PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Anchoring effects in a wide binding pocket: The molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from *B. megaterium*

Ricardo J.F. Branco,[1] Alexander Seifert,[2] Michael Budde,[2] Vlada B. Urlacher,[2] Maria J. Ramos,[1] and Jürgen Pleiss[2]*

[1] REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, P 4169-007 Porto, Portugal

[2] Institute of Technical Biochemistry, Universitaet Stuttgart, Allmandring 31, D 70569 Stuttgart, Germany

ABSTRACT

The molecular basis of regioselectivity of cytochrome P450 monooxygenases from *Bacillus megaterium* (CYP102A1) with its flexible and widely opened active site is still not well understood. In the present work (−)-α-pinene bound complexes with two triple mutants were modeled to elucidate the contribution of the three major factors that mediate selectivity: active site shape, protein flexibility, and chemical reactivity of the substrate. For the triple mutant A74G F87V L188Q (GVQ), one stable, productive conformation of the substrate (conformation I) was identified by multiple molecular dynamics simulations. The model predicts pinene epoxide as a major product (42% pinene oxide, 23% verbenol) which is in agreement with the experimental product profile (70% pinene oxide, 20% verbenol). In contrast, for the triple mutant A74G F87G L188Q (GGQ) two stable productive substrate conformations were identified (conformations IIa and IIb), and verbenol was predicted as major product (81% verbenol, 16% myrtenol), which is in agreement with experimental results (77% verbenol, 10% myrtenol). The effect of chemical reactivity of the substrate was demonstrated by comparison of (−)-α-pinene to its regioisomer (−)-β-pinene, where the product profile is shifted from 68% pinocarveol and 32% myrtanal in mutant GVQ, to 40% pinocarveol and 60% myrtanal in mutant GGQ. Our results strongly suggest a major role of residue 87 in anchoring (−)-α-pinene during substrate binding which provides a simple and elegant rationalization of the dynamic structure of this enzyme-substrate complex.

Proteins 2008; 73:597–607.

Key words: molecular dynamics simulation; pinene; CYP102A1 monooxygenase; mutants.

INTRODUCTION

The selective oxidation of hydrocarbons using molecular oxygen which leads to valuable chemical intermediates is still a challenge for traditional chemical synthesis. Oxyfunctionalization of a vast variety of organic molecules by cytochrome P450 monooxygenases (CYPs) represents an attractive route to alcohols and epoxides due to mild reaction conditions and the potentially high regio- and stereoselectivity. To increase activity and improve selectivity of recombinant CYPs, protein engineering tools such as directed evolution and rational design have been successfully applied[1–3] in recent years.

The active site of CYPs contains a heme moiety where the fifth iron ligand is formed by a cysteinate. Introduction of oxygen atom into C—H bonds occurs via a radical mechanism which involves the abstraction of the hydrogen by an oxyferryl species followed by a rebound between the substrate and the hydroxyl-iron center.[4] High-resolution X-ray crystal structures,[5,6] solution NMR structures,[7] as well as density-functional calculations[8,9] provided important insights into the catalytic mechanism of P450 monooxygenases.[10]

The CYP102A1 monooxygenase is a most efficient medium- to long-chain (C_{12} to C_{22}) fatty acid hydroxylase[11,12] which also epoxidizes unsaturated fatty acids,[13] however the wild-type enzyme has no oxidizing activity toward terpenes like (−)-α- and (−)-β-pinene. CYP102A1 mutants have been described, which are able to oxidize alkanes,[14] polycyclic aromatic hydrocarbons,[15]

Additional Supporting Information may be found in the online version of this article.
Abbreviations: CYPs, cytochrome P450 monooxygenases; MD, molecular dynamics; RMSD, root mean square deviation.

Grant sponsor: German Research Foundation; Grant sponsor: Fundação para a Ciência e a Tecnologia (FCT); Grant number: SFRH/BD/10598/2002.
*Correspondence to: Jürgen Pleiss, Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, D 70569 Stuttgart, Germany. E-mail: juergen.pleiss@itb.uni-stuttgart.de
Received 1 August 2007; Revised 3 March 2008; Accepted 20 March 2008
Published online 12 May 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22083

monoterpenes,[16] and sesquiterpenes.[17,18] Oxidation of low value terpenes to higher value derivatives has been recognized for some time as an attractive opportunity for synthetic chemistry and biocatalysis,[19] for example (−)-α-pinene, a waste product from pulp processing, whose oxygenated products are valuable compounds for the flavoring, fragrance,[20,21] and pharmaceutical industries.

To better understand and predict regio- and stereoselectivity of the enzyme, it is crucial to know the preferred orientations and conformations of the substrate in the active site.[22,23] Since the first crystal structure of the heme domain of soluble, self-sufficient CYP102A1 from *Bacillus megaterium* (also referred to as $P450_{BM3}$),[24] a total of 20 further crystal structures of this enzyme have been solved, but only few in complex with a substrate.[5,6] On the basis of structure information, key residues of the catalytic cycle were identified[25] and verified by site-directed mutagenesis.[26–29] It has been shown that the active site residue F87 is highly conserved and mutations at this position will affect the activity and stereoselectivity of the enzyme.[13,30]

However, the X-ray structures of the complexes between the P450 and substrate exist often in a nonproductive conformation: either the substrate binds at a large distance from the heme[31] or the orientation of the substrate cannot explain the product profile.[32]

Previously, it has been demonstrated that molecular dynamics simulations is suitable to study the molecular basis of selectivity in CYP102A1.[33,34] To explore the molecular mechanism of substrate recognition and predict regioselectivity, we performed multiple MD simulations on two CYP102A1 mutants in complex with (−)-α-pinene. To elucidate the contribution of chemical reactivity of the substrate to the regioselectivity of the reaction, we compared this substrate to its regioisomer (−)-β-pinene. The model based on MD simulations on the triple mutant CYP102A1 with (−)-α-pinene successfully predicted the regioselectivity of both mutants towards the regioisomer (−)-β-pinene.

EXPERIMENTAL METHODS

Materials

Yeast extract and tryptone/peptone from caseine were purchased from Roth (Karlsruhe, Germany). All other chemicals were purchased from Fluka (Buchs, Switzerland) or Sigma (Deisenhofen, Germany) and were of analytical grade or higher.

Mutants expression

CYP102A1 and its mutants A74G L188Q (GQ), A74G F87V L188Q (GVQ) and CYP102A1 A74G F87G L188Q (GGQ) were heterologously expressed in *E. coli* BL21 (DE3) using the pET28+ expression system as reported previously.[35] Mutations were introduced by PCR using the QuickChange™ site-directed mutagenesis kit from Stratagene (La Jolla, CA) according to the manufacturer's protocol.

Enzyme activity measurements

The CYP concentration was calculated from CO-difference spectra using an extinction coefficient of 91 mM^{-1} cm^{-1} as described elsewhere.[36] Initial activity of the mutants towards (−)-α- and (−)-β-pinene was determined using an NADPH oxidation assay. A 1-mL cuvette was used containing 50 m*M* potassium phosphate buffer pH 7.5, 2% (v/v) DMSO, 0.6 m*M* (−)-α- or (−)-β-pinene, and 0.5 to 0.9 μ*M* CYP. Reaction was started by addition of 100 μL 1.2 m*M* NADPH solution. Absorption decrease at 340 nm was measured and the initial activity was determined from the slope of the first 30 s.

GC-MS analysis

The aqueous reaction mixture was extracted twice with 500 μL of methyl-*tert*-butyl ether and the collected organic phase was dried over magnesium sulphate. Analyses of (−)-α- and (−)-β-pinene and their reaction products were performed on a Shimadzu QP2010 GC/MS with EI-ionization; the GC was equipped with a FS-supreme-5 capillary column (length: 30 m, internal diameter: 0.25 mm, film thickness: 0.25 μm). The GC temperature program was as follows: 90°C, 5 min iso; 15°C/min to 250°C, 1 min iso-; injector temperature 250°C. The products were identified using authentic samples of (−)-α-pinene-oxide, (*S*)-*cis*-verbenol, and (−)-*trans*-pinocarveol. Other products were identified by comparison of their characteristic mass fragmentation pattern with the NIST mass spectrometry data base.[37]

Substrate docking

The X-ray structure of the substrate-free CYP102A1 was obtained from the Protein Data Bank[38] (PDB code 1BU7, chain A[39]) and used as initial structure. Two triple mutants GVQ and GGQ, and one double mutant GQ were designed using the visualization program PyMOL v0.99.[40]

The geometry of (−)-α-pinene was optimized at a DFT level using the UB3LYP functional and the 6-31G basis set as implemented in the program GAUSSIAN 98.[41] The substrate was docked into the crystallographic structure with the docking software FlexX version 2.0[42] assuming O^{Fe} as a target and a default overlapping volume of 3.5 Å. The best docking solutions obtained for both mutants were manually clustered according to the orientation of the substrate, and redundant solutions were discarded. This procedure resulted in nine solutions of (−)-α-pinene for each mutant, each of which was taken as a starting configuration for molecular dynamic simulations of the complex.

Molecular dynamics simulations

Particle Mesh Ewald molecular dynamics simulations were carried out using the software package AMBER 8[43] and the all-atom AMBER force field ff99.[44] The SHAKE algorithm[45] was applied to all C—H bonds. A time step of 1 fs was used, and the system was coupled to constant temperature and pressure.[46]

The protonation states of all histidines, glutamic and aspartic acids at pH 7 were calculated by the program TITRA.[47] All the histidines except H266 were unprotonated. The remaining titratable amino acids were protonated by the xleap module of the AMBER 8 software. The RESP fitting charges[48] for (−)-α-pinene were calculated using the GAUSSIAN 98 software, and force field parameters were derived using the antechamber module of AMBER 8. The heme group was modeled in its oxyferryl state (Fe^{3+}=O), the activated compound I.[49]

The enzyme-substrate complex was immersed in a truncated octahedral box of TIP3P water molecules[50] with a minimal distance of 12.0 Å between the box boundary and the protein, and a closeness factor of 0.42 Å was used. The formal net charge of the enzyme was neutralized by the addition of 14 Na^+ counter ions to the system. The protein-substrate-water-ion systems of mutants GVQ, GGQ, and GQ consist of 56,103, 56,100, and 56,110 atoms, respectively.

The 9 docking solutions of (−)-α-pinene for each triple mutant were energy minimized for 2000 steps (1000 using steepest descent and 1000 conjugate gradient algorithm). Subsequently, the system was heated to 300 K and equilibrated for 60 ps. During this phase the C^α atoms were restrained using a harmonic potential gradient with a decreasing force constant from 10 to 0.1 kcal/mol. After heating, the system was equilibrated during 2 ns until the root mean square deviation (RMSD) became constant. During the production phase, the conformational space was sampled at 0.5 ps intervals for further structural analysis. All MD simulations were carried out assuming different initial random velocity distribution by using different IG values. The structural analysis was done with the ptraj module of AMBER 8. The RMSD was calculated assuming the minimized structure as a reference.

Two additional MD simulations of 4 ns each were performed on the (−)-α-pinene-bound CYP102A1 A74G L188Q (F87) (GQ) double mutant complex, following the same procedure as described previously for the mutants GVQ and GGQ, keeping the bulky wild-type phenylalanine residue at the position 87.

Prediction of regioselectivity

The chemical reactivity of C—H and C=C bonds in the substrate were ranked using the PETRA program version 2.6[51] (Table III). In addition, a cutoff distance of 4 Å between a carbon of the substrate and the O^{Fe} of the heme was used as a pre-requisite for a centre being hydroxylated.[52,53] It has been shown that different cutoff criteria for the distance between the ferryl oxygen and the hydrogen atom used to predict regioselectivity of CYP101 with camphor analogs (2.7 and 3.4 Å) resulted in similar product profiles, even though the number of MD conformations is higher for longer cutoff distance.[54] After an equilibration phase of 2 ns, all snapshots of the trajectory where at least one of the substrate's carbon atom met the distance criterion were counted as a productive conformation in the product profile calculation. If this criterion is fulfilled by more than one atom, we assumed that oxidation occur at the most reactive centre, i.e. the center with the lowest C—H bond dissociation energy.

RESULTS

Molecular dynamics simulation analysis

For each of the two triple mutants GVQ (A74G F87V L188Q) and GGQ (A74G F87G L188Q), 9 MD simulations of 2.5 ns were performed starting with initial structures derived from docking solutions, and the trajectories were analyzed. After 2 ns all systems were in equilibrium. However, in four simulations of each mutant the substrate did not show any stable conformation, as defined previously. These simulations were discarded based on the random movement of the substrate in the active site characterized either by a translation or by a rotation around its centre of mass (see Supporting Figs. 2S and 3S). The five remaining simulations per mutant were continued up to 10 ns of simulation time. For each mutant only one simulation led to a stable conformation of the substrate close to the O^{Fe} of the heme. In these two cases, the trajectories were extended to 20 ns. The distance between substrate carbon atoms and the heme O^{Fe} were monitored along the entire trajectory and analyzed during the production phase, the last 18 ns of the simulation (Fig. 1).

For the mutant GVQ, (−)-α-pinene is predominantly found in one stable and productive conformation (during 14.6 ns, i.e. 81% of the production phase), henceforth called conformation I [Fig. 1(a)]. In this conformation, (−)-α-pinene atom C8 is the farthest, and C3, C4, C10 are the closest atoms to the O^{Fe}, with C3, C4, and C10 being within cutoff distance of 4 Å in 23, 42, and 28% of the time (Table I). The RMSD of the protein backbone (C^α, C, N) from the minimized structure during the 20 ns simulation of the GVQ mutant was 1.8 Å.

