Solvent Coordinates and Free Energy Surfaces in an Enzymatic S_N2 Reaction. Elucidating the Role of Protein Structure, Flexibility and Dynamics

Rafael García-Meseguer[‡], Sergio Martí[§], J. Javier Ruiz-Pernía[§], Vicente Moliner[§], Iñaki Tuñón[‡*]*

Departament de Química Física, Universitat de València, 46100 Burjassot, (Spain);
Departament de Química Física i Analítica; Universitat Jaume I, 12071 Castellón (Spain)

‡Universitat de València
§ Universitat Jaume I
*to whom correspondence should be addressed

ignacio.tunon@uv.es

pernia@qfa.uji.es

Abstract

Conformational changes drive the protein through the catalytic cycle allowing, for example, substrate binding or product release. However, the impact of protein motions in the chemical step is a very controversial issue. It has been proposed that Transition State Theory could be insufficient to account for the catalytic effect of enzymes and that protein motions should be dynamically described. We propose the use of free energy surfaces, obtained as a function of a chemical and an environmental coordinate, as an efficient way to elucidate the role of protein structure and motions during the reaction. This study shows that the protein structure provides the adequate environment for the progress of the reaction, although certain degree of flexibility is needed to attain the full catalytic effect. However, these motions do not introduce significant dynamical corrections to the rate constant and can be described as equilibrium fluctuations.

One of the most intriguing characteristics of enzymes is their flexibility. It has been stressed that, to function, enzymes must be stable enough to retain their three-dimensional structure but flexible enough to permit the evolution of the protein among the different conformations relevant at each step of the full catalytic process.[1,2] For example, conformational changes are known to be a requisite for substrate binding and product release in many cases. In fact, the multidimensional free energy surface corresponding to an enzymatic process is very rugged, containing multiple minima that appear along the multiple conformational coordinates available for the protein.[3] Transitions between different conformational substrates of the macromolecule and their population distribution are governed by motions happening in a broad range of timescales, covering from milliseconds to femtoseconds.[2,4]

However, the impact of protein flexibility on the rate constant of the chemical step remains as the subject of a long-standing debate in scientific literature.[5-15] Even if the active site is designed to catalyze the reaction and then to accommodate the charge distribution of the transition state (TS),[16] some protein motions are still needed to evolve from the reactants state (RS) to the TS.[1,5,15,17] In other words, certain degree of protein reorganization is needed to attain the TS.[15] This reorganization would play a similar role to solvent polarization in Marcus theory for electron transfer in solution,[18] although much slower conformational components could be involved in enzymatic catalysis.[12,19] In general, the coordinate that describes the transformation of the system from reactants to products is a collective coordinate that involves the degrees of freedom not only of the solute/substrate but also of the environment (the solvent and/or the enzyme).

Theoretical descriptions of chemical reactions in enzymatic environments are usually based in the selection of a distinguished reaction coordinate, typically defined in terms of the substrate or solute coordinates (i. e. some valence coordinates related to the bonds to be broken and/or formed). Within this description, some protein motions can be coupled to the advance of the system along the reaction coordinate. For example, correlated motions within the protein have been found to promote tunneling in hydride transfer reaction[20,21] and compressive local motions within the active site facilitate the approach between the donor and the acceptor atoms in transfer reactions.[11,22] The participation of protein motions in the reaction progress is thus a well-established fact that has been invoked using different terminology: protein reorganization,[15] coupled motions[9] or promoting vibrations.[8,22] A more controversial question is the way in which these motions must be described and if the current theoretical frameworks are adequate or not to explain the rate of enzymatic reactions.[14,22]

