HAL
open science

Coupled Valence-Bond State Molecular Dynamics Description of an Enzyme-Catalyzed Reaction in a Non-Aqueous Organic Solvent

Elise Duboue-Dijon, Eva Pluhařová, Dominik Domin, Kakali Sen, Aoife C. Fogarty, Nicolas Chéron, Damien Laage

To cite this version:

Elise Duboue-Dijon, Eva Pluhařová, Dominik Domin, Kakali Sen, Aoife C. Fogarty, et al.. Coupled Valence-Bond State Molecular Dynamics Description of an Enzyme-Catalyzed Reaction in a Non-Aqueous Organic Solvent. Journal of Physical Chemistry B, 2017, 10.1021/acs.jpcb.7b03102 . hal-01558260

HAL Id: hal-01558260

https://hal.sorbonne-universite.fr/hal-01558260

Submitted on 10 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coupled Valence-Bond State Molecular Dynamics Description of an Enzyme-Catalyzed Reaction in a Non-Aqueous Organic Solvent

Elise Duboué-Dijon,[†,‡,‖] Eva Pluhařová,[†,‖] Dominik Domin,[†] Kakali Sen,[†,¶] Aoife C. Fogarty,[†,§] Nicolas Chéron,[†] and Damien Laage[*,†]

†*École Normale Supérieure - PSL Research University, Département de Chimie, Sorbonne Universités - UPMC Univ Paris 06, CNRS UMR 8640 PASTEUR, 24 rue Lhomond, 75005 Paris, France*

‡*Present address: Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nam. 2, Prague 16610, Czech Republic*

¶*Present address: School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK and Scientific Computing Department, STFC Daresbury Laboratory, Warrington, Cheshire, WA4 4AD, UK*

§*Present address: Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

‖*Contributed equally to this work*

E-mail: damien.laage@ens.fr

Phone: +33 (0)144322418

Abstract

Enzymes are widely used in non-aqueous solvents to catalyze non-natural reactions. While experimental measurements showed that the solvent nature has a strong effect on the reaction kinetics, the molecular details of the catalytic mechanism in non-aqueous solvents have remained largely elusive. Here we study the transesterification reaction catalyzed by the paradigm subtilisin Carlsberg serine protease in an organic apolar solvent. The rate-limiting acylation step involves a proton transfer between active-site residues and the nucleophilic attack of the substrate to form a tetrahedral intermediate. We design the first coupled valence-bond state model that simultaneously describes both reactions in the enzymatic active site. We develop a new systematic procedure to parameterize this model on high-level ab initio QM/MM free energy calculations that account for the molecular details of the active site and for both substrate and protein conformational fluctuations. Our calculations show that the reaction energy barrier changes dramatically with the solvent and protein conformational fluctuations. We find that the mechanism of the tetrahedral intermediate formation during the acylation step is similar to that determined in aqueous conditions, and that the proton transfer and nucleophilic attack reactions occur concertedly. We identify the reaction coordinate to be mostly due to the rearrangement of some residual water molecules close to the active site.

Introduction

Enzymes are remarkably efficient catalysts but were long considered to require mild conditions to function. It therefore came as a surprise when some enzymes were found to remain active in non-aqueous organic solvents, where they can catalyze new chemical reactions that are not possible in water.[1,2] This has opened a broad range of applications in synthetic chemistry, and for example enzymes in organic solvent are now widely used in the chemical and pharmaceutical industry to prepare molecules with a high enantioselectivity.[1]

One of the most widely used enzyme classes in organic solvents are serine proteases, which include e.g. trypsin, chymotrypsin and subtilisin. In aqueous conditions, these enzymes hydrolyze peptide bonds. Their molecular structure and their catalytic mechanism in water have been extensively studied, both experimentally[3–5] and computationally.[6–9] These studies revealed the key role played by a conserved catalytic triad, consisting of serine, histidine and aspartate residues. They further showed that the catalytic mechanism involves successive acylation and deacylation steps, which both proceed via the formation of a tetrahedral intermediate.[4] In contrast, in non-aqueous solvents, serine protease enzymes catalyze a transesterification reaction and while the mechanism still involves acylation and deacylation steps,[10] kinetic measurements have shown that the nature of the solvent has a dramatic effect on the rate constants.[11] For ester substrates, the acylation step was measured[12] to be rate-limiting in organic solvents with very few residual water molecules, while the deacylation step is rate-limiting in aqueous conditions.[4] In addition, the overall catalytic rate constant was found to drop by more than one order of magnitude when the amount of residual water molecules in the solvent decreases.[13] However, despite the great practical importance of non-aqueous solvents for these enzymes, the molecular details of serine protease catalysis are comparatively far less well understood in non-aqueous solvents than in water.

We therefore present here a theoretical and computational study of the paradigm transesterification reaction catalyzed by subtilisin Carlsberg in an apolar organic solvent, which has been extensively studied experimentally.[13–15] To model chemical reactions in enzymes, hybrid QM/MM methods combining a quantum description of the active-site region with a classical description of the rest of the enzyme and of the solvent are widely employed. However, a series of recent studies[16–19] have underlined the need for a proper sampling of the protein configurations, since different conformations may lead to different catalytic rates. We will show here that for subtilisin Carlsberg the reaction energy barrier exhibits dramatic changes of several tens of kcal/mol with the enzyme conformation. Since QM/MM approaches based on a DFT or even a semiempirical description of the QM region would

be too expensive to properly sample the protein conformations, we describe the electronic structure of the reactants with coupled valence bond (VB) states. This technique has already been successfully applied to describe a broad range of chemical reactions,[20–25] including many enzyme-catalyzed reactions.[26–30] However, the quality of this description largely depends on the proper parameterization of the VB states, which remains the main difficulty with this approach's implementation. Different strategies have been suggested (see, e.g., refs. 26,30–32), and here we present a new, simple and systematic procedure to parameterize the VB states on high level – here DFT QM/MM – free energy calculations in the presence of the enzyme that are shown to yield an excellent description of both the reaction free energy barrier and of the reaction equilibrium constant. We then show how this method can be applied to gain an improved understanding of the catalytic mechanism of serine proteases in organic solvents, and we contrast our results with prior studies in aqueous conditions.

The outline of the remainder of this paper is the following. We first describe the computational methodology, including the different types of quantum, classical and hybrid calculations that were performed. We then show that conformational fluctuations have a dramatic impact on the reaction energy barrier, and that the proton transfer and nucleophilic attack reactions involved in the rate-limiting step proceed in a concerted fashion. We describe a new procedure to parameterize the coupled valence-bond state model on high-level ab initio QM/MM free energy calculations, and we present the resulting free energy profiles, discuss the reaction transition state and give a molecular interpretation of the reaction coordinate. Finally, we offer some concluding remarks.

Methods

System

We study the transesterification reaction of the N-acetyl phenylalanine ethyl ester (APEE) substrate catalyzed by subtilisin Carlsberg in an organic solvent. Our simulated system

mimics the experimental conditions and is composed of the enzyme, one APEE molecule, a monolayer of water and some surfactant molecules used to solubilize the enzyme, immersed in hexane used as a model apolar organic solvent. The coordinates of the subtilisin Carlsberg protein and of crystallographic water molecules were taken from the PDB structure 1VSB.[33] The bound L-para-chlorophenyl-1-acetamido boronic acid inhibitor present in this structure was replaced with the L-phenylalanine-N-acetyl-ethyl-ester (APEE) substrate, using the inhibitor coordinates to place the APEE molecule in the active site. The titrable residues (Asp, Glu, Lys and Arg) were kept in their standard protonation state at pH=7 and the histidine residues were all taken to be neutral. This resulted in a total charge of -1 e. Experimental studies[13] used the AOT surfactant (dioctyl sodium sulfosuccinate) to extract the enzyme from the aqueous phase and avoid enzyme aggregation. Enzymes are then present as single molecules in organic solutions with the positively charged amino acid residues ion-paired with the negatively charged AOT sulfonate headgroups.[34] In our simulations, six AOT molecules were thus added next to the positively charged residues at the protein surface, avoiding the charged residues that were otherwise engaged in salt bridge interactions. Seven sodium ions were added next to negatively charged surface residues far from the active site to neutralize the system. A total of 856 water molecules, including the crystallographic water molecules, were required to hydrate the protein with a monolayer of water. The whole system was finally solvated in a cubic box with 1671 hexane molecules.

Calculations

The different aspects of this study require several levels of description and types of simulations, which are detailed now.

Classical MD

Classical molecular dynamics simulations are employed both for the initial preparation of the system and to obtain equilibrated protein conformations for different electronic structures

of the substrate and active site, corresponding to different steps of the reaction. These simulations are done with the NAMD[35] and Amber12[36] programs.

We used the CHARMM27 forcefield with CMAP corrections[37,38] to describe the enzyme, the SPC/E potential for water molecules,[39] parameters from the literature[40] for AOT, and a united atom force field[41] compatible with the CHARMM all atom protein force field for hexane.

During its stepwise construction, the system is minimized after each new component addition. The system is finally equilibrated, first during 1.5 ns in the NPT ensemble and then for 2 ns in the NVT ensemble, where the volume is fixed to its average value during the NPT run. This equilibration is performed with the substrate in its product state, which is very close to the crystal structure of the covalently bound inhibitor. All classical molecular dynamics simulations are performed at 300 K using a Langevin thermostat with a frequency of 0.1 ps^{-1} and a time step of 1 fs. Bonds between hydrogen and heavy atoms are constrained using the SHAKE and SETTLE algorithms.[42,43] Periodic boundary conditions are used together with a Particle Mesh Ewald treatment[44] of long-range electrostatic interactions, and a 16 Å cutoff for non-bonded interactions.

A classical force-field was designed for each of the limiting resonance forms of the substrate electronic structure which exhibits different bonding patterns (see Fig 2). In the A state, the bonding pattern corresponds to the reactant arrangement, i.e. before proton transfer and nucleophilic attack; it consists of the δ-singly protonated His64 and standard Ser221, both described by the standard CHARMM force field, and of the APEE substrate. In the product-like B state, i.e. after proton transfer and nucleophilic attack, the substrate is covalently bonded to the active-site serine, while His64 is in its standard doubly protonated form. Finally, the C state describes a putative proton transfer state (after proton transfer and before nucleophilic attack), where the substrate is the same as in state A, His64 is in its doubly protonated B-state form, and Ser221 is deprotonated. Parameters and charges for the APEE substrate and the residues in their different states were designed following the

standard CHARMM procedure, based on existing fragments and a previous parameterization of such a tetrahedral intermediate[45] (see SI).

Valence-bond state simulations

Two different types of VB state simulations were performed in this study, using the Amber12 software.[36]

First, in the mapping potential technique,[46] the electronic structure of the reactive system is fixed, i.e., the coefficients of the VB states are constant throughout the simulation and the potential is

$$V_\lambda = (1-\lambda)\, V_A + \lambda V_B \ , \tag{1}$$

with λ ranging from 0 (pure state A) to 1 (pure state B). The mapping potential is used in this work to generate environment configurations equilibrated to different electronic structures of the reactants, and to sample the configurations in order to construct the reaction free energy profile.

In contrast, in the second approach, the weights of the two VB states change at each time step and are determined by the instantaneous energies of the two VB states. The trajectory is propagated with the adiabatic ground state of the coupled VB Hamiltonian eq. 5. This is used e.g. to propagate trajectories from starting configurations in the vicinity of the transition state.

In all mapping potential simulations, the cartesian coordinates of the protein backbone atoms, AOT sulfur atoms and Na^+ ions were harmonically restrained around their initial positions with a soft 0.05 kcal mol^{-1} Å^{-2} force constant, in order to prevent aggregation of AOT and sodium ions. Experimentally, no aggregation occurs since the surfactant concentration is below the critical micelle concentration and surfactants remain ion-paired with the positively charged protein residues.[34] Experiments[47] further showed that no backbone conformational change occurs in non-aqueous solvents, and the soft restraints on protein backbone atoms used here thus do not hinder important conformational changes.

To calculate eight independent free energy profiles, eight distinct starting configurations are produced by a common 2 ns equilibration with $\lambda = 0.5$, followed by separate 1 ns runs restarted with different velocities. Each free energy profile is computed from 39 successive λ simulations, with a $\delta\lambda = 0.025$ increment. Each λ window is equilibrated during 10 ps before the 50 ps production run. The average free energy profile is constructed from the 312 distinct windows and aggregates more than 15 ns of sampling.

Gas phase *ab initio* and DFT

Ab initio calculations are used to determine the potential energy profiles for the elongation of the three covalent bonds being broken and formed during the reaction, *i.e.*, the Ser221 O-H bond, the His64 N-H bond, and the C-O bond between Ser221 and APEE. We successively consider the Ser221 residue, the protonated His64 residue and the covalently bonded Ser221-APEE product isolated in the gas phase. In each case, the two peptide bonds are cut by forming an unprotonated amine group on the N-terminus side and cutting the C-N amide bond and capping it with a hydrogen atom on the C-terminus side. The geometry is first optimized at the MP2/6-31+G** level of theory, consistently with the standard CHARMM procedure to parameterize a novel residue. Potential energy scans are then performed at the BMK/6-311+G** level. The Ser O-H and His N-H bond lengths are varied from 0.7 Å to 3.0 Å in 0.05 Å intervals by moving the hydrogen atom along the (heavy atom)-H vector. The C-O distance is varied from 1.21 Å to 2.96 Å in 0.05 Å intervals, while the internal geometries of the Ser and APEE fragments are kept frozen. These calculations were performed with the Gaussian09 program package.[48]

QM/MM

Mixed quantum/classical calculations (QM/MM) are used both to calculate two-dimensional reactive potential energy surfaces in selected configurations of the environment and to obtain reference energies from which the valence bond state parameters will be calibrated. The QM

region consists of the Asp32, His64, Ser125, Asn155, Ser221 active-site residues and the APEE substrate (see Figure S2), described at the BMK/6-311+G** level of theory. The location of the QM/MM boundary was verified to have very little effect on our results by repeating some calculations with a different QM region containing the same residues, but whose boundary goes through C-C bonds only (see Figure S3). For both definitions the bonds cut by the QM boundary are capped with hydrogen atoms.

The BMK[49] functional was selected because it yields a less than 2 kcal/mol mean absolute deviation for a selection of reaction barrier heights while retaining small mean absolute deviations for geometries and harmonic frequencies of ground state molecules.[49] The rest of the system is described with the CHARMM27[37,38] molecular mechanics force field. Calculations are performed with the QM/MM AMBER[36]-Gaussian[48] interface, using the default electrostatic embedding scheme. Since periodic boundary conditions cannot be described by this QM/MM implementation, each configuration is first shifted to center the enzyme–substrate complex in the simulation box.

Two-dimensional QM/MM potential energy surface (2D PES) on a 2D grid of discrete values of the proton transfer and nucleophilic attack coordinates (resp. the N-H and C-O distances) were obtained for four typical configurations of the environment. Each of the four configurations results from a classical MD equilibration of at least 250 ps where the active-site and the substrate are respectively described with the A-, B- and C-state force fields, and with an equal mixture of A- and B-state force fields (using the mapping potential approach) as an approximation of the transition state.

Each 2D PES scan is calculated in a frozen environment, i.e., all atoms are fixed except for the proton to be transferred and the APEE substrate. The ranges of values for each distance are chosen so that the minima of the PES can be determined and thus vary slightly from one environment configuration to another. The N-H distance range is approximately 0.9 to 1.8 Å and that of the C-O distance 1.4 to 3.2 Å. The N-H and C-O distances are scanned with a 0.1 Å and a 0.2 Å step respectively. To avoid steric clashes, for a given set of

fixed C-O and N-H distances, a geometry optimization of the remaining degrees of freedom of APEE and of the proton orientation is performed, using the ground state of the coupled valence bond state description, i.e., the proton was not restricted to move along the N-O axis. The valence bond state description was chosen for this minimization because of its low computational cost and to have a consistent set of parameters throughout this study. We stress that the actual method used for minimization does not play an important role, since it only serves to relax the APEE degrees of freedom with the C-O and N-H distances constrained.

Quantum vibrational energies

The quantum vibrational energy levels for the stretch mode of the transferred proton are calculated by solving the vibrational Schrödinger equation in a given frozen configuration. Following the approach described in ref. 50, a discrete variable representation (DVR)[51] with a sinc-function basis set[52] and the iterative Lanczos algorithm[53] are used to obtain the vibrational eigenvalues and eigenfunctions. We employ a 40 kcal/mol energy cut-off of the vibrational potential and a $\simeq$0.02 Å constant spacing grid along the $\delta = d_{OH} - d_{NH}$ proton coordinate.

Committor and transmission coefficient

The committor $p(\mathbf{x})$ is defined for a given configuration $\mathbf{x}$ as the fraction of trajectories initiated from that configuration with random velocities drawn from a Maxwell-Boltzmann distribution that fall first in the reactant well rather than in the product well,

$$p(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \theta \left(-\Delta E_i \left(T \right) \right) , \tag{2}$$

where the sum runs over the N different sets of initial velocities, θ is the Heaviside function and $\Delta E_i \left(T \right)$ is the value of the collective energy gap coordinate ΔE (see eq 12) along

trajectory i after a delay T.

The transmission coefficient κ is the correction factor applied to the Transition State Theory rate constant k_{TST} to account for transition state surface recrossing, $k = \kappa\ k_{TST}$. It is calculated as[54]

$$\kappa = \frac{\sum_{i=1}^{N} \left| \dot{\widehat{\Delta E}}(0) \right| Q_i}{\sum_{i=1}^{N} \left| \dot{\widehat{\Delta E}}(0) \right|} \ , \tag{3}$$

where the sum runs over N pairs of trajectories propagated forward and backward in time from their starting transition state configuration, $\dot{\widehat{\Delta E}}(0)$ is the initial velocity along the collective reaction coordinate taken to be positive, Q_i is 1 for a trajectory that goes from reactant to product, -1 for product to reactant, and 0 for reactant to reactant or product to product. We note that no Jacobian correction is needed here because the initial configurations were taken from a restricted ensemble, and not a constrained ensemble.[55]

The committor and the transmission coefficient are calculated from a selection of 372 configurations from a trajectory propagated with an equal-weight mixture of the A and B VB states which satisfy $|\Delta E| < 20$ kcal/mol, i.e. which are close to the reaction transition state. From each configuration, 100 pairs of trajectories are propagated in the NVE ensemble without any constraints. For each pair of trajectories, a set of initial velocities $+\mathbf{v}$ is randomly determined from a canonical distribution at 300 K, and two short T=300 fs runs are performed on the VB adiabatic ground state: one with $+\mathbf{v}$ initial velocities, and the other one using opposite initial velocities $-\mathbf{v}$. The committor calculation considers each trajectory separately, while the transmission coefficient involves the pairs of trajectories with opposite velocities.

Results

Reactive PES in different enzyme conformations

We study the acylation step of the APEE transesterification catalyzed by subtilisin Carlsberg in an apolar solvent. This step is known to be rate-limiting in these conditions.[12] This acylation reaction itself proceeds in two elementary steps, which are first the covalent binding of the ester substrate to the enzyme to form a tetrahedral intermediate, and second the rearrangement of this intermediate which leads to the acylenzyme.[4] The formation of the high-energy tetrahedral intermediate is rate-limiting for the acylation,[4] and we will focus on this step.

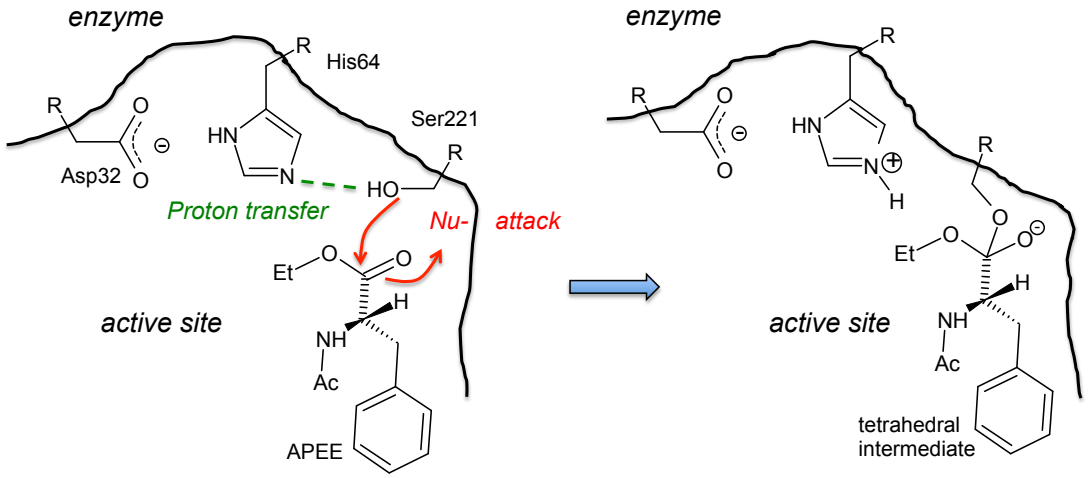


Figure 1: First step in the transesterification reaction of N-acetyl phenylalanine ethyl ester (APEE) by subtilisin Carlsberg. It involves the proton transfer from Ser221 to His64 and the nucleophilic attack of the Ser221 oxygen on the APEE substrate ester carbon atom.

As shown in Figure 1, the formation of the tetrahedral intermediate involves a proton transfer (PT) from Ser221 to His64 and a nucleophilic attack (NA) of the substrate carbonyl by Ser221. It could occur following one of two possible mechanisms, involving the limiting structures defined in Figure 2: (i) the PT and NA could occur in a concerted fashion, so that the reaction pathway directly connects the reactant (structure A) and the product (structure

B), i.e. the acylation tetrahedral intermediate, or (ii) the mechanism could be sequential, with first a PT from state A leading to a putative intermediate structure C, followed by the NA which yields the product structure. We stress that this putative intermediate C is not to be confused with state B which is the product of the reaction step we study here but which is usually designated in the literature as the "tetrahedral intermediate" during the acylation process.

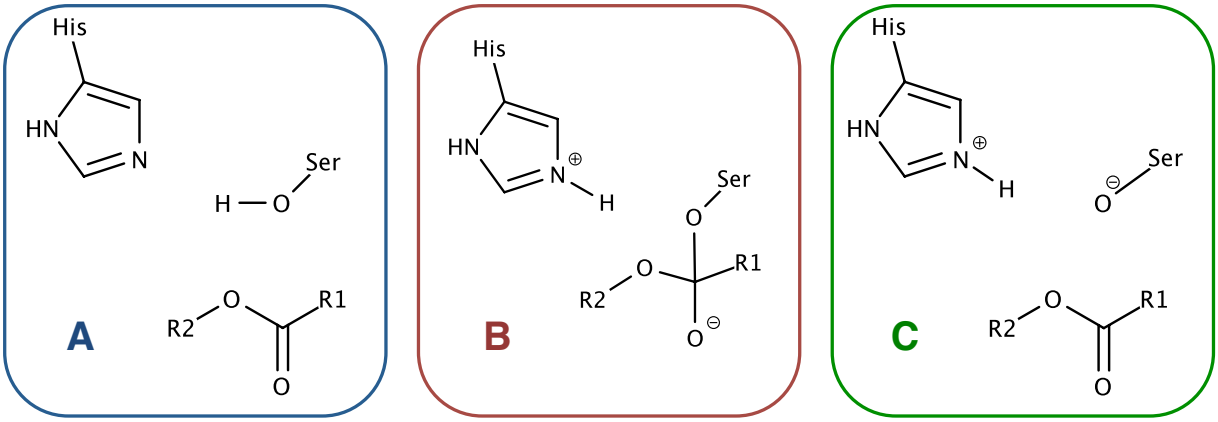


Figure 2: Structures of the three limiting states for the substrate and active-site residues: state A before proton transfer and nucleophilic attack, state B after proton transfer and nucleophilic attack, state C after proton transfer, but before nucleophilic attack.

We calculate the potential energy surface (PES) for the catalyzed PT and NA reactions (Figure 1) in different enzyme conformations and solvent arrangements. In what follows, we will collectively designate the enzyme and the solvent as the environment, and we will demonstrate that environment fluctuations have a major impact on the reaction free energy barrier. We consider two geometric coordinates: the transferred proton location is monitored *via* its NH distance to the His nitrogen acceptor, and the nucleophilic attack is followed *via* the CO distance between the Ser221 attacking oxygen atom and the APEE carbon atom. Energies are calculated with the QM/MM approach described in Methods.

In traditional QM/MM studies, the substrate geometric reaction coordinate is scanned along the PES and the rest of the enzyme is optimized at each step (see e.g. refs[8,56]). This approach aims at finding the enzyme structures that will contribute the most to the free en-

ergy profile along the substrate coordinate. Relaxing the protein structure in this fashion at every value of, e.g., the proton coordinate is equivalent to the equilibrium solvation picture for chemical reactions in solution.[57] However, a clear timescale separation exists between the very slow protein conformational motions and the very fast motion of the transferred proton (and to some extent of the C-O distance). Along a reactive trajectory, it is thus the proton that adapts almost instantaneously to the protein configuration, and not the opposite. Defining the transition state with the proton coordinate will therefore lead to many transition state recrossing events and a low transmission coefficient, due to the configurations where the enzyme conformation has not properly rearranged to allow the reaction to occur.[58] A better reaction coordinate should describe these necessary rearrangements of the enzyme (in an enzymatic context, see e.g. refs 18,29 for collective reaction coordinates respectively defined within an empirical valence-bond approach and a molecular-orbital based QM/MM approach). An additional difficulty with the traditional enzyme minimization approach is that the protein energy landscape is very rugged, and the protein conformation can be easily trapped in local minima instead of the global energy minimum. Many studies of chemical reactions in solution and in enzymes have shown that the environment rearrangements can have a dramatic impact on the reaction barrier and are a key part of the reaction coordinate (see, e.g., refs[18,21,27,59,60]). In this section, we illustrate this for the present enzymatic reaction: we consider a series of frozen environment conformations (where only the APEE substrate configuration is locally relaxed to avoid steric clashes, see Methods), and for each of them we determine the PES along the proton transfer and nucleophilic attack geometric coordinates. In the subsequent sections of this paper, the protein and solvent conformational fluctuations will be a key part of the reaction coordinate and their dynamics will be explicitly considered.

We performed two-dimensional PES calculations for a series of distinct environment configurations, taken from equilibrium trajectories where the substrate and active site are in different electronic states (see Figure 2) along the reaction pathway. We successively consid-

ered cases where the environment is equilibrated to the reactant-like state A (environment configuration $\mathbf{E}_A$), the product-like state B (configuration $\mathbf{E}_B$), the putative proton transfer intermediate state C (configuration $\mathbf{E}_C$), and finally a structure whose charge distribution and bonding pattern are an equal mixture of the A and B states (using the mapping potential defined by eq. 1 (see Methods) with λ =0.5), which approximates the transition state structure for the concerted mechanism (configuration $\mathbf{E}_{AB}$).

The resulting 2D PES (Figure 3) and the resulting minimum energy profiles (Figure 4) are found to change dramatically with the environment configuration. Both the relative stabilities of the reactant and product and the reaction potential energy barrier vary by several tens of kcal/mol between the different environment configurations. For example, in configuration $\mathbf{E}_A$, the reactant structure is 35 kcal/mol more stable than the product. In contrast, in configuration $\mathbf{E}_B$, it is the product that is 20 kcal/mol more stable than the reactant. The potential energy barrier for the PT and NA reactions also exhibits considerable changes with the environment configuration and decreases from more than 35 kcal/mol in the typical reactant $\mathbf{E}_A$ environment configuration to less than 7 kcal/mol in the $\mathbf{E}_{AB}$ transition-state like environment configuration (Figs. 3-4).

Since the reaction potential energy barrier changes so dramatically with the environment, the computational effort in order to obtain the free energy barrier should be carefully balanced between the chosen quantum chemistry level, which determines the accuracy of the potential energy barrier in a given configuration, and the proper sampling of the protein conformations and solvent arrangements. For example, the DFT-based QM/MM approach that we used for this series of configurations would be prohibitively expensive to sample the slow, at least nanosecond, protein rearrangements, and is therefore not adequate.

We thus decided to use a description of the substrate and active site electronic structure based on a linear combination of electronically coupled valence bond (VB) states. Each valence bond state is a diabatic state with fixed bond pattern and charge distribution, and it is described by a classical force field. The low computational cost of such calculations thus

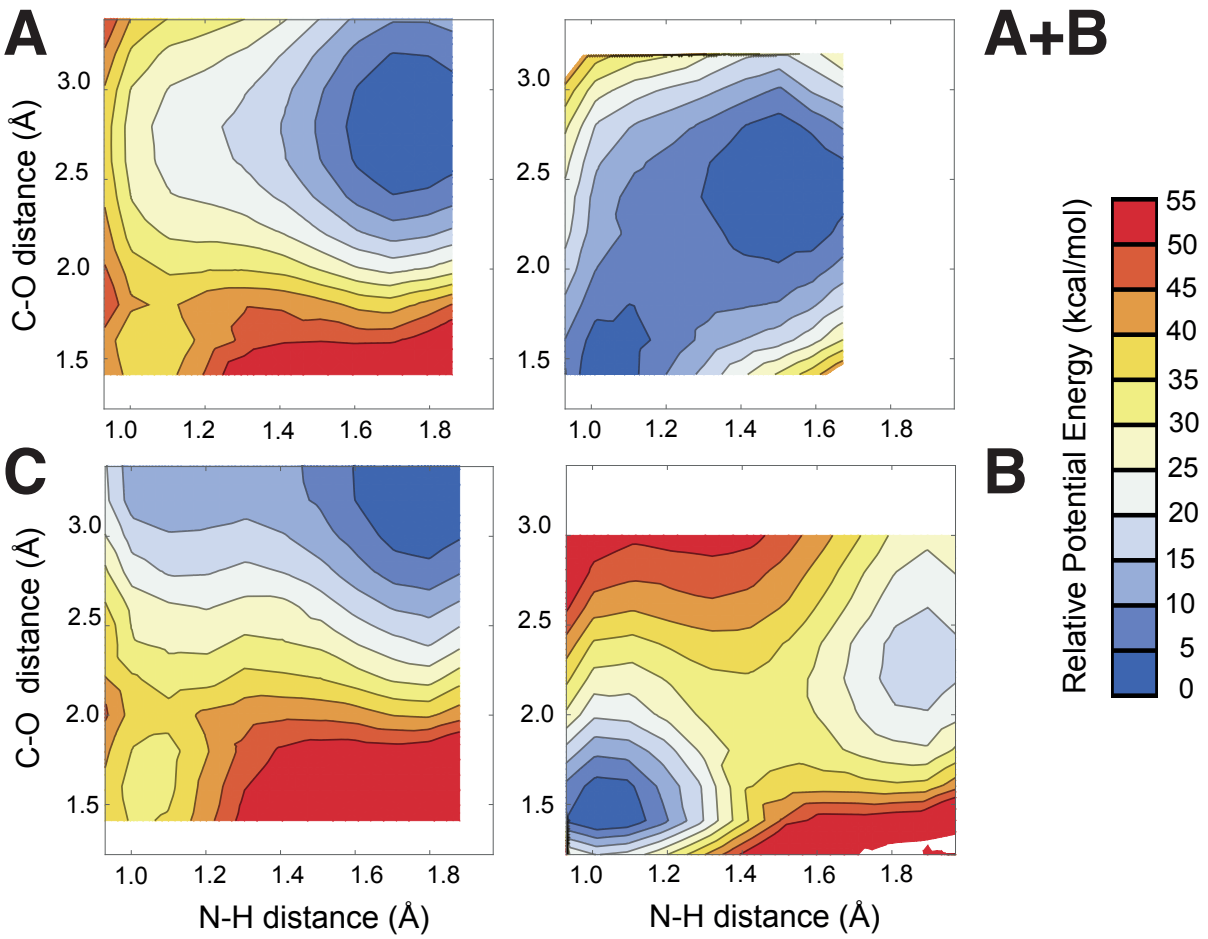


Figure 3: 2D potential energy surfaces along the proton transfer coordinate N-H and nucleophilic attack coordinate C-O in a series of frozen environment configurations respectively equilibrated around the active site and substrate in the reactant-like state A ($\mathbf{E}_A$), the product-like state B ($\mathbf{E}_B$), the putative proton transfer intermediate state C ($\mathbf{E}_C$) and an equal mixture of states A and B ($\mathbf{E}_{AB}$) (see states in Figure 2). Long N-H and C-O distances (top right-hand corner of each plot) correspond to the substrate reactant geometry, short N-H and C-O distances (bottom left-hand corner of each plot) to the product geometry.

allows the system to be extensively sampled. The coupled VB state description has been successfully employed for a broad range of chemical reactions in clusters, in solution and catalyzed by enzymes (see e.g. refs 21–25,27–30).