(−)-α-pinene in conformation I is located between atoms C1A and C4D of the heme A and D pyrrole rings, with atoms C8 and C3 pointing to the meso carbon CHA and to the O^{Fe} of the heme, respectively [see Supporting Fig. 1S(a)]. The side chain methyl group of

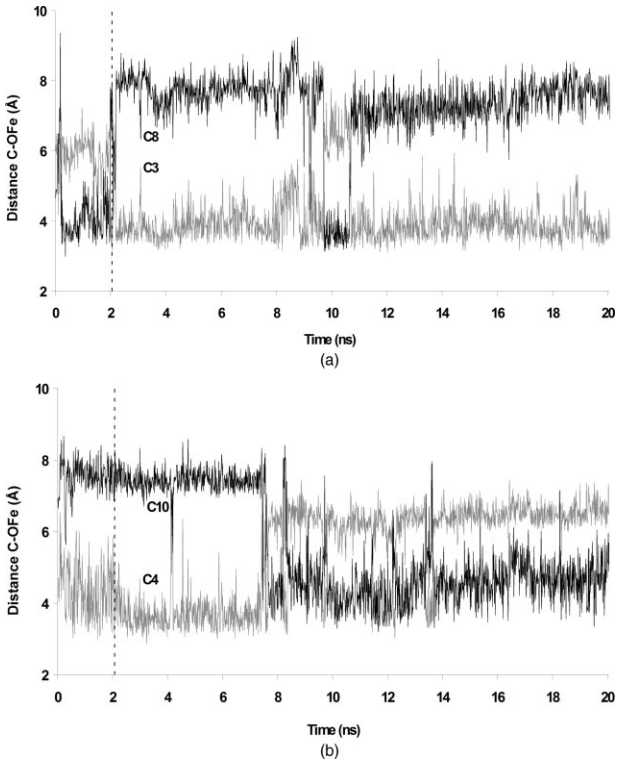

Figure 1

MD trajectories of the simulated comlexes: (**a**) Distance between atoms C3 (grey) and C8 (black) of (−)-α-pinene and ferryl oxygen (O^{Fe}) for the GVQ triple mutant; (**b**) Distance between atoms C4 (grey) and C10 (black) of (−)-α-pinene and heme oxygen for the GGQ triple mutant. The dashed line indicates the beginning of the production phase after an equilibration period of 2 ns.

V87 is in close interaction with atoms C1, C8, and C10 (3.4, 3.6, and 3.9 Å, respectively). O^{Fe} is the only heme atom in close interaction with atom C3 of (–)-α-pinene (3.6 Å) (Fig. 2). Atom CB of A264 and CD1 of L437 are close to C10 and C7 (3.6 and 3.7 Å, respectively). The side chains of T438, A328, and L75 interact weakly with C3, C4, and C8 atoms at a distance of 4 Å (see Supporting Table 1S).

The MD simulation of the GGQ mutant in complex with (–)-α-pinene reveals two distinct, stable, and productive conformations IIa and IIb (5.2 ns and 7.0 ns, respectively, which account for 68% of the production phase) [Fig. 1(b)]. In conformation IIa atom C10 is the farthest, while C4 and C5 are the closest atoms to O^{Fe} (81 and 13% of the time in which at least one atom of the substrate is close to O^{Fe}, 5.5 ns) [Fig. 3(b)].

The conformation IIa is related to the conformation I by a 180° rotation around the axis defined by atoms C1 and C4 of (–)-α-pinene [Fig. 2(b)]. The substrate, like in conformation I, is located between atoms C1A and C4D of the heme A and D pyrrole rings. Atoms C4, C7, and C8 are pointing to the meso carbon CHA, O^{Fe}, and C2D of the heme, respectively [see Supporting Fig. 1S(b)]. Atom CA of G87 is in close interaction only with atom C8 (3.8 Å). The heme has three additional interactions with (–)-α-pinene as compared to conformation I: O^{Fe} is within cutoff distance to atoms C4 and C5 (3.6 and 3.4 Å, respectively), atom CHA is in interaction with C7 (3.4 Å), and C2D with C8 (3.9 Å). The side chain atoms CD2 of L437, CG2 of T438, and CB of A328 interact with the atoms C9, C3, and C4 (3.7, 3.9, and 3.7 Å, respectively). The side chains of residues L75, K69, and A264 are not in contact with (–)-α-pinene (see Supporting Table 1S).

In conformation IIb atom C4 is the farthest, C1 and C10 are the closest atoms to O^{Fe} (40 and 16% of the time in which at least one atom of the substrate is within the cutoff distance of 4 Å from the O^{Fe}, 11.3 ns) [Fig. 3(c)]. Conformation IIb is related to conformation IIa by a 180° rotation around the axis perpendicular to the plane defined by (–)-α-pinene atoms C1, C3, and C5 [Fig. 2(c)]. The substrate is located essentially above the D pyrrole ring of the heme, with atoms C1 and C10 pointing directly to O^{Fe} and C2D, respectively [see Supporting Fig. 1S(c)]. Atom CA of G87 interacts with C3 (3.5 Å), the carbonyl oxygen of G87 with C3 (3.9 Å) and C4 (3.6 Å). The heme has three interactions with (–)-α-pinene: O^{Fe} and C4D interact with C1 (4.0 and 3.8 Å, respectively) and C2D with C10 (3.5 Å). CD1 of L75 and CD2 of L437 are in contact with atoms C4 and C9 (3.9 and 3.8 Å, respectively). There is an additional interaction between the oxygen of the backbone carbonyl group of T259 and atom C10 (3.8 Å). The side chains of residues K69, T438, and A264 are not in contact with (–)-α-pinene (see Supporting Table 1S). The RMSD of the protein backbone (C^{α},C,N) of the GGQ mutant was 1.9 Å (Fig. 2).

Biochemical characterization of the mutants

Investigation of the wild type CYP102A1 revealed that the enzyme has no oxidizing activity toward (–)-α- and (–)-β-pinene. The initial activity of the mutants GQ, GVQ, and GGQ toward (–)-α- and (–)-β-pinene was determined by measuring the NADPH oxidation rate in

Table I

Product Profile Predicted Based on the Population Distribution Analysis and Observed Experimental Results for GVQ and GGQ Mutants in %

Conformation	C_4-O^{Fe}	C_3-O^{Fe}	$C_{10}-O^{Fe}$	Nonreactive atoms
I	23	42	28	7
Experimental GVQ mutant	20	70	10	—
IIa	81	1	0	18
IIb	2	4	16	78
Experimental GGQ mutant	77	13	10	—

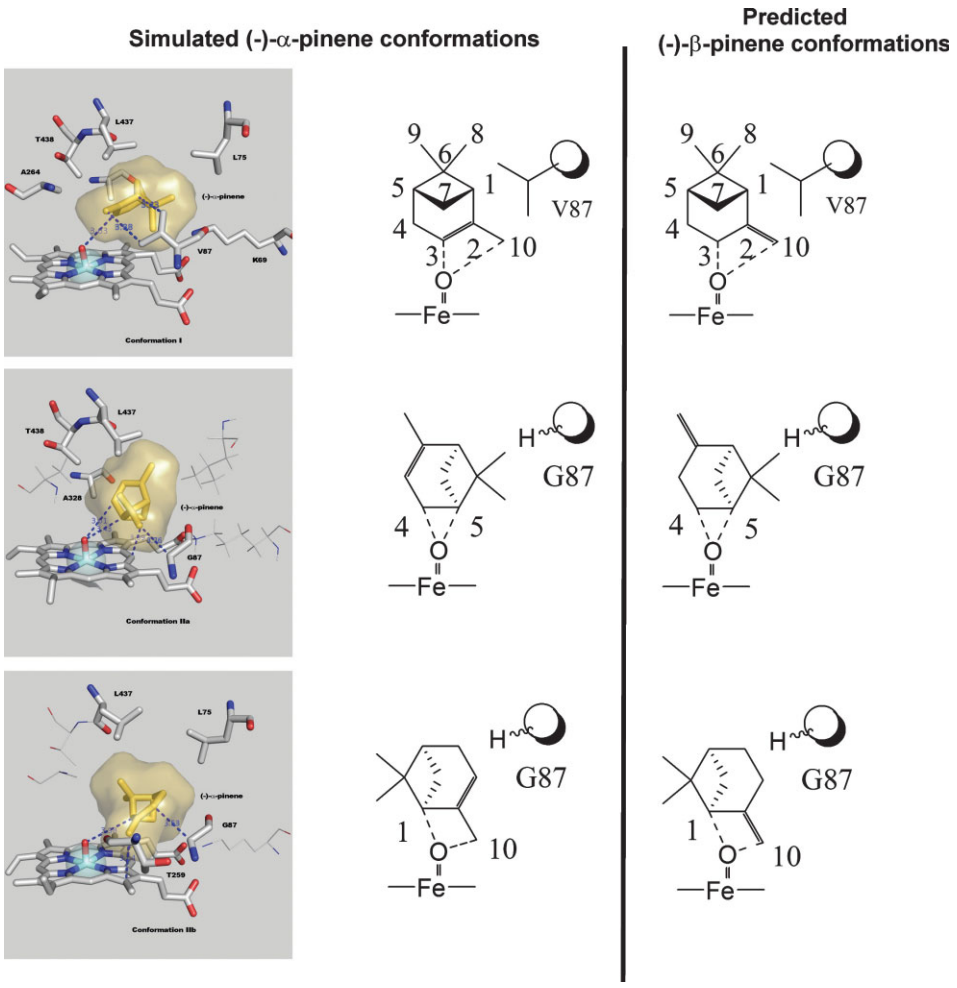

Figure 2

Stable productive conformations: Atomic contacts of the substrate with the closest residues side chains (sticks coulored by atom type), found during the simulation of (−)-α-pinene-bound (colored in gold) to the GVQ mutant (conformation I) and to the GGQ mutant (conformations IIa and IIb). Distances within 4 Å distance are depicted with dashed blue lines.

the first 30 s of the reaction and refers to the amount of enzyme used (Table II). Mutant GVQ has a slightly higher activity (113 min^{-1} towards (–)-α-pinene; 80 min^{-1} towards (–)-β-pinene) than the GGQ mutant (72 min^{-1} towards (–)-α-pinene; 79 min^{-1} towards (–)-β-pinene). Regioselectivity was determined from the gas chromatograms by integrating the corresponding peaks.

The main product of (–)-α-pinene oxidation catalyzed by mutant GVQ was pinene oxide (70%, epoxidation of the C2═C3 double bond); additionally verbenol (20%, hydroxylation at the secondary allylic position C4) and myrtenol (10%, hydroxylation at the primary allylic position C10) were found. The regioisomer (–)-β-pinene was converted to pinocarveol (68%, hydroxylation at the secondary allylic position C3) and myrtanal (32%, terminal epoxidation at C10) by this mutant. In contrast, the main product of (–)-α-pinene hydroxylation by GGQ mutant was verbenol (77%), and α-pinene oxide (13%), and myrtenol (10%) were minor products. Its regioisomer (–)-β-pinene was converted to myrtanal (60%) and pinocarveol (40%) (Fig. 4). Myrtanal is an aldehyde formed exclusively during hydroxylation of β-pinene, which may arise by acid-catalyzed rearrangement of the C1—C10 epoxide intermediate or by direct formation from a cationic intermediate in the epoxidation process.[55]

DISCUSSION

Activity of CYP102A1 mutants toward pinene conversion

Molecular dynamics simulations of human CYP2C9 indicated that the narrow access channel to the heme group mediates regioselectivity of the enzyme.[49] It has

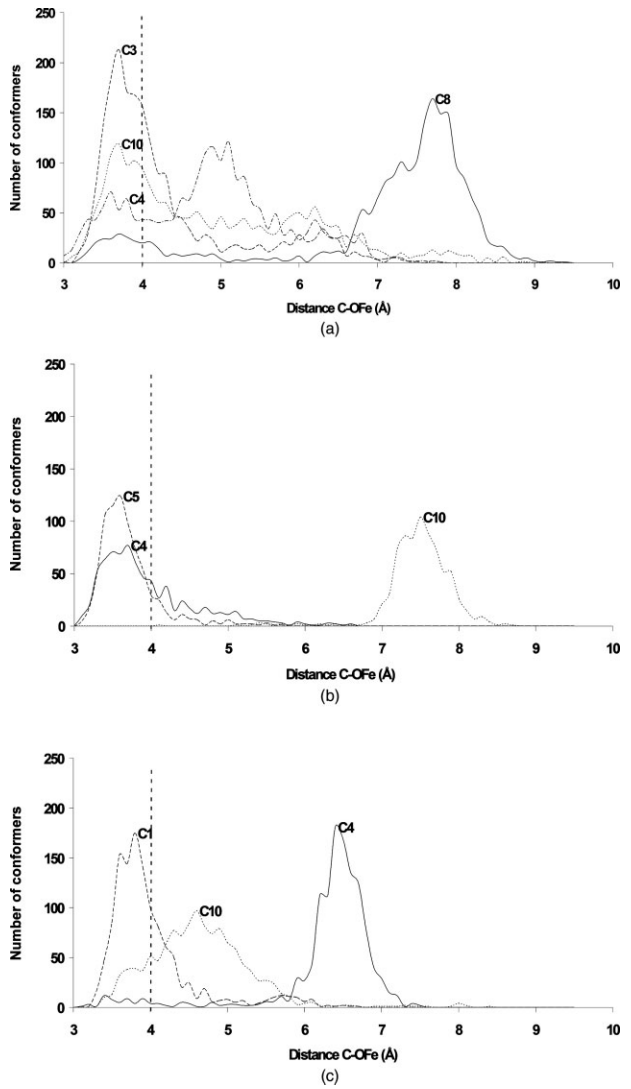

Figure 3

The number of conformations with a given distance between (−)-α-pinene atoms and O^{Fe}: (**a**) Conformation I; (**b**) Conformation IIa; (**c**) Conformation IIb. Distances were sampled at 0.1 Å intervals from the total of simulated snapshots after equilibration of 2 ns. The dashed line indicates the cutoff distance of 4 Å.

been suggested that a wide binding site of CYP102A1 is related to its low selectivity due to lack of geometric constraint.[18] Therefore, a successful strategy to increase enzyme selectivity by mutagenesis consists in narrowing down the access to the active site by increasing the bulkiness of binding site residues, or to stabilize the substrate over the heme by electrostatic interactions with polar active site residues, if it is functionalized with polar groups.[56] However, many CYPs lack a simple relation between selectivity and size of the binding site.

The CYP102A1 wild-type enzyme shows no oxidizing activity toward α-(–)- and β-(–)-pinene, which has been explained either by a limited substrate access channel (A74, L188), or by a blocked access to the heme (F87). As previously shown for other bulky substrates,[57] exchange of A74 and L188 by glycine and glutamine, respectively, is prerequisite to activity toward pinene. MD simulations of the double mutant A74G L188Q (GQ) have shown that the substrate is blocked from accessing the heme oxygen by the side chain of the F87, when it adopts the conformation found in the X-ray structure. However, during the simulations the F87 side chain swung open (see Supporting Fig. 5S), thus allowing (–)-α-pinene to approach the heme at a distance of less than 4 Å (see Supporting Fig. 6S) with either C3/C10 or C9 pointing toward the ferryl oxygen. Since the C9 is a nonreactive centre and the epoxidation at the C2═C3 double bond is more favorable than the hydroxylation at the primary C10 atom, pinene oxide is expected to be the major product, however with a low turnover rate which is in accordance with experimental results (Table II). The catalytic implications of the observed gating mechanism of F87 are still under investigation and will not be discussed in details here.

Regioselectivity control by position 87

The introduction of a smaller amino acid at position 87 results in a more active monooxygenase and a different product profile for the oxidation of substrate (–)-α-pinene. The product profile switches for a ratio of pinene oxide to verbenol from 70:21 to 13:77, and depending on valine (mutant GVQ) or glycine (mutant GGQ) being introduced at position 87. The same is true for (–)-β-pi-

Table II
Products Distribution for CYP102A1 Mutants

	Substrates	NADPH oxidation [1/min]	pinene oxidation [1/min]	coupling [%]	Products in %				
					pinene oxide	verbenol	myrtenol	pino-carveol	myrtanal
WT	(−)-α-pinene	10 ± 2	n.d.	—	—	—	—	—	—
GQ	(−)-α-pinene	38 ± 4	12 ± 2	32	85	15	—	—	—
GVQ	(−)-α-pinene	113 ± 13	32 ± 4	28	70	20	10	—	—
	(–)-β-pinene	80 ± 5	25 ± 2	31	—	—	—	68	32
GGQ	(–)-α-pinene	72 ± 2	36 ± 1	50	13	77	10	—	—
	(–)-β-pinene	79 ± 6	38 ± 2	48	—	—	—	40	60

n.d. stands for nondetected.

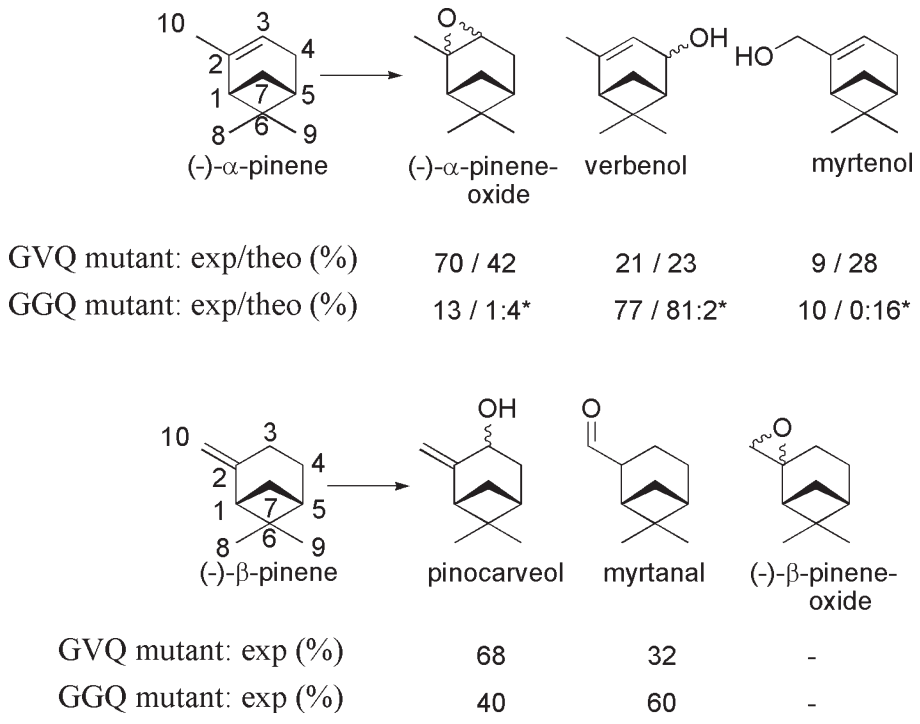

Figure 4

Structures of the products formed in the oxidation of (−)-α- and (−)-β-pinene by CYP102A1 mutants. (*) Prediction of the product formation for conformation IIa and IIb.

nene which was converted by both mutants to a mixture of pinocarveol and myrtanal: 68 and 32% for mutant GVQ and 40 and 60% for mutant GGQ.

The heme access channel is widely open in these two mutants, as concluded from the distance of 13.4 Å between the C^{α} atoms of residues T268 and S332, and the substrate cannot be fixed by polar interactions like hydrogen bonds, as observed for the complex of CYP101 (also known as P450cam) and camphor.

The active site residues have been previously shown to play a major role in mediating regioselectivity in two bacterial P450 monooxygenases, CYP102A1[30,58] and CYP101.[56] Upon hydroxylation of (+)-α-pinene catalyzed by CYP101, replacement of the wild-type residue at position 87 by tryptophan (F87W), in addition to mutations Y96F and L244A, led to an increase of the (+)-*cis*-verbenol ratio from 31% for the wild type to 86% for the triple mutant.[56] The superposition of the structure

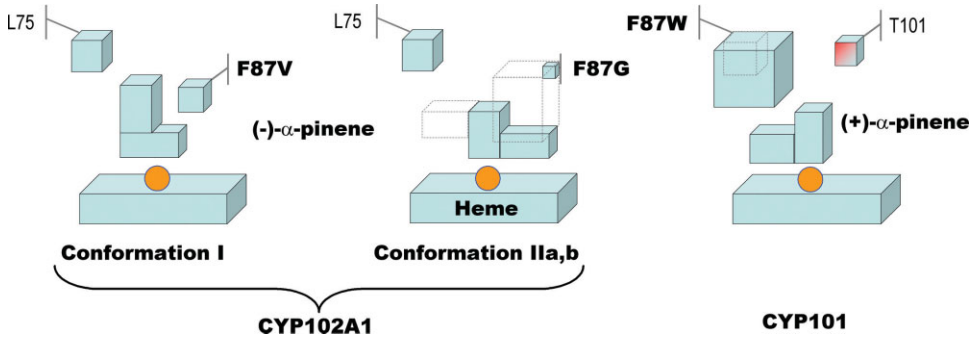

Figure 5

Simplified model to rationalize the regioselectivity observed for the CYP102A1 GVQ and GGQ mutants based on the anchoring effect of the residue at position 87 (comparison with the effect of the W87 in the CYP101).

of a CYP101 triple mutant in complex with (+)-α-pinene (PDB entry 1MPW) with the substrate complexes of CYP102A1 mutants GVQ and GGQ, allows us to compare the orientations of the substrates and shape of the binding sites in both enzymes (see Supporting Fig. 1S). It has been proposed that W87 in CYP101 fits better the substrate shape, interacting directly with pinene.[56] On the other hand, residues T101 in CYP101 and L75 in CYP102A1 which are structurally equivalent to positions F87 in CYP102A1 and W87 in CYP101, respectively, are too far from the substrate to control its orientation. Nevertheless, in mutant GVQ fixation of the substrate is done by valine at position 87, which is in contact with the substrate. Thus, in both enzymes there is at least one specific hydrophobic residue that mediates the substrate orientation by reducing its mobility near to the active site, when complementarity between this anchoring residue and the substrate in a stable, productive orientation is achieved. V87 thus serves as an anchor for pinene in conformation I, but not in conformation IIa and IIb. In contrast, glycine in position 87 lacks the capability of anchoring the pinene in conformation I, and the substrate interconverts between conformations IIa and IIb (Fig. 5). The impact of specific anchoring residues on the oxidation of α-pinene by CYP101 and CYP102A1 could not be explained only by structural constraints of the active site, since the substrate binding pocket in CYP102A1 is much wider than in CYP101.

Regioselectivity prediction for a structurally related substrate

A productive conformation is characterized by three properties: (1) closeness of the substrate to the heme; (2) suitable side chain size to induce a stable substrate conformation; (3) chemical reactivity of the nearest substrate position. MD simulations of mutant GVQ in complex with (−)-α-pinene led to a single productive conformation I, while mutant GGQ stabilizes the substrate in two distinct conformations IIa and IIb (Fig. 2). Conformation IIa has the major contribution to the product profile for the mutant GGQ. In conformation IIb, the atom C10 was rarely closer to the heme (16%) than C1 (40% of production phase). This explains why this conformation has such a small contribution to the product profile, even though it is stable for a longer simulation time of 7 ns.

The different reactivity of the carbon centers of the substrate were accessed from the C—H bond dissociation energy analysis (Table III) which showed that the homolytic breaking of C=C bonds and allylic C—H bonds are thermodynamically favored. Although the C1—H has the lowest bond dissociation energy (74.2 kcal/mol), the formation of a radical species on atom C1 is destabilized by the steric constraints imposed by the [3,1,1] bridge. There are no steric constraints in the case of C4 and C10 where the co-planarity between the single occupied orbital and the double bond in C2=C3 allows a radical delocalization and consequently promote the formation of the intermediate.[59] Therefore, C4 and C10 are expected to be the most reactive positions to be oxidized. The C3—H bond has the highest bond dissociation energy, 24.0 kcal/mol higher than the C1—H bond, and the formation of radical species on a sp^2 carbon is also energetically penalized. Therefore, the electrophilic attack of the heme oxygen to the double bond toward the formation of the epoxide product (the C2=C3 bond has a dissociation energy 7.3 kcal/mol higher than C1—H), is favorable and the hydroxylation of C3 is not observed.

Table III
C—H Bond Dissociation Energies

Substrate	Bond	*E* (kcal/mol)
(–)-α-pinene	**C10—H**	**80.0**
	C2=C3	**81.5**
	C3—H	98.2
	C4—H	**75.9**
	C1—H	74.2[a]
	C5—H	90.1
	C8,9—H	96.7
(–)-β-pinene	C10—H	98.2
	C2=C10	**81.5**
	C3—H	**78.8**
	C4—H	93.8
	C1—H	75.6[a]
	C5—H	89.4
	C8,9—H	96.7

Reactive bonds are indicated in bold.
[a]Bond is not reactive despite its low bond dissociation energy.

The regioisomer (−)-β-pinene has a double bond external to the six-membered ring, between atoms C2 and C10. Therefore the allylic positions C1 and C3, as well as the double bond C2=C10 have a high reactivity. The bond dissociation energies for C3—H and C2=C10 are similar to C4—H and C2=C3 in (−)-α-pinene (3.2 and 6.0 kcal/mol higher than C1—H bond in (−)-β-pinene). Thus, pinocarveol is expected to be the major product of hydroxylation, if the selectivity would be mediated by chemical reactivity only.

We compared the intrinsic hydrogen bond dissociation energy calculated by PETRA with more accurate quantum mechanical calculations for aliphatic carbon atoms in a general drug candidate.[60] The qualitative estimation of the reactivity among allylic/nonallylic and primary/secondary/tertiary C—H bonds is similar in both methods, with the exception of sterically constrained positions, like the C1—H bond which seems to be underestimated. However, one should be aware that C1—H and C5—H are structurally and chemically equivalent and that the virtual stabilization of the formed radical at C1 by the conjugated double bond, as it was assumed by PETRA, is unrealistic, since the antiparallel alignment of the corresponding orbitals is physically not possible, due to the

constraints imposed by the bridge. The PETRA estimate of the C5—H bond dissociation energy should be more realistic to describe C1—H reactivity (Table III). Thus, the unlikely formation of a tertiary bridged radical provides a reliable thermodynamic criterion for a preliminary discrimination filter of regioselectivity,[54] making the identification of putative hydroxylation targets in pinene much easier.

