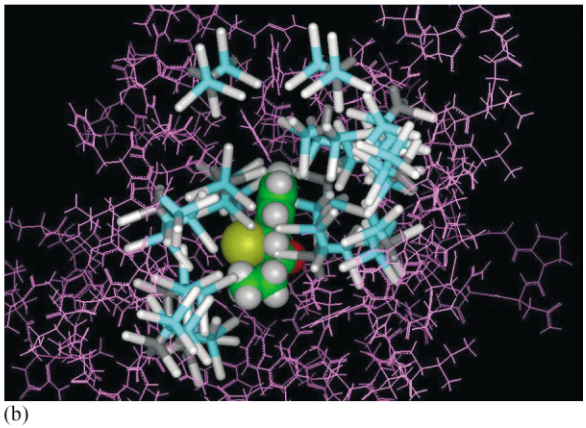

(b)

Figure 6. Network of hydrophobic bulky residues. (a) Superimposition of the enzyme/substrate complexes for two subsequent snapshots [points C and D; $D_{(\text{B1–B2})}$: 8.25 and 8.55 Å, respectively]. Several side chain orientations of residues Leu287 and Leu17. (b) Representation of their potentially pivoting side chains with stick bonds. The substrate is shown in CPK representation.

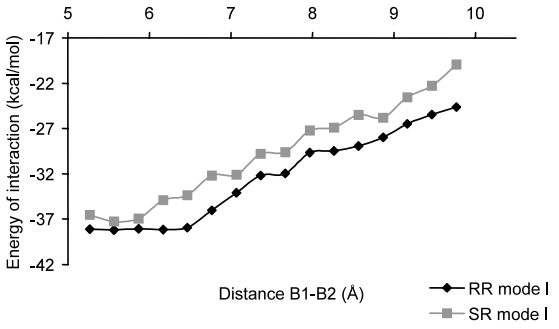


Figure 7. Curves of corrected interaction energy between the enzyme and the substrate versus $D_{(B1–B2)}$.

2.2.2.3. Substrate guiding system. Leu17 is not the only residue in charge of guiding the substrate from outside toward the catalytic site. From a static point of view, examination of the access path of *P. cepacia* lipase reveals a dense network of bulky hydrophobic residues such as valine, leucine or isoleucine. As shown in Figure 6b, Leu17, Leu167, Leu248, Val266, Leu287, Leu293 are well distributed all around the path and could act similarly to Leu17 with their pivoting side chain. From a dynamic point of view, the turnstile-like motions of their side chain are illustrated in Figure 8 by superimposing several snapshots ($\sim$5 Å$<D_{(B1–B2)}<\sim$8 Å). For the sake of clarity, only some of them are shown. Figure 8 has the same orientation as Figure 2a and the path axis (or B1–B2 axis) is roughly perpendicular to its plane. Finally, along this axis, the side chain motions occur at different levels. For example Leu17 and Val266 interact with the substrate at 5–7 Å on the $D_{(B1–B2)}$ abscissa, and Leu248, Leu287 and Leu293 at 7–12 Å.

More interestingly, these hydrophobic residues, which are very close to each other, must have concerted side chain rotations to avoid intra-molecular steric hindrances in addition to inter-molecular ones with the substrate. Furthermore, all these local motions must be in phase to monitor the substrate displacement through the access path. In other words, even if it is difficult to precisely describe the action of all the elementary pieces of this machinery, it clearly appears that this network of hydrophobic side chains acts in a concerted step–tooth gear motion. This common network architecture must be a sophisticated but above all a very efficient dynamic guiding system.

This simultaneous rotation of several side chains considerably perturbs the minimisation algorithm. It fully explains the frustrated rotation of one member of this network when the relaxation process is technically insufficiently diffused to the other members yielding a local increase of energy.