The quality of the VB description depends on the proper identification of the relevant VB states and on the parameterization of these VB states. While a very large basis set of covalent and ionic VB states could be considered, we focus here on the three most relevant VB states

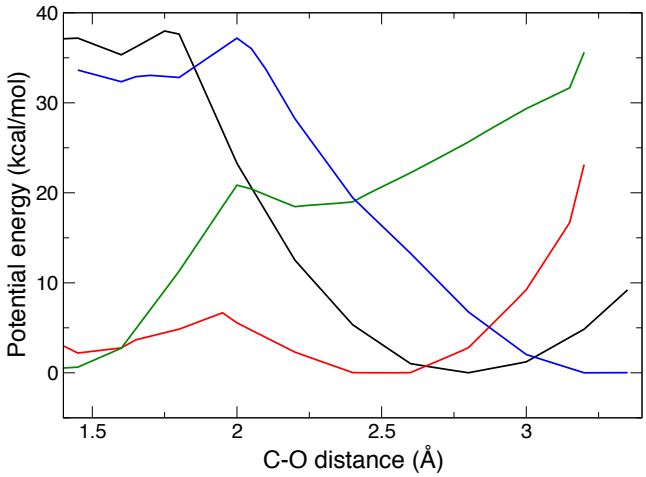


Figure 4: Minimum potential energy profiles obtained from the 2D surfaces in Figure 3 along the C-O coordinate when the N-H coordinate is relaxed, for different frozen protein-solvent environments taken from trajectories with the active site and substrate in state A (black), state B (green), state C (blue) and an equal mixture of states A and B (red).

for the PT and NA reactions leading to the formation of the tetrahedral intermediate. They include the resonance forms describing the reactant and product bonding patterns, together with the putative PT intermediate state that would be relevant if the PT and NA steps were sequential (Figure 2). An early theoretical study of the reaction catalyzed by subtilisin[61] suggested that such a PT intermediate (state C in Figure 2) is stable and is formed by a fast protonation equilibrium with the initial state A, before the rate-limiting NA reaction yields state B. In contrast, subsequent theoretical studies[8,9,62] of serine proteases in water found that PT and NA reactions occur concertedly, without formation of a stable PT intermediate.

As described now, our present calculations show that the PT and NA reactions are concerted in non-aqueous conditions, and that state C can be neglected in our coupled VB state description. First, Figure 3 shows that while the 2D PES for the $\mathbf{E}_A$, $\mathbf{E}_B$ and $\mathbf{E}_{AB}$ protein conformations are very different, all the minimum energy paths approximately follow the diagonal along the N-H and C-O coordinates, without any sign of a local minimum for a PT intermediate (short N-H and long C-O distances, upper left-hand corner of each plot

in Figure 3). This is further supported by the 2D PES in the $\mathbf{E}_C$ conformation obtained after a long equilibration around this putative intermediate (Figure 3) where the reactant arrangement is the global energy minimum. This implies that even in this environment configuration which is most favorable for the C state, if the proton is transferred from Ser221 to His64, it will immediately be transferred back to Ser221 and the PT state is unstable. Our results thus show that PT and NA reactions occur concertedly, and along the course of the reaction the substrate electronic structure can be well approximated as a combination of only two VB states, respectively the A and B forms. This result regarding the concerted character of the reaction is in agreement with recent QM/MM studies on other serine proteases in aqueous solution.[8,9,62] However, it contrasts with a pioneering EVB study on subtilisin,[61] where the NA step was assumed to occur after a fast initial protonation equilibrium, and which had thus developed VB states focusing exclusively on the NA step, while our present VB description – informed by QM/MM calculations – includes both PT and NA reactions.

Coupled Valence Bond State Model

Hamiltonian

Within a two coupled VB state description, the adiabatic ground state is the lower energy eigenstate of the two-state Hamiltonian

$$H\left(\mathbf{R},\dot{\mathbf{R}}\right)=\begin{pmatrix} K\left(\dot{\mathbf{R}}\right)+V_A\left(\mathbf{R}\right) & \beta \\ \beta & K\left(\dot{\mathbf{R}}\right)+V_B\left(\mathbf{R}\right) \end{pmatrix}, \tag{4}$$

where $V_{A,B}\left(\mathbf{R}\right)$ are the energies of VB states A and B in configuration $\mathbf{R}$, K is the kinetic energy associated with the velocities $\dot{\mathbf{R}}$, and β is the electronic coupling. V_B includes a constant energy offset ΔV^0 which accounts for the gas-phase energy shift between the two

VB states. The adiabatic ground state potential is

$$V_{\mathrm{VB}}(\mathbf{R}) = \frac{V_A(\mathbf{R}) + V_B(\mathbf{R})}{2} - \frac{1}{2}\sqrt{\left(V_A(\mathbf{R}) - V_B(\mathbf{R})\right)^2 + 4\beta^2}\ . \tag{5}$$

Individual VB States

Each individual VB state is described by a classical force field (see Methods). For the covalent bonds being broken and formed, the usual harmonic potential is replaced with a Morse potential. This applies to the Ser221 O-H bond in the A state, and the His64 N-H and Ser-APEE C-O bonds in the B state.

Potential energy profiles for the elongation of these three bonds are determined by gas-phase DFT calculations (see Methods) and fitted with a Morse potential $V_{Morse}(r) = D_e\left(1 - \mathrm{e}^{-a(r-r_e)}\right)^2$ (see SI). The resulting parameters are reported in Table 1.

Table 1: Morse potential parameters for the three bonds being broken and formed during the reaction. The error on the fit is provided in parenthesis, and the hyphen indicates that the estimated error bars are much smaller than the last digit given for the value.

bond	D_e (kcal/mol)	r_e (Å)	a (Å^{-1})
O-H	175.3 (1.0)	0.972 (0.003)	1.75 (0.02)
N-H	174.3 (0.7)	1.032 (0.001)	1.72 (0.01)
C-O	63.998 (0.005)	1.515 (-)	1.715 (-)

Offset and Coupling

The central difficulty in the parameterization of VB models and reactive force fields lies in the proper description of the energetic evolution when the system evolves from one resonant state to the other. Here, the two key ingredients are the ΔV^0 diagonal offset between the resonant state energies and the β off-diagonal electronic coupling (eq. 4): in each instantaneous enzyme configuration, the weights of the two VB forms in the ground electronic state depend on the energy gap between these two states (which depends on ΔV^0) and on the strength of their electronic coupling β.

Different strategies have already been employed for the determination of ΔV^0 and β. Many VB simulations use empirical kinetic and thermodynamic data[26,29] to fit these quantities in order to reproduce the reaction rate constant. Other studies on small systems determine these parameters so as to reproduce ab initio gas phase potential energy surfaces along some selected geometric coordinates,[20,22,32,63–66] and recent works have included a solvation free energy contribution due to the condensed phase environment, described as a dielectric continuum.[30,67–69] A different, promising but computationally intensive technique is the paradynamics procedure[31,70,71] where the VB description is iteratively refined and subsequently corrected with higher-level QM/MM calculations. Here we describe a different approach, in which the key ΔV^0 and β parameters are determined to reproduce high-level QM/MM reaction free energies in the presence of the full protein and solvent environment, and which explicitly accounts for the molecular details of the active site and for the substrate and protein conformational fluctuations. We now describe this procedure and we will contrast it with prior approaches in the following subsection.

Our goal is to find offset and coupling values such that the VB description reproduces the reaction free energy (*i.e.*, the free energy difference between reactants and products $\Delta G^{\mathrm{rxn}} = G^{\mathrm{P}} - G^{\mathrm{R}}$ which governs the reaction equilibrium constant) and the reaction free energy barrier (*i.e.*, the free energy difference between reactants and transition state $\Delta G^{\ddagger} = G^{\mathrm{TS}} - G^{\mathrm{R}}$) determined with reference *ab initio* QM/MM calculation. The main difficulty is that the direct *ab initio* QM/MM calculation of these free energies is computationally prohibitive, and these free energy values are unknown. The key point of our procedure is thus to use a free energy perturbation approach to determine the free energy differences between the VB and *ab initio* descriptions in three key locations of the reaction free energy profile, namely in the reactant, at the transition state and in the product. When the free energy shift between the VB and *ab initio* QM/MM descriptions is the same in these three states, then the VB description will reproduce the (unknown) *ab initio* QM/MM ΔG^{rxn} and $\Delta G^{\ddagger}$. While other studies have used free energy perturbation to perform an *a posteriori*

correction of EVB energies with *ab initio* calculations[72] or have iteratively refined an initial EVB description,[31,70] here we directly parameterize the VB model to reproduce the key QM/MM free energy differences, while avoiding the expensive direct evaluation of these free energy differences at the QM/MM level. We use the common assumption that the β coupling is constant but the present approach could be straightforwardly extended to include a dependence of β on some geometric coordinate, see e.g. refs 32,63,65,66 where the coupling is Gaussian.

ΔV^0 and β thus minimize the differences

$$\begin{aligned} \left|\Delta\Delta G^{\ddagger}\right| &= \left|\Delta G^{\ddagger}_{\mathrm{ai}} - \Delta G^{\ddagger}_{\mathrm{VB}}\right| \\ \left|\Delta\Delta G^{\mathrm{rxn}}\right| &= \left|\Delta G^{\mathrm{rxn}}_{\mathrm{ai}} - \Delta G^{\mathrm{rxn}}_{\mathrm{VB}}\right| \ , \end{aligned} \tag{6}$$

where $\Delta G^{\ddagger,\mathrm{rxn}}_{\mathrm{ai,VB}}$ are the free energy differences calculated respectively with the *ab initio* QM/MM and VB Hamiltonians. In the two free energy differences eqs. 6 to be minimized, the terms relative to the same state can be grouped as follows,

$$\begin{aligned} \Delta\Delta G^{\ddagger} &= \left(G^{\mathrm{TS}}_{\mathrm{ai}} - G^{\mathrm{TS}}_{\mathrm{VB}}\right) - \left(G^{\mathrm{R}}_{\mathrm{ai}} - G^{\mathrm{R}}_{\mathrm{VB}}\right) = \Delta G^{\mathrm{TS}}_{\mathrm{ai-VB}} - \Delta G^{\mathrm{R}}_{\mathrm{ai-VB}} \\ \Delta\Delta G^{\mathrm{rxn}} &= \left(G^{\mathrm{P}}_{\mathrm{ai}} - G^{\mathrm{P}}_{\mathrm{VB}}\right) - \left(G^{\mathrm{R}}_{\mathrm{ai}} - G^{\mathrm{R}}_{\mathrm{VB}}\right) = \Delta G^{\mathrm{P}}_{\mathrm{ai-VB}} - \Delta G^{\mathrm{R}}_{\mathrm{ai-VB}} \ . \end{aligned} \tag{7}$$

The free energy differences $\Delta G^{\mathrm{R,TS,P}}_{\mathrm{ai-VB}}$ between the *ab initio* QM/MM and VB descriptions respectively in the reactant, transition and product states are determined with a Bennett acceptance ratio scheme.[73,74] This approach is more reliable and efficient than standard free energy perturbation calculations, since it combines forward and backward calculations of the free energy difference between the two states and it minimizes the variance of the free energy difference for a given data set.[74] Each free energy difference $\Delta G^{x}_{\mathrm{ai-VB}}$ (with x=R, TS, P) satisfies[74]

$$\langle f\left[\left(V_{\mathrm{VB}} - V_{\mathrm{ai}} + \Delta G^{x}_{\mathrm{ai-VB}}\right)/k_B T\right]\rangle_{\mathrm{ai}} = \langle f\left[\left(V_{\mathrm{ai}} - V_{\mathrm{VB}} - \Delta G^{x}_{\mathrm{ai-VB}}\right)/k_B T\right]\rangle_{\mathrm{VB}} \ , \tag{8}$$

where f is the Fermi function

$$f(x) = 1/(1+e^{x}) \ , \tag{9}$$

$V_{\mathrm{ai,VB}}$ are the *ab initio* QM/MM and VB (eq. 5) adiabatic ground state potential energies, and $\langle \cdots \rangle_{\mathrm{ai,VB}}$ are thermal equilibrium averages with the *ab initio* QM/MM and ground state VB Hamiltonians.

However, these thermal averages cannot be calculated directly because phase space sampling with the *ab initio* QM/MM Hamiltonian would be too expensive and the VB Hamiltonian depends on ΔV^0 and β which are still undefined. We therefore determine them from a series of enzyme-substrate configurations in the reactant, transition and product states, selected from three trajectories propagated with the mapping potential (see Methods) $V_\lambda = (1-\lambda)\, V_A + \lambda V_B$ with λ_{R} =0.05, λ_{TS} =0.5 and λ_{P} =0.95 for the reactant, transition state and product (these values correspond to typical weights of state B in the reactant, transition and product states). The terms in eq. 8 are then calculated with a perturbative approach correcting for the biasing potential $V^b = V_\lambda - V_{\mathrm{ai,VB}}$,

$$\langle f \rangle_{\mathrm{ai,VB}} = \frac{\langle f e^{V^b/k_BT} \rangle_\lambda}{\langle e^{V^b/k_BT} \rangle_\lambda} \ , \tag{10}$$

where $\langle \cdots \rangle_\lambda$ is the equilibrium average with the λ mapping potential Hamiltonian.

