# CHEMBIOCHEM

# A Structure-Controlled Investigation of Lipase Enantioselectivity by a Path-Planning Approach

David Guieysse,[a] Juan Cortés,[b] Sophie Puech-Guenot,[a] Sophie Barbe,[a] Vincent Lafaquière,[a] Pierre Monsan,[a] Thierry Siméon,*[b] Isabelle André,[a] and Magali Remaud-Siméon*[a]

A novel approach based on efficient path-planning algorithms was applied to investigate the influence of substrate access on Burkholderia cepacia lipase enantioselectivity. The system studied was the transesterification of 2-substituted racemic acid derivatives catalysed by B. cepacia lipase. In silico data provided by this approach showed a fair qualitative agreement with experimental results, and hence the potential of this computational method for fast screening of racemates. In addition, a collision detector algorithm used during the pathway searches enabled the rapid identification of amino acid residues hindering the displacement of substrates along the deep, narrow active-site pocket of B. cepacia lipase and thus provided valuable information to guide the molecular engineering of lipase enantioselectivity.

## Introduction

With the increasing demand for enantiopure compounds, the transformation of a racemate into a single stereoisomer has become a key step in the development of fine chemicals for the agrochemical or pharmaceutical industries.[1,2] Among the existing enantioselective processes, biocatalysis based on the enzymatic kinetic resolution of racemates indisputably stands out.[3–9] It has been proven to be efficient and it is more and more appreciated. Enzyme enantioselectivity can be modulated by the reaction conditions, such as the temperature or the solvent employed. However, much research shows that enantioselectivity is mainly governed by the existence of appropriate fitting between the enzyme and each enantiomer, this being responsible for the preferred kinetic transformation of one of them.[10] With the advent of genome sequencing and structural biology, more and more enzymes are available to catalyse the kinetic resolution of a given racemate. In this context, the development of efficient computational methods that will enable fast evaluation of enantioselectivity is of prime interest for the development of the optimal biocatalyst or to suggest possible chemical transformation of a given substrate.

To contribute to the development of such methods, we have focused our work on the understanding of lipase enantioselectivity towards 2-substituted racemic acids such as (R,S)-bromophenylacetic acid ethyl ester.[11–13] These compounds are important chemical building blocks,[11] and their resolution was carried out through transesterification reactions catalysed by lipases in organic media. Transesterification is the result of a two-step mechanism with an acylation leading to the formation of an acyl-enzyme intermediate, which is then followed by a deacylation step. In 2-substituted racemic acids, the chiral centre is located on the acyl part of the substrate. Consequently, enantioselectivity can occur during the steps leading to the formation of the first tetrahedral intermediate and/or during the deacylation step involving the nucleophilic attack of an al-

cohol.[14] Among the lipases tested, *Burkholderia cepacia* lipase, classified as belonging to the GXSXG type, was the most enantioselective.[12] Several alcohols have been used in the deacylation step.[13] Their structures had little influence on enantioselectivity, therefore indicating that the enantiopreference occurs mostly during the step leading to the formation of the first acyl-enzyme. In addition, variation in water activity or solvents did not affect the enzyme enantioselectivity much.[15] To achieve better understanding of the enantiopreference at a molecular level, the reaction tetrahedral intermediates formed with each enantiomer were modelled by classical molecular mechanics.[13,16,17] However, the conformations obtained both with the more rapidly reacting enantiomers and with their more slowly reacting counterparts had equivalent energies, and most of them fulfilled required hydrogen-bonding interactions required for catalysis. It was thus not possible to differentiate the diastereomeric complexes in their tetrahedral intermediate forms or to explain enantiopreference reasonably on the basis of their energetics and structural analysis.[13]

Given the architecture of the *B. cepacia* lipase active site, which is located at the bottom of a narrow pocket 17 Å in length, it was suggested that enzyme ligand affinity—and, by extension, enantioselectivity—could be affected by the accessi-

[a] Dr. D. Guieysse, S. Puech-Guenot, Dr. S. Barbe, V. Lafaquière, Prof. P. Monsan, Dr. I. André, Prof. M. Remaud-Siméon
UMR5504, UMR792 Ingénierie des Systèmes Biologiques et des Procédés
CNRS, INRA, INSA
31400 Toulouse (France)
Fax: (+33) 5-61-55-94-00
E-mail: magali.remaud@insa-toulouse.fr

[b] Dr. J. Cortés, Dr. T. Siméon
LAAS—CNRS
7, avenue du Colonel-Roche, 31077 Toulouse (France)
Fax: (+33) 5-61-33-64-55
E-mail: nic@laas.fr

bility of the substrate to the catalytic site and the difficulty encountered by the substrate in adopting a productive conformation at the reaction site. Preliminary pseudo-molecular dynamics studies under constraints[13] indicated structure-based discrimination possibly mediated by hydrophobic amino acids with pivoting side chains might be encountered prior to catalysis during the access. This molecular-mechanics-based study suggested that the more rapidly reacting enantiomer encountered less hindrance than the more slowly reacting one in accessing the catalytic site. However, modelling of substrate pathways toward reaction sites remains a difficult and time-consuming task, in particular when molecular flexibility is taken into account. Most classical molecular modelling techniques consider only static properties of binding, such as the final bound ligand–receptor complex, due to the high computational costs involved in simulating ligand access from the protein surface into a narrow, deep active site. These methods can thus often describe a local and stationary state reasonably well, but are unable to predict larger molecular behaviour when dynamical aspects are fundamentally involved. The technical bottleneck is often the huge number of degrees of freedom that have to be managed in order to provide a dynamic overview of the whole phenomenon. Therefore, new original methodologies such as the robotics-based approach used in this work allow a more global mechanistic vision, notably by filtering the main degrees of freedom in molecular systems, and rely on new and highly efficient conformational search algorithms originating from recent robotics research. Such approaches have been successfully applied for studying various molecular motion problems such as ligand docking[18] and accessibility pathways[19] in flexible receptors, folding pathways in proteins[20] and DNA,[21] or conformational changes due to protein loop motions[22] and protein domain motions.[23] The key advantage of the mechanistic approach[19] that we use for computing enantiomer trajectories into the enzyme active site pocket is that it combines the efficacy of a geometric treatment of the main molecular constraints (for example, steric clash avoidance and structural constraints acting on the molecular chain model) with the performance of sampling-based path-planning algorithms,[24,25] enabling fast conformational searching. Such a combination makes the approach well adapted for handling molecular flexibility and large motions in a computationally efficient way.

In a continuation of our efforts,[13,19] the work reported here aims at further exploitation of the potential of these path-planning algorithms to tackle the influence of substrate access on *B. cepacia* lipase enantioselectivity. A set of 2-substituted racemic acids was examined with an improved version of the software prototype BioMove3D.[19,25] The large set of in silico data provided by BioMove3D was then analysed and compared with experimental results. Atom distance information collected during the path searches by the collision detector BioCD[26] was also used for fast identification of amino acid residues hindering the displacement of (*R,S*) enantiomers along the narrow and deep active site pocket.

Our results provide new insights into the interdependency between enantiomer structures and the dynamic nature of the

combined motion of both the enzyme and the substrate during the accessibility step. The role of this synergistic interaction occurring prior to catalysis in the enzyme selectivity is discussed. The potential of our novel modelling approach for use as a pre-filtering stage to select a catalyst or to accelerate the engineering of a given catalyst for a given racemate resolution is examined.

## Results and Discussion

### Enantioselectivity of *B. cepacia* lipase towards a set of (*R,S*)-2-substituted acids

A set of enantiomer pairs derived from (*R,S*)-bromophenylacetic acid ethyl ester was first synthesized. They were all subjected to a transesterification reaction in octane with octanol as second substrate and the *B. cepacia* lipase as catalyst (Scheme 1).
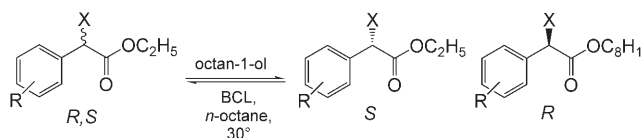


**Scheme 1.** Transesterification reactions of 2-substituted acetic acid derivatives catalysed by *B. cepacia* lipase.

A preference for the *R* compound was always observed, with enantioselectivity values ranging from 1.5 to 57 depending on the substrate (Table 1). For substrates **1–4**, which differ

**Table 1.** Enantioselectivity of free *B. cepacia* lipase towards ethyl α-X-arylacetate: influence of X and R substituents.

| Compound number | Substrate X | R | *E* value (vi *R*/vi *S*)[b] | Conversion [%] |
|---|---|---|---|---|
| **1**[a] | Br | H | 57 ($\pm$7) | 47 (84 h) |
| **2**[a] | Br | *ortho*-CH$_3$ | 4.3 ($\pm$0.5) | 15 (120 h) |
| **3**[a] | Br | *meta*-CH$_3$ | 59 ($\pm$2) | 46 (120 h) |
| **4**[a] | Br | *para*-CH$_3$ | 50 ($\pm$3) | 41 (120 h) |
| **5** | Cl | H | 26 ($\pm$1) | 48 (72 h) |
| **6** | F | H | 1.5 ($\pm$0.1) | 54 (49 h) |
| **7** | CH$_3$ | H | 35 ($\pm$3) | 29 (175 h) |

[a] Refs. [12,13]. [b] vi *R*, vi *S*: initial rates.

only in the position of the methyl group on the aromatic ring, experimentally determined *E* values showed that the enzyme is enantioselective toward the unsubstituted and the *meta*- and *para*-substituted substrates (**1**, **3**, **4**), with *E* values of 57, 59 and 50, respectively, but does not discriminate the *ortho*-substituted substrate (**2**), for which the conversion rate is notably slowed down.

The experimental results also indicate that the lipase enantioselectivity progressively decreases with the size of the halogen at the chiral centre. Indeed, the enantioselectivity drops

from an *E* value of 57 obtained with bromine (**1**), to 26 and 1.5 with chlorine (**5**) and fluorine (**6**), respectively. To evaluate the contribution of the halogen electronic effect on *B. cepacia* lipase enantioselectivity, the bromine atom ($r_{vdW} = 1.90$ Å) in **1** was substituted by a neutral methyl group (**7**; $r_{vdW} = 2.0$ Å). Although the reaction rate decreased for substrate **7**, the lipase remained selective with an *E* value of 35, suggesting that lipase enantioselectivity is governed by steric hindrance of the X-substituent.

In an attempt to evaluate the influence of substrate access and positioning in the active site on *B. cepacia* lipase enantioselectivity, a computational approach was applied to simulate and compare the enantiomer pathways.

## Computation of substrate access/exit pathways

With the goal of analysing the relationship between the structural suitability of the active site access channel and the substrate topology and lipase enantioselectivity, pathway computation was performed by using all-atom, hard-sphere models of the molecules. Both the lipase and the substrate are modelled as polyarticulated mechanisms, the motions of which are restricted by geometric constraints such as steric clash avoidance between spherical atoms with van der Waals radii. Groups of rigidly bonded atoms form the bodies, and the articulations between bodies correspond to bond torsions. In this study, the substrate and the lipase side-chains are considered flexible, while the protein backbone is kept rigid. The key interest of such mechanistic molecular models is to enable very fast pathway computation from a conformational search based on geometric constraint satisfaction, thus avoiding costly energy calculations and long computation time spared to explore the local minima of the energy landscape.

For a given substrate–enzyme pair, a path-planning technique is used to compute the substrate pathway for going from a "productive" conformation in the active site to the protein surface. We can assume that the geometric difficulty encountered during the exit is similar to that during the entry. The path-planning technique is a variant of the Rapidly-exploring Random Trees (RRT) algorithm[25, 27] specially designed for solving molecular disassembly problems. RRT is a randomized path-search algorithm that incrementally constructs a tree structure by expanding branches towards unexplored regions of the search-space, while satisfying motion constraints (for example, collision avoidance). Such behaviour makes it tend to explore the space reachable from a given initial conformation very rapidly (Figure 1).

Importantly, the computing time spent by the RRT algorithm in finding a solution pathway is representative of the geometric difficulty of the problem. When the topologies of the substrate and the active site access cavity are well suited, the RRT algorithm will compute the solution pathway with a small number of iterations. However, it will need more time to find a pathway for a substrate of a structure that makes the access/exit more constrained.

A first stage, before computation of access/exit pathways, is to determine the substrate conformation in the enzyme active
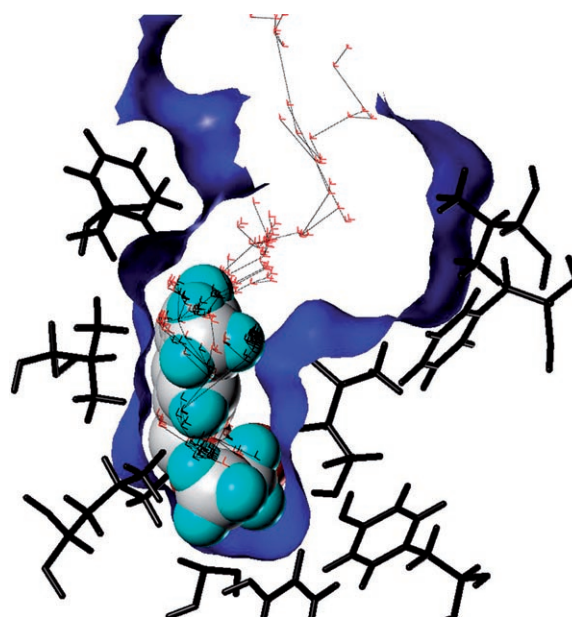


**Figure 1.** Three-dimensional projection of an RRT search tree on the position parameters of the substrate's centre of mass. The small red frames are the nodes, corresponding to feasible ensemble protein–substrate conformations. The edges (black lines) correspond to possible transitions. The tree projection tends to cover the volume accessible to the substrate from the "docking" position. It is computed with consideration of the flexibility of the substrate and the protein side-chains.

site prior to reaction, which is the starting point for the RRT algorithm. The tetrahedral intermediates mimicking the reaction transition state were constructed for each enantiomer by a previously described molecular-mechanics-based procedure.[13] An example of the models obtained for the *R* and *S* enantiomers of substrate **3** is shown in Figure 2. This conformational search allowed us to select conformations of the tetrahedral intermediates satisfying the conditions required for catalysis: 1) hydrogen-bonding pattern needed to assist the catalytic reaction, and 2) an acceptable lipase/substrate complex energy. The group found at the stereocentres (Br, Cl, F, CH$_3$) of both enantiomers clearly adopted two different orientations, pointing either towards Leu167 or towards Val267 for the *R* and the *S* enantiomer, respectively. The covalent bond between the catalytic serine (Ser87) and the substrate was then broken to create two separate molecular entities.

The robotics-based conformational search planner integrated in Biomove3D software was then used to sample the space of all possible paths that substrates **1**–**7** can take to go from the bottom (catalytic site) of *B. cepacia* lipase to the protein surface (Figure 3).

## Substrate pathway analysis

Because of the stochastic nature of the RRT algorithm, both the solution pathway and the computing time differ slightly from one run to another, so the algorithm has to be run several times in order to compute a more accurate estimate of the substrate access/exit difficulty. To shorten the time required for
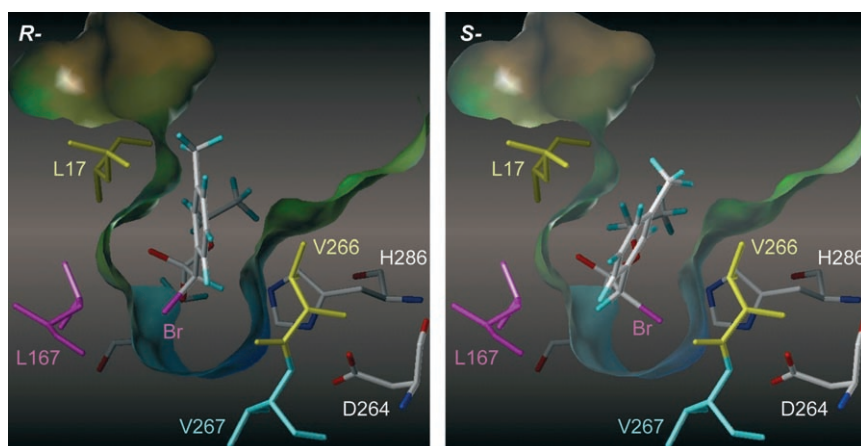
**Figure 2.** Tetrahedral intermediates of substrate **3**: the bromine atom is oriented towards Leu167 in the case of the *R* enantiomer (left), and towards Val267 in that of the *S* enantiomer (right). Val266 and Leu17, coloured in yellow, define a bottleneck in the active site.
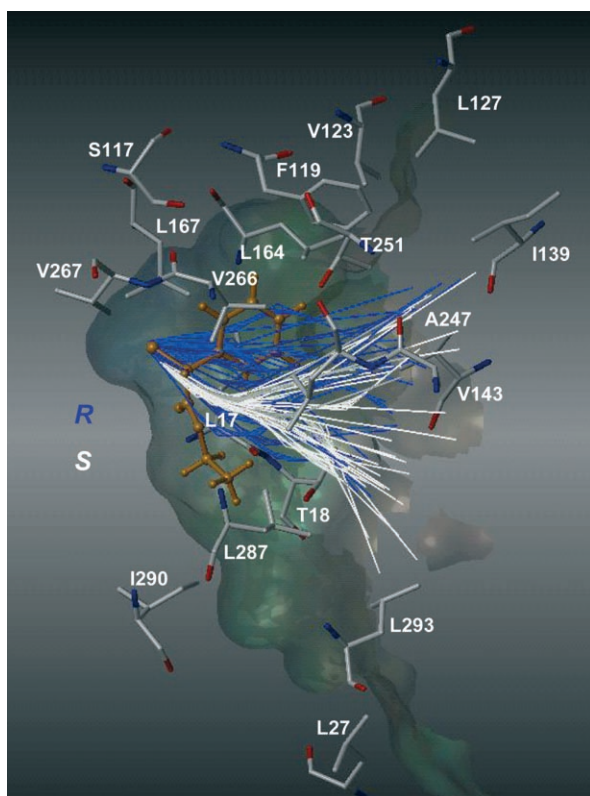


**Figure 3.** Representation of the exit trajectories computed for ethyl (*R*)- and (*S*)-bromo-*m*-tolyl-acetate (50 computed trajectories are represented for each enantiomer). One enantiomer is shown in catalytic position (orange) for reference purposes. In this illustration, trajectories display the displacement of the enantiomer centre of mass along the computed path. The distribution obtained for the *R* enantiomer (blue) appears to be clearly larger and less constrained, in particular at the bottleneck formed by V266 and L17, than in the case of the *S* enantiomer (white). *B. cepacia* lipase is shown as a Connolly channel surface mapped with the lipophilic potential calculated by the MOLCAD module implemented in Sybyl7.3.[28]

the tests, the algorithm reports a "failure" if a pathway reaching the protein surface is not found after a time limit.
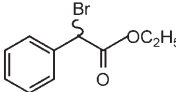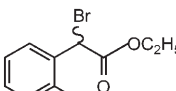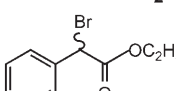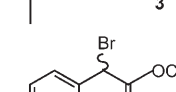
Tables 2 and 3 report the average CPU times and the number of failures for each pair of enantiomers listed in Table 1 obtained after five series of 100 calculation runs. Both parameters to some extent reflect the difficulty of the problem to be solved. Significant variations of the average CPU times are observed for the different racemates. The acyl part of the substrate has a clear influence on the geometric exploration and, by extension, on the trajectory search. Nonetheless, the average CPU time required to compute *S* trajectories (3.0–104.9 s) remains consistently larger than the CPU time necessary to compute *R* trajectories (1.0–3.0 s). For example, Figure 4 illustrates the conformational changes occurring along the pathway for the *R* and *S* enantiomers of *m*-tolyl derivatives (**3**). Both the substrate and the protein side chains have to adopt different conformations to go through the bottleneck formed by Val266 and Leu17. Indeed, the cavity narrowing at this position severely hinders the access of the *S* enantiomer, while the access of the *R* enantiomer is barely obstructed. This is also clearly illustrated in Figure 3, where the distribution of the computed trajectories for the *S* enantiomer is significantly narrower at the bottleneck.

The difficulty experienced by the path planner in computing the displacement of *S* enantiomers is clearly more pronounced for substrates **1**, **3** and **4**, as illustrated by the number of computing failures encountered in finding a solution (Table 2). In parallel, experimental results indicate a preference for the *R* enantiomer. Consequently, we can suggest that a short CPU time with very few failures reflects a more accessible path, which also correlates with the experimentally observed enantiopreference. In contrast, a longer CPU time and/or a large number of failures due to more difficult access for the substrate can be correlated to lower reactivity. On this basis, CPU time descriptor (and failures) thus appears to be a good qualitative indicator of enzyme enantioselectivity.
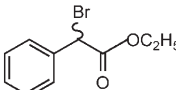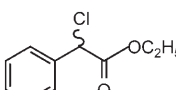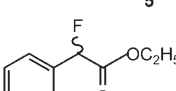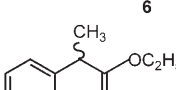
The CPU(*S*)/CPU(*R*) ratios for all enantiomer pairs listed in Table 1 were also determined and compared with experimentally determined *E* values. For substrates **2**–**4**, the ratios of the computing times follow the same trend as observed for experimentally determined *E* values (Table 2). These substrates differ only in the presence or the position of the methyl group on the aromatic ring. Notably, the motion planner never failed to compute the trajectories of the *ortho*-substituted substrates (**2**) whereas for the other three *S* enantiomers it failed several times. The presence of the methyl group at the *ortho*-position (**2**) slows down the computation of *R* enantiomer pathways (larger CPU time), but considerably speeds up the CPU time needed to search *S* enantiomer pathways, thus reducing the CPU(S)/CPU(R) ratio.

**Table 2.** Simulation of substrate trajectories: substrates differing in the R substituent on the phenyl ring.

| Substrate | Experimentally determined $E$ value (vi $R$/vi $S$)[b] | Average computing time [s] ($\pm$ s.d.) | | In silico CPU ($S$)/ CPU ($R$) ratio | Number of computing failures | |
| --- | --- | --- | --- | --- | --- | --- |
| | | $R$ | $S$ | | $R$ | $S$ |
| **1** (Br, OC$_2$H$_5$, O) | 57 ($\pm$7)[a] | 1.7 ($\pm$0.2) | 24.0 ($\pm$1.3) | 14 | 0 | 2 |
| **2** (Br, OC$_2$H$_5$, O) | 4.3 ($\pm$0.5)[a] | 3.0 ($\pm$0.7) | 7.6 ($\pm$1.0) | 3 | 0 | 0 |
| **3** (Br, OC$_2$H$_5$, O) | 59 ($\pm$2)[a] | 1.7 ($\pm$0.4) | 104.9 ($\pm$5.5) | 61 | 0 | 197 |
| **4** (Br, OC$_2$H$_5$, O) | 50 ($\pm$3)[a] | 1.0 ($\pm$0.4) | 39.0 ($\pm$3.4) | 40 | 0 | 11 |

[a] Refs. [12, 13]. [b] vi $R$, vi $S$: initial rates.

**Table 3.** Simulation of substrate trajectories: substrates differing in the X substituent at the chiral centre.

| Substrate | Experimentally determined $E$ value (vi $R$/vi $S$)[b] | Average computing time [s] ($\pm$ s.d.) | | In silico CPU ($S$)/ CPU ($R$) ratio | Number of computing failures | |
| --- | --- | --- | --- | --- | --- | --- |
| | | $R$ | $S$ | | $R$ | $S$ |
| **1** (Br, OC$_2$H$_5$, O) | 57 ($\pm$7)[a] | 1.7 ($\pm$0.2) | 24.0 ($\pm$1.3) | 14 | 0 | 2 |
| **5** (Cl, OC$_2$H$_5$, O) | 26 ($\pm$1) | 1.7 ($\pm$0.2) | 6.1 ($\pm$0.4) | 4 | 0 | 0 |
| **6** (F, OC$_2$H$_5$, O) | 1.5 ($\pm$0.1) | 1.7 ($\pm$0.1) | 3.0 ($\pm$0.7) | 2 | 0 | 0 |
| **7** (CH$_3$, OC$_2$H$_5$, O) | 35 ($\pm$3) | 1.6 ($\pm$0.2) | 8.1 ($\pm$0.3) | 5 | 0 | 0 |

[a] Refs. [12, 13]. [b] vi $R$, vi $S$: initial rates.

A relationship between CPU($S$)/CPU($R$) ratios and experimentally determined $E$ values is also observed for the series of substrates differing in the X group at their stereocentres (Table 3). Analysis of the trajectories also showed that the substrate with the largest group at the stereocentre, such as a bromine atom (**1**) or a methyl group (**7**), are more retained by the bottleneck, and so it takes more time for the motion planner to find a solution for the $S$ enantiomer. However, the CPU times ratio cannot be directly taken as quantitative estimates of $E$ values since they include some bias relating to the computational efficiency of the path-planning technique. In a prior report,[19] in which a less evolved version of Biomove3D was used, differ-ent CPU ratio values were obtained for compounds **1**, **5** and **6** but the same tendencies were observed, higher values being obtained for the larger halogen atom.

Overall, our results indicate that the time necessary to find a trajectory solution is a good indicator for predicting enzyme enantiopreference. Classification of enzyme enantioselectivity towards racemates differing only in the size of one substituting group is also possible, and in silico predictions based on CPU time ratio calculations correlate reasonably well with the enantioselectivity. If the overall catalytic sequence is considered to be describable as $E + S \rightleftarrows ES \rightarrow P$ and the equation of enantioselectivity is defined as $E = (k_{cat}^R/K_M^R)/(k_{cat}^S/K_M^S)$, by inclusion of the
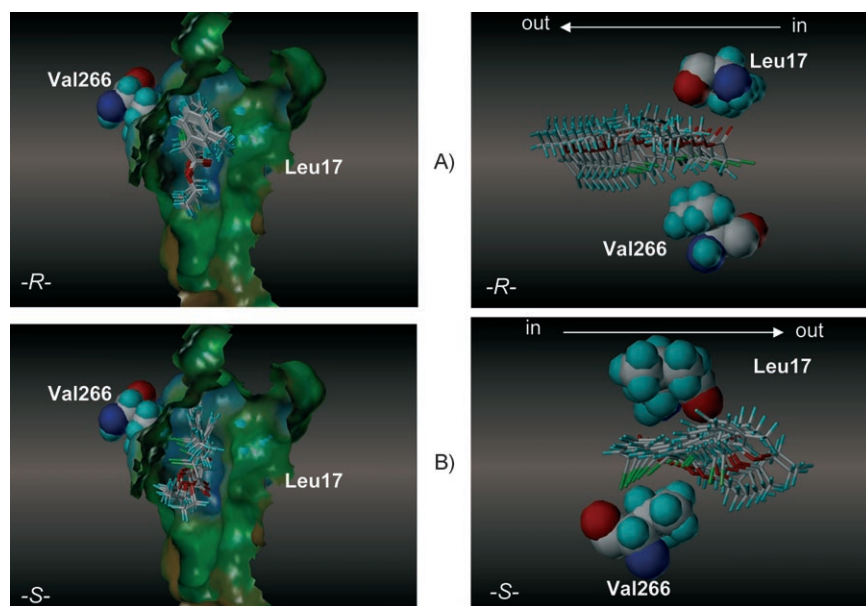
**Figure 4.** Conformational changes occurring along the paths for ethyl A) (*R*)- and B) (*S*)-bromo-*m*-tolyl-acetate (**3**). Left: View of the active site. *B. cepacia* lipase is shown as a Connolly channel surface mapped with the lipophilic potential calculated by the MOLCAD module implemented in Sybyl7.3.[28] The enzyme active site surface is colour-coded (brown indicates more lipophilic regions, whereas blue codes for more polar ones). Right: Side views of the conformational snapshots taken along the trajectory. For purposes of clarity, only conformational changes observed for the substrate are shown.