As a fundamental approach to describe the reaction rate of chemical reactions, Transition State Theory (TST) provides the tools for analyzing also the rate of enzymatic reactions.[23] The basic assumption in TST is that the selected reaction coordinate (ζ) is separable from the rest of coordinates of the system in such a way that the averaged dynamical behavior of the trajectories evolving from reactants to products can be represented as the equilibrium flux across the dividing surface. In this case the rate constant can be simply related to the free energy difference between the TS and the reactants ($\Delta G^{\ddagger}$). For a unimolecular process one has:

$$k_r = \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^{\ddagger}(\zeta)}{RT}} \qquad (1)$$

where k_B, h and R are the Boltzmann, Planck and gas constants, and T the absolute temperature. This approach means that the remaining degrees of freedom can be considered at equilibrium at any value of the reaction coordinate and then described as a Boltzmann distribution. If the reaction coordinate is defined exclusively in terms of the degrees of freedom of the substrate (or 'solute coordinate') then the protein degrees of freedom (or 'solvent coordinate'), would be assumed to be at equilibrium at all the stages of the reaction path. Obviously, protein dynamics spans a hierarchy of timescales[2,4,10] and only those motions much faster than the reaction coordinate can be considered at equilibrium. Thus, this approach breaks for those motions coupled to the solute coordinate and taking place in similar or slower timescales. In such a case an explicit treatment of vibrational dynamics in the enzyme could be required. This has been suggested for some cases covering both fast vibrational motions[22] and slow conformational changes.[19] However, other analysis have stressed that the role of protein motions can be satisfactorily incorporated in the description of the chemical step as equilibrium fluctuations and thus explicit dynamical treatments of protein motions would be not really necessary for modeling enzymatic catalysis.[5,9,13,14,24]

TST offers a convenient framework to incorporate non-statistical effects of protein motions. A transmission coefficient (κ) can be calculated from time-dependent simulations and incorporated to the TST expression of the rate constant:[25]

$$k_r = \kappa(\zeta) \cdot \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^{\ddagger}(\zeta)}{RT}} \qquad (2)$$

Since the transmission coefficient takes values lower than unity the equilibrium approach (equation (1)) provides an upper limit to the rate constant. There is no an unambiguous way

to separate the effects of the enzyme appearing on the activation free energy (or equilibrium effects) from those appearing on the transmission coefficient (non-equilibrium effects) because both depend on the choice of the reaction coordinate.[23] If this is defined exclusively on the basis of the coordinates of the substrate, deviations from the equilibrium distribution of protein motions should be reflected in the transmission coefficient. Assuming that these effects are important in the vicinity of the dynamic bottleneck, the transmission coefficient can be evaluated from the frictional force exerted on the reaction coordinate at the TS[26] or by means of rare event simulations that count the number of recrossings across the dividing surface.[27] Up to date, simulations performed on enzymatic reactions have shown that the transmission coefficient deviates only modestly from unity for reasonable choices of the reaction coordinate. Typical values usually range between 0.5-0.9[23,24,28] reflecting a modest participation of protein motions in the passage of the system over the barrier top.

A convenient way to analyze quantitatively the role of protein motions during the chemical transformation is to project the multidimensional free energy surface (FES) of the enzymatic reaction in a two-dimensional model obtained as a function of a solute and a solvent coordinate.[29,30] Such a FES allows estimating, at quantitative level, the timing and coupling between the solute and solvent motions along the reaction path. Furthermore, the transmission coefficient can be derived from the differences between the dividing surface defined on the FES obtained under the equilibrium solvation approximation and the one obtained on the two-dimensional FES (reflecting the coupling between the solute and solvent coordinates).

A prototypical example where solvent effects can play an important role is the S_N2 reaction. Studies in aqueous solution already established that the reaction proceeds with

large rearrangement of water molecules around the nucleophile and the leaving group.[30,31] Thus, the change in the electrostatic potential created by the environment on the nucleophile and the leaving group is a good coordinate to follow the evolution of the environment along the reaction process,[32] while the distances associated to the bonds to be broken and formed provide an adequate solute coordinate. In this work, we have studied the S_N2 nucleophilic reaction between dichloroethane and Asp124 in DhlA, a Haloalkane Dehalogenase from *Xanthobacter Autotrophicus* (see Figure 1),[33,34] an enzymatic reaction that has been theoretically analyzed by several groups.[35-38] We also modeled the counterpart S_N2 reaction in aqueous solution, where a molecule of acetate is employed as the nucleophile. The comparison between the enzymatic and in solution process offers an excellent opportunity to analyze the role of structure, flexibility and dynamics of the environment on a fundamental class of chemical reactions.

Results

The PM3/MM FESs corresponding to the nucleophilic attack in aqueous solution and in the active site of DhlA traced as a function of the solute (ξ) and solvent (s) coordinates are presented in Figure 2. The solute coordinate (ξ=d(CCl)-d(CO)) evolves from negative to positive values as the reaction proceeds. The solvent coordinate is obtained from the electrostatic potential created by the environment on the leaving group and the nucleophile ($s=V_{Cl}-V_O$; see Methods) and evolves from negative values at the RS (where the nucleophile atom bearing the negative charge is stabilized by electrostatic interactions with the environment) to positive values at the product state (where the electrostatic potential takes larger positive values on the leaving group that now is negatively charged). This

figure also displays the minimum free energy paths (MFEPs), obtained from the free energy gradient, as continuous lines on the FESs. The free energy paths obtained using the solute coordinate and assuming that the solvent coordinate equilibrates at each value of the former (equilibrium free energy paths or EFEPs) are also shown on Figure 2 as dashed lines.

Analysis of Free Energy Surfaces

The free energy differences between the saddle points and the reactant minima located on the FESs are 27.4 and 37.1 kcal·mol^{-1} for the catalyzed and uncatalyzed processes, respectively. The activation free energy deduced from the experimental rate constant of the enzymatic reaction at 298 K is 15.3 kcal·mol^{-1},[34] while the barrier estimated for the process in aqueous solution is 26 kcal·mol^{-1} [36] (a value of 29.9 kcal·mol^{-1} has been reported for the reaction in solution at 373K).[35,39] The overestimation observed in our theoretical values is due to the use of the PM3 hamiltonian.[35,37] The M06-2X-corrected free energy barriers (see Methods) for the enzymatic and in solution processes are 16.5 and 27.4 kcal·mol^{-1} respectively, in better quantitative agreement with the experimental values. In any case the PM3/MM calculations provide a correct estimation of the catalytic effect, defined as the difference between the in solution and the enzymatic free energy barriers. The PM3/MM difference is 9.7 kcal·mol^{-1}, in good agreement with the difference derived from the M06-2X values, 10.9 kcal·mol^{-1}, and from the experimental values, 10.7 kcal·mol^{-1}.

The FESs show noticeable differences between the reaction in solution and in the enzyme. At the RS the protein structure provides a much more adequate environment for the progress of the reaction than the solution. In solution, the RS is found at a value of the solvent coordinate of about -100 kcal·mol^{-1}·|e|$^{-1}$, while the enzymatic RS is found at about -25 kcal·mol^{-1}·|e|$^{-1}$. This latter value of the solvent coordinate is much closer to the value

needed to reach the TS ($s \approx -10$ kcal·mol^{-1}·|e|$^{-1}$ in both environments). At the Michaelis complex the protein is already organized from the electrostatic point of view to favor the reaction, while in aqueous solution the environment needs to be largely reorganized to facilitate the reaction.[30,31,38]

Nevertheless, the protein structure does not behave as a rigid scaffold where the reaction takes place. The reaction would be significantly more difficult in a frozen-protein environment where the solvent coordinate remains unchanged from RS to TS. A straight line from the enzymatic RS at a constant value of the solvent coordinate represents the rigid-environment path. This path (white arrow in Figure 2) shows a PM3/MM free energy barrier of 31.5 kcal·mol^{-1}; ~4 kcal·mol^{-1} higher than the value observed for the flexible protein. The rate constant corresponding to a hypothetical rigid protein would be about 10^3 times smaller, at room temperature, than in the real enzyme. This means that protein flexibility plays a role in catalysis and that the environment needs to be rearranged when going from the Michaelis complex to the TS in order to reach a maximum reduction in the activation free energy.

Flexibility can be quantified by means of the force constant associated to the solvent coordinate. The force constants obtained from a parabolic fit of the free energy change along the solvent coordinate in solution and in DhlA are given in Table 1. The force constant obtained in the enzyme is about 4.2 times larger than that obtained in solution. The protein structure is stiffer than the structure of water, which is related to the existence of a network of covalent bonds in the former. However, it must be pointed out that, as stated above, the change needed in the solvent coordinate to reach the TS from the RS is much smaller in the enzyme than in solution. The final result is that the work to be done on the

solvent coordinate to reach the TS is significantly smaller in the enzyme than in solution. According to the MFEP traced on the FESs, the free energy difference between the TS and the RS can be approximately written as:

$$\Delta G_i^{\ddagger} \approx \Delta G_{s,i}(s_i^{RS} \rightarrow s_i^{\ddagger}; \xi_i^{RS}) + \Delta G_{\xi,i}(\xi_{enz}^{RS} \rightarrow \xi_{enz}^{\ddagger}; s_i^{\ddagger}) \qquad (3)$$

where i stands for the environment (enzyme or the water solution). The first term of the right hand side of eq. 3 represents the work to be done on the solvent coordinate and the second one the work to be done on the solute coordinate. The values corresponding to the solvent coordinate are about 11 and 3 kcal·mol^{-1}, in aqueous solution and in DhlA, respectively. Thus, the free energy cost associated to the change along the solvent coordinate in the enzyme is substantially smaller than in solution and this difference represents 80% of the catalytic effect. It is interesting to note also that, in the enzyme, the probability of sampling configurations of the environment conductive to the reaction ($s^{\ddagger}$) is much larger than the probability of sampling adequate values of the solute coordinate ($\xi^{\ddagger}$). The work associated to the reorganization of the environment in the enzymatic process represents only 11% of the total free energy barrier.

Another important aspect to be analyzed is the timing between the solute and solvent motions. According to the MFEPs obtained in the two environments, the change in the solvent coordinate precedes the change in the solute coordinate. From a dynamic point of view this means that the solvent coordinate is slower than the solute coordinate. A significant difference in the time scales of these two motions could result in important dynamical effects due the delay between them. In order to characterize the time evolution of the environment we computed the characteristic frequencies associated to the motion along the solvent coordinate in the two environments using the force constants and the

effective masses deduced from the equipartition principle. These frequencies are provided in Table 1. As observed, not only the force constant but also the effective mass associated to the solvent coordinate is larger in the enzyme than in solution. Both effects cancel out and, as a result, the frequency associated to the motion along the solvent coordinate in the enzyme (410 cm^{-1}) is very similar to the value obtained in solution (480 cm^{-1}). These values essentially correspond to the reorientation of hydrogen bond donors around the nucleophile and the leaving group; motions that take place in the picosecond time scale or faster.[38] So, from the dynamical point of view there are no significant differences in the participation of the environment during the reaction progress in aqueous solution or in DhlA. According to the MFEPs, the motions along the solvent coordinate, breaking of hydrogen bonds to the nucleophile and forming of new ones to the leaving group, take place before or after the motion along the solute coordinate and the timescales associated to these movements are very similar in the two media. Interestingly we didn't find any evidence of slow conformational motions of the protein affecting the chemical step. These motions can obviously exist but either they do not have consequences on the electrostatic coordinate (and then on the energetics of the reaction) or they do not happen in the close neighborhood of the Michaelis complex in the free energy landscape.

Once analyzed the results obtained using explicitly the solute and the solvent coordinates, we can compare them with those obtained assuming equilibrium solvation at any value of the solute coordinate. This is the usual approximation employed to analyze chemical reactions in condensed environments. The EFEPs, presented also in Figure 2, go from the reactant to the product minima through the saddle point. Then, due to the fact that the free energy is a state function, one-dimensional profiles traced along the solute coordinate provide almost the same activation free energies than two-dimensional surfaces obtained as

a function of the solute and solvent coordinates (the origin of small differences is discussed below). However, while the free energy differences are correct, the equilibrium solvation approach is unable to describe properly the timing between the solute and solvent coordinates. Effectively, as observed in Figure 2, in the equilibrium treatment one pulls along the solute coordinate and the solvent coordinate abruptly changes in the vicinities of the TS, while in the MFEPs, both in solution and in the enzyme, solvent motions precede the changes along the solute coordinate. In any case, this limitation does not affect the estimation of the reaction rate constant, because this is mostly determined by the free energy difference between the TS and the RS.

Evaluation of the transmission coefficient

While the MFEPs and the EFEPs coincide at the RS and the TS, there is a small difference in the activation free energies estimated from the two-dimensional (or non-equilibrium) and the one-dimensional (or equilibrium) treatments. The origin of this difference is in the definition of the TS ensemble obtained in each treatment. As shown in Figure 3, the dividing surface in the two-dimensional description contains the saddle point and goes through the ridges that separate the reactants and products valleys. In the one-dimensional treatment only the solute coordinate is employed and thus the dividing surface is defined just as $\xi=\xi^{\ddagger}$ (represented as dashed lines in Figure 3). The difference corresponds to a rotation of the dividing surface in variational TST.[40] The TS well along the non-equilibrium dividing surface is narrower than the TS well along the equilibrium dividing surface (see Figure 3c) and then the frequencies associated to the motion of the TS along the former

dividing surface ($\nu_{s,\xi}$) are larger than for the latter (ν_{ξ}). This can be translated into a entropic difference in the TS ensembles and consequently in the activation free energies:[29]

$$\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}(s,\xi) - \Delta G^{\ddagger}(\xi) = RT\, ln\frac{\nu_{s,\xi}}{\nu_{\xi}} \qquad (4)$$

From our FESs we estimated that the ratio between frequencies are about 1.3 and 1.1 and using these values we estimate that $\Delta\Delta G^{\ddagger}$ at 298K is about 0.2 and 0.1 kcal·mol^{-1}, in solution and in the enzyme, respectively. These free energy differences are quite small, below the statistical uncertainty of typical free energy simulations, and thus non-equilibrium effects make a very small contribution to the activation free energies.

Obviously, a smaller activation free energy is translated into a larger rate constant. This overestimation should be then compensated by the consideration of a transmission coefficient smaller than unity (see equation (2)). The origin of this value can be understood considering the dividing surfaces obtained in the non-equilibrium and the equilibrium descriptions. The dividing surface obtained from the non-equilibrium treatment is defined in such a way that any trajectory arriving to that surface from the reactant side will continue to the products region because the free energy continuously decrease in that direction and then the transmission coefficient is equal to unity for that surface. However, using the equilibrium dividing surface, some trajectories going from reactants to products find a free energy barrier after crossing this surface and they could return to the reactant side (see Figure 3a). This means that in this description the transmission coefficient would be lower than unity. Obviously, using equation (2), the final estimation of the rate constant obtained from the non-equilibrium and the equilibrium approaches would be the same if the transmission coefficient is obtained as:[29]

$$\kappa = e^{\frac{\Delta\Delta G^{\ddagger}}{RT}} \qquad (5)$$

Using the free energy differences given above the transmission coefficients for the reaction in solution and in the enzyme obtained using the solute coordinate as the distinguished reaction coordinate are 0.8 and 0.9, respectively. Obviously this procedure leads to a very crude estimation of the transmission coefficient because of the statistical errors associated to the free energies and due to the non-explicit treatment of all the degrees of freedom. A more accurate estimation shows that the transmission coefficients for this reaction are about 0.6 and 0.8 in aqueous solution and in the enzyme, respectively.[38] Note that the transmission coefficients, evaluated at the TS, depend on the participation of the solvent coordinate in the barrier crossing event and this is quite small for this reaction. However, it must be stressed again that this does not mean that the protein remains unchanged during the chemical step. Protein motions occur first facilitating the motion along the solute coordinate; but at the TS they can be considered to be in equilibrium. In other words, while reactions involve protein motions coupled to the chemical transformation, the probability that these motions take the system to the TS is determined mainly by the activation free energy.[1,5,15]

Discussion

The use of an explicit solvent coordinate can be a useful strategy to dissect the role of structure, flexibility and dynamics in catalysis. Comparison of the FESs obtained in solution and in the enzyme for a prototypic S_N2 reaction shows that the origin of the catalytic efficiency of the enzyme is due to a protein structure that provides the adequate

environment for the progress of the reaction. However, the protein is not completely rigid and some motions contribute to reduce the free energy barrier. These motions, that take place in the picosecond timescale or faster, dynamically precede the change in the solute coordinate. We have also shown that these protein motions can be treated as equilibrium fluctuations and then the rate constant is mainly determined by the free energy difference between the TS and the RS. The activation free energies estimated under the assumption that the solvent coordinate is in equilibrium with the solute coordinate are only slightly underestimated with respect to those obtained in the non-equilibrium description. These differences can be taken into account by means of the inclusion of a transmission coefficient in the TST rate constant.

We believe that the analysis made on this reaction can be easily extended to other enzymatic systems providing an adequate framework for a quantitative discussion on the role of protein motions in catalysis.

Methods

We used a Quantum Mechanics/Molecular Mechanics (QM/MM) computational scheme where dichloroethane and the side chain of residue Asp 124 are described using the PM3 semiempirical Hamiltonian.[41] The rest of the enzyme and/or water molecules form the MM subsystem described by means of the all- atoms Optimized Potential for Liquid Simulation (OPLS) for the enzyme[42] and a flexible TIP3P potential for water molecules.[43] The Lenard-Jones parameters for the QM/MM interactions are also taken form the OPLS potential except for the chlorine atoms, for which we used those from ref [44].

For the system in aqueous solution we placed dichloroethane and acetate in a pre-equilibrated box of water molecules of 55.8 Å of side, deleting all those water molecules whose oxygen atoms were found at less than 2.8 Å from any non-hydrogen atom of the solute fragments. For the enzyme-substrate system the X-ray crystal structure coordinates were taken from Protein Data Bank (code 2DHC)[45]. The protonation state of titrable residues was determined using PropKa program.[46] The whole system was placed in a pre-equilibrated cubic box of water molecules of 79.5 Å of side. 16 sodium ions were added to neutralize the charge of the protein, so both in solution and in the enzymatic system the total charge was -1. The initial coordinates for the TSs in both environments were taken from our previous work.[37,38] The systems were further equilibrated by means of 200 ps of Molecular Dynamics simulation in the NVT ensemble at the reference temperature of 298 K using the Langevin integrator with a time step of 1 fs and Periodic Boundary Conditions. A cutoff radii switched between 12.5 and 15 Å was applied for all kind of interactions.

In this work we obtained FESs using two different coordinates: a solute coordinate (ξ) and a solvent coordinate (s). The FES can be expressed as:

$$W(\xi,s) = C' - kT\,ln\int \rho(\boldsymbol{x}^N)\delta(\xi(\boldsymbol{x}^N - \xi_0) - r)\delta(s(\boldsymbol{x}^N) - s_0)d\boldsymbol{x}^N \qquad (6)$$

where $\rho(\mathbf{x}^N)$ is the probability density of finding the system at the configuration $\mathbf{x}^N$. The solute coordinate (ξ) is in this case the antisymmetric combination of the distances of the outgoing chloride and the incoming oxygen to the carbon atom (ξ = d(ClC) − d(OC)). The solvent coordinate selected was the antisymmetric combination of the electrostatic potential created by the environment on the outgoing chlorine atom and the incoming oxygen atom, the electrostatic potential created by the MM environment on the chlorine atom.

$$s = V_{Cl}(\boldsymbol{x}^N) - V_O(\boldsymbol{x}^N) = \sum_{j=1}^{M}\frac{q_j}{|\boldsymbol{x}_j - \boldsymbol{x}_{Cl}|} - \sum_{j=1}^{M}\frac{q_j}{|\boldsymbol{x}_j - \boldsymbol{x}_O|} \qquad (7)$$

where the sums run on the M sites of the environment with point charges q_j. This is a collective coordinate involving all the MM atoms that have an electrostatic influence on the donor or acceptor atoms.

The FESs corresponding to the reactions in aqueous solution and the enzyme were obtained using umbrella sampling[47] applying parabolic constraints to the solute and the solvent coordinates:

$$V_r = \frac{1}{2}K_\xi(\xi - \xi_o)^2; \qquad V_s = \frac{1}{2}K_s(s - s_o)^2 \qquad (8)$$

Molecular Dynamics will preferentially explore the most probable configurations of the system around the reference values ξ_o and s_o The joint probability distributions of the two coordinates were obtained by means of the weighted histogram analysis method (WHAM).[48] To save computational cost simulations were performed keeping frozen any atom beyond 25 Å of dichloroethane. A total of 5454 simulation windows consisting in 5

ps of equilibration and 50 ps of production were employed to trace the FESs in solution, while in the enzyme 3100 windows were needed. The force constants employed to keep the system at the reference values of the solute and solvent coordinates were 2500 kJ · mol^{-1} $Å^{-2}$ and 0.01 · kJ^{-1} · mol · $|e|^2$, which provided a good control of the coordinates.[32]

It must be noticed that the PM3/MM Hamiltonian results in systematically overestimated energy barriers, but the geometries obtained for the RS and TS are good enough to have reasonable estimations of Kinetic Isotope Effects.[35] We then corrected the systematic error in the activation free energies by means of single-point calculations at higher theoretical levels. With this purpose we optimized ten TS structures starting from different configurations selected from the corresponding simulation. After IRC calculation the energy barrier was obtained at the PM3/MM level and by means of single-point calculations at the M06-2X/6-311+G(2df,2p)/MM level.[49] The correction energy term was evaluated as the averaged difference between the semiempirical and M06-2X energy barriers.

References

1 Hammes, G. G., Benkovic, S. J. & Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* **50**, 10422-10430 (2011).

2 Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916 (2007).

3 Benkovic, S. J., Hammes, G. G. & Hammes-Schiffer, S. Free-Energy Landscape of Enzyme Catalysis†. *Biochemistry* **47**, 3317-3321 (2008).

4 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).

5 Garcia-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **303**, 186-195 (2004).

6 Gao, J. *et al.* Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **106**, 3188-3209 (2006).

7 Olsson, M. H. M., Parson, W. W. & Warshel, A. Dynamical Contributions to Enzyme Catalysis: Critical Tests of A Popular Hypothesis. *Chem. Rev.* **106**, 1737-1756 (2006).

8 Antoniou, D., Basner, J., Núñez, S. & Schwartz, S. D. Computational and Theoretical Methods to Explore the Relation between Enzyme Dynamics and Catalysis. *Chem. Rev.* **106**, 3170-3187 (2006).

9 Nashine, V. C., Hammes-Schiffer, S. & Benkovic, S. J. Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.* **14**, 644-651 (2010).

10 Ramanathan, A. & Agarwal, P. K. Evolutionarily Conserved Linkage between Enzyme Fold, Flexibility, and Catalysis. *PLoS Biol* **9**, e1001193 (2011).

11 Zhang, J. & Klinman, J. P. Enzymatic Methyl Transfer: Role of an Active Site Residue in Generating Active Site Compaction That Correlates with Catalytic Efficiency. *J. Am. Chem. Soc.* **133**, 17134-17137 (2011).

12 Bhabha, G. *et al.* A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **332**, 234-238 (2011).

13 Adamczyk, A. J., Cao, J., Kamerlin, S. C. L. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. A. Sci. USA* **108**, 14115-14120 (2011).

14 Glowacki, D. R., Harvey, J. N. & Mulholland, A. J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nature Chem.* **4**, 169-176 (2012).

15 Kamerlin, S. C. L. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339-1375 (2010).

16 Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210-3235 (2006).

17 Kurplus, M & McCammon, J A. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem.* **52**, 263-300 (1983).

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

19 Kosugi, T. & Hayashi, S. Crucial Role of Protein Flexibility in Formation of a Stable Reaction Transition State in an α-Amylase Catalysis. *J. Am. Chem. Soc.* **134**, 7045-7055 (2012).

20 Pang, J., Pu, J., Gao, J., Truhlar, D. G. & Allemann, R. K. Hydride Transfer Reaction Catalyzed by Hyperthermophilic Dihydrofolate Reductase Is Dominated by Quantum Mechanical

Tunneling and Is Promoted by Both Inter- and Intramonomeric Correlated Motions. *J. Am. Chem. Soc.* **128**, 8015-8023 (2006).

21 Kanaan, N. *et al.* Temperature Dependence of the Kinetic Isotope Effects in Thymidylate Synthase. A Theoretical Study. *J. Am. Chem. Soc.* **133**, 6692-6702 (2011).

22 Hay, S. & Scrutton, N. S. Good vibrations in enzyme-catalysed reactions. *Nature Chem.* **4**, 161-168 (2012).

23 Pu, J., Gao, J. & Truhlar, D. G. Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chem. Rev.* **106**, 3140-3169 (2006).

24 Boekelheide, N., Salomón-Ferrer, R. & Miller, T. F. Dynamics and dissipation in enzyme catalysis. *Proc. Natl. A. Sci. USA* **108**, 16159-16163 (2011).

25 Garrett, B. C. & Truhlar, D. G. in *Theory and Applications of Computational Chemistry* (eds Clifford E. Dykstra, Gernot Frenking, Kwang S. Kim, & Gustavo E. Scuseria) 67 - 87 (Elsevier, 2005).

26 Grote, R. F. & Hynes, J. T. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.* **73**, 2715-2732 (1980).

27 Bergsma, J. P., Gertner, B. J., Wilson, K. R. & Hynes, J. T. Molecular dynamics of a model SN2 reaction in water. *J. Chem. Phys.* **86**, 1356-1376 (1987).

28 Ruiz-Pernía, J. J., Tuñón, Iñaki, Moliner, Vicente, Hynes, James T. & Roca, Maite. Dynamic Effects on Reaction Rates in a Michael Addition Catalyzed by Chalcone Isomerase. Beyond the Frozen Environment Approach. *J. Am. Chem. Soc.* **130**, 7477-7488 (2008).

29 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.* **86**, 1377-1386 (1987).

30 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.* **110**, 5297-5311 (1988).

31 Gertner, B. J. , Wilson, K. R. & Hynes, J. T. . Nonequilibrium solvation effects on reaction rates for model SN2 reactions in water. *J. Chem. Phys.* **90**, 3537-3558 (1989).

32 Ruiz-Pernía, J. J., Martí, S., Moliner, V. & Tuñón, I. A Novel Strategy to Study Electrostatic Effects in Chemical Reactions: Differences between the Role of Solvent and the Active Site of Chalcone Isomerase in a Michael Addition. *J. Chem. Theory Comput.* **8**, 1532-1535 (2012).

33 Janssen, D. B., Scheper, A., Dijkhuizen, L. & Witholt, B. Degradation of Halogenated Aliphatic-Compounds by Xanthobacter-Autotrophicus Gj10. *Appl. Environ. Microbiol.* **49**, 673-677 (1985).

34 Schanstra, J. P., Kingma, J. & Janssen, D. B. Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem.* **271**, 14747-14753 (1996).

35 Devi-Kesavan, L. S. & Gao, J. Combined QM/MM Study of the Mechanism and Kinetic Isotope Effect of the Nucleophilic Substitution Reaction in Haloalkane Dehalogenase. *J. Am. Chem. Soc.* **125**, 1532-1540 (2003).

36 Shurki, A., Štrajbl, M., Villà, J. & Warshel, A. