Exploration of swapping enzymatic function between two proteins: A simulation study of chorismate mutase and isochorismate pyruvate lyase

Alexandra Choutko, Andreas P. Eichenberger, Wilfred F. van Gunsteren, and Jožica Dolenc*

Physical Chemistry, Swiss Federal Institute of Technology, ETH, Zürich 8093, Switzerland

Received 6 March 2013; Accepted 31 March 2013

DOI: 10.1002/pro.2264

Published online 18 April 2013 proteinscience.org

Abstract: The enzyme chorismate mutase EcCM from *Escherichia coli* catalyzes one of the few pericyclic reactions in biology, the transformation of chorismate to prephenate. The isochorismate pyruvate lyase PchB from *Pseudomonas aeruginosa* catalyzes another pericyclic reaction, the isochorismate to salicylate transformation. Interestingly, PchB possesses weak chorismate mutase activity as well thus being able to catalyze two distinct pericyclic reactions in a single active site. EcCM and PchB possess very similar folds, despite their low sequence identity. Using molecular dynamics simulations of four combinations of the two enzymes (EcCM and PchB) with the two substrates (chorismate and isochorismate) we show that the electrostatic field due to EcCM at atoms of chorismate favors the chorismate to prephenate transition and that, analogously, the electrostatic field due to PchB at atoms of isochorismate favors the isochorismate to salicylate transition. The largest differences between EcCM and PchB in electrostatic field strengths at atoms of the substrates are found to be due to residue side chains at distances between 0.6 and 0.8 nm from particular substrate atoms. Both enzymes tend to bring their non-native substrate in the same conformation as their native substrate. EcCM and to a lower extent PchB fail in influencing the forces on and conformations of the substrate such as to favor the other chemical reaction (isochorismate pyruvate lyase activity for EcCM and chorismate mutase activity for PchB). These observations might explain the difficulty of engineering isochorismate pyruvate lyase activity in EcCM by solely mutating active site residues.

Keywords: enzyme; engineering; molecular dynamics simulation; catalytic activity

Introduction

Elucidation of the driving forces and mechanisms of catalysis by enzymes is one of the intriguing

challenges in biochemistry with practical consequences.^{1,2} Obviously, the observed speed-ups of the catalyzed chemical reactions compared to the same reaction in the gas phase or solution phase are due to the environment of the reactant or substrate bound in the active site of the enzyme. In the catalytic process, four phases may be distinguished: (i) Binding of the substrate in the active site of the enzyme, thereby possibly changing its own or the enzyme's conformation as to favor the reaction. (ii) The reactant mounting the activation barrier of the transition state (TS). The activation barrier may be decreased compared to the reaction in vacuo or in aqueous solution by the electrostatic field due to the

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Center of Competence in Research (NCCR) in Structural Biology; Grant sponsor: Swiss National Science Foundation; Grant number: 200020-137827. Grant sponsor: National Competence Center for Research (NCCR) in Structural Biology; Grant sponsor: European Research Council; Grant number: 228076.

*Correspondence to: Jožica Dolenc; Physical Chemistry, Swiss Federal Institute of Technology, ETH, Zürich 8093, Switzerland. E-mail: jozi@igc.phys.chem.ethz.ch

various residues of the enzyme and by structural changes in the substrate and enzyme. The probability of reaching TS conformations may be enhanced by positional fluctuations of atoms of the enzyme or solvent that induce fluctuations in the forces on the atoms of the substrate, for example, by a fluctuating electrostatic field or through short-range covalent or van der Waals interactions, or originating from motional correlations in the enzymes. (iii) Formation of the product. (iv) Release of the product from its binding pocket in the enzyme.

Experimentally, only a few observable quantities characterizing the catalytic process can be measured: (i) Overall reaction rates. (ii) The free energy or enthalpy of binding of ligand molecules, in particular transition state analogs (TSA), to the enzymes. The binding free energy of a particular substrate A to an enzyme may in some cases be obtained indirectly by measuring the product formation of a substrate B and the inhibition of this process by the presence of substrate A or by inhibiting the conversion of substrate A in one way or the other. (iii) Atomic structure of the enzyme in a particular crystalline form as derived from X-ray or neutron diffraction patterns, or to much less precision in aqueous solution as derived from NMR spectroscopic measurements, but again in general only for the apoenzyme or the enzyme with a ligand bound that is not the substrate, unless the conversion of the substrate can be prevented by inducing particular conditions. These three types of quantities can be measured or derived as a function of (i) the thermodynamic state point, that is, temperature, pressure, pH, ionic strength, or (ii) the composition of the enzyme which is varied in mutant enzymes, enzymes from different organisms, minimum amino acid alphabet enzymes, deuterated enzymes, etc., or (iii) the substrate or ligand composition. As these data are much lower in number than the number of degrees of freedom of an enzyme system, the detailed driving forces and mechanism of the catalytic process can not be unambiguously determined by experimental means. This has led to a variety of descriptions, models, and theories of enzymatic reactions that cannot unambiguously be falsified or repudiated based on experimental data.

From a theoretical and computational point of view, statistical quantum mechanics offers the basic model by which catalytic processes should be explained. In computational practice, though severe simplifications and approximations of the statistical quantum-mechanical (QM) equations have to be invoked in order to obtain a result, even using the fastest computers available. For reviews of the level, we refer to Refs. 3–5. Time-independent QM models for the electronic degrees of freedom of the substrate and part of the enzyme are used to obtain the energy of the system along a chosen pathway or reaction coordinate connecting the reactant state (RS)

configuration with the postulated TS configurations of the enzyme and the ligand. If entropic contributions are estimated in one way or the other, the energy profile becomes a free energy profile. Because of the statistical-mechanical nature of an enzyme in solution, a proper sampling of the configurational Boltzmann ensembles is required to include entropic contributions to the free energy differences that drive the reaction. Due to the severity of the required assumptions and approximations both at the quantum electronic and the classical atomic levels of modelling, the current computational models of enzymatic reactions are rather of qualitative than of quantitative accuracy. On the quantum-chemical side, empirical valence bond, semiempirical, and density-functional methods are rather approximate, while higher level *ab initio* methods are too expensive to allow for a proper sampling of the many relevant degrees of freedom. On the classical statistical molecular-mechanical (MM) side, biomolecular force fields have a limited accuracy and although the sampling can reach time scales of nanoseconds to microseconds, this may not be sufficient to cover the slow conformational relaxation in enzymes. In addition, in view of the limited set of experimentally observable quantities a proper validation of theoretical-computational models of enzyme catalysis is problematic. The reproduction of a few experimental data for a high-dimensional model of an enzyme that contains many parameters to be chosen does not necessarily imply the correctness of the model or observed mechanism, unfortunately.

The variety of computational approaches to investigate enzyme catalysis can be illustrated by considering this type of work on the enzymes of the chorismate mutase (CM) family.^{6,7} Because hybrid quantum-mechanical molecular-mechanical (QM/MM) simulation of this enzyme in aqueous solution currently only covers tens of picoseconds, too short to surmount the activation barrier, in QM/MM studies of CM the enzyme plus substrate is energy minimized or simulated in conformations along a one⁸- or two⁹- dimensional reaction pathway connecting the RS with a hypothetical TS.^{10,11} From the generated configurations, an energy or free energy profile or potential of mean force (PMF) along the pathway can be calculated. This can be done for the reaction inside the enzyme, in solution, and in the gas phase, which allows the analysis of the energetic and geometric differences in these different environments. Such an analysis addresses the rate enhancement by the enzyme over a gas or solvent environment. In this type of approach, it is not well possible to determine the contribution of the different amino acid residues to the catalysis without making assumptions on the insensitivity of the pathway to residue variation, unless the PMF calculation is repeated for each mutant of the enzyme.^{12–15}

Figure 1. The simulated enzymes. Representative X-ray structures for the EcCM-chr (left) and PchB-ichr (middle) systems together with the substrates. On the right, the CM and IPL enzymes are superimposed. EcCM is shown in orange and PchB in yellow. For the two substrates, carbon atoms are depicted in cyan, oxygen atoms in red, and hydrogen atoms in white.

An alternative type of computational approach in which configurational sampling can be much enhanced is to perform classical molecular dynamics (MD) simulations of the enzyme with substrate, TSA, or product and then evaluate the energy, the structural ensembles, or the forces on atoms of interest, for example, to determine whether the conformation of the substrate is driven toward that of a TS.^{16,17} The averages of different quantities such as forces on atoms or distances between atoms over the generated Boltzmann ensemble are of lesser interest than the distributions of these quantities. The tails of such distributions may contain configurations or forces that are rarely present but may induce the reaction, that is, the electronic rearrangement. This type of computational approach has the advantage that contributions to the forces from the different amino acid residues can be readily determined from a single simulation.

All approaches used so far to elucidate enzyme reaction mechanisms suffer from more or less severe

assumptions or limitations. In the quantum-chemical model, these reside in the choice of Hamiltonian basis set and limited number of degrees of freedom. In the MM model, these reside in the choice of force field and degree of sampling of the degrees of freedom. In QM/MM model studies, approximations in the way the QM/MM boundary is treated are additional sources of inaccuracy. Yet, the intriguing catalytic properties of enzymes nevertheless beg for computational investigation, be it only on a qualitative level. In the present article, we address a question that is different from the ones addressed in most QM/MM or MD studies of enzyme reactions.

Two enzymes that adopt very similar folds consisting of α -helices, CM, and isochorismate pyruvate lyase (IPL; Fig. 1) catalyze the transformation of different substrates (Fig. 2), that is, chorismate (chr) and isochorismate (ichr), respectively. This difference must be due to the different forces exerted on the atoms of the bound substrates by the different protein atoms in the similar folds. This raises the

Figure 2. Schematic diagrams illustrating the transformations of chr to prephenate (CM-catalyzed reaction 1) and ichr to salicylate (IPL-catalyzed reaction 2).

Table I. Overview of the MD Simulations

Simulation Name	Enzyme	Number of residues	Substrate	Number of water molecules
EcCM-chr	<i>E. coli</i> CM	2 × 101	chr	42186
EcCM-ichr	<i>E. coli</i> CM	2 × 101	ichr	42179
PchB-chr	<i>P. aeruginosa</i> IPL	2 × 101	chr	46753
PchB-ichr	<i>P. aeruginosa</i> IPL	2 × 101	ichr	46743

question whether such a difference in forces can be detected when modeling the enzyme substrate complexes.