2.2.2.4. Selecting role of the hydrophobic network. According to binding mode I, the (*R*)- and (*S*)-enantiomers must approach the catalytic site with the bromine atom directed towards the right and left sides, respectively. Access for the (*R*)-enantiomer seems much easier than for the (*S*) one. Due to its relative orientation during the final steps just before covalent bond formation, there is enough room for the displacement of the (*R*)-substrate without additional significant internal distortions of both entities. As seen in Figure 8a, except for the imposed translation toward the active site, only small global rotations of the substrate were observed with limited fluctuations of the hydrophobic side chains on the enzyme. On the contrary, the bottleneck formed by Leu17 and Val266 seemed to hinder the access of the (*S*)-enantiomer more severely. As seen in Figure 8b, to pass through this obstacle with the bromine atom pointed toward Val266, the (*S*)-enantiomer trajectory must shift left but much closer to Leu17. This overcrowded context led to more severe conflicts with the enzyme, obviously yielding higher interaction energies that were only partially reduced by larger enzyme side chain motions (especially for Leu17 as seen in Figure 8b) and also by more significant substrate internal distortions (although not taken into account in the interaction energy curve). These two features are clearly seen in Figure 8b and the steric obstacle corresponding to Val266 and Leu17 seems to

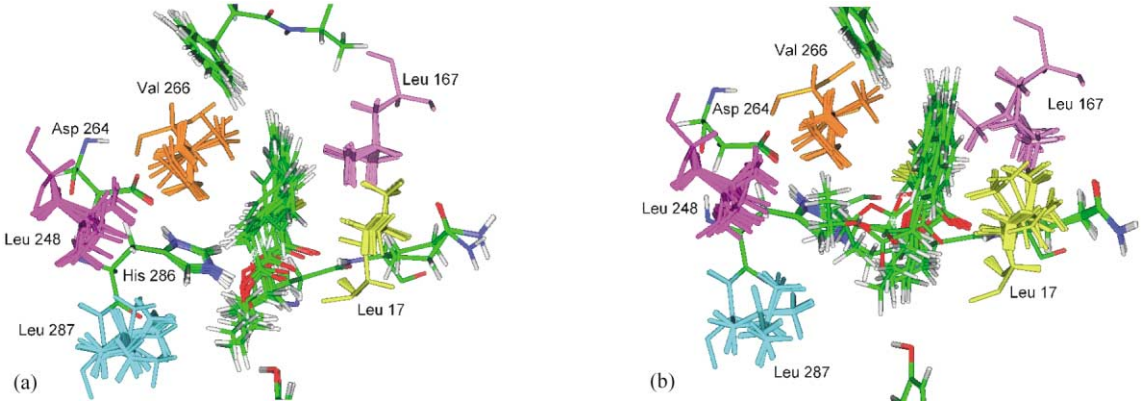


Figure 8. Superimposition of the different structures of *P. cepacia* lipase obtained for the corrected trajectories for the B1–B2 distance of 4.75–7.95 Å. Left: the (*R*)-enantiomer, right the (*S*)-enantiomer.

play a key role in the discrimination of the (*R*)- and (*S*)-enantiomers. Thus, in addition to their guiding role, some of these hydrophobic residues can also have a selecting role and control which substrate is allowed to access to the inner catalytic site.

2.2.3. Enantioselectivity of the Ps-FVL mutant. To further highlight the role of the concerted motion of hydrophobic residues with pivoting side chains and the importance of the barrier created by Leu17 and Val266 in the enantiomer discrimination, we tested the enantioselectivity of the mutant of *P. cepacia* lipase: Ps-Phe221Leu-Val266Leu-Leu287Ile.[49] Residue Phe221 is very far from the active site thus, it sounds reasonable to assume that the F221L mutation has no influence on the enantioselectivity. On the other hand, the Val266Leu and Leu287Ile mutations are much more important because of the substitution of the initial hydrophobic residues by more bulky ones. This is especially true for Leu266 facing Leu17 in this geometrical context. Consequently, the size of the bottleneck was significantly reduced yielding a dramatic change in enantioselectivity. The Ps-FVL enantioselectivity toward (*RS*)-2-bromophenyl acetic acid ethyl ester was found to be exceptionally high (>200). This result is thus fully consistent with our explanations and supports the possible role of the hydrophobic residues with their pivoting side chains in the discrimination of the (*R*)- and (*S*)-enantiomers.

3. Conclusions and perspectives

The aim of this study was to understand the enantioselectivity of *P. cepacia* for (*RS*)-2-bromophenyl acetic acid ethyl ester, and to propose a methodology to explain and eventually predict the enantioselectivity of *P. cepacia* lipase using a molecular modelling approach.