The combination of eqs. 8 and 10 shows that the free energy differences $\Delta G^{\mathrm{R,TS,P}}_{\mathrm{ai-VB}}$ are the solutions of the following equation,

$$\frac{\langle f\left[\left(V_{\mathrm{VB}} - V_{\mathrm{ai}} + \Delta G^{\mathrm{x}}_{\mathrm{ai-VB}}\right)/k_BT\right] e^{(V_{\lambda_{\mathrm{x}}} - V_{\mathrm{ai}})/k_BT} \rangle_{\lambda_{\mathrm{x}}}}{\langle e^{(V_{\lambda_{\mathrm{x}}} - V_{\mathrm{ai}})/k_BT} \rangle_{\lambda_{\mathrm{x}}}} = \frac{\langle f\left[\left(V_{\mathrm{ai}} - V_{\mathrm{VB}} - \Delta G^{\mathrm{x}}_{\mathrm{ai-VB}}\right)/k_BT\right] e^{(V_{\lambda_{\mathrm{x}}} - V_{\mathrm{ai}})/k_BT} \rangle_{\lambda_{\mathrm{x}}}}{\langle e^{(V_{\lambda_{\mathrm{x}}} - V_{\mathrm{VB}})/k_BT} \rangle_{\lambda_{\mathrm{x}}}} \ , \tag{11}$$

with x=R, TS, P.

We generated a total of 2130 configurations, including respectively 918, 596 and 616 configurations in the reactant, transition and product states. The offset and coupling are optimized iteratively. At each iteration step, the adiabatic ground state V_{VB} energy eq. 5 of

each configuration is recalculated with the new trial ΔV^0 and β values and the differences in reaction free energy and free energy barrier between the VB and *ab initio* QM/MM descriptions are recalculated with eqs. 7 and 11. As shown in the SI (Fig S7a), the offset and coupling values reach a plateau beyond a total of 800 configurations and are fully converged with our total set of 2130 configurations, leading to the values of ΔV^0=83.3 kcal/mol and β=31.3 kcal/mol.

We assessed the error on these ΔV^0 and β values by dividing the ensemble of configurations in 3 independent sets and performing the iterative parameter optimization on each set. The error bars defined as one half of the Student 90% confidence interval are respectively 3 and 5 kcal/mol for ΔV^0 and β. These error bars come from the uncertainty on the ab initio QM/MM free energies in the reactant, transition state and product, due to the limited size of each individual data set. We note that these VB parameter uncertainties are usually not accessible with other approaches fitting the parameters on experimental values or on single point ab initio free energies of optimized structures.

We stress that this set of parameters provides an excellent description of the QM/MM reaction free energy ΔG^{rxn} and $\Delta G^{\ddagger}$ free energy barrier, since after minimization the residual errors on $\Delta\Delta G^{\mathrm{rxn}}$ and $\Delta\Delta G^{\ddagger}$ are below 0.06 kcal/mol for both (Fig S7). In addition, while the error on the energy of an instantaneous configuration is less relevant than these free energies, we note that the potential energies from our VB model are extremely well correlated with the ab initio QM/MM reference energies, with a root mean square error of less than 0.10 kcal/mol/atom, which is for example at least as good as the results of a recent state-of-the-art neural network approach[75] (see SI).

Our approach thus provides a computationally efficient VB parameterization, which minimizes the errors on ΔG^{rxn} and $\Delta G^{\ddagger}$ relative to the values obtained with a reference ab initio QM/MM description. It offers an excellent compromise, since a converged $\Delta G^{\ddagger}$ determination at the QM/MM level would be prohibitively expensive, and an approximate QM/MM $\Delta G^{\ddagger}$ estimate from a small number of configurations would lead to large error bars that may

reach several tens of kcal/mol, as shown above.

Comparison with other approaches

Some recent studies used different strategies to parameterize the EVB description on *ab initio* free energy calculations. We now briefly review these approaches and underline the key differences and similarities with the present approach.

In a first approach,[30,67–69] the reactant, transition state and product free energies are determined as the sum of the single-point QM/MM potential energy of the reactive system in its optimized geometry plus a free energy correction based on a frequency analysis, and a solvation free energy term, employing a dielectric continuum description of the environment. The VB parameters are then optimized to reproduce the free energy barrier and reaction free energy. While this approach has the great advantage of its simplicity, it relies on a continuum solvation approximation. The typical solvation free energy errors of these methods are on the order of 5-10 kcal/mol[76] for charged systems, and the continuum approximation is probably very crude for heterogeneous environments like our present protein immersed in a mixture of water and hexane, where the local dielectric constant is not homogeneous. In addition, as will be shown below, a key role is played by some specific solvent molecules and by the hydrogen-bonds that they can form, which cannot be described by a continuum picture. In contrast, in our approach the molecular structure of the environment is explicitly described and treated in the same way during the parameter optimization and the subsequent dynamical trajectory calculations. In addition, we use a thermal sampling of hundreds of configurations in the reactant, product and transition states instead of a single optimized geometry.

Our approach bears some similarities with the recently introduced paradynamics sampling,[31,70,77] where an initial EVB parameterization is iteratively refined against *ab initio* potential energies and the final EVB description is employed as a reference sampling potential to calculate the free energy profile; in this technique, the QM/MM free energy profile is obtained from the EVB profile using either the linear response approximation or more

computationally demanding free energy perturbation calculations. In both paradynamics and our approach the molecular structure of the local environment is explicitly described, and a statistical ensemble of configurations is used to optimize the VB parameters. However, the goals and the computational costs of the two methods differ: while paradynamics uses EVB as an intermediate sampling potential from which the QM/MM free energy profile will be obtained, our method aims at efficiently optimizing a VB description to minimize the VB – QM/MM difference on key free-energy differences. The main differences are thus the following: a) our approach is computationally cheap since we exclusively optimize the offset and coupling (and not the force fields of the resonant VB states since as shown in the SI the force-field rms error on these states' energies is very small) and QM/MM calculations are only required for this optimization but not for the subsequent free energy profile calculations; b) our approach aims at reproducing the QM/MM free energies, while paradynamics considers the rms error on the potential energies; c) our approach recognizes the importance of the product configurations to determine the VB offset energy; d) we employ the BAR estimator to determine the free energy differences between the EVB and QM/MM descriptions in the R, TS and P states, which does not require the energy differences between the VB and QM/MM descriptions to be small with respect to the thermal energy as assumed in the linear response approach (which was shown in the SI not to be satisfied here), and which is less computationally demanding than free energy integration. While paradynamics is certainly a promising and general approach for free energy calculations and evaluation of free-energy profiles at the QM/MM level,[78] our present technique thus has the advantage of its simplicity and the obtained optimized VB description was shown to yield the QM/MM free energy barrier within less than 0.06 kcal/mol.

Reaction free energy profile

We now use our coupled VB state description calibrated on high level *ab initio* QM/MM calculations to investigate the reaction catalyzed by subtilisin and we first calculate the

reaction free energy profile.

Collective reaction coordinate

The reaction free energy barrier is determined from the free energy profile along a collective reaction coordinate analogous to the solvent coordinate in Marcus theory for electron transfer.[79] In contrast to the purely geometric coordinates, this collective coordinate probes the rearrangements of the environment (*e.g.*, enzymatic active site, solvent molecules, ligand conformation) that are necessary for the NA and PT reactions to take place. It is defined here as the energy difference between the A and B VB states in a given environment configuration.

While it would be desirable to have an environment coordinate that does not depend on the chemical coordinate (*e.g.*, the position of the transferred proton) and thus considers the VB states in their equilibrium geometries,[80] this would dramatically increase the computational effort, since it would require additional geometry optimization to the VB state equilibrium geometries every time the environment coordinate value was to be calculated, i.e. at every simulation timestep. We therefore follow here prior works (see e.g. refs. 27,59) and use the instantaneous energy gap between the two VB states in a given configuration $\mathbf{R}$, that can be more easily computed from the molecular dynamics trajectory,

$$\Delta E\left(\mathbf{R}\right) = V_A\left(\mathbf{R}\right) - V_B\left(\mathbf{R}\right) \ . \tag{12}$$

Free energy and uncertainties

A number of computational techniques exist to calculate free energy profiles when the free energy barrier exceeds the thermal energy.[81] Umbrella sampling[82] is a popular approach to drive the system along the reaction coordinate, involving biasing potentials along the reaction coordinate. However, this technique can lead to artifacts in the present system, where several coordinates with very different timescales contribute to the instantaneous ΔE

value, including e.g. the ultrafast motion of the transferred proton and the slower active site rearrangements. When a constraint is applied to ΔE, the fast proton coordinate adapts first to satisfy the target ΔE value, and the constraint does not induce any subsequent rearrangement of the slower protein coordinates, leading to a potentially incorrect environment polarization at the transition state.[64] We therefore adopt the mapping potential approach[46] in which the substrate electronic structure is progressively switched from that of a pure reactant-like A state to that of a pure product-like B state, $V_\lambda = (1-\lambda)V_A + \lambda V_B$, where λ is changed gradually from 0 to 1.

The free energy profile is reconstructed from independent trajectories run at a series of different λ values using an adapted version of the weighted histogram analysis method (WHAM).[83] While WHAM is usually employed in conjunction with umbrella sampling, it can be easily generalized to mapping potential simulations by defining the biasing potential as

$$w_\lambda(\Delta E) = V_\lambda(\Delta E) - V_{\mathrm{VB}}(\Delta E) = \Delta E\left(\frac{1}{2} - \lambda\right) + \frac{1}{2}\sqrt{\Delta E^2 + 4\beta^2}\ . \tag{13}$$

We use the improved WHAM implementation suggested in ref. 84 where convergence is accelerated by maximizing a target function instead of the traditional iterative approach. We note that a different approach combining mapping potential simulations with an umbrella integration approach was recently suggested.[85] However, this method assumes that the ΔE reaction coordinate fluctuations are Gaussian in every window, a condition that we found not to be satisfied in the trajectories close to the transition state (λ=0.5).

The uncertainties on the resulting free energy profile, and especially on the free energy barrier, are calculated by generalizing the approach previously described[84] for the combination of WHAM and umbrella sampling. The uncertainties stem both from statistical errors on the ΔE probability distribution in each λ window and from the free energy shift between independent windows. In the umbrella sampling approach where the system can be confined within a narrow range of ΔE values by a stiff biasing potential, the latter term is clearly dominant. In the empirical mapping approach, although the system cannot be arbitrarily

confined along ΔE, we verified that the ΔE probability distribution converges rapidly within each window and we use the same approximation.

Extending the result of ref. 84, the variance of the free energy difference between two mapping potential windows λ_j and λ_k is

$$\mathrm{var}\left[G\left(\Delta E_k\right) - G\left(\Delta E_j\right)\right] = \left(\Delta\lambda\right)^2 \left[\frac{\mathrm{var}\left(\overline{\Delta E_j}\right) + \mathrm{var}\left(\overline{\Delta E_k}\right)}{4} + \sum_{i=j+1}^{k-1} \mathrm{var}\left(\overline{\Delta E_i}\right)\right] , \qquad (14)$$

where $\Delta\lambda$ is the difference between successive λ values and $\mathrm{var}\left(\overline{\Delta E_i}\right)$ is the variance of the average ΔE in window λ_i, computed by block averages[84] within this window. The $\overline{\Delta E_i}$ variance in each window is independent of the number of windows, and the sum in the right-hand side of eq. 14 scales with the number of windows, and thus inversely with $\Delta\lambda$. The variance of the free energy difference in eq. 14 therefore scales linearly with $\Delta\lambda$, and this shows that the uncertainty on $\Delta G^{\ddagger}$ can be systematically decreased by decreasing the $\Delta\lambda$ increment between successive mapping potential windows.

Reaction free energy profile

We used the above-described mapping potential approach and obtained 8 independent free energy profiles generated from 8 uncorrelated starting configurations. We propagated 39 distinct windows per profile, for a total sampling time of more than 15 ns (see Methods). We stress again that this extensive sampling is allowed by the efficient coupled VB state description and would not be accessible with ab initio or DFT-based QM/MM methods.

The resulting 8 reaction free energy profiles together with the average profile are shown in Figure 5. We find the free energy barrier to be $\Delta G^{\ddagger} = 15.7 \pm 0.15$ kcal/mol (where the error bar is the standard deviation determined from the cumulated variance eq. 14). We now show that this value is in fair agreement with estimates from available experimental measurements. Experimentally, the overall rate constant for the transesterification reaction catalyzed by subtilisin Carlsberg in isooctane with approximately a monolayer of water was

measured[13] to be $k_{cat} = 81.8$ s^{-1}. Since the formation of the first tetrahedral intermediate studied is rate-limiting at low hydration,[12] this rate can be converted into an approximate free energy barrier using Transition State Theory with the usual $k_B T/h$ prefactor, and yields a 15 kcal/mol value, in good agreement with the present results. This determination is approximate but it supports the suitability of our approach and the quality of our VB parameterization, since none of our VB parameters were fitted on empirical values.

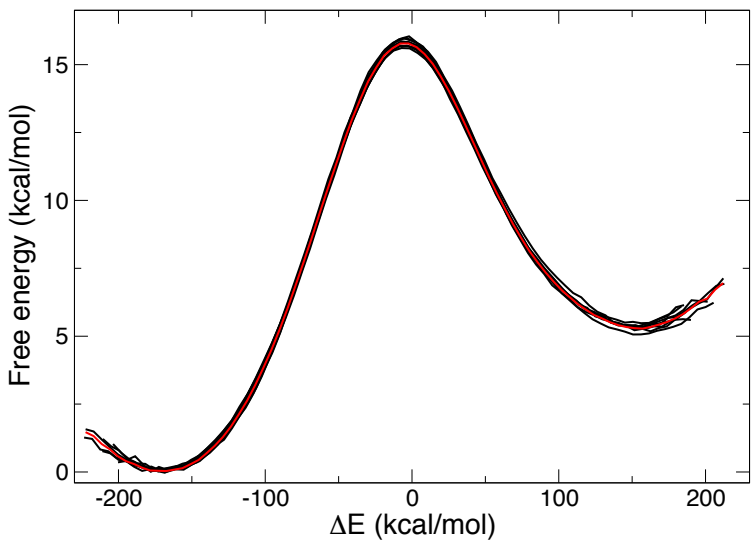


Figure 5: Free energy profile for the concerted NA and PT reactions as a function of the collective reaction coordinate ΔE. The eight independent profiles generated from distinct configurations are shown in black, and the average profile in red.

In addition, we note that our calculated free energy barrier lies within the range of values obtained in previous QM/MM-MD studies of the acylation step tetrahedral intermediate formation catalyzed by serine proteases in water: $\Delta G^{\ddagger} = 16.4 - 17.8 \pm 0.3$ kcal/mol for trypsin[8,62] and $\Delta G^{\ddagger} = 12.4 \pm 0.3$ for acetylcholinesterase.[9] However, no direct comparison is possible since these calculations were performed on enzymes in aqueous solution and not in an apolar solvent, and these studies did not explicitly consider the importance of the enzyme conformational sampling.

We find that the reaction is endergonic, with a reaction free energy $\Delta G^{\mathrm{rxn}} = +5$ kcal/mol. Our result thus suggests that the endergonic character of the tetrahedral intermediate formation is common to the apolar hexane solvent and to the aqueous solvent, for which a prior experimental study on a serine protease had concluded that the acylation step transition state resembles the tetrahedral intermediate product[86] and previous calculations had found reaction free energies ranging from +4 kcal/mol[87] to +11.3 kcal/mol[9] with acetylcholinesterase, and from +7 kcal/mol[8] to +14.9 kcal/mol[62] with trypsin.

Transition state analysis

We now study the reaction transition state configuration. We first examine whether the collective ΔE coordinate provides a good description of the transition state. Trajectories initiated from the transition state should have equal probabilities to end in the reactant and product wells. This is confirmed by the committor analysis, in which we determine the fraction of trajectories initiated from a given set of transition-state configurations that fall first in the reactant well rather than in the product well (see Methods for further details). The results are shown in Figure 6a. As expected, the committor approaches 1 on the reactant side of the barrier ($\Delta E << 0$), where trajectories fall in the reactant well, and conversely tend to 0 on the product side ($\Delta E >> 0$). The transition state is located at the ΔE value where the average committor is equal to 0.5, leading to $\Delta E^{\ddagger} = 2.2 \pm 1.4$ kcal/mol, where the uncertainty is determined from the Student 95% confidence interval calculated by dividing the data in 5 independent blocks. We note here that while the transition state is expected to be found when the energies of the two VB states are approximately equal, i.e. $\Delta E \simeq 0$, this assignment is only approximate since for example the two VB states may have different entropic contributions.

The ability of the ΔE coordinate to define the transition state can be further examined by evaluating the transmission coefficient κ, which corrects the Transition State Theory rate constant for transition state recrossing[88] and ranges from 0 to 1. Transmission coeffi-

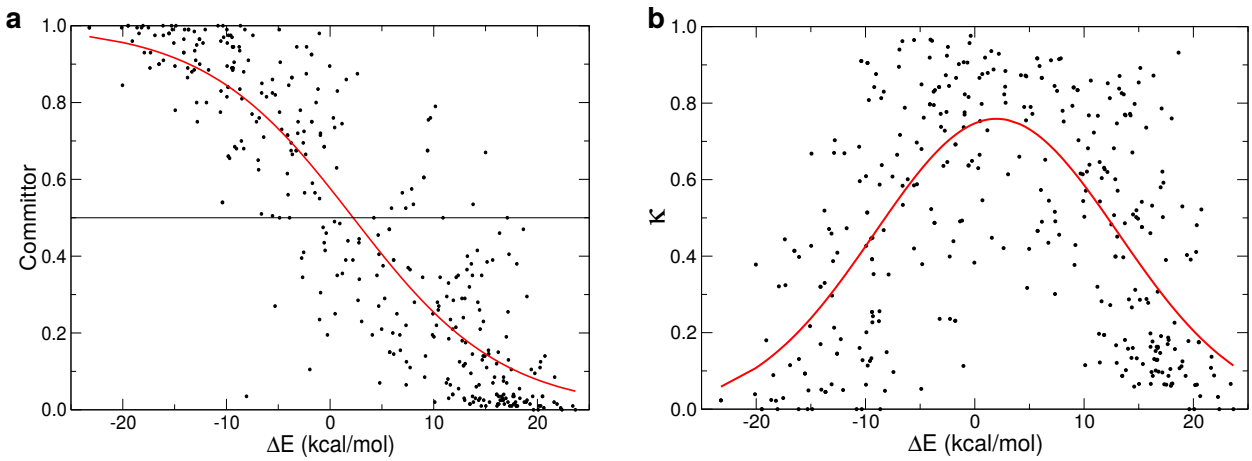


Figure 6: a) Committor values for 372 configurations in the transition state vicinity, plotted against the collective reaction coordinate ΔE and fitted with a switch function (plain line). b) Transmission coefficient calculated for the same configurations, fitted with a Gaussian function (red line). The range of ΔE values used here correspond to the top of the free energy barrier, as can be seen in Figure 5.

cients were determined for a series of initial configurations close to the transition state (see Methods). Since the free energy profile is locally parabolic in the vicinity of the transition state, $G(\Delta E) = G^{\ddagger} - 1/2\ \eta\left(\Delta E - \Delta E^{\ddagger}\right)^2$, the transmission coefficient for a given initial configuration exhibits a Gaussian dependence on the reaction coordinate ΔE,

$$\kappa(\Delta E) = \kappa_0 \exp\left(-\frac{\eta\left(\Delta E - \Delta E^{\ddagger}\right)^2}{2k_B T}\right) . \tag{15}$$

Figure 6b shows that κ reaches a maximum value of $\kappa = 0.76 \pm 0.06$ at the transition state. This fairly high value means that the transition state surface recrossing is very limited. This supports our choice of the collective reaction coordinate ΔE to follow the key rearrangements of the enzyme local environment for the reaction.

We now turn to the description of the transition state geometry for the substrate and active site. Table 2 summarizes the key geometric parameters obtained from our Valence-Bond approach and compares them with the results of previous calculations on other serine proteases in water. Our results suggest that the transition state geometry for the reaction in a non-aqueous solvent is similar to that obtained in water, and that our coupled VB

approach yields results consistent with those obtained with other QM/MM methods.

Table 2: Selected transition state distances in Å, with standard deviation given in parenthesis, respectively from our coupled VB approach on subtilisin Carlsberg in hexane (averaged over typical transition state geometries obtained from a $\lambda = 0.5$ mapping potential trajectory) and from previous calculations in aqueous conditions, either with fully optimized geometries at the QM/MM-MD HF-6-31+G$**$ level for trypsin[8] or from averaged geometric parameters at the QM/MM-MD B3LYP-6-31+G$*$ level for trypsin[62] and for acetylcholinesterase (AChE).[9]

bond	subtilisin Carlsberg (this work)	Trypsin (from Ref. 8)	Trypsin (from Ref. 62)	AChE (from Ref. 9)
Ser O-H	1.29 (0.07)	1.447	1.66 (0.06)	1.37 (0.07)
His N-H	1.21 (0.06)	1.090	1.06 (0.03)	1.17 (0.05)
O(Ser)-N(His)	2.48 (0.04)	2.76		
O(Ser)-C(substrate)	1.89 (0.07)	2.0	1.56 (0.05)	1.82 (0.07)

Finally, we investigated whether the proton transfer reaction from Ser221 to His64 proceeds *via* tunneling or not.[89] We calculated the potential energy profile experienced by the transferred proton when all the other (slower) coordinates are in their transition state arrangement. From transition state geometries selected from a trajectory generated using the mapping potential eq. 1 with $\lambda = 0.5$, we performed rigid scans of the proton position, moving it within the frozen enzyme environment along a straight line from the serine oxygen atom towards the histidine nitrogen atom, using single point QM/MM energies (see Methods). Figure 7 shows representative proton potential energy profiles at the transition state, plotted against the antisymmetric proton transfer coordinate $\delta = d_{OH} - d_{NH}$. They show that when the environment has reorganized to reach the transition state, the proton is approximately equally stabilized on the donor and acceptor sides, and it experiences either a broad single well or a low barrier (<1.5 kcal/mol) (quasi-)symmetric double well. To determine whether the PT reaction requires tunneling, we computed the quantum vibrational zero-point energy (ZPE) level in each configuration (see Methods). Figure 7 shows that the barrier along the PT coordinate is either non-existent or much lower than the ZPE. This shows that while the proton vibration should be considered as a quantum because its ZPE

is much larger than the thermal energy $k_B T$, the PT reaction does not involve tunneling through the barrier. In the reactant state and away from the transition state region, the proton lies on the donor side and the energy of its ground vibrational state is much lower than that any proton vibrational energy level with the proton on the acceptor side (Figs 3A-4) and no tunneling is possible. At the transition state, the proton is largely delocalized between the donor and acceptor.

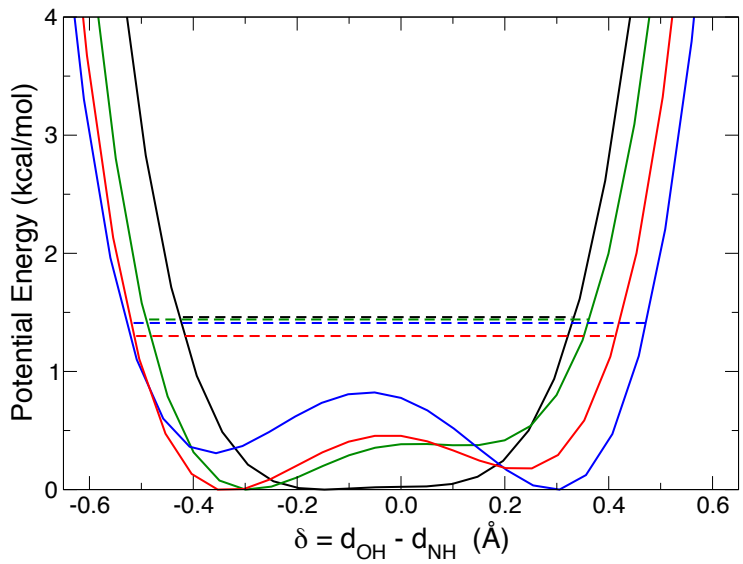


Figure 7: Typical potential energy profiles (solid lines) along the antisymmetric PT coordinate $\delta = d_{OH} - d_{NH}$ and ZPE levels (dashes) calculated in four different configurations in the transition state vicinity.

Molecular interpretation of collective reaction coordinate

We finally provide a molecular interpretation of the reaction coordinate that governs the reaction. With respect to simple geometric coordinates, the energy gap coordinate ΔE used here has the great advantage that it can probe the collective rearrangements of the environment that must precede the actual bond breaking and making and that are thus rate-limiting. However, its molecular interpretation is often less intuitive than that of geometric

distances. We therefore remedy this limitation by analyzing which molecular groups and which motions govern the change in ΔE.

The change in ΔE between the reactant well ($\Delta E \simeq$-175 kcal/mol) and the product well ($\Delta E \simeq$+160 kcal/mol) is extremely high, because it includes contributions from the VB state internal energy, which can become very large when the instantaneous geometry being considered is the stable configuration of the other VB state (e.g. the dissociation energy of a covalent bond that exists in one state but not in the other). However, the relevant part of ΔE is the relative stabilization of the two VB states by the electrostatic polarization created by the environment (including the rest of the enzyme and the solvent), i.e. the difference in electrostatic interaction energies between the two VB states and their environment,

$$\Delta E^{\text{elec}}(\mathbf{R}) = V_A^{\text{elec}}(\mathbf{R}) - V_B^{\text{elec}}(\mathbf{R}) \ . \tag{16}$$

Our calculations show that the total change in ΔE^{elec} between the reactant and the product is $\simeq$+21 kcal/mol, which can be decomposed as a dominant +14 kcal/mol contribution due to the interaction with the surrounding water molecules and a smaller +7 kcal/mol component due to the interaction with (non-reacting) enzyme residues (see details in SI). We now examine which reactive groups are stabilized by which parts of the environment.

The +7 kcal/mol change in the VB-enzyme term arises from the stabilization of the growing negative charge on the substrate oxyanion site by the NH groups of the Ser221 and Asn155 residues forming the oxyanion hole (+1.5 kcal/mol) and from the stabilization of the growing positive charge on His64 by the rest of the enzyme (+3.5 kcal/mol, including an important +2.5 kcal/mol contribution from Asp32). This is visible in the change in the enzyme conformation described in Figure 8; when going from the reactant to the transition state and the product, the distance from the APEE carbonyl O atom to the Ser221 and Asn155 residues decreases, and the His64 and Asp32 residues move closer.

Regarding the large +14 kcal/mol VB-water term, it is decomposed in approximately

equal contributions due to the stabilization of the growing negative charge on the oxyanion and on the neighboring APEE atoms (+7 kcal/mol) and to the water-His64 interaction (+7 kcal/mol). The changes in the number and in the orientation of water molecules around the His64 residue and the APEE oxyanion site in the reactant, transition state and the product are supported by the geometric analysis in Figure 8.

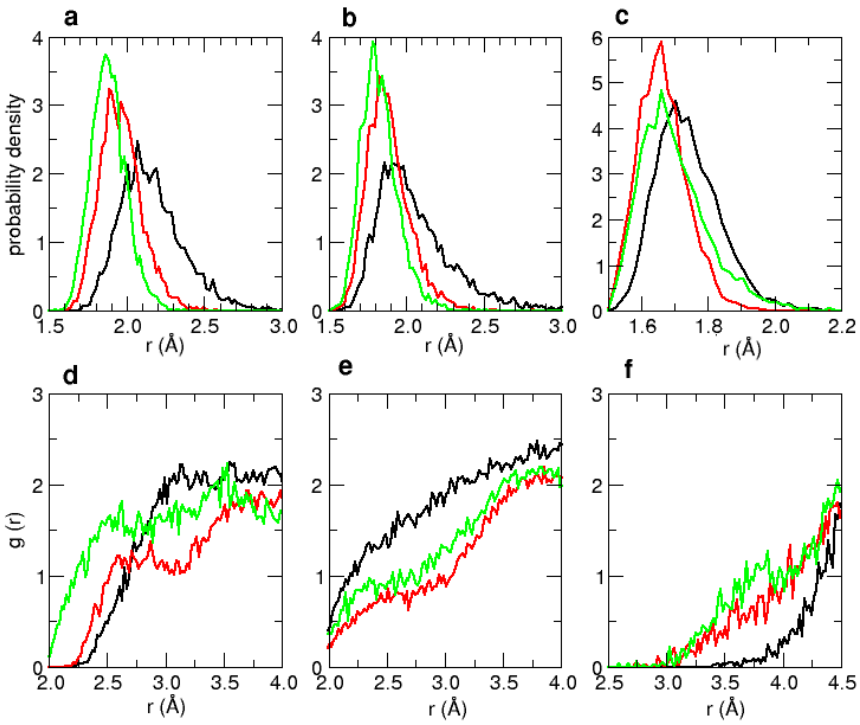


Figure 8: Changes in the active-site geometry when going from reactant (black) to transition state (red) and product (green): distributions of distances respectively between APEE carbonyl O atom and Ser221 amide H atom (a), APEE carbonyl O atom and Asn155 amide H (b) and between His64 H and Asp32 carboxylic O atoms (c). Radial distribution functions between the His64 side chain atoms and the water O atoms (d) and H atoms (e), and between the APEE carbonyl O atom and the water H atoms (f).

An important conclusion is that although the presence of the key catalytic residues (catalytic triad and oxyanion hole) is essential to lower the reaction barrier, their motions do not bring an important contribution to the reaction coordinate, i.e. the reaction is not driven by their motions. Both the Asp residue, which is part of the catalytic triad, and the oxyanion hole are conserved in all serine proteases, and their electrostatic stabilization of the product

charge distribution is the main source of the catalytic activity and of the acceleration of the reaction with respect to conditions without the enzyme.[4,6] However, our analysis reveals that this electrostatic stabilization changes only little during the reaction and it is not the motions of these residues that need to be activated to reach the transition state. In contrast, we find that the rearrangement of the very few water molecules remaining in the vicinity of the active site (even though the enzyme is immersed in hexane) is the main cause of the destabilization of the reactant charge distribution and stabilization of the product charge distribution, leading to the reaction. Although the enzyme conformation is an important factor which affects the reaction barrier, this shows that the solvent configuration also plays a major role.

Concluding remarks

We have studied the transesterification reaction catalyzed by subtilisin Carlsberg in hexane. Our calculations have shown that the reaction energy barrier is extremely sensitive to the protein conformation and solvent configuration and can exhibit fluctuations that exceed several tens of kcal/mol. The calculation of a converged free energy barrier therefore requires the extensive sampling of the protein conformations and solvent arrangements. A particularly well-suited approach is offered by the description of the reactive system electronic structure by a combination of coupled valence bond states, and we have developed a new rigorous procedure to parameterize this description on high-level *ab initio* QM/MM free energy calculations. Our results show that the reaction mechanism of the rate-limiting formation of the tetrahedral intermediate during the acylation step is similar to that in an aqueous solvent. The proton transfer and nucleophilic attack reactions occur concertedly and the proton transfer does not require tunneling. While the enzyme catalytic triad and oxyanion hole residues are essential to lower the reaction free energy barrier with respect to the uncatalyzed situation, we determined that the remaining barrier in the catalyzed

reaction is mostly overcome by rearrangements of the very few water molecules present in the vicinity of the active site, which drive the reaction. Future work will apply this coupled valence-bond state description to determine the origin of the dramatic dependence of the experimentally-measured[13] catalytic rate constant on the hydration level.

Supporting Information Available

The following files are available free of charge. Further details about valence-bond state force-field parameters, definition of the QM region, convergence of VB parameters, and ΔE electrostatic decomposition.

Acknowledgement

We gratefully thank I. Tuñón, J. Ruiz-Pernía and J. Aranda (Univ. Valencia, Spain) for discussions on QM/MM methods and Kristen Marino for her help with the force-field parameterization. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC Grant Agreement No. 279977 and was performed using HPC resources from GENCI-TGCC (Grant 2014-077156).

References

(1) Klibanov, A. M. Improving Enzymes by Using Them in Organic Solvents. *Nature* **2001**, *409*, 241–246.

(2) Liszka, M. J.; Clark, M. E.; Schneider, E.; Clark, D. S. Nature Versus Nurture: Developing Enzymes That Function Under Extreme Conditions. *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3*, 77–102.

(3) Kraut, J. Serine Proteases: Structure and Mechanism of Catalysis. *Annu. Rev. Biochem.* **1977**, *46*, 331–358.

(4) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4524.

(5) Wilmouth, R. C.; Edman, K.; Neutze, R.; Wright, P. A.; Clifton, I. J.; Schneider, T. R.; Schofield, C. J.; Hajdu, J. X-Ray Snapshots of Serine Protease Catalysis Reveal a Tetrahedral Intermediate. *Nat. Struct. Biol.* **2001**, *8*, 689–694.

(6) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. How Do Serine Proteases Really Work? *Biochemistry* **1989**, *28*, 3629–3637.

(7) Topf, M.; Várnai, P.; Richards, W. G. Ab Initio QM/MM Dynamics Simulation of the Tetrahedral Intermediate of Serine Proteases: Insights Into the Active Site Hydrogen-Bonding Network. *J. Am. Chem. Soc.* **2002**, *124*, 14780–14788.

(8) Ishida, T.; Kato, S. Theoretical Perspectives on the Reaction Mechanism of Serine Proteases: The Reaction Free Energy Profiles of the Acylation Process. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048.

(9) Zhou, Y.; Wang, S.; Zhang, Y. Catalytic Reaction Mechanism of Acetylcholinesterase Determined by BornOppenheimer Ab Initio QM/MM Molecular Dynamics Simulations. *J. Phys. Chem. B* **2010**, *114*, 8817–8825.

(10) Yang, Z.; Russell, A. J. *Enzymatic Reactions in Organic Media*; Springer Netherlands, 1996; pp 43–69.

(11) Chatterjee, S.; Russell, A. J. Determination of Equilibrium and Individual Rate Constants for Subtilisin-Catalyzed Transesterification in Anhydrous Environments. *Biotechnol. Bioeng.* **1992**, *40*, 1069–1077.

(12) Wangikar, P. P.; Graycar, T. P.; Estell, D. A.; Clark, D. S.; Dordick, J. S. Protein and Solvent Engineering of Subtilisin BPN' in Nearly Anhydrous Organic Media. *J. Am. Chem. Soc.* **1993**, *115*, 12231–12237.

(13) Hudson, E. P.; Eppler, R. K.; Beaudoin, J. M.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. Active-Site Motions and Polarity Enhance Catalytic Turnover of Hydrated Subtilisin Dissolved in Organic Solvents. *J. Am. Chem. Soc.* **2009**, *131*, 4294–4300.

(14) Eppler, R. K.; Komor, R. S.; Huynh, J.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. Water Dynamics and Salt-Activation of Enzymes in Organic Media: Mechanistic Implications Revealed by NMR Spectroscopy. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5706–5710.

(15) Eppler, R. K.; Hudson, E. P.; Chase, S. D.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. Biocatalyst Activity in Nonaqueous Environments Correlates With Centisecond-Range Protein Motions. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15672–15677.

(16) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. Taking Ockham's Razor to Enzyme Dynamics and Catalysis. *Nat. Chem.* **2012**, *4*, 169–176.

(17) Hammes-Schiffer, S. Catalytic Efficiency of Enzymes: A Theoretical Analysis. *Biochemistry* **2013**, *52*, 2012–2020.

(18) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. Studying the Role of Protein Dynamics in an SN2 Enzyme Reaction Using Free-Energy Surfaces and Solvent Coordinates. *Nat. Chem.* **2013**, *5*, 566–571.

(19) Masgrau, L.; Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res* **2014**, *48*, 431–438.

(20) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. Simulation of Free Energy Relationships and Dynamics of SN2 Reactions in Aqueous Solution. *J. Am. Chem. Soc.* **1988**, *110*, 5297–5311.

(21) Ando, K.; Hynes, J. T. HCl Acid Ionization in Water: A Theoretical Molecular Modeling. *J. Mol. Liq.* **1995**, *64*, 25–37.

(22) Thompson, W. H.; Hynes, J. T. Frequency Shifts in the Hydrogen-Bonded OH Stretch in HalideWater Clusters. The Importance of Charge Transfer. *J. Am. Chem. Soc.* **2000**, *122*, 6278–6286.

(23) Vuilleumier, R.; Borgis, D. An Extended Empirical Valence Bond Model for Describing Proton Mobility in Water. *Isr. J. Chem.* **1999**, *39*, 457–467.

(24) Schmitt, U. W.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.

(25) Glowacki, D. R.; Rose, R. A.; Greaves, S. J.; Orr-Ewing, A. J.; Harvey, J. N. Ultrafast Energy Flow in the Wake of Solution-Phase Bimolecular Reactions. *Nat. Chem.* **2011**, *3*, 850–855.

(26) Åqvist, J.; Warshel, A. Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chem. Rev.* **1993**, *93*, 2523–2544.

(27) Hammes-Schiffer, S.; Billeter, S. R. Hybrid Approach for the Dynamical Simulation of Proton and Hydride Transfer in Solution and Proteins. *Int. Rev. Phys. Chem.* **2001**, *20*, 591–616.

(28) Hammes-Schiffer, S.; Soudackov, A. V. Proton-Coupled Electron Transfer in Solution, Proteins, and Electrochemistry extdagger. *J. Phys. Chem. B* **2008**, *112*, 14108–14123.

(29) Kamerlin, S. C. L.; Warshel, A. The EVB as a Quantitative Tool for Formulating Simulations and Analyzing Biological and Chemical Reactions. *Faraday Discuss.* **2009**, *145*, 71–106.

(30) Kazemi, M.; Himo, F.; Åqvist, J. Enzyme Catalysis by Entropy Without Circe Effect. *Proc. Natl. Acad. Sci. U S A* **2016**, *113*, 2406–11.

(31) Plotnikov, N. V.; Kamerlin, S. C.; Warshel, A. Paradynamics: An Effective and Reliable Model for Ab Initio QM/MM Free-Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115*, 7950–7962.

(32) Hartke, B.; Grimme, S. Reactive Force Fields Made Simple. *Phys Chem Chem Phys* **2015**, *17*, 16715–8.

(33) Stoll, V. S.; Eger, B. T.; Hynes, R. C.; Martichonok, V.; Jones, J. B.; Pai, E. F. Differences in Binding Modes of Enantiomers of 1-Acetamido Boronic Acid Based Protease Inhibitors: Crystal Structures of Γ-Chymotrypsin and Subtilisin Carlsberg Complexes. *Biochemistry* **1998**, *37*, 451–462.

(34) Paradkar, V. M.; Dordick, J. S. Mechanism of Extraction of Chymotrypsin Into Isooctane at Very Low Concentrations of Aerosol OT in the Absence of Reversed Micelles. *Biotech. Bioeng.* **1994**, *43*, 529–540.

(35) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics With NAMD. *J. Comput. Chem.* **2005**, *16*, 1781–1802.

(36) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M. et al. Amber 12. *University of California, San Francisco* **2012**,

(37) MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(38) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(39) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(40) Abel, S.; Sterpone, F.; Bandyopadhyay, S.; Marchi, M. Molecular Modeling and Simulations of AOT-water Reverse Micelles in Isooctane: Structural and Dynamic Properties. *J. Phys. Chem. B* **2004**, *108*, 19458–19466.

(41) Hénin, J.; Shinoda, W.; Klein, M. L. United-Atom Acyl Chains for CHARMM Phospholipids. *J. Phys. Chem. B* **2008**, *112*, 7008–7015.

(42) Ryckaert, J. P. Special Geometrical Constraints in the Molecular Dynamics of Chain Molecules. *Mol. Phys.* **1985**, *55*, 549–556.

(43) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(44) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(45) Otte, N.; Bocola, M.; Thiel, W. Force-Field Parameters for the Simulation of Tetrahedral Intermediates of Serine Hydrolases. *J. Comput. Chem.* **2009**, *30*, 154–162.

(46) Warshel, A.; Sussman, F.; Hwang, J. K. Evaluation of Catalytic Free Energies in Genetically Modified Proteins. *J. Mol. Biol.* **1988**, *201*, 139–159.

(47) Fitzpatrick, P. A.; Steinmetz, A. C.; Ringe, D.; Klibanov, A. M. Enzyme Crystal Structure in a Neat Organic Solvent. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8653–8657.

(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G. J.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian 09, Revision A.02. 2009.

(49) Boese, A. D.; Martin, J. M. L. Development of Density Functionals for Thermochemical Kinetics. *J. Chem. Phys.* **2004**, *121*, 3405–3416.

(50) Thompson, W. H. Mixed Quantumclassical Simulation of Vibrational Frequency Modulations of a Diatomic Molecule in a Rare Gas Fluid. *Chem. Phys. Lett.* **2001**, *350*, 113–118.

(51) Light, J. C.; Hamilton, I. P.; Lill, J. V. Generalized Discrete Variable Approximation in Quantum Mechanics. *J. Chem. Phys.* **1985**, *82*, 1400–1409.

(52) Colbert, D. T.; Miller, W. H. A Novel Discrete Variable Representation for Quantum Mechanical Reactive Scattering via the Smatrix Kohn Method. *J. Chem. Phys.* **1992**, *96*, 1982–1991.

(53) Lanczos, C. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. *J. Res. Nat. Bur. Stand.* **1950**, *45*, 255–282.

(54) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. Molecular Dynamics of a Model SN2 Reaction in Water. *J. Chem. Phys.* **1987**, *86*, 1356–1376.

(55) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events. *Chem. Phys. Lett.* **1989**, *156*, 472–477.

(56) Daggett, V.; Schroeder, S.; Kollman, P. Catalytic Pathway of Serine Proteases: Classical and Quantum Mechanical Calculations. *J. Am. Chem. Soc.* **1991**, *113*, 8926–8935.

(57) Gertner, B. J.; Bergsma, J. P.; Wilson, K. R.; Lee, S.; Hynes, J. T. Nonadiabatic Solvation Model for SN2 Reactions in Polar Solvents. *J. Chem. Phys.* **1987**, *86*, 1377–1386.

(58) Tuñón, I.; Laage, D.; Hynes, J. T. Are There Dynamical Effects in Enzyme Catalysis? Some Thoughts Concerning the Enzymatic Chemical Step. *Arch Biochem Biophys* **2015**, *582*, 42–55.

(59) Warshel, A. Dynamics of Reactions in Polar Solvents. Semiclassical Trajectory Studies of Electron-Transfer and Proton-Transfer Reactions. *J. Phys. Chem.* **1982**, *86*, 2218–2224.

(60) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. The Role of Collective Solvent Coordinates and Nonequilibrium Solvation in Charge-Transfer Reactions. *The Journal of Physical Chemistry B* **2001**, *105*, 9672–9685.

(61) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. Hybrid Ab Initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin. *J. Phys. Chem. B* **1998**, *102*, 2293–2301.

(62) Zhou, Y.; Zhang, Y. Serine Protease Acylation Proceeds With a Subtle Re-Orientation of the Histidine Ring at the Tetrahedral Intermediate. *Chem. Comm.* **2011**, *47*, 1577–1579.

(63) Chang, Y. T.; Miller, W. H. An Empirical Valence Bond Model for Constructing Global Potential Energy Surfaces for Chemical Reactions of Polyatomic Molecular Systems. *J. Phys. Chem* **1990**, *94*, 5884–5888.

(64) Muller, R. P.; Warshel, A. Ab Initio Calculations of Free Energy Barriers for Chemical Reactions in Solution. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(65) Schlegel, H. B.; Sonnenberg, J. L. Empirical Valence-Bond Models for Reactive Potential Energy Surfaces Using Distributed Gaussians. *J Chem Theory Comput* **2006**, *2*, 905–911.

(66) Sonnenberg, J. L.; Wong, K. F.; Voth, G. A.; Schlegel, H. B. Distributed Gaussian Valence Bond Surface Derived From Ab Initio Calculations. *J Chem Theory Comput* **2009**, *5*, 949–961.

(67) Isaksen, G. V.; Hopmann, K. H.; Åqvist, J.; Brandsdal, B. O. Computer Simulations Reveal Substrate Specificity of Glycosidic Bond Cleavage in Native and Mutant Human Purine Nucleoside Phosphorylase. *Biochemistry* **2016**, *55*, 2153–2162.

(68) Kazemi, M.; Åqvist, J. Chemical Reaction Mechanisms in Solution From Brute Force Computational Arrhenius Plots. *Nat. Comm.* **2015**, *6*, 7293.

(69) Isaksen, G. V.; Åqvist, J.; Brandsdal, B. O. Thermodynamics of the Purine Nucleoside Phosphorylase Reaction Revealed by Computer Simulations. *Biochemistry* **2017**, *56*, 306–312.

(70) Plotnikov, N. V.; Warshel, A. Exploring, Refining, and Validating the Paradynamics QM/MM Sampling. *J. Phys. Chem. B* **2012**, *116*, 10342–10356.

(71) Lameira, J.; Kupchencko, I.; Warshel, A. Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J. Phys. Chem. B* **2016**, *120*, 2155–2164.

(72) Strajbl, M.; Hong, G.; Warshel, A. Ab Initio QM/MM Simulation With Proper Sampling:first Principle Calculations of the Free Energy of the Autodissociation of Water in Aqueous Solution. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.

(73) Bennett, C. H. Efficient Estimation of Free Energy Differences From Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(74) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

(75) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations With Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.

(76) Cramer, C. J.; Truhlar, D. G. A Universal Approach to Solvation Modeling. *Acc Chem Res* **2008**, *41*, 760–8.

(77) Lameira, J.; Kupchencko, I.; Warshel, A. Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J Phys Chem B* **2016**, *120*, 2155–64.

(78) Duarte, F.; Amrein, B. A.; Blaha-Nelson, D.; Kamerlin, S. C. Recent Advances in QM/MM Free Energy Calculations Using Reference Potentials. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850*, 954–965.

(79) Marcus, R. A. Chemical and Electrochemical Electron-Transfer Theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155–196.

(80) Ando, K.; Hynes, J. T. Molecular Mechanism of HCl Acid Ionization in Water: Ab Initio Potential Energy Surfaces and Monte Carlo Simulations. *J. Phys. Chem. B* **1997**, *101*, 10464–10478.

(81) Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer, 2007; Vol. 86.

(82) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(83) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(84) Zhu, F.; Hummer, G. Convergence and Error Estimation in Free Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2012**, *33*, 453–465.

(85) Chakravorty, D. K.; Kumarasiri, M.; Soudackov, A. V.; Hammes-Schiffer, S. Implementation of Umbrella Integration Within the Framework of the Empirical Valence Bond Approach. *J. Chem. Theory Comput.* **2008**, *4*, 1974–1980.

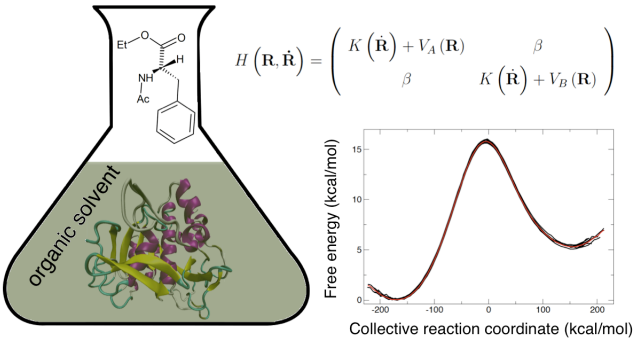
(86) Malany, S.; Sawai, M.; Sikorski, R. S.; Seravalli, J.; Quinn, D. M.; Radić, Z.; Taylor, P.; Kronman, C.; Velan, B.; Shafferman, A. Transition State Structure and Rate Determination for the Acylation Stage of Acetylcholinesterase Catalyzed Hydrolysis of (Acetylthio)choline. *J. Am. Chem. Soc.* **2000**, *122*, 2981–2987.

(87) Fattebert, J.-L.; Lau, E. Y.; Bennion, B. J.; Huang, P.; Lightstone, F. C. Large-Scale First-Principles Molecular Dynamics Simulations With Electrostatic Embedding: Application to Acetylcholinesterase Catalysis. *J. Chem. Theory Comput.* **2015**, *11*, 5688–5695.

(88) Truhlar, D. G.; Hase, W. L.; Hynes, J. T. Current Status of Transition-State Theory. *J. Phys. Chem.* **1983**, *87*, 2664–2682.

(89) Kiefer, P.; Hynes, J. T. Adiabatic and Nonadiabatic Proton Transfer Rate Constants in Solution. *Solid State Ionics* **2004**, *168*, 219–224.

TOC Graphic


Et
O
O
H
NH
Ac
organic solvent
$H\left(\mathbf{R},\dot{\mathbf{R}}\right) = \begin{pmatrix} K\left(\dot{\mathbf{R}}\right) + V_A\left(\mathbf{R}\right) & \beta \\ \beta & K\left(\dot{\mathbf{R}}\right) + V_B\left(\mathbf{R}\right) \end{pmatrix}$
Free energy (kcal/mol)
15
10
5
0
-200
-100
0
100
200
Collective reaction coordinate (kcal/mol)

Coupled Valence-Bond State Molecular Dynamics Description of an Enzyme-Catalyzed Reaction in a Non-Aqueous Organic Solvent

Supporting Information

Elise Duboué-Dijon,[†,¶] Eva Pluhařová,[†,¶] Dominik Domin,[†] Kakali Sen,[†] Aoife C. Fogarty,[†,‡] Nicolas Chéron,[†] and Damien Laage[*,†]

†*École Normale Supérieure - PSL Research University, Département de Chimie, Sorbonne Universités - UPMC Univ Paris 06, CNRS UMR 8640 PASTEUR, 24 rue Lhomond, 75005 Paris, France*

‡*Present address: Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

¶*Contributed equally to this work*

E-mail: damien.laage@ens.fr

Phone: +33 (0)144322418

System preparation

Charges and atom types for the APEE substrate and for the tetrahedral intermediate are summarized in Table S1, using the atom labels defined in Fig. S1. Atoms of Ser221 that are not explicitly assigned new charges for the tetrahedral intermediate bear the same charge as in the standard CHARMM serine residue. The charges and parameters for APEE were taken from CHARMM22 database by analogy, utilizing the transferability property of this forcefield. In the product state, Ser221 forms a covalent bond leading to formation of a tetrahedral intermediate. A set of parameters was determined in ref 1 for the tetrahedral intermediate for serine hydrolases that share the same catalytic triad: esterases, proteases and lipases. The tetrahedral intermediate was modeled as $(CH_3\text{-}O)_2\text{-}CO\text{-}CH_3$, a transferable core among all the members of this family. The H-atoms of the $\text{-}O\text{-}CH_3$ group are exchanged with the specific ester substrate used in the study. We adopted the same strategy for our study. We used the parameters and charges directly as described in ref 1 for the transferable core unit of the tetrahedral intermediate. We adjusted the charges of the exchangeable ester groups by analogy to similar groups in CHARMM database. Most of the groups maintained the same charges as used for APEE in VB state A. The charges of the immediate adjacent groups to the TI were adjusted manually. While adjusting the charges, the charges of the aliphatic H atoms were maintained to 0.09 and the sum of groups added up to zero or integral values.

Charges for the deprotonated serine used for equilibrations of the system around the potential intermediate (state C) are summarized in Table S2. They should not be regarded as an optimized force field for deprotonated serine, since they were only meant to reasonably describe the change in charges between protonated and deprotonated serine. They were only used to equilibrate the environment around the putative intermediate (state C) – after proton transfer and before nucleophilic attach – and test the stability of this state when the environment favors it. The charges were obtained by manually distributing the -1 charge on the CB, HB1, HB2 and OG atoms, without touching backbone charges.

Table S1: Charges and atom types for the APEE substrate and for the tetrahedral intermediate.

	APEE (state A)		Tetrahedral Intermediate (state B)	
Atom Name	Atom type	Charge	Atom type	Charge
CB			CIL2	-0.18
HB1			HA	0.09
HB2			HA	0.09
OG			OS	-0.4
CG	CA	0	CA	0
CD1	CA	-0.115	CA	-0.115
HD1	HP	0.115	HP	0.115
CE1	CA	-0.115	CA	-0.115
HE1	HP	0.115	HP	0.115
CZ	CA	-0.115	CA	-0.115
HZ	HP	0.115	HP	0.115
CD2	CA	-0.115	CA	-0.115
HD2	HP	0.115	HP	0.115
CE2	CA	-0.115	CA	-0.115
HE2	HP	0.115	HE2	0.115
N	NH1	-0.47	NH1	-0.47
HN	H	0.31	H	0.31
CA3	CT1	0.17	CIL3	0.07
HA4	HA	0.09	HA	0.09
CB3	CT2	-0.18	CT2	-0.18
HB3	HA	0.09	HA	0.09
HB4	HA	0.09	HA	0.09
CA1	C	0.51	C	0.51
OA	O	-0.51	O	-0.51
CA2	CT3	-0.27	CT3	-0.27
HA1	HA	0.09	HA	0.09
HA2	HA	0.09	HA	0.09
HA3	HA	0.09	HA	0.09
C	CD	0.63	CTET	0.55
O	OB	-0.52	OC	-0.75
OS	OS	-0.34	OS	-0.4
CS1	CT2	-0.05	CIL2	-0.18
HS1	HA	0.09	HA	0.09
HS2	HA	0.09	HA	0.09
CS2	CT3	-0.27	CT3	-0.27
HS3	HA	0.09	HA	0.09
HS4	HA	0.09	HA	0.09
HS5	HA	0.09	HA	0.09

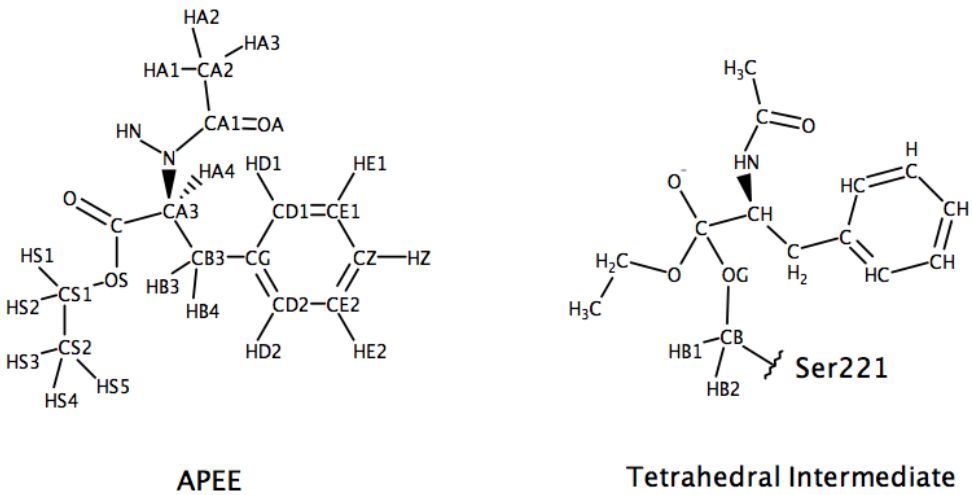


Figure S1: N-acetyl-L-phenylalanine-ethylester (APEE) and tetrahedral intermediate structures, with definition of the atom labels used in Table S1. Tetrahedral intermediate atoms already present in APEE are assigned the same atom names.

Table S2: Charges and atom types for the deprotonated serine in state C.

Atom Name	Atom type	Charge
N	NH1	-0.47
HN	H	0.31
CA	CT1	0.07
HA	HB	0.09
CB	CT2	-0.18
HB1	HA	0.04
HB2	HA	0.04
OG	OH1	-0.9
C	C	0.51
O	O	-0.51

Definition of the QM region used in QM/MM calculations.

The QM region used in mixed quantum/classical calculations (QM/MM) is composed of Asp32, His64, Ser125, Asn155, Ser221 active-site residues and the APEE substrate. Two different boundaries were compared in order to ensure that the location of the boundary

does not affect our results (see Figs S2-S3).

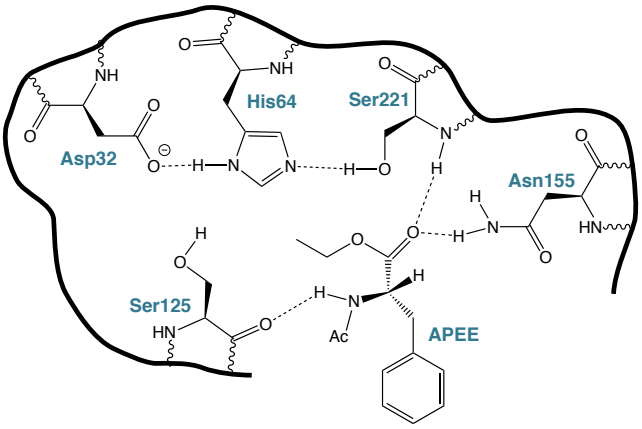


Figure S2: Definition of the QM region used in QM/MM calculations for the optimization of VB state parameters and in the proton potentials in fig 7.

The great similarity between the two-dimensional potential energy surfaces obtained for the same enzyme-substrate configurations but with the two different QM/MM boundaries (Fig 3 with the boundary in fig S3 and fig S4 with the boundary shown in fig S2) shows that the QM/MM boundary is sufficiently far from the place where the reaction occurs and does not affect our results.

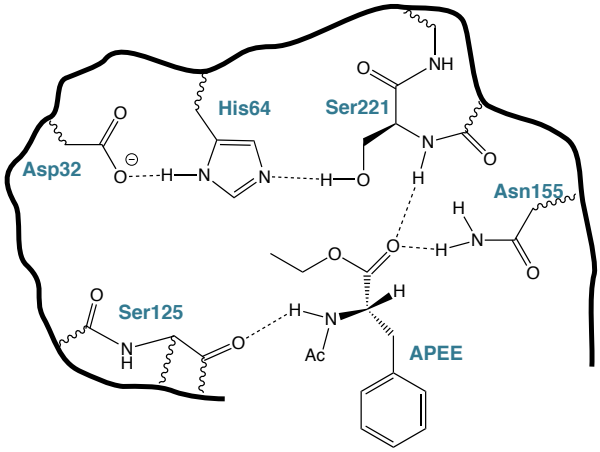


Figure S3: Definition of the QM region with the QM-MM boundary passing only through C-C bonds.

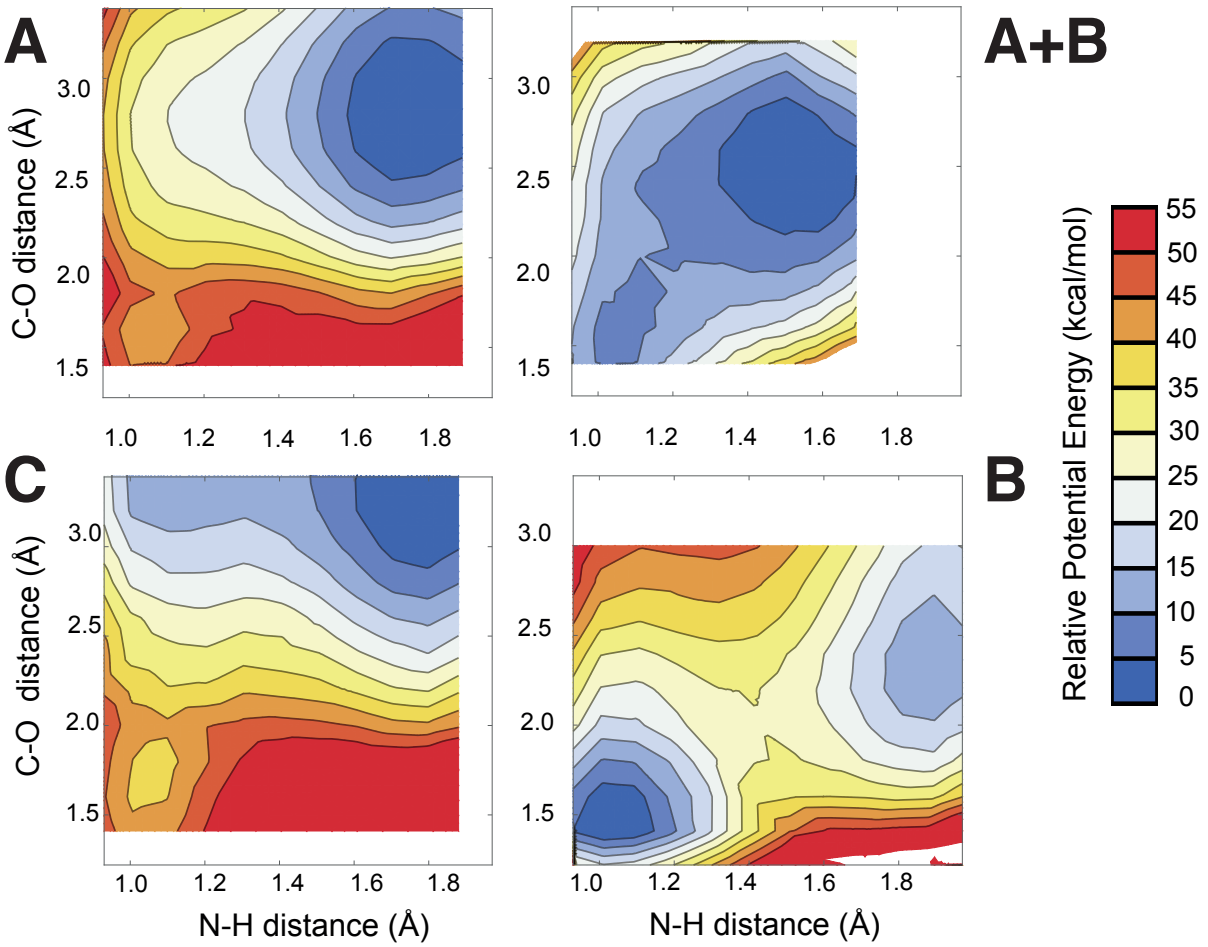


Figure S4: 2D potential energy surfaces along the proton transfer coordinate N-H and nucleophilic attack coordinate C-O in a series of frozen environment configurations respectively equilibrated around the active site and substrate in the reactant-like state A ($\mathbf{E}_A$), the product-like state B ($\mathbf{E}_B$), the putative proton transfer intermediate state C ($\mathbf{E}_C$) and an equal mixture of states A and B ($\mathbf{E}_{AB}$) (see states in Figure 2). Long N-H and C-O distances (top right-hand corner of each plot) correspond to the substrate reactant geometry, short N-H and C-O distances (bottom left-hand corner of each plot) to the product geometry. QM/MM calculations were performed with the boundary defined in fig S2.

Morse potentials in VB states

In the two VB state Hamiltonian, the Ser221 O-H bond (A state) and the His64 N-H and Ser-APEE C-O bonds (B state) are described by a Morse potential $V_{Morse}(r) = D_e\left[1 - \exp\left(-a(r - r_e)\right)\right]^2$. Each Morse potential is fitted on the potential energy profile for the elongation of the bond, as determined by gas-phase DFT calculations (see Methods section). Figure S5 shows for

each one of the three bonds the very good agreement between the elongation potential energy profile and the Morse potential fit.

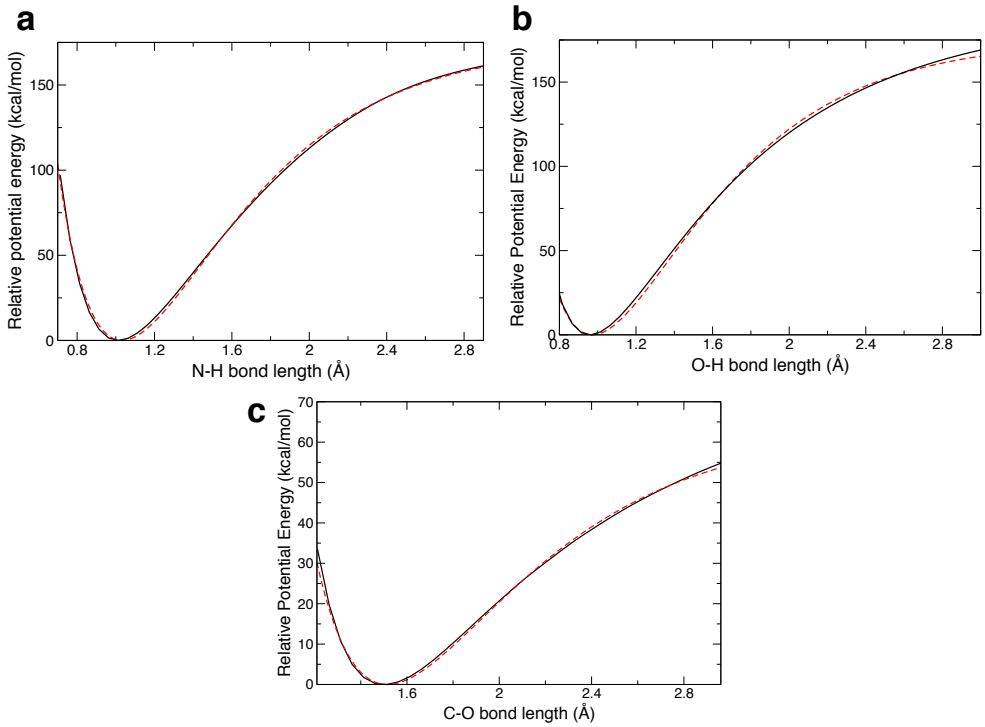


Figure S5: Bond elongation potential energy curves for the protonated histidine N-H (a), the serine O-H (b) and the tetrahedral intermediate C-O (c) bonds obtained at the BMK/6-311+G** level of theory, respectively in the gas phase (black), together with the fit used for the final determination of the Morse potentials (red dashes).

We further compare the bond elongation potential energy profiles obtained with our coupled VB Hamiltonian with those obtained with a QM/MM description of the system, the active site residues being treated at the BMK/6-311+G** level, as described in the methodology section. Figure S6 shows the comparison of the Serine O-H bond elongation profiles at the QM/MM and coupled VB level for two distinct reactant conformations and of the Histidine N-H bond elongation profiles at the QM/MM and coupled VB level for two distinct product conformations. We show that the QM/MM bond elongation profile strongly depends on the chosen conformation, which makes it unpractical to calibrate the Morse potential on QM/MM elongation profiles. We also note that around the equilibrium geometry

(bottom of the well) the QM/MM elongation profiles are reasonably well reproduced with our coupled VB description. Stronger deviations occur at larger distances, which are probably due to the presence of residues around the elongated bond that can stabilize the proton. Such effects are not captured by our coupled VB description.

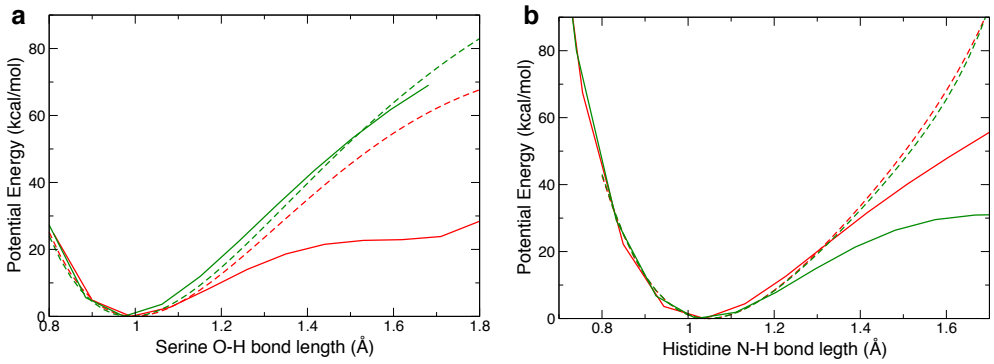


Figure S6: Bond elongation potential energy curves for the the serine O-H (a) and the protonated histidine N-H (b) bonds obtained at the QM/MM BMK/6-311+G**/CHARMM level of theory (plain lines) and with the coupled VB Hamiltonian (dashes), respectively on two distinct reactant and product conformations as denoted by two different colors.

Convergence of VB Hamiltonian parameters

Figure S7a gives the evolution of the optimized VB parameters ΔV^0 and β when the number of configurations in the reactant, transition and product states increases. It clearly shows that the values are well converged. This is further supported by the evolution of the difference between the values obtained from the QM/MM calculations and from the coupled VB description respectively for the reaction free energy ΔG^{rxn} and for the reaction free energy barrier $\Delta G^{\ddagger}$ in fig S7b, which shows that the residual error on these free energies is less than 0.06 kcal/mol.

Since our goal is to construct a VB description that correctly describes the reaction kinetics and thermodynamics, we accordingly optimized the VB parameters to reproduce the ΔG^{rxn} and $\Delta G^{\ddagger}$ free energies. The root mean square error on the potential energies of

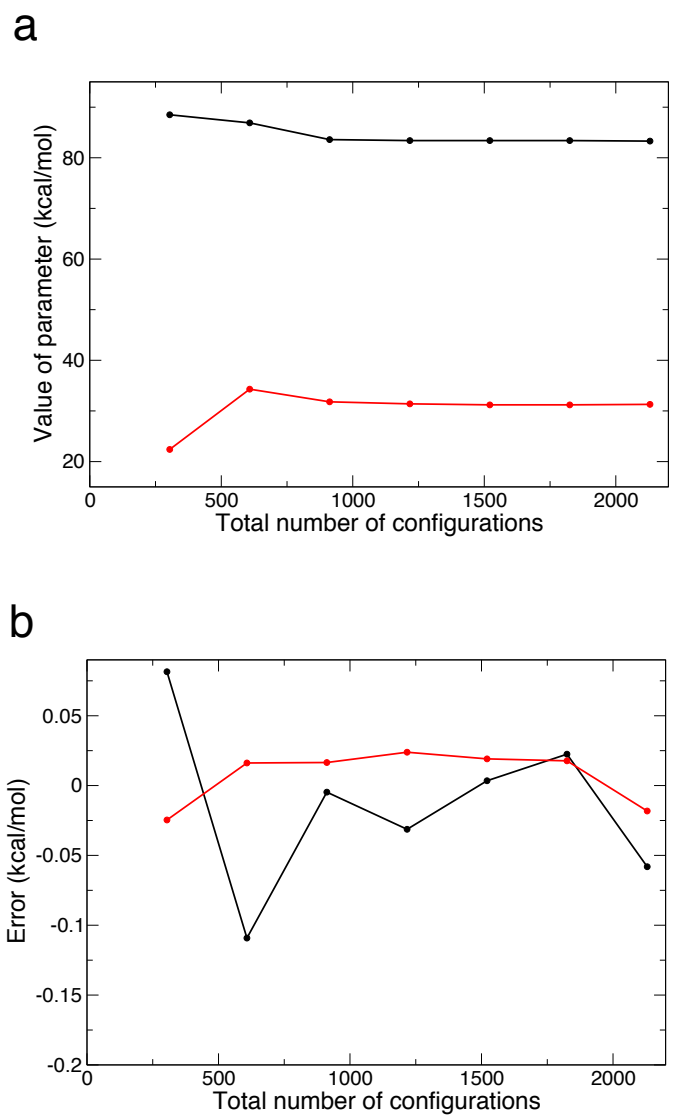


Figure S7: a) Convergence of the ΔV^0 (black) and β (red) parameters of the VB Hamiltonian with the total number of reactant, transition state and product configurations used for their determination. b) Differences between the values of the reaction free energy $\Delta\Delta G^{\mathrm{rxn}}$ (black) and free energy barrier $\Delta\Delta G^{\ddagger}$ (red) eq 4 obtained from the QM/MM calculation and from the coupled VB description, as a function of the total number of reactant, transition state and product configurations used to determine the coupled VB Hamiltonian parameters.

all configurations used to parameterize the VB Hamiltonian is less relevant, since it does not include the appropriate weighting to emphasize the importance of low-energy configurations. However, fig S8 shows that our VB description displays an excellent correlation with the QM/MM energies for all reactant, transition and product state configurations. The resulting

potential energy root mean square errors are respectively 11.8, 10.1 and 10.5 kcal/mol for the reactant, transition and product state configurations. The rms error is thus below 0.10 kcal/mol per atom, which compares very well with the 0.12-0.19 kcal/mol/atom errors recently obtained with state-of-the-art neural network approaches.[2]

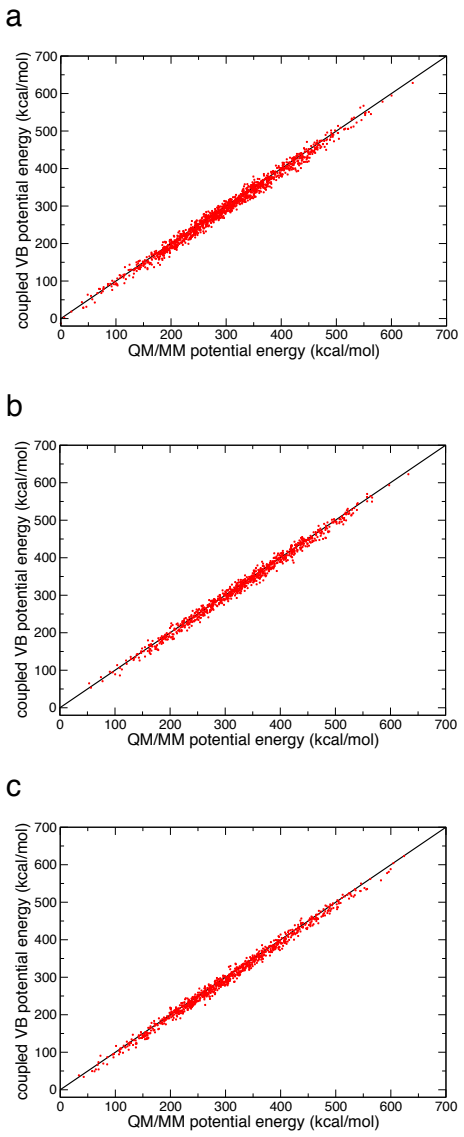


Figure S8: Correlation between the QM/MM and coupled VB potential energies for the reactant (a), transition state (b) and product (c) configurations used to determine the VB Hamiltonian parameters.

Electrostatic decomposition analysis.

We provide here the details of the analysis of the molecular groups governing the change in the electrostatic part of the reaction coordinate ΔE^{elec} between reactants (R) and products (P), $\Delta\Delta E^{\mathrm{elec}}$.

$$\Delta\Delta E^{\mathrm{elec}} = \Delta E^{\mathrm{elec}}\left(\mathbf{P}\right) - \Delta E^{\mathrm{elec}}\left(\mathbf{R}\right) \ . \tag{1}$$

The analysis is performed on snapshots extracted from the mapping potential windows used for the PMF calculations. Snapshots taken from the 8 windows generated with a mapping potential $0.0 \leq \lambda \leq 0.05$ are considered as "reactant" geometries , while "product" snapshots are taken from the 8 windows generated with a mapping potential $0.95 \leq \lambda \leq 1.0$. The numbers provided for the electrostatic decomposition are averaged over all the snapshots coming from the selected windows.

The system is divided between the reactive region (composed of the reactive residues His64, Ser221 and the APEE), and the different components of the environment: rest of the enzyme, water, ions, surfactants (AOTs). Table S3 provides the results of this analysis.

Table S3: Decomposition of $\Delta\Delta E^{\text{elec}}$ between different regions of the system: the reactive region (APEE, Ser221, His64, and the transferred H which is counted separately from the two residues Ser221, His64), the rest of the enzyme (labeled "enzyme" for simplicity), the water molecules, the AOT surfactants and the sodium ions.

Regions considered for the electrostatic interaction		Contribution to $\Delta\Delta E^{\text{elec}}$ (kcal/mol)
reactive region	**Full Environment**	**20.9**
reactive region	**water**	**14.1**
His64	water	7.25
Ser221	water	-0.05
transferred H	water	0.0
APEE	water	6.9
reactive region	**enzyme**	**6.85**
His64	enzyme *(incl. Asp32)*	3.4 *(2.6)*
Ser221	enzyme	-0.6
transferred H	enzyme	0.05
APEE	enzyme *(incl. anionic hole)*	4.0 *(1.6)*
reactive region	**AOT**	**-0.3**
reactive region	**sodium ions**	**0.2**

References

(1) Otte, N.; Bocola, M.; Thiel, W. Force-Field Parameters for the Simulation of Tetrahedral Intermediates of Serine Hydrolases. *J. Comput. Chem.* **2009**, *30*, 154–162.

(2) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations With Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.