CM catalyzes one of the few pericyclic reactions in biology and is a key enzyme in the biosynthetic pathway leading to the aromatic amino acids phenylalanine and tyrosine in bacteria, fungi, and plants.⁶ The transformation of chr to prephenate (Fig. 2), formally a [3,3]-pericyclic rearrangement, also known as a Claisen rearrangement, proceeds in the absence of enzymatic catalysis as a concerted but asynchronous process via a chair-like TS.¹⁸ Natural CMs accelerate the rearrangement of chr more than a million-fold.⁷ The structure of the all-helical CM from *Escherichia coli* (EcCM) has been determined by X-ray crystallography at 2.2 Å resolution.¹⁹

The IPL from *Pseudomonas aeruginosa* (PchB) physiologically catalyzes the elimination of the enol-pyruvyl side chain from ichr to make salicylate for incorporation into the siderophore pyochelin²⁰ (Fig. 2). The IPL reaction is a concerted but asynchronous hydrogen transfer, by a proposed [1,5]-sigmatropic reaction mechanism with a pericyclic TS.²¹ The structure of PchB has been determined at 1.95 and 2.35 Å resolution,²² revealing that this enzyme is a structural homolog of EcCM, despite the low sequence identity of 21%.^{20,21} The similarity of their folds is clear from Figure 1. This structural similarity possibly accounts for residual CM activity of PchB, albeit with considerably lower efficiency than EcCM.

PchB has been engineered by mutagenesis to catalyze more efficiently the chr to prephenate rearrangement. CM catalytic activity was improved 43-fold, while IPL catalytic activity was lost.²³ The PchB enzyme's promiscuity in catalyzing the chr to prephenate rearrangement has been the source of inspiration to engineer by mutagenesis a CM into an enzyme catalyzing the transformation of ichr into salicylate. Nevertheless, not a trace of this activity has been detected in CM variants mutated to more closely resemble the IPL active site. It has been recently suggested²⁴ that unique dynamic properties of chr-utilizing enzymes might be partly responsible for their enzymatic activities. Investigation of the active site participants alone might, therefore, not

be enough to understand catalysis in those enzymes. Hence, regions further from the binding pocket might influence catalytic activity in EcCM as well.

Can a model calculation detect reasons for the differential catalytic activity of the enzymes EcCM and PchB with respect to the substrates chr and ichr despite the similar fold of these two enzymes? The key must lie in the different side chains surrounding the active sites. To delineate the differences in the contributions to the forces from these residue side chains, we simulated four enzyme-substrate systems, each of the two enzymes with each of the two substrates, in aqueous solution. As we are interested in the distribution of the forces on the atoms of the substrates originating from residues close to and far from the active site, we restricted the calculations to classical MD simulations of the four enzyme-substrate systems and analyzed the configurational Boltzmann ensembles in terms of particular features of substrate geometry and of electrostatic forces on particular atoms of the substrate due to shells or layers of protein atoms around the substrates.

We do not intend to explain catalytic activity of near attack configurations^{17,25,26} or differences in (free) energy of the transition versus RS between the enzyme and the solution or gas phase.^{27,28} To that end, we would have to involve QM modeling and to apply umbrella sampling along a hypothetical reaction coordinate leading to a hypothetical TS. We only investigate the four enzyme-substrate complexes to obtain an indication of the effect of the different environments on the substrates and whether such effects can be traced to particular amino acid residues. Our work differs in a number of aspects from other computational studies of CM catalytic activity: (i) we consider IPL from PchB and CM from EcCM, that is, CM from *E. coli* as in Ref. 17,27,28, while most computational studies^{5,8,10–13,16,25,26,29–31} consider CM from *Bacillus subtilis*; (ii) we investigate the difference in substrate binding to IPL and CM, while most other studies aim at explaining CM catalysis, for example, Ref. 9; (iii) we do not calculate a PMF along a reaction pathway using QM/MM methodology,^{27,28} but perform long MD simulations of enzymes-substrate complexes and analyze not only the configurations¹⁷ of the substrates but also the electrostatic forces on particular atoms of the substrates.

Results

Table I gives an overview over the four 10-ns simulations performed. The *E. coli* CM (left side of Fig. 1) was simulated with its own substrate chr and with the substrate of the IPL catalyzed reaction, ichr. The *P. aeruginosa* IPL (middle panel of Fig. 1) was simulated with its own substrate ichr and with the substrate of the CM catalyzed reaction, chr.

PchB can catalyze both reactions depicted in Figure 2: ichr to salicylate (reaction 2) and much less chr to prephenate (reaction 1). Although both enzymes have an almost identical fold, EcCM, unlike PchB, can only catalyze reaction 1.

Chorismate and ichr are similar in their configurations, they have the same number and types of atoms, the enolpyruvyl side chain at carbon C₃ is the same in both the molecules; the main difference resides in the hydroxyl substitution of the carbon C₂ in ichr and of the carbon C₄ in chr, see Figure 2.

Both enzymes are simulated as dimers (Fig. 1). The two ligand-binding pockets are surrounded by residues of both monomers, which are different in the two binding sites.^{19,22} This implies that the forces on the substrates in the two binding sites may be different too and induce differences in substrate geometry.

Structural analysis of the enzymes

The atom-positional root-mean-square deviations (RMSD) of the protein backbone atoms from their respective energy-minimized X-ray structures are shown as a function of simulation time in Supporting Information Figure 1 for the four MD simulations. All simulations display an RMSD below 0.3 nm for the C_α backbone atoms that are part of an α-helix except for simulation PchB-ichr, where the RMSD lies slightly above 0.5 nm. The simulations involving EcCM show an all C_α backbone RMSD value below 0.3 nm as well. The PchB-chr structural deviations from the PchB X-ray structure are mainly due to structural changes in residues of the loop secondary structure (green curve). The largest RMSDs are found for the protein PchB and the substrate ichr.

The occurrence of regular secondary structure as a function of time is presented in Supporting Information Figure 2. Each protein monomer consists of three α-helices (see Fig. 1), resulting in six helices per dimer simulation. Those six helices are mostly conserved during the course of the MD simulations. Sometimes parts of the α-helices are identified as ₃₁₀- and/or π-helices (in terms of the Kabsch and Sander definitions of secondary structure) in the PchB and to a lower extent in the EcCM simulations.

Conformational analysis of the substrates using atom-atom distances

For the IPL catalyzed reaction 2 to occur, the C₃–O₇ bond of ichr has to be broken and a hydrogen at position C₂ has to be abstracted by carbon C₉ (see Fig. 2). For the CM catalyzed reaction 1 to occur, the C₃–O₇ bond of chr has to be broken as well and a bond has to be formed between carbon atoms C₁ and C₉. The C₃–O₇ bond cleavage has to occur for both reactions to take place and is therefore not a determinant for the differential catalysis of the two enzymes. To gain insights into the difference

between the stabilization and transformation of the substrates by the two enzymes, we took a closer look at the distribution of distances C₁–C₉ and C₂–C₉ during the course of the MD simulations (Fig. 3). The C₁–C₉ distance in chr (first column), which should be smaller for reaction 1 to occur, is in the EcCM simulation with mean values of 0.46 and 0.43 nm indeed smaller than in the PchB simulation with mean values of 0.54 and 0.53 nm. Even a short time period during which the C₁ and C₉ atoms come close to each other might be enough for electron transfer to occur and for a new bond to be formed. In other words, the occurrence of a close contact is of greater importance than the duration of it. Consequently, it is also of interest to take note of the smallest distance between the atoms of interest in the simulation. EcCM brings the substrate chr in a conformation where its transformation to the product prephenate is more likely to occur, whereas PchB does much less so.

The C₂–C₉ distance in ichr (fourth column), which should be smaller for the hydrogen on carbon C₂ to be abstracted and for reaction 2 to occur is, however, not smaller in the PchB simulation with mean values of 0.42 and 0.42 nm than in the EcCM simulation with mean values of 0.38 and 0.37 nm. The conformation of the substrate ichr seems less sensitive to the enzyme surroundings than that of the substrate chr.

The second column shows that EcCM reduces the C₁–C₉ distance in ichr, as it does in chr, whereas PchB does not favor a smaller C₁–C₉ distance in ichr. However, PchB favors distances C₂–C₉ in chr (third column) comparable to those in ichr (fourth column), whereas EcCM slightly reduces the C₂–C₉ distance in chr compared to ichr. Both enzymes tend to bring the other substrate, ichr for EcCM and chr for PchB, in a similar conformation to that of their native substrate, chr for EcCM and ichr for PchB. When bound to EcCM, the mean distance C₁–C₉ in chr is with 0.46 and 0.43 nm larger than the mean distance C₂–C₉ in ichr with values of 0.38 and 0.37 nm. When bound to PchB, the mean distance C₂–C₉ in ichr is with 0.42 and 0.42 nm smaller than the mean distance C₁–C₉ in chr with values of 0.54 and 0.53 nm.

Conformational analysis of the substrates using torsional angles

Two dihedral angles C₂–C₃–O₇–C₈ and C₃–O₇–C₈–C₉ of the substrates were monitored (Fig. 4) during the course of the four MD simulations. The distributions of the dihedral angle C₂–C₃–O₇–C₈ in chr bound to EcCM are different from the distributions of that angle in chr bound to PchB (first column). When simulated with EcCM, chr displays a preferred conformation with a mean value of about 70° for C₂–C₃–O₇–C₈ and 270° for C₃–O₇–C₈–C₉. When

Figure 3. Distribution of the distances between atoms C_1 – C_9 (first two columns) and C_2 – C_9 (last two columns) for the substrates chr (first and third columns) and ichr (second and fourth columns) in the four simulations. For each enzyme, two sets of distances are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values (in nm) are given in the top right corners.

simulated with PchB, chr displays a different conformational preference (mean values about 260° for C_2 – C_3 – O_7 – C_8 and 275° for C_3 – O_7 – C_8 – C_9).

Similar observations can be made concerning the PchB-ichr and EcCM-ichr systems. In the PchB simulation, ichr displays values between 200° and 300° for the C_2 – C_3 – O_7 – C_8 dihedral-angle (second column), whereas in the EcCM simulation, it displays a narrower distribution around 70° . The ichr C_3 – O_7 – C_8 – C_9 dihedral angle distributions in the EcCM and PchB simulations (fourth column) are not too different except for a second peak around 75° for the first binding site in the PchB simulation.

The overall picture is, as did emerge from the atom–atom distance analysis, that each enzyme tends to put the non-native substrate in a conformation similar to that of its native substrate.

Analysis of the electrostatic field at sites in the substrate

To gain further insights into the difference between the stabilization and transformation of the substrates by the enzymes, we calculated according to Eq. (1) the projection of the electrostatic field due to

the protein environment at atoms C_1 and C_9 in chr and C_2 and C_9 in ichr, on the line connecting atoms C_1 or C_2 with atom C_9 . As the electrostatic force on an atom is equal to its partial charge times the electrostatic field, the direction of this force will depend on the sign of the partial charge of the atom. The average partial charges on the C_1 , C_2 , and C_9 atoms of the substrates chr and ichr were calculated from QM/MM simulations of the same systems using a supramolecular coarse-grained solvent.³² These are given in Table II. They differ between chr and ichr in particular for atom C_2 , which has a negative partial charge in chr and a positive one in ichr. The distributions of the projections of the electrostatic fields at atoms C_1 , C_2 , and C_9 according to the Eq. (1) for the four enzyme-substrate complexes are shown in Figures 5–11 for different sets of protein atoms. As atoms C_1 , C_2 , and C_9 can be assumed to possess a negative partial charge (Table II) except for atom C_2 in ichr, a positive electrostatic field strength will result in a force in the direction of the other atom involved in the reaction, except for atom C_2 in ichr (third columns in Figs. 5–11) for which a negative electrostatic field strength yields a force in the

Figure 4. Distribution of the torsional angles $C_2-C_3-O_7-C_8$ (first two columns) and $C_3-O_7-C_8-C_9$ (last two columns) for the substrates chr (first and third columns) and ichr (second and fourth columns) in the four simulations. For each enzyme, two sets of angles are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows).

direction of the other atom (atom C_9 in this case) involved in the reaction.

Considering field contributions from all protein atoms within 1.4 nm from the site at which the field is to be determined (Fig. 5), one observes that the C_1 atom in chr (first column) is pushed more toward the C_9 atom in EcCM than in PchB. However, the C_9 atom in chr (second column) is in the first binding site pushed more toward the C_1 atom in PchB than in EcCM. The overall effect of the electrostatic forces on the distance between atoms C_1 and C_9 in chr can be obtained by adding the forces on both atoms. This yields for chr in EcCM mean values of

Table II. Partial Charges (in e) on the Atoms C_1 , C_2 , and C_9 of the Substrates chr and ichr in EcCM and PchB as Obtained from QM/MM Simulations³² of the Four Enzyme-Substrate Complexes

Atom name	chr		ichr	
	In EcCM	In PchB	In EcCM	In PchB
C_1	-0.066	-0.076	-0.146	-0.153
C_2	-0.289	-0.305	0.188	0.173
C_9	-0.621	-0.569	-0.630	-0.597

+187 and +323 $\text{kJ mol}^{-1} \text{nm}^{-1} e^{-1}$ for the two binding sites and for PchB the mean values of +219 and -118 $\text{kJ mol}^{-1} \text{nm}^{-1} e^{-1}$. The larger values for EcCM compared to PchB indicate that the electrostatic field in EcCM is more favorable than in PchB for the chr reaction. Because of the assumed positive charge of the C_2 atom in ichr (Table II), the overall effect of the electrostatic forces on the distance between atoms C_2 and C_9 in ichr can be obtained by subtracting the force on atom C_2 from that on atom C_9 . This yields for ichr in EcCM mean values of -262 and -115 $\text{kJ mol}^{-1} \text{nm}^{-1} e^{-1}$ for the two binding sites and for PchB the mean values of +105 and -38 $\text{kJ mol}^{-1} \text{nm}^{-1} e^{-1}$. The larger values for PchB compared to EcCM indicate that the electrostatic field in PchB is more favorable than in EcCM for the ichr reaction to occur.

In Figure 6, the corresponding electrostatic field distributions due to the protein side chains are displayed. The distributions are rather similar to those in Figure 5. This shows that the electrostatic field at the substrate is dominated by contributions from the side chains of the enzymes. The contributions from the backbone atoms (Supporting Information Fig. 3)

Figure 5. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms closer than 1.4 nm to the substrates for the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{nm}^{-1} \text{e}^{-1}$) in the top right corners.

are much smaller and more similar between the enzymes. The latter is not surprising considering the structural similarity of the backbone folds of both enzymes (Fig. 1).

The next question is which regions in the enzyme do contribute most to the differences in electrostatic field strengths at the substrates. In Figures 7–11, the contributions of atoms (i.e., charge groups) that lie in shells of different radii around the substrate atoms C_1 , C_2 , and C_9 are displayed in the same way as for the sum of these six shells in Figure 5. Atoms very close to the selected substrate atoms contribute little to the fields (Fig. 7). The shells 0.4–0.6 nm (Fig. 8) and 0.6–0.8 nm (Fig. 9) show the largest contributions and the latter shell shows the largest differences between the enzymes. The shells beyond 0.8 nm show narrower distributions and smaller mean values, as is expected due to the longer distances from the atoms to the substrate. This analysis suggests that residue side chains at distances between 0.6 and 0.8 nm are primarily responsible for the difference in the catalytic activity between EcCM and PchB.

Discussion

We have performed MD simulations of four enzyme-substrate complexes. CM from *E. coli* was simulated with its substrate chr (chr) and with the substrate ichr (ichr) of the IPL catalyzed reaction. Analogously, the IPL from *P. aeruginosa* was simulated with its substrate ichr and with the substrate chr of the CM catalyzed reaction. The structures determined by X-ray crystallography were not destabilized in the simulations which maintained the protein folds and secondary structure. In an effort to learn more about the differential stabilization of the two substrates chr and ichr by the two enzymes, we performed an analysis of the substrate conformations, in terms of interatomic distances between, torsional angles involving and electrostatic fields acting on particular atoms of interest in the substrates. The geometric analysis shows that the conformations adopted by the substrates bound to the different enzymes are different. The analysis of the electrostatic field strengths shows that EcCM influences chr in a way which renders the chr to prephenate reaction more likely, whereas PchB does that to a lesser extent. Conversely, PchB

Figure 6. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's side-chain atoms closer than 1.4 nm to the substrates for the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of electrostatic field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{nm}^{-1} \text{e}^{-1}$) in the top right corners.

influences ichr in a way which renders the ichr to salicylate reaction more likely, whereas EcCM does not. Despite their very similar secondary structure, folds (see Fig. 1), and active sites, the enzymes influence the substrates in different ways. This is reflected in the experimental finding that the conversion by mutagenesis of an IPL into a functional CM was possible while all trials of converting a CM into an IPL have failed. The electrostatic field analysis shows that the backbone atoms do not significantly contribute to the differences between the enzymes, and that in particular side-chain atoms of residues located at distances between about 0.6 and 0.8 nm from the substrates display the largest differences in electrostatic field strengths at atoms of the substrates between EcCM and PchB. A next step towards elucidating more accurately the forces driving the reaction could be a detailed investigation of reaction pathways using QM/MM^{5,32–34} simulation methodology.

Methods

Table I gives an overview over the simulations performed. All simulations were carried out using the

GROMOS05^{35,36} simulation software and the GROMOS 45A4 force field.³⁷

Molecular model

Four MD simulations of two enzymes, EcCM (101 residues) and PchB (101 residues) were performed. Each enzyme was simulated in its naturally occurring dimeric form with one pair of the two substrates chr or ichr. Each enzyme subunit was simulated with a substrate in its binding pocket, resulting in dimeric enzyme simulations with two (identical) substrates per simulation. The two enzymes are graphically depicted in Figure 1, the two substrates and the reactions catalyzed by the enzymes are shown in Figure 2.

For the EcCM simulations, the X-ray structure (Protein Data Bank entry 1ECM)¹⁹ was taken as starting structure. For the PchB simulations the X-ray structure (Protein Data Bank entry 3HGX)³⁸ was taken as starting structure. The ionizable groups were set to their protonated or deprotonated state according to the standard pK_a values of amino acids and a pH of 6.8. Thus, the histidine residues

Figure 7. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms closer than 0.4 nm to the substrates for the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of electrostatic field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{nm}^{-1} \text{e}^{-1}$) in the top right corners.

were protonated at N_δ for histidines 66(2), 67(1), 71(1,2), and 95(1) and at N_ϵ for histidines 43(1,2), 66(1), 67(2), and 95(2) of EcCM and at N_ϵ for histidines 84(1,2) of PchB, the lysine and arginine side chains and the N-terminus were protonated, whereas the aspartic and glutamic acid side chains and the C-terminus were deprotonated, resulting in a net charge of 0 e for the EcCM dimer and of $-4 e$ for the PchB dimer. The GROMOS 45A4 force field parameters for the substrates are specified in Supporting Information Tables I and II. The net charge of the substrates is $-2 e$. Water molecules were modeled as rigid three point molecules using the simple point charge (SPC) water model.³⁹

Simulation setup

The folded conformations of the dimeric proteins determined by X-ray spectroscopy were used as starting structures. The EcCM structure had been determined when crystallized together with a TSA. The chr and ichr substrate atom positions were mapped onto the positions of the TSA atoms and the substrate was then energy minimized while keeping

the protein atom positions restrained. All position restraints were performed with a force constant of $2.5 \times 10^4 \text{ kJ mol}^{-1} \text{nm}^{-1}$. The PchB structure had been crystallized together with pyruvic acid and 2-hydroxybenzoic acid, the products of the IPL reaction. The ichr and chr substrates atom positions were mapped onto the positions of the product atoms and the substrate was then energy minimized while keeping the protein atom positions restrained. For all four systems, in a first energy minimization, SHAKE⁴⁰ was turned off, the bond forces were calculated and the dihedral angle forces were set to zero. In a second energy minimization, SHAKE was turned on, the dihedral angle forces were accounted for while the protein atoms were positionally restrained. The resulting protein-substrate systems were energy minimized once more to improve substrate-protein contacts without any position restraints.

The systems were then solvated in a rectangular periodic box using a pre-equilibrated box of SPC water resulting in a system size of approximately 130,000 atoms for the simulations (Table I). In a

Figure 8. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column) and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms at a distance between 0.4 and 0.6 nm to the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of electrostatic field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{nm}^{-1} \text{e}^{-1}$) in the top right corners.

Figure 9. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms at a distance between 0.6 and 0.8 nm to the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{nm}^{-1} \text{e}^{-1}$) in the top right corners.

Figure 10. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms at a distance between 0.8 and 1.0 nm to the substrates for the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{ nm}^{-1} \text{ e}^{-1}$) in the top right corners.

Figure 11. Distribution of the projected electrostatic field strengths C_1-C_9 (first column), C_9-C_1 (second column), C_2-C_9 (third column), and C_9-C_2 (fourth column), see Eq. (1), exerted by the enzyme's atoms at a distance between 1.0 and 1.4 nm to the substrates for the substrates chr (first and second columns) and ichr (third and fourth columns) in the four simulations. For each enzyme, two sets of field strengths are given, as there are two substrates, one in each catalytic pocket of the dimeric unit, from top to bottom for the two binding sites of EcCM (first two rows) and of PchB (last two rows). The average values are given (in $\text{kJ mol}^{-1} \text{ nm}^{-1} \text{ e}^{-1}$) in the top right corners.

further energy minimization process, water molecules were relaxed with the atoms of the substrates and the proteins positionally restrained.

The equilibration process was started by taking initial velocities from a Maxwellian distribution at 60 K, the atoms of the enzymes and substrates were positionally restrained using a harmonic restraining force with a force constant of $2.5 \times 10^4 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. During 300 ps, the force constant of the restraining force was step-wise reduced to zero whereas the temperature was increased to 300 K.

All simulations were carried out for 10 ns at a constant temperature of 300 K and a constant pressure of 1 atm using the weak coupling algorithm.⁴¹ The temperature-coupling time was set to 0.1 ps and the pressure coupling time to 0.5 ps, an isothermal compressibility of $4.575 \times 10^{-4} (\text{kJ mol}^{-1} \text{ nm}^{-3})^{-1}$ was used.⁴²

All bond lengths were kept rigid at ideal bond lengths using the SHAKE algorithm,⁴⁰ allowing a time step of 2 fs in the leap-frog algorithm to integrate the equations of motion. Nonbonded interactions were calculated using a triple-range cutoff scheme with cutoff radii of 0.8/1.4 nm. Interactions within 0.8 nm were evaluated every time step. The intermediate range interactions were updated every fifth time step, and the long-range electrostatic interactions beyond 1.4 nm were approximated by a reaction field force⁴³ representing a dielectric continuum with a dielectric permittivity of 61 for the water model.⁴⁴

Analysis

The analyses were performed on the ensembles of system configurations extracted at 0.5 ps time intervals from the simulations. Atom-positional RMSDs were calculated after translational superposition of the solute centers of mass and least-squares rotational fitting of atomic positions, using all backbone C_α atoms, or backbone C_α atoms that are part of an α -helix, or backbone C_α atoms that are part of a loop, or all atoms.

The secondary structure assignment for the set structures extracted from the simulation trajectories was done according to the DSSP rules proposed by Kabsch and Sander.⁴⁵

The projection E_{i-j} of the electrostatic field $\vec{E}_i$ at an atom i along the direction to atom j , was calculated as

$$E_{i-j} = \frac{\vec{E}_i \cdot \vec{r}_{ij}}{|\vec{r}_{ij}|}, \quad (1)$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$.

The contributions of the protein atoms to the electrostatic field at the atoms C_1 , C_2 , and C_9 of the substrates were calculated by selecting all atom charge groups^{36,42} through their centers of geometry within

a given distance of the substrate atom of interest (spheres) or between two such distances (shells).

References

1. Warshel A (1991) Computer modeling of chemical reactions in enzymes and in solutions. New York: Wiley.
2. Aqvist J, Warshel A (1993) Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem Rev* 93:2523–2544.
3. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 48:1198–1229.
4. Kamerlin SCL, Warshel A (2010) At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? *Proteins: Struct Funct Bioinf* 78:1339–1375.
5. Ranaghan KE, Mulholland AJ (2010) Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *Int Rev Phys Chem* 29:65–133.
6. Haslam E (1993) Shikimic acid: metabolism and metabolites. New York: Wiley.
7. Andrews PR, Smith GD, Young IG (1973) Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. *Biochemistry* 12:3492–3498.
8. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2003) Preorganization and reorganization as related factors in enzyme catalysis: the chorismate mutase case. *Chem Eur J* 9:984–991.
9. Martí S, Andres J, Moliner V, Silla E, Tunon I, Bertran J (2009) Mechanism and plasticity of isochorismate pyruvate lyase: a computational study. *J Am Chem Soc* 131:16156–16161.
10. Claeysens F, Ranaghan KE, Manby FR, Harvey JN, Mulholland AJ (2005) Multiple high-level qm/mm reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chem Commun* 40: 5068–5070.
11. Crespo A, Marti MA, Estrin DA, Roitberg AE (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J Am Chem Soc* 127: 6940–6941.
12. Lee YS, Worthington SE, Krauss M, Brooks BR (2002) Reaction mechanism of chorismate mutase studied by the combined potentials of quantum mechanics and molecular mechanics. *J Phys Chem B* 106:12059–12065.
13. Szeftczyk B, Mulholland AJ, Ranaghan KE, Sokalski WA (2004) Differential transition-state stabilization in enzyme catalysis: quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J Am Chem Soc* 126: 16148–16159.
14. Ishida T (2008) Probing protein environment in an enzymatic process: all-electron quantum chemical analysis combined with ab initio quantum mechanical/molecular mechanical modeling of chorismate mutase. *J Chem Phys* 129:125105.
15. Ishida T (2010) Effects of point mutation on enzymatic activity: correlation between protein electronic structure and motion in chorismate mutase reaction. *J Am Chem Soc* 132:7104–7118.
16. Guo H, Cui Q, Lipscomb WN, Karplus M (2001) Substrate conformational transitions in the active site of chorismate mutase: their role in the catalytic mechanism. *Proc Natl Acad Sci USA* 98:9032–9037.
17. Bruice TC (2002) A view at the millennium: the efficiency of enzymatic catalysis. *Acc Chem Res* 35: 139–148.

18. Gorisch H (1978) Mechanism of chorismate mutase reaction. *Biochemistry* 17:3700–3705.
19. Lee AY, Karplus PA, Ganem B, Clardy J (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J Am Chem Soc* 117:3627–3628.
20. Gaille C, Kast P, Haas D (2002) Salicylate biosynthesis in *Pseudomonas Aeruginosa* - purification and characterization of PchB, a novel bifunctional enzyme displaying isochorismate pyruvate-lyase and chorismate mutase activities. *J Biol Chem* 277:21768–21775.
21. Kunzler D, Sasso S, Gamper M, Hilvert D, Kast P (2005) Mechanistic insights into the isochorismate pyruvate lyase activity of the catalytically promiscuous PchB from combinatorial mutagenesis and selection. *J Biol Chem* 280:32827–32834.
22. Zaitseva J, Lu J, Olechowski KL, Lamb AL (2006) Two crystal structures of the isochorismate pyruvate lyase from *Pseudomonas Aeruginosa*. *J Biol Chem* 281:33441–33449.
23. Künzler DE (2006) Taking advantage of the catalytic promiscuity of isochorismate pyruvate lyase from *Pseudomonas Aeruginosa*: mechanistic studies and evolution into an efficient chorismate mutase. PhD Thesis, ETH Zurich, Switzerland.
24. Lamb AL (2011) Pericyclic reactions catalyzed by chorismate-utilizing enzymes. *Biochemistry* 50:7476–7483.
25. Strajbl M, Shurki A, Kato M, Warshel A (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc* 125:10228–10237.
26. Ranaghan KE, Mulholland AJ (2004) Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chem Commun* 10:1238–1239.
27. Roca M, Messer B, Hilvert D, Warshel A (2008) On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc Natl Acad Sci USA* 105:13877–13882.
28. Roca M, Vardi-Kilshtain A, Warshel A (2009) Toward accurate screening in computer-aided enzyme design. *Biochemistry* 48:3046–3056.
29. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2001) Transition structure selectivity in enzyme catalysis: a QM/MM study of chorismate mutase. *Theor Chem Acc* 105:207–212.
30. Ishida T, Fedorov DG, Kitaura K (2006) All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction. *J Phys Chem B* 110:1457–1463.
31. Claeysens F, Ranaghan KE, Lawan N, Macrae SJ, Manby FR, Harvey JN, Mulholland AJ (2011) Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions. *Org Biomol Chem* 9:1578–1590.
32. Meier K, Choutko A, Dolenc J, Eichenberger AP, Riniker S, van Gunsteren WF (2013) Multi-resolution simulation of biomolecular systems: a review of methodological issues. *Angew Chem Int Ed* 52:2–17.
33. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103:227–249.
34. Lin H, Truhlar DG (2007) QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc* 117:185–199.
35. Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GROMOS software for biomolecular simulation: GROMOS 2005. *J Comput Chem* 26:1719–1751.
36. Available at <http://www.gromos.net>.
37. Schuler LD, Xaura D, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem* 22:1205–1218.
38. Luo Q, Olucha J, Lamb AL (2009) Structure function analyses of isochorismate pyruvate lyase from *Pseudomonas Aeruginosa* suggest differing catalytic mechanisms for the two pericyclic reactions of this bifunctional enzyme. *Biochemistry* 48:5239–5245.
39. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Interaction models for water in relation to protein hydration. In: Pullman B, Ed. (1981) *Intermolecular Forces*. Dordrecht: Reidel, pp. 331–342.
40. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341.
41. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
42. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tirion IG (1996) *Biomolecular simulation: the GROMOS96 manual and user guide*. Zürich, Switzerland: Vdf Hochschulverlag AG, ETH Zürich.
43. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102:5451–5459.
44. Heinz TN, van Gunsteren WF, Hünenberger PH (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J Chem Phys* 115:1125–1136.
45. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.