Combining the three major contributing factors on the selectivity of CYP102A (structural guidance of the substrate orientation by the anchor residue at position 87, chemical reactivity of the C—H bonds of the substrate, and active site shape and flexibility), it was possible to predict the product profiles of (–)-α-pinene oxidation which were validated by experimental results. Under the assumption that conformations I, IIa, and IIb identified for the (–)-α-pinene complex would also occur for (–)-β-pinene, we expect that pinocarveol should be the major product for mutant GVQ upon conversion of (–)-β-pinene. In contrast, for mutant GGQ oxidation products on C2=C10 will be preferentially formed, while hydroxylation at C4 and C5 should not occur, since these atoms have a low chemical reactivity. The experimental validation of the predicted product profile for the conversion of (–)-β-pinene (Table II) confirmed the formation of 68% and 32% of pinocarveol and myrtanal, respectively, for the GVQ mutant. This ratio was shifted to 40 and 60% for the GGQ mutant, which suggests that conformation IIb should be preferred.

Extensive sampling of substrate conformational space

It is widely assumed that docking solutions with the highest score lead to appropriate initial structures to further explore the potential energy surface of a flexible enzyme-substrate complex by molecular dynamics simulations.[61] For widely open binding sites, there are many local energy minima accessible to the substrate, most of them however being unproductive. To find a productive conformation of a substrate in a flexible enzyme, it is not sufficient to start from a single docking solution. We therefore performed multiple MD simulations, each starting with a different docking position of the substrate in the binding site and with different initial velocities.[49] These docking solutions differ essentially in the orientation of the substrate relative to the heme iron, while the distance between the ferryl oxygen and the closest atom of the substrate is about 4 Å in all initial structures. Interestingly, no interconversion between the initial conformations occurred during the MD simulation. Thus, extensive sampling by multiple simulations is necessary to explore the conformational space of small, hydrophobic substrates in large, flexible substrate binding sites, like in CYP102A1. As a result of this extensive sampling, one stable and productive substrate conformation was identified for each mutant, where the side chains, especially of the anchoring residue, were appropriately oriented to stabilize the substrate near the heme. Extending the simulation of the productive conformation to 20 ns demonstrated that the substrate is not fixed throughout the entire trajectory, but left this conformation for a short time interval and immediately returned to the stable and productive conformation (Fig. 1). In most of the MD simulations, however, the substrate moved away from the heme or showed a tumbling motion near the heme (see Supporting Fig. 3S), which is a consequence of the large conformational space available to the substrate in this class of CYPs. There is a second consequence of such a large conformational space: although one stable conformation was identified for each mutant, which is in agreement with experimental data, the existence of additional productive conformations cannot be excluded.

CONCLUSIONS

The modeling approach used here to rationalize and predict the regioselectivity of the CYP102A1 mutants toward (–)-α- and (–)-β-pinene, revealed three distinct substrate binding modes near the heme. Furthermore, modeling and experimental results demonstrated that a widely open binding site does not necessarily result in low regioselectivity. Here, we show that a single residue induces selectivity by stabilizing the unpolar substrate in defined binding conformations. The use of molecular docking and MD simulations to identify the binding modes in combination with a ranking by the chemical reactivity of the substrate proved to be an effective and valid approach to elucidate the structural basis of selectivity and to predict the preferred products.

ACKNOWLEDGMENTS

This work was performed within the Collaborative Research Centre SFB 706 (Selective Catalytic Oxidations Using Molecular Oxygen; Stuttgart). M.B. thanks the Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V. and the Bundesministerium für Wirtschaft und Technologie.

REFERENCES

1. Cirino PC, Arnold FH. Protein engineering of oxygenases for biocatalysis. Curr Opin Chem Biol 2002;6:130–135.
2. Miles CS, Ost TWB, Noble MA, Munro AW, Chapman SK. Protein engineering of cytochromes P-450. Biochim Biophys Acta-Protein Struct Mol Enzyme 2000;1543:383–407.
3. Tee KL, Schwaneberg U. Directed evolution of oxygenases: Screening systems, success stories and challenges. Comb Chem High Throughput Screen 2007;10:197–217.
4. Jiang YY, He X, de Montellano PRO. Radical intermediates in the catalytic oxidation of hydrocarbons by bacterial and human cytochrome P450 enzymes. Biochemistry 2006;45:533–542.

5. Li HY, Poulos TL. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat Struct Biol 1997;4:140–146.
6. Haines DC, Tomchick DR, Machius M, Peterson JA. Pivotal role of water in the mechanism of P450BM-3. Biochemistry 2001;40: 13456–13465.
7. Oliver CF, Modi S, Sutcliffe MJ, Primrose WU, Lian LY, Roberts GCK. A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation. Biochemistry 1997;36:1567–1572.
8. Kumar D, Hirao H, de Visser SP, Zheng JJ, Wang DQ, Thiel W, Shaik S. New features in the catalytic cycle of cytochrome P450 during the formation of compound I from compound 0. J Phys Chem B 2005;109:19946–19951.
9. Shaik S, Kumar D, de Visser SP, Altun A, Thiel W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. Chem Rev 2005;105:2279–2328.
10. Li HY, Darwish K, Poulos TL. Characterization of recombinant *Bacillus megaterium* cytochrome-P-450BM-3 and its 2 functional domains. J Biol Chem 1991;266:11909–11914.
11. Truan G, Komandla MR, Falck JR, Peterson JA. P450BM-3: absolute configuration of the primary metabolites of palmitic acid. Arch Biochem Biophys 1999;366:192–198.
12. Munro AW, Leys DG, McLean KJ, Marshall KR, Ost TWB, Daff S, Miles CS, Chapman SK, Lysek DA, Moser CC, Page CC, Dutton PL. P450BM3: the very model of a modern flavocytochrome. Trends Biochem Sci 2002;27:250–257.
13. GrahamLorence S, Truan G, Peterson JA, Falck JR, Wei SZ, Helvig C, Capdevila JH. An active site substitution. F87V, converts cytochrome p450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase. J Biol Chem 1997;272:1127–1135.
14. Glieder A, Farinas ET, Arnold FH. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat Biotechnol 2002;20:1135–1139.
15. Li QS, Ogawa J, Schmid RD, Shimizu S. Engineering cytochrome P450BM-3 for oxidation of polycyclic aromatic hydrocarbons. Appl Environ Microbiol 2001;67:5735–5739.
16. Watanabe Y, Laschat S, Budde M, Affolter O, Shimada Y, Urlacher VB. Tetrahedron 2007;63:9413–9422.
17. Urlacher VB, Makhsumkhanov A, Schmid RD. Biotransformation of β-ionone by engineered cytochrome P450BM-3. Appl Microbiol Biotechnol 2006;70:53–59.
18. Sowden RJ, Yasmin S, Rees NH, Bell SG, Wong LL. Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450(cam) and P450(BM-3). Org Biomol Chem 2005;3:57–64.
19. Dordick JS, Khmelnitsky YL, Sergeeva MV. The evolution of biotransformation technologies. Curr Opin Microbiol 1998;1:311–318.
20. Yoo SK, Day DF, Cadwallader KR. Bioconversion of α- and β-pinene by *Pseudomonas sp* strain PIN. Process Biochem 2001;36:925–932.
21. Farooq A, Tahara S, Choudhary MI, Atta ur R, Ahmed Z, Baser KHC, Demirci F. Biotransformation of (−)-α-pinene by *Botrytis cinerea*. Z Naturforsch C 2002;57:303–306.
22. Poulos TL. Cytochrome P450. Curr Opin Struct Biol 1995;5:767–774.
23. Poulos TL. Ligands and electrons and haem proteins. Nat Struct Biol 1996;3:401–403.
24. Ravichandran KG, Boddupalli SS, Hasemann CA, Peterson JA, Deisenhofer J. Crystal-structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450s. Science 1993;261:731–736.
25. Noble MA, Miles CS, Chapman SK, Lysek DA, Mackay AC, Reid GA, Hanzlik RP, Munro AW. Roles of key active-site residues in flavocytochrome P450 BM3. Biochem J 1999;339:371–379.
26. Ost TWB, Miles CS, Munro AW, Murdoch J, Reid GA, Chapman SK. Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450BM3. Biochemistry 2001;40:13421–13429.
27. Ost TWB, Clark J, Mowat CG, Miles CS, Walkinshaw MD, Reid GA, Chapman SK, Daff S. Oxygen activation and electron transfer in flavocytochrome P450BM3. J Am Chem Soc 2003;125:15010–15020.
28. Clark JP, Miles CS, Mowat CG, Walkinshaw MD, Reid GA, Simon NDA, Chapman SK. The role of Thr268 and Phe393 in cytochrome P450BM3. J Inorg Biochem 2006;100:1075–1090.
29. Yeom H, Sligar SG, Li HY, Poulos TL, Fulco AJ. The Role of Thr268 in oxygen activation of cytochrome P450(BM-3). Biochemistry 1995;34:14733–14740.
30. Li QS, Ogawa J, Schmid RD, Shimizu S. Residue size at position 87 of cytochrome P450BM-3 determines its stereoselectivity in propylbenzene and 3-chlorostyrene oxidation. FEBS Lett 2001;508:249–252.
31. Ravindranathan KP, Gallicchio E, Friesner RA, McDermott AE, Levy RM. Conformational equilibrium of cytochrome P450BM-3 complexed with *N*-palmitoylglycine: A replica exchange molecular dynamics study. J Am Chem Soc 2006;128:5786–5791.
32. Jovanovic T, Farid R, Friesner RA, McDermott AE. Thermal equilibrium of high- and low-spin forms of cytochrome P450BM-3: repositioning of the substrate? J Am Chem Soc 2005;127:13548–13552.
33. Paulsen MD, Manchester JI, Ornstein RL. Using molecular modeling and molecular dynamics simulation to predict P450 oxidation products. Cytochrome P450, Part B, Vol. 272, Methods in enzymology. San Diego: Academic Press Inc; 1996. pp 347–357.
34. Chang YT, Loew GH. Molecular dynamics simulations of P450BM3-examination of substrate-induced conformational change. J Biomol Struct Dyn 1999;16:1189–1203.
35. Maurer SC, Schulze H, Schmid RD, Urlacher V. Immobilisation of P450BM-3 and an NADP(+) cofactor recycling system: towards a technical application of heme-containing monooxygenases in fine chemical synthesis. Adv Synth Catal 2003;345:802–810.
36. Omura T, Sato RJ. The Carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem 1964;239:2370–2378.
37. NIST 05 Mass Spectral Library. National Institute of Standards and Technology, Gaithersburg, Maryland, USA.
38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
39. Sevrioukova IF, Li HY, Zhang H, Peterson JA, Poulos TL. Structure of a cytochrome P450-redox partner electron-transfer complex. Proc Natl Acad Sci USA 1999;96:1863–1868.
40. Delano WL. The PyMOL molecular graphics system. San Carlos, CA, USA: DeLano Scientific LLC.
41. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski JW, Petersson GA, Ayala PY, Liu G, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Lui G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzales C, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98, Revision A. 6. Pittsburgh PA: Gaussian, Inc.; 1998.
42. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
43. Case DA, Darden TA, Cheatham TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA. AMBER 8. San Francisco, CA: University of California; 2004.
44. Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conforma-

tional energies of organic and biological molecules? J Comput Chem 2000;21:1049–1074.
45. van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 1977;34:1311–1327.
46. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
47. Martel P, Baptista A, Petersen SB. Protein electrostatics. In: El-Gewely MR, editor. Biotechnology Annual Review; Elsevier; 1996. pp 315–372.
48. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A Well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the Resp model. J Phys Chem 1993;97:10269–10280.
49. Seifert A, Tatzel S, Schmid RD, Pleiss J. Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin. Proteins 2006;64:147–155.
50. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
51. Gasteiger J, Marsili M, Hutchings MG, Saller H, Low P, Rose P, Rafeiner K. Models for the representation of knowledge about chemical-reactions. J Chem Inf Comput Sci 1990;30:467–476.
52. Altun A, Guallar V, Friesner RA, Shaik S, Thiel W. The effect of heme environment on the hydrogen abstraction reaction of camphor in P450(cam) catalysis: a QM/MM study. J Am Chem Soc 2006;128:3924–3925.
53. Altun A, Shaik S, Thiel W. Systematic QM/MM investigation of factors that affect the cytochrome P450-catalyzed hydrogen abstraction of camphor. J Comput Chem 2006;27:1324–1337.
54. Harris D, Loew G. Prediction of regiospecific hydroxylation of camphor analogs by cytochrome-P450(cam). J Am Chem Soc 1995; 117:2738–2746.
55. Fruetel JA, Collins JR, Camper DL, Loew GH, Demontellano PRO. Calculated and experimental absolute stereochemistry of the styrene and β-methylstyrene epoxides formed by cytochrome-P450(Cam). J Am Chem Soc 1992;114:6987–6993.
56. Bell SG, Chen XH, Sowden RJ, Xu F, Williams JN, Wong LL, Rao ZH. Molecular recognition in (+)-α-pinene oxidation by cytochrome P450(cam). J Am Chem Soc 2003;125:705–714.
57. van Vugt-Lussenburg BMA, Damsten MC, Maasdijk DM, Vermeulen NPE, Commandeur JNM. Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450BM3. Biochem Biophys Res Commun 2006;346:810–818.
58. Graham-Lorence S, Peterson JA. P450s: Structural similarities and functional differences. FASEB J 1996;10:206–214.
59. Ruchardt C. Relations between structure and reactivity in free-radical chemistry. Angew Chem Int Ed Engl 1970;9:830–843.
60. Olsen L, Rydberg P, Rod TH, Ryde U. Prediction of activation energies for hydrogen abstraction by cytochrome P450. J Med Chem 2006;49:6489–6499.
61. Keizers PHJ, de Graaf C, de Kanter FJJ, Oostenbrink C, Feenstra KA, Commandeur JNM, Vermeulen NPE. Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-*N*-alkylamphetamines: in silico predictions and experimental validation. J Med Chem 2005;48:6117–6127.