The concerted side chain rotations of hydrophobic residues organized in a specific network could be a general scheme for the enantioselectivity of some other lipases because of its sophisticated architecture and at the same time, its efficiency. As a matter of fact, except for *R. miehei* lipase, this structural feature also concerns lipases from *C. antarctica B* and *C. rugosa*. Indeed, these lipases also possess a hindered active site with a contiguous funnel-like or tunnel-like access path. As in PCL, a very dense network of hydrophobic residues lines the sides of the path. Similarly, these close-packed residues could play a major role in activity and enantioselectivity with their pivoting side chains. The protocol described here, once adapted to each catalytic environment, should enable a straightforward examination of the role of these hydrophobic residues. More generally, if such a dynamic scheme was finally found to be an essential key of enantiorecognition, it would be necessary to reanalyse all lipase cases, even those for which static modelling of the tetrahedral intermediate previously gave results consistent with experiment.

4. Materials and methods

4.1. Biotransformation

The synthesis of (*RS*)-2-bromophenyl acetic acid ethyl ester and the kinetic resolution by *P. cepacia* lipase were performed as previously described.[11]

4.1.1. Determination of enantiomeric excess (e.e.), conversion rate and enantioselectivity (*E*). HPLC analyses (carried out as previously described[11]) were used to determine the rate of production of each enantiomer at various concentrations of racemic. From these analyses, enantiomeric excess (e.e.) was calculated as defined below: $\text{e.e.}_s = \{[R]-[S]\}_s/\{[R]+[S]\}_s$ (s = substrate) and the conversion rate: $C = 1-[(R+S)_t/(R+S)_{t=0}]\times 100$. Enantioselectivity value was determined as the ratio of the initial rate of (*R*)-enantiomer production (viR) versus the initial rate of (*S*)-enantiomer production (viS): $E = (viR/viS)$. Initial rates were determined, before 10% of substrate conversion, by linear regression over at least five points.

4.2. Molecular modelling

4.2.1. Equipment and software. Calculations, molecular constructions and graphic displays were performed on O2 (R10000) Silicon Graphics workstations. All calculations were done within the InsightII environment (Accelrys, San Diego, California) using the Discover package. Potential energies were expressed with the CFF91 force field implemented in Discover (version 2.97). All minimisations were performed with the steepest descent algorithm.

4.2.2. Initial lipase geometry. The X-ray structure of *Pseudomonas cepacia* lipase complexed with diethyl phosphate inhibitor were kindly provided by M. Cygler and J. Schrag (NRC Biotechnology Research Center, Montreal). The enzyme coordinates were compared to those of the native *P. cepacia* lipase[23] (Brookhaven Data Base code: 3LIP). Since no significant differences were observed, the coordinates of the inhibitor/lipase complex were used as template for the construction of the models.

4.2.3. Construction of the starting geometry. The crystallographic water molecules were deleted from the X-ray structure, and the hydrogen atoms were added using the Biopolymer module. The catalytic histidine (His286) was protonated ($His^{\oplus}$). A covalent bond was created between the Ser87 oxygen atom and the phosphorus atom of the inhibitor. The hydrogen atom positions were first refined (all the other atoms fixed; derivative convergence of 0.01 kcal/Å mol). Then heavy atom positions of the enzyme were refined using the "tether" option of Discover (constraint force of 200 kcal/mol; derivative convergence of 0.5 kcal/Å mol).

To spare CPU time, calculations were further performed with a subset of amino acids (Table 3) according to the following rules: (i) All the 43 residues delimiting the active site are kept. (ii) If one of these 43