rate constants *E* can be rearranged as $E = [k_{+1}^R k_{+2}^R (k_{-1}^S + k_{+2}^S)]/[k_{+1}^S k_{+2}^S (k_{-1}^R + k_{+2}^R)]$. For each substrate, enantioselectivity is thus dependant on the various rate constants. On this basis, our modelling approach would be appropriate to predict enantioselectivity for the cases in which $k_{-1} \ll k_{+2}$ and the equation of enantioselectivity can be rearranged as $E = k_{+1}^R / k_{+1}^S$, if it is assumed that the CPU time found for each trajectory is proportional to $k_{+1}$. These assumptions give us the boundaries of our model. Indeed, they imply that the access and binding would be slower than the chemical step, and thus be rate limiting. This may be valid only in some cases, in particular for enzymes featuring highly constrained active sites and reacting with bulky substrates. In addition, the enzyme enantioselectivity is the result of a subtle balance between thermodynamic and kinetic control contributions. Quantitative predictions have not yet been achieved, and the introduction of some energetic calculations during the trajectory search is now being considered, in particular to take electrostatic attraction or repulsion into account in order to improve our approach further.

### Collision mapping—a tool for enzyme engineering

BioCD,[26] the algorithm for collision detection and distance computation integrated in Biomove3D, was used for fast identification of amino acid residues hindering the displacement of (*R*,*S*) enantiomers along the active site pocket of *B. cepacia* lipase. The distances between substrate and protein atoms were measured along each solution pathway, and "contact-pairs" were reported when these distances were below 85 % of the van der Waals equilibrium distances. The relative frequency of contacts between the enzyme amino acid residues and (*R*,*S*)

racemates along the computed pathways is shown in the form of a collision histogram (Figure 5). A contact with a frequency of 100 % for a given residue means that atoms of the substrate and of this residue are, at one moment, very close in all the computed pathways. Ten amino acid residues display a higher frequency of contact with the (*R*,*S*) enantiomers. These amino acids can be classified into three groups (Figure 6): 1) amino acids of the active site (such as the catalytic His286) and residues forming the bottleneck (Leu17 and Val266), 2) the amino acids right after the bottleneck, located in the intermediate region of the funnel, 5–8 Å from the bottom of the pocket (Thr18, Tyr29 and Leu287), and 3) the amino acids of the catalytic pocket located closer to the protein surface (Tyr23, Phe146, Phe119 and Leu293). All these amino acids are in proximity, which implies that the motion of one residue involves a gear-like motion of others. Contacts are mainly detected (with the exception of Leu17) with amino acid side chains attached to backbone atoms that contribute greatly to the topology of the pocket. Of the residues listed in Figure 5, some appear in a marginal way, such as Leu27, Thr291, Ala247, Val123, Phe142, Leu248, Thr251 and Phe146. These lower collision frequencies again reflect here the location of these residues in the pocket, closer to the entrance of the pocket.

Analysis of the histogram of Figure 5 indicates that overall, the more slowly reacting *S* enantiomers are in contact with residues of the active site pocket more often than the *R* enantiomers. For each racemate, the frequency of contacts observed is in good agreement with the average computing times needed by the motion planner, as well as with the experimentally observed enantiopreference determined for *B. cepacia* lipase. These remarks support the in silico predictions of the relative difficulty of accessibility encountered by the *S* enantiomers relative to their *R* counterparts quite well and highlight the differences that exist between an active site exposed to a solvent and a buried one surrounded by amino acid walls. The number of contacts is also believed to be related to the difficulty of positioning of a given substrate in a productive conformation at the catalytic site.

A more detailed structural analysis reveals that the accessibility of the active site pocket and the positioning of the substrate in it is basically governed mainly by four amino acids (His286, Val266, Thr18 and Leu17). Indeed, the *R* and *S* enantiomers have to go through the bottleneck (Val266, Leu17, and neighbouring Thr18) in order to gain access to the core of the
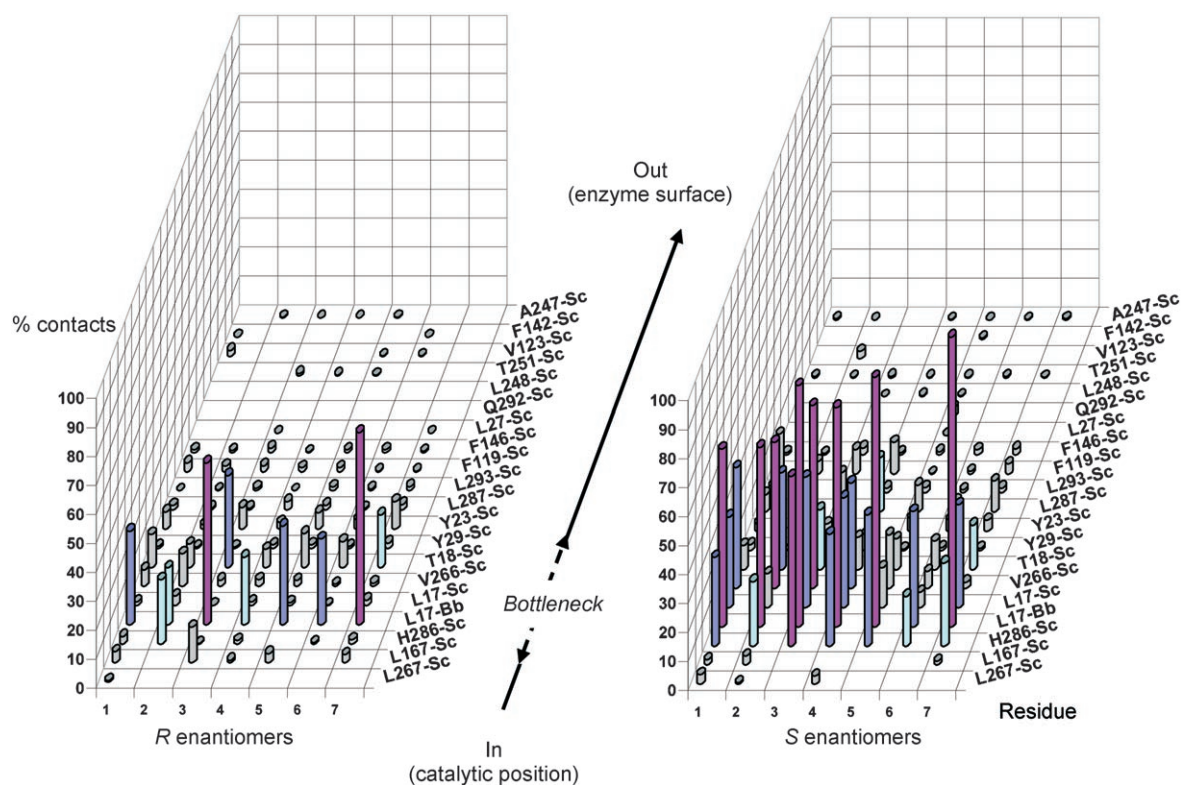
**Figure 5.** Histogram representing the relative frequency of contacts between the enzyme amino acid residues and (*R,S*) racemates along the computed paths (computed over 50 runs). Contacts over 50 % are coloured in magenta, between 30–50 % in purple, and between 15–30 % in cyan. (Bb = Backbone, Sc = Side chain of amino acid residues.)
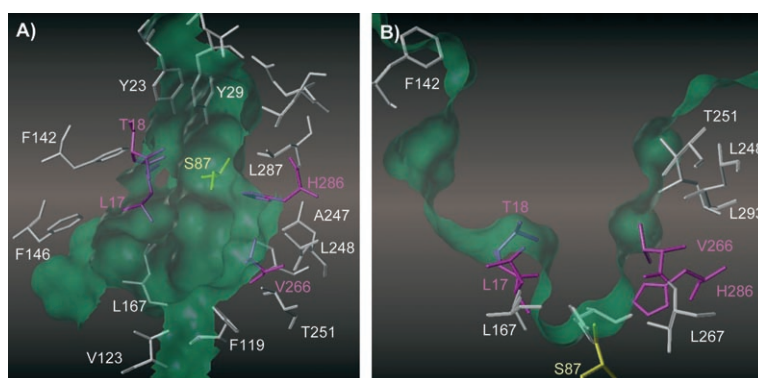


**Figure 6.** Representation of the main amino acid residues identified by the collision detector. A) Front view of the *B. cepacia* lipase active site pocket. B) Cross-section view of the active site, showing a funnel shape, with Val266 and Leu17 forming a bottleneck. The catalytic triad (Ser87, His286, Asp264) and the oxyanion hole (Gln88, Leu17) are located at the bottom of the pocket. The four amino acid residues presenting the highest frequency of contacts (≥ 50 % contacts) with substrates are coloured in magenta (Leu17, Thr18, Val266 and His286). For reference, the catalytic Ser87 is coloured in yellow.

mutations would be expected to produce drastic geometric changes in the active site space, although the effects of single mutations at these four positions were not investigated. The accordance of these results demonstrates the interest of the development of tools, such as Biomove3D,[24] to guide enzyme engineering.

More and more, computational and experimental studies have revealed the importance of enzyme dynamics for catalysis[31–34] and, for example, the connection between enzyme active site motion and substrate turnover[35–37] or the effect of muta-

catalytic pocket to form the tetrahedral intermediate. Interestingly, four of the amino acid residues identified by the collision detector (Leu17, Phe119, Leu167 and Leu266 (equivalent to Val266 in our *B. cepacia* lipase) have been selected for mutation in previous investigations into *B. cepacia* enantioselectivity engineering.[29,30] Combinations of amino acid substitutions at these four positions were found to be very effective for changing the enantiopreference from the *S* form substrate to the *R* form in the case of ethyl 3-phenylbutyrate derivatives. These

tions distal to the active site.[38–40] With regard to lipase enantioselectivity, theoretical analyses are generally based on descriptions of the reactions by transition state theory. Thermodynamic investigations have shown that differences between enantiomers in activation free energy are due to variations of both the enthalpic and the entropic terms,[41–43] but in most reports the role of motions in the reaction environment has been neglected. This is due to the difficulty involved in experimentally demonstrating the gate-keeping and rate-controlling

role of the molecular transport and motion from the enzyme surface to the reactive centre, which requires data over a wide range of timescales and the use of sophisticated experimental methods. In the case of the kinetic resolution catalysed by lipases with deep and narrow active sites, modelling of the substrate pathways while accounting for molecular flexibility is also a difficult task for currently available techniques. Molecular dynamics simulation could be applied to model substrate pathways but will provide results only in the nanosecond time range. Monte-Carlo algorithms enable the computation of large motions but failed to provide a continuous chronological description of the motion. The pseudo-molecular dynamics under constraint that we had previously proposed needed manual corrections to orientate protein side chains correctly and avoid steric conflicts; it was also biased by a tendency of falling in local minima of the energy landscapes and was much too long for routine use.

The new approach proposed in this article, based on efficient path-planning algorithms and fast geometric operators specially designed for complex articulated chains, is well adapted to and very efficient for exploring high-dimensional conformational spaces. In addition, our simulations indicate a qualitative correlation between the ratio of the average times needed to compute the enantiomer pathways and the enantioselectivity value, thus revealing their potential for fast screening of racemates. In addition, analysis of the computed pathways also enables rapid identification of the amino acid(s) constraining access and movement of the enantiomers through the narrow and deep active site pocket.

With regard to *B. cepacia* lipase, these results may help to drive directed mutagenesis experiments to enhance lipase enantioselectivity further. At a more general level, this fast technique could be used as a pre-screening procedure to provide highly valuable information on mutation sites relatively remote from the enzyme reaction site that cannot be easily explored by classical molecular modelling techniques. Indeed, replacement of such a residue might not only alter the accessibility pathway of a substrate by making it more or less constrained, but could also induce conformational changes in the substrate along the path, which could differ depending on the enantiomer and thus exhibit modified enantioselectivity. This strategy was applied to *B. cepacia* lipase to identify amino acid positions important for enantioselectivity and to define the best combination of mutations to introduce into the catalyst in order to enhance its enantioselectivity with respect to a given racemate of interest. Construction and screening of mutant libraries targeted at these positions are currently being completed in our laboratory. Results will be reported in a future communication.

## Experimental Section

All reagents were of commercial quality and were purchased from Sigma/Aldrich. *n*-Octane was dried over molecular sieves (3 Å) before use. Free lipase from *Burkholderia cepacia* (Chirazyme L-1, lyo) was purchased from Roche Diagnostics (Germany).

**General procedure for the preparation of 2-substituted acids and their ester derivatives**: The synthesis of compounds **1–4** has been described in a prior report.[12]

(±)-**Ethyl 2-chlorophenylacetate (5)**: Chlorophenylacetyl chloride (43.6 mmol), ethanol (100 mL) and *p*-toluenesulfonic acid (0.2 g) were stirred under reflux (4–5 h). Ethanol was evaporated at 35 °C (waterpump vacuum). The residual oil was dissolved in dichloromethane (25 mL) and washed with a saturated sodium bicarbonate solution (3×10 mL) and finally with distilled water (10 mL). The dichloromethane solution was dried over magnesium sulfate, filtered and then evaporated to dryness at 35 °C (water pump vacuum).

(±)-**Ethyl 2-fluorophenylacetate (6)**: Mandelic acid (commercial) was first transformed into ethyl mandelate by the same procedure as described above for the compound **5**. A solution of ethyl mandelate (11.1 mmol) in dichloromethane (2 mL) was added slowly to a solution of DAST [(diethylamino)sulfur trifluoride, 11.1 mmol] in dichloromethane (5 mL), cooled to −78 °C. The reaction mixture was allowed to warm to room temperature and mixed with cold water. The lower layer was separated, washed with water, dried with MgSO₄, and evaporated to dryness at 35 °C (water pump vacuum). The obtained yellow oil was purified on silica gel with *n*-hexane/ethyl acetate (95:5, *v/v*) as eluent to provide a colourless liquid.

(±)-**Ethyl 2-phenylpropionate (7)**: Compound **7** was synthesized by the same procedure as described above for compound **5**, by starting from (±)-2-phenylpropionic acid.

**Spectroscopic data**: Infrared spectra were recorded on a Perkin–Elmer 1310 infrared spectrophotometer. $^1$H and $^{13}$C NMR spectra were recorded on a Bruker AC 200.1 ($^1$H 200.1 MHz and $^{13}$C 50.3 MHz) spectrometer.

(±)-**Ethyl 2-chlorophenylacetate (5)**: Yield: 73%; $^1$H NMR (200 MHz, CDCl₃, 25 °C): $\delta = 1.21$–1.28 (t, $^3J_{H,H} = 7.1$ Hz, 3H; CH₃), 4.15–4.28 (qd, $^3J_{H,H} = 2.7$ and 7.1 Hz, 2H; CH₂), 5.35 (s, 1H; CH), 7.35–7.53 ppm (m, 5H; ArH); $^{13}$C NMR (50 MHz, CDCl₃, 25 °C), $\delta = 14.01$ (CH₃), 59.19 (CH), 62.55 (CH₂), 127.99 (×2), 128.90 (×2), 129.30, 135.93, 168.41 ppm (COO); IR (neat): $\tilde{v} = 1750$ and 1730 (C–O), 1600 and 1475 (C–C), 1280–1140 cm$^{-1}$ (C–O); elemental analysis calcd (%) for C₁₀H₁₁O₂Cl: C 60.59, H 5.60; found: C 60.41, H 5.75%.

(±)-**Ethyl 2-fluorophenylacetate (6)**: Yield: 61%; $^1$H NMR (200 MHz, CD₃COCD₃, 25 °C), $\delta = 1.16$–1.23 (t, $^3J_{H,H} = 7.1$ Hz, 3H; CH₃), 4.14–4.27 (qd, $^3J_{H,H} = 3.4$, 7 Hz, 2H; CH₂), 5.84 and 6.08 (d, $^2J_{H,F} = 47.5$ Hz, 1H; CHF), 7.42–7.52 ppm (m, 5H; ArH); $^{13}$C NMR (50 MHz, CD₃COCD₃, 25 °C), $\delta = 14.32$ (CH₃), 62.12 (CH₂), 88.25 and 91.85 (d, CHF), 127.77, 127.88, 129.64 (×2), 130.43, 135.64 and 136.04, 168.77 and 169.33 ppm (d, COO); IR (neat): $\tilde{v} = 1750$ and 1730 (C–O), 1600 and 1475 (C–C), 1280–1140 cm$^{-1}$ (C–O); elemental analysis calcd (%) for C₁₀H₁₁O₂F: C 65.91, H 6.09; found: C 65.58, H 7.73%.

(±)-**Ethyl 2-phenylpropionate (7)**: Yield: 80%; $^1$H NMR (200 MHz, CDCl₃, 25 °C): $\delta = 1.11$–1.18 ppm (t, $^3J_{H,H} = 7.1$ Hz, 3H; CH₃), 1.41–1.45 (d, $^3J_{H,H} = 7.1$ Hz, 3H; CH₃), 3.69–3.79 (q, $^3J_{H,H} = 7.1$ Hz, 1H; CH), 4.01–4.13 (qd, $^3J_{H,H} = 2.4$ and 7.1 Hz, 2H; CH₂), 7.23–7.33 ppm (m, 5H; ArH); $^{13}$C NMR (50 MHz, CDCl₃, 25 °C): $\delta = 14.43$ (CH₃), 19.17 (CH), 46.03 (CH), 61.00 (CH₂), 127.77, 128.24 (×2), 129.38 (×2), 142.06, 174.62 ppm (COO); IR (neat): $\tilde{v} = 1750$ and 1730 (C–O), 1600 and 1475 (C=C), 1280–1140 cm$^{-1}$ (C–O); elemental analysis calcd (%) for C₁₁H₁₄O₂: C 74.12, H 7.92; found: C 73.71, H 7.75%.

**Enzyme-catalysed transesterification reactions**: A typical transesterification was carried out in *n*-octane (5 mL) containing the ester (0.25 mmol, 50 mM), octan-1-ol (0.75 mmol, 150 mM) and lipase (5 mg mL$^{-1}$). The water activity was set to 0.04, and the temperature was maintained at 30 °C. The mixture was shaken at 500 rpm for the time indicated in Table 1. The progress of the reaction was followed by sampling the reaction at regular intervals.

**HPLC analysis**: The chiral HPLC instrument was equipped with a chiral column [Chiralpack AD, AS or OJ (25 cm×4.6 mm), Daicel Chemical Industries Ltd, Japan] connected to a UV detector (at 254 nm). The conditions were: *n*-hexane/isopropanol 90:10, *v/v* for **1** and **5**–**9** with OJ column, and *n*-hexane/isopropanol 99.8:0.2, *v/v* for **2**–**4** with AD column. Each transesterification reaction sample was diluted 8–10 times in the mobile phase, filtered before injection (20 μL), and analysed at a flow rate of 1.0 mL min$^{-1}$ at RT. The retention times ($t_R$/min) were as follows: **1**: 15.40 (*S*), 17.80 (*R*); **2**: 7.01 (*S*), 7.44 (*R*); **3**: 7.14 (*S*), 7.50 (*R*); **4**: 8.68 (*S*), 9.42 (*R*); **5**: 13.30 (*S*), 14.30 (*R*); **6**: 9.10 (*S*), 10.60 (*R*); **7**: 6.0 (*S*), 9.10 (*R*).

**Determination of conversion rates and enantioselectivities (*E* values)**: The conversion rate was calculated from HPLC results by $C = 1 - [(R+S)t/(R+S)t = 0] \times 100$. The enantioselectivity value was the ratio of the initial rate of *R* enantiomer production (vi *R*) versus the initial rate of *S* enantiomer production (vi *S*): *E* value = (vi *R*/vi *S*). The initial rates were determined, before 10% of substrate conversion, by linear regression over at least five points.

## Computational Methods

**Construction of starting geometries**: Starting models used to map the enantiomer trajectories into the *B. cepacia* active site were generated by the procedure described in an earlier report. Firstly, a systematic conformational search around two dihedral angles—$\Phi_x$ (C1-C2-C7-C8) and $\Phi_Y$ (C2-C7-C8-O9)—was performed on a model of the substrate covalently bound to a catalytic serine (Ser87). Geometry of low-energy conformers was then fully relaxed. Subsequently, the minima solutions obtained in vacuum were placed within the enzyme active site environment. To do this, we used the X-ray structure of *B. cepacia* in complexation with a diethylphosphate inhibitor, kindly provided by M. Cygler and J. Schrag (NRC Biotechnology Research Center, Montreal). For purposes of consistency with our earlier paper we used the same truncated form of the enzyme in this work.[13] After an additional energy optimization of the covalent intermediate complex models, only solutions satisfying the hydrogen bonding pattern required for catalysis and having an acceptable energy were kept.

**Calculation of enantiomer trajectories**: To allow the displacement of the substrates along the active site path, the covalent bond between the catalytic serine and the carbon of the substrate carbonyl function was broken to create two separate molecular entities. These structures, in which the given substrate is in a catalytic position but not bound to the enzyme, were subsequently minimized in order to correct the hybridization of the carbonyl function from sp$^3$ to sp$^2$ and to generate the molecular models used to start the search for trajectories. From the starting docked position, enantiomer trajectories were computed with use of the Biomove3D integrated path planner,[24] going from the bottom towards the entrance of the active site. During the search, the ligand was considered flexible, as well as the side chains of 17 amino acid residues bordering the access channel to the active. In total, 68 degrees of freedom of the entire enzyme–substrate complex were taken into account (11 for the ligand and 57 for the protein side-chains). Atoms were modelled at 80% of their van der Waals radii.

To ensure that the geometric space of the active site was fully explored during the trajectory search, a preliminary test series of an increasing number of trajectories (25, 50, 100, 500 and 1000) was calculated for a given pair of enantiomers. The CPU time was then averaged over the total number of tests performed. Comparison of the average time needed for computing trajectories indicated very similar results whenever series of 50, 100, 500 or 1000 random tests were performed. On this basis, the choice was made to run five series of 100 random tests to compute the trajectories for each pair of enantiomers listed in Table 1. The standard deviation was thus calculated from these five series of 100 tests to improve the estimate of the imprecision over the computing times.

**Equipment**: Calculations, molecular constructions and graphic displays were performed on an Intel Pentium4 PC with a 3.2 GHz processor and an O2 (R10 000) Silicon Graphics workstation.

[1] J. S. Carey, D. Laffan, C. Thomson, M. T. Williams, *Org. Biomol. Chem.* **2006**, *4*, 2337–2347.

[2] M. Breuer, K. Ditrich, T. Habicher, B. Hauer, M. Keßeler, R. Stürmer, T. Zelinski, *Angew. Chem.* **2004**, *116*, 806–843; *Angew. Chem. Int. Ed.* **2004**, *43*, 788–824.

[3] A. Ghanem, *Tetrahedron* **2007**, *63*, 1721–1754.

[4] V. Gotor-Fernández, R. Brieva, V. Gotor, *J. Mol. Catal. B* **2006**, *40*, 111–120.

[5] A. Liljeblad, L. T. Kanerva, *Tetrahedron* **2006**, *62*, 5831–5854.

[6] A. P. Osborne, D. Brick, G. Ruecroft, I. N. Taylor, *Org. Process Res. Dev.* **2006**, *10*, 670–672.

[7] J. Paetzold, J. E. Backvall, *J Am Chem. Soc* **2005**, *127*, 17620–17621.

[8] A. J. Straathof, S. Panke, A. Schmid, *Curr. Opin. Biotechnol.* **2002**, *13*, 548–556.

[9] U. T. Bornscheuer, R. J. Kazlauskas, *Hydrolases in Organic Synthesis. Regio- and Stereoselective Biotransformations*, Wiley, Weinheim, **1999**.

[10] A. Mezzetti, J. D. Schrag, C. S. Cheong, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 427–437.

[11] D. Guieysse, C. Salagnad, P. Monsan, M. Remaud-Simeon, *Tetrahedron: Asymmetry* **2001**, *12*, 2473–2480.

[12] D. Guieysse, C. Salagnad, P. Monsan. M. Remaud-Simeon, *Tetrahedron: Asymmetry* **2003**, *14*, 317–323.

[13] D. Guieysse, C. Salagnad, P. Monsan, M. Remaud-Simeon, V. Tran, *Tetrahedron: Asymmetry* **2003**, *14*, 1807–1817.

[14] M. Cygler, P. Grochulski, R. J. Kazlauskas, J. D. Schrag, F. Bouthillier, B. Rubin, A. N. Serreqi, A. K. Gupta, *J. Am. Chem. Soc.* **1994**, *116*, 3180–3186.

[15] P. L. A. Overbeeke, J. A. Jongejan, J. J. Heinen, *Biotechnol. Bioeng.* **2000**, *70*, 278–290.

[16] V. Léonard, L. Fransson, S. Lamare, K. Hult, M. Graber, *ChemBioChem* **2007**, *8*, 662–667.

[17] S. Mittal, S. Khanna, A. Roy, P. V. Bharatam, H. P. S. Chawla, *Enzyme Microb. Technol.* **2005**, *36*, 232–238.

[18] M. S. Apaydin, D. L. Brutlag, C. Guestrin, D. Hsu, J. C. Latombe, C. Varma, *J. Comput. Biol.* **2003**, *10*, 257–281.

[19] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, V. Tran, *Bioinformatics* **2005**, *21*, i116–i125.

[20] N. M. Amato, K. A. Dill, G. Song, *J. Comput. Biol.* **2003**, *10*, 239–255.

[21] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, N. M. Amato, *Proc. RECOMB* **2004**, 252–261.

[22] J. Cortés, T. Siméon, M. Remaud-Siméon, V. Tran, *J. Comput. Chem.* **2004**, *25*, 956–967.

[23] S. Kirillova, J. Cortés, A. Stéphaniu, T. Siméon, *Proteins* **2008**, *70*, 131–143.

[24] L. Jaillet, T. Siméon, *IEEE Int. Conf. on Int. Robots and Systems* **2004**.

[25] J. Cortés, L. Jaillet, T. Siméon, *IEEE Int. Conf. on Int. Robots and Systems* **2007**.

[26] V. Ruiz de Angulo, J. Cortes, T. Siméon, *Robotics: Science and Systems* **2005**.

[27] S. M. Lavalle, J. J. Kuffner in *Algorithmic and Computational Robotics: New Directions* (Eds.: B. R. Donald, K. Lynch, D. Rus), A. K. Peter, Boston, **2001**, pp. 293–308.

[28] Sybyl7.3, Tripos, St Louis, MO, USA.

[29] Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585–592.

[30] R. Kato, H. Nakano, H. Konishi, K. Kato, Y. Koga, T. Yamane, T. Kobayashi, H. Honda, *J. Mol. Biol.* **2005**, *351*, 683–692.

[31] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2006**, *312*, 208–209.

[32] S. Hammes-Schiffer, S. J. Benkovic, *Annu. Rev. Biochem.* **2006**, *75*, 519–541.

[33] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196–1202.

[34] P. K. Agarwal, S. R. Billeter, P. T. Rajagopalan, S. J. Benkovic, S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2794–2799.

[35] P. K. Agarwal, *Microb. Cell Fact.* **2006**, *5*, 2.

[36] P. K. Agarwal, *J. Am. Chem. Soc.* **2005**, *127*, 15248–15256.

[37] E. Z. Eisenmesser, D. A. Bosco, M. Akke, D. Kern, *Science* **2002**, *295*, 1520–1523.

[38] L. Wang, N. M. Goodey, S. J. Benkovic, A. Kohen, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15753–15758.

[39] L. Wang, N. M. Goodey, S. J. Benkovic, A. Kohen, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1307–1315.

[40] L. Wang, S. Tharp, T. Selzer, S. J. Benkovic, A. Kohen, *Biochemistry* **2006**, *45*, 1383–1392.

[41] J. Ottosson, L. Fransson, K. Hult, *Protein Sci.* **2002**, *11*, 1462–1471.

[42] J. Ottosson, L. Fransson, J. W. King, K. Hult, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2002**, *1594*, 325–334.

[43] J. Ottosson, J. C. Rotticci-Mulder, D. Rotticci, K. Hult, *Protein Sci.* **2001**, *10*, 1769–1774.