Table 3. Definition of the truncated *Pseudomonas cepacia* lipase

Segments of residues lining the active cavity (catalytic site+access path)	Amino acids selected[a]
*G16, L17, T18, Y23, L27, E28, Y29	**-14**-15-…-21-**22-23**-24-25-26-**27-28**-29-…-32-**33**
*F52	**-46**-47-…-60-**61-62**
*H86, S87, Q88	**-85-86**-87-88-**89-90-91**
*G111, T112, P113, G116, S117, F119, A120, V123, L127, I139, F142, V143, F146, T150	**-110**-111-…-117-**118-…-127**-128-…-133-**134-…-150**
*A160, L161, A163, L164, L167	**-160-…-166**-167-168-**169-170**
*P243, S244, A247, L248, T251, V266, V267	**-241**-242-243-**244-…-257**-258-…-268-**269**
*H286, L287, I290, Q292, L293, L294	**-282**-283-…-287-**288-289-290**-291-…-303-**304**

[a] See the rules defined in the text; the backbone of the amino acids belonging to the secondary element (in bold) is kept fixed during the calculations.

residues belongs to a secondary element, one or two adjacent residues along the primary structure are also selected. (iii) If the amino acids forming the active site are localized inside a loop, all the residues of this loop are conserved as well as the extremities of the loop belonging to the adjacent secondary elements. In total, this subset represents 46% of the whole protein structure. Consequently, this artefactual truncation imposed, for all following calculations, a basic set of constraints with fixed backbone atoms for residues belonging to secondary elements to preserve the global folding of the protein, regardless of other additional constraints necessary for the selected strategy.

4.2.4. First protocol: modelling of the first tetrahedral intermediate. Basically, the tetrahedral intermediate geometries were built in two distinct stages. Primarily, a systematic conformational search of the core entity (the catalytic serine Ser87 covalently bound with (*R*)- and (*S*)-2-bromophenyl acetic acid ethyl ester) was performed, regardless of the enzyme active site context. Then, all acceptable geometries found in this former stage were 'transposed' within the active site environment. The goal of this protocol is to describe, as systematically as possible, all conformations with regards to energy as well as catalytic topography criteria.

First stage. The diethyl phosphate inhibitor coordinates and those of S87 in the complex crystallographic structure were initially used to build the catalytic core. The formation of the tetrahedral intermediate leads to the generation of a second stereogenic carbon, thus with each enantiomers, two configurations can be obtained. A total of four combinations are thus possible and were referred as: *RR*, *RS*, *SR* and *SS* in the text. The first letter (*R* or *S*) refers to the initial stereogenic centre of the substrate and the second letter to the stereogenic centre created by the intermediate formation. For each of these combinations, dihedral angles Φ_X(C1–C2–C7–C8) and Φ_Y(C2–C7–C8–O9) represent the two major degrees of flexibility (Scheme 1). Therefore, they were systematically scanned by 10° increments, fixing all the other internal parameters (ROTOR option of Discover) to roughly locate energy minima. These minima solutions were further fully refined (no constraints; derivative convergence of 0.01 kcal/Å mol) in vacuo and without any steric consideration of the active site topography.

Second stage. All these core geometries were then "transposed" into the active site by superimposing the Ser87 cartesian coordinates of modelled cores to those of the crystallized complex to obtain the truncated tetrahedral intermediate geometries. Clearly, this brutal technical transposition generally provokes initial huge (and somewhat artificial) steric clashes that could be solved in some cases by the local rearrangement of a limited number of atoms. Therefore, each tetrahedral intermediate geometry was submitted to three successive minimisations with a decreasing number of constraints combined with an increasing number of iterations to manage, as properly as possible, the side chain atom relaxation. The first minimisation consisted of 2000 iterations with all heavy atoms fixed except those of the catalytic core. In so doing, it was expected (if necessary or possible) to reposition the core slightly. During the second minimisation (10,000 iterations), the protein backbone was fixed and the enzyme side chains were allowed to vary in order to remove local conflicts between the protein and the branched core. Finally, during the third minimisation (50,000 iterations), only the backbone of the residues belonging to secondary elements was fixed. This corresponds to the largest relaxation preserving the general shape of this truncated system. Finally, after the successive minimisation only two solutions were kept according the geometric features and the energy criterion.

4.2.5. Second protocol: modelling of the substrate path in the catalytic cavity. *Starting geometries*. As in the first approach, the truncated tetrahedral intermediates previously built were used to spare CPU time. Firstly, it was necessary to break the covalent bond between the catalytic serine and the carbon of the substrate carbonyl function to create two molecular entities with distinct topologies: the truncated enzyme (encompassing Ser87) and the substrate. Two virtual points (local barycenters) were defined to represent these entities. The first one, **B1**, based on the backbone atoms of the catalytic serine is attached to the heaviest entity (truncated enzyme). The second one **B2**, is based on atoms (C2, C7, C8, O9) and represents the substrate (Scheme 1).

Trajectory protocol. PCL active site is a well defined contiguous channel. The particular shape of this channel, like a funnel, yielded a specific trajectory protocol.

Formally, the absolute location of the substrate in the enzyme reference should need 6 parameters (e.g. 3

rotations and 3 translations). But in the present context, the simplest way to monitor the relative displacement of these two entities was to impose regular incremental variations of B1–B2 distance ($D_{B1–B2}$) all through a series of calculations to describe the accessibility of the substrate to the active site. It must be pointed out that, in this scheme, the 6 positioning parameters were replaced by $D_{B1–B2}$. This drastic reduction of the conformation space to explore is acceptable only because of the narrow enzyme path, especially near the catalytic site. The use of a similar dynamic protocol in another geometrical context (e.g. a large access path with enough room for substrates to have internal rotations) would certainly lead to useless and random results.

In our protocol, only the final coordinates are precisely known (formation of the tetrahedral intermediate) which corresponds to the shortest $D_{(B1–B2)}$ value (4.55 Å). Due to the large entrance of the access path, any arbitrary choice of initial substrate orientation would undoubtedly guide the subsequent trajectory performed in a chronological model (from outside into the catalytic site) without any guarantee that such simulations are the most significant. Thus, it was technically easier to generate "reverse" substrate trajectories by progressively increasing $D_{(B1–B2)}$. From visual considerations of the enzyme access path, we chose a maximum distance of 12.55 Å corresponding to a substrate position inside the access path with enough room to potentially avoid steric conflicts with the enzyme. In fact, the first steps of this increasing $D_{(B1–B2)}$ protocol are much more physically meaningful (up to a certain distance to be defined further) since it can specifically reveal the residues responsible for the steric conflicts. Then, due to the funnel-like shape of the access path, steric hindrances will progressively vanish for larger $D_{(B1–B2)}$ values.

Within the total range of variation (7.5 Å), regular increments of 0.1 Å were imposed. This increment was chosen short enough to temporarily relax the system until the following increment step. In relation with this short substrate displacement, a sophisticated elementary sequence of calculations was chosen to relax, as much as possible, the total molecular system. At each distance constraint, molecular dynamics simulations were performed (300K during 1 ps) and the enzyme/substrate entity coordinates were stored every 100 fs. These calculations were used as an exploration technique of all internal parameters of the system (annealing simulations but at room temperature). Finally, the best of each 10 conformations was selected for subsequent minimisation (10,000 iterations). During the dynamics simulations and the minimisation stage, $D_{(B1–B2)}$ was highly constrained with a specific harmonic term $a(r-r_0)^2$ with a constraint force of 1000 kcal/Å^2. Initially, only the enzyme side chains were allowed to fluctuate (to preserve the general shape of the access path) while no constraint was imposed on the substrate. Unfortunately, spurious substrate deformations were always observed whose origin is the mass ratio between the enzyme and the substrate which artificially provokes huge deformations of the smallest entity during the calculations (dynamics simulations and minimisation as well). Therefore, additional constraints on dihedrals Φ_X and Φ_Y were also imposed to preserve the global shape of the substrate and to technically counterbalance the high mass of the enzyme and the constraints imposed on it.

All along the trajectory, the interaction energy between the enzyme and the substrate was the first criterion taken into account to explain the enantioselectivity as in the following section. However, the potential energy variations of the enzyme and substrate were constantly checked to ensure a quite realistic representation of the trajectory. In practice, a momentary increase of one of these two terms would generally mean an artificial deformation of the corresponding entity due to insufficient relaxation of the whole system. Finally, for the sake of clarity, the interaction energies were reported only every 0.3 Å although calculations were performed with increments of 0.1 Å.

Acknowledgements

The authors thank Dr. Hirose (Amano Co., Japan) for providing the Ps-FVL mutant of *Pseudomonas cepacia*; Dr. Miroslaw Cygler and Dr. Joseph Schrag (NRC Biotechnology Research Center, Montreal) for access to the X-ray crystallographic coordinates of PCL complex; and Dr. Lorenzo Di Bari and Dr. Gennaro Pescitelli (Dipartimento di Chimica e Chimica Industriale, Pisa, Italy) for the determination of the absolute configurations of α-bromo phenyl acetic acid by the Circular Dichroism Exciton method.

References

1. Bornscheuer, U. T.; Bessler, C.; Srinivas, R.; Krishna, S. H. *Trends Biotechnol.* **2002**, *20*, 433–437.
2. Reetz, M. T. *Curr. Opin. Chem. Biol.* **2002**, *6*, 145–150.
3. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. *Nature* **2001**, *409*, 258–268.
4. Klibanov, A. M. *Nature* **2001**, *409*, 241–246.
5. Berglund, P. *Biomol. Eng.* **2001**, *18*, 13–22.
6. Schmidt, R. D.; Verger, R. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1608–1633.
7. Kamal, A.; Sandbhor, M.; Ramana, K. V. *Tetrahedron: Asymmetry* **2002**, *13*, 815–820.
8. Salunkhe, M. M.; Nair, R. V. *Enzyme Microb. Technol.* **2001**, *28*, 333–338.
9. Pchelka, B. K.; Loupy, A.; Plenkiewicz, J.; Petit, A.; Blanco, L. *Tetrahedron: Asymmetry* **2001**, *12*, 2109–2119.
10. Lundell, K.; Raijola, T.; Kanerva, L. T. *Enzyme Microb. Technol.* **1998**, *22*, 86–93.
11. Guieysse, D.; Salagnad, C.; Monsan, P.; Remaud-Simeon, M. *Tetrahedron: Asymmetry* **2003**, *14*, 317–323.
12. Haughton, L.; Williams, J. M. J. *Synthesis* **2001**, *6*, 943–946.
13. Tranel, F.; Haufe, G. *Tetrahedron: Asymmetry* **2000**, *11*, 889–896.

14. Miyazawa, T.; Kurita, S.; Ueji, S.; Yamada, T. *Biocatal. Biotrans*. **2000**, *17*, 459–473.
15. Sundholm, O.; Kanerva, L. T. *Models Chem*. **1998**, *135*, 625–640.
16. Jones, M. M.; Williams, M. J. *Chem. Commun*. **1998**, 2519–2520.
17. Palomer, A.; Cabre, M.; Ginesta, J.; Mauleon, D.; Carganico, G. *Chirality* **1993**, *5*, 320–328.
18. O'Hagan, D.; Rzepa, H. S. *J. Chem. Soc., Perkin Trans. 2* **1994**, 3–4.
19. Kalaritis, P.; Regenye, R. W.; Partridge, J. J.; Coffen, D. L. *J. Org. Chem*. **1990**, *55*, 812–815.
20. Kazlauskas, R. J. *Science* **2001**, *293*, 2277–2279.
21. Kazlauskas, R. J. *Curr. Opin. Chem. Biol*. **2000**, *4*, 81–88.
22. Kim, K. K.; Song, H. K.; Shin, D. H.; Hwang, K. Y.; Suh, S. W. *Structure* **1997**, *5*, 173–185.
23. Schrag, J. D.; Li, Y.; Cygler, M.; Lang, D.; Burgdorf, T.; Hecht, H.-J.; Schmid, R.; Schomburg, D.; Rydel, T. J.; Oliver, J. D.; Strickland, L. C.; Dunaway, C. M.; Larson, S. B.; Day, J.; McPherson, A. *Structure* **1997**, *5*, 187–202.
24. Grochulski, P.; Bouthillier, F.; Kazlauskas, R. J.; Serreqi, A. N.; Schrag, J. D.; Ziomek, E.; Cygler, M. *Biochemistry* **1994**, *33*, 3494–3500.
25. Uppenberg, J.; Ohrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T. A. *Biochemistry* **1995**, *34*, 16838–16851.
26. Uppenberg, J.; Hansen, M. T.; Patkar, S.; Jones, T. A. *Structure* **1994**, *2*, 293–308.
27. Brady, L.; Brzozowski, A. M.; Derewenda, Z. S.; Dodson, E.; Dodson, G.; Tolley, S.; Turkenburg, J. P.; Christiansen, L.; Huge-Jensen, B.; Norskov, L.; Thim, L.; Menge, U. *Nature* **1990**, *343*, 767–770.
28. Hu, C. H.; Brinck, T.; Hult, K. *Int. J. Quantum Chem*. **1997**, *69*, 89–103.
29. Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, *28*, 3629–3637.
30. Chapus, C.; Sémériva, M. *Biochemistry* **1976**, *15*, 4988–4991.
31. Chapus, C.; Sémériva, M.; Bovier-Lapierre, C.; Desnuelle, P. *Biochemistry* **1976**, *15*, 4980–4987.
32. Luic, M.; Tomic, S.; Lescic, I.; Ljubovic, E.; Sepac, D.; Sunjic, V.; Vitale, L.; Saenger, W.; Kojic-Prodic, B. *Eur. J. Biochem*. **2001**, *268*, 3964–3973.
33. Lang, D. A.; Mannesse, M. L. M.; De Haas, G. H.; Verheij, H. M.; Dijkstra, B. W. *Eur. J. Biochem*. **1998**, *254*, 333–340.
34. Cygler, M.; Grochulski, P.; Kazlauskas, R. J.; Schrag, J. D.; Bouthillier, F.; Rubin, B.; Serreqi, A. N.; Gupta, A. K. *J. Am. Chem. Soc*. **1994**, *116*, 3180–3186.
35. Kazlauskas, R. J.; Weissfloch, A. N. E.; Rappaport, A. T.; Cuccia, L. A. *J. Org. Chem*. **1994**, *56*, 2656–2665.
36. Ahmed, S. N.; Kazlauskas, R. J.; Morinville, A. H.; Grochulski, P.; Schrag, J. D.; Cygler, M. *Biocatalysis* **1994**, *9*, 209–225.
37. Haeffner, F.; Norin, T.; Hult, K. *Biophys. J*. **1998**, *74*, 1251–1262.
38. Orrenius, C.; van Heusden, C.; van Ruiten, J.; Overbeeke, P. L. A.; Kierkels, H.; Duine, J. A.; Jongejan, J. A. *Protein Eng*. **1998**, *11*, 1147–1153.
39. Norin, M.; Haeffner, F.; Adnane, A.; Norin, T.; Hult, K. *Protein Sci*. **1994**, *3*, 1493–1503.
40. Schulz, T.; Pleiss, J.; Schmid, R. D. *J. Mol. Model*. **2001**, *7*, 265–270.
41. Schulz, T.; Pleiss, J.; Schmid, R. D. *Protein Sci*. **2000**, *9*, 1053–1062.
42. Berglund, P.; Vallikivi, I.; Fransson, L.; Dannacher, H.; Holmquist, M.; Martinelle, M.; Björkling, F.; Parve, O.; Hult, K. *Tetrahedron: Asymmetry* **1999**, *10*, 4191–4202.
43. Berglund, P.; Holmquist, M.; Hult, K. *J. Mol. Catal. B: Enzym*. **1998**, *5*, 283–287.
44. Holmquist, M.; Haeffner, F.; Norin, T.; Hult, K. *Protein Sci*. **1996**, *5*, 83–88.
45. Tuomi, W. V.; Kazlauskas, R. J. *J. Org. Chem*. **1999**, *64*, 2638–2647.
46. Tafi, A.; van Almsick, A.; Corelli, F.; Crusco, M.; Laumen, K. E.; Schneider, M. P.; Botta, M. *J. Org. Chem*. **2000**, *65*, 3659–3665.
47. Haeffner, F.; Norin, T. *Chem. Pharm. Bull*. **1999**, *47*, 591–600.
48. Yagnik, A. T.; Littlechild, J. A.; Turner, N. I. J. *J. Computer-Aided Mol. Design* **1997**, *11*, 256–264.
49. Hirose, Y.; Kariya, K.; Nakanishi, Y.; Kurono, Y.; Achiwa, K. *Tetrahedron Lett*. **1995**, *36*, 1063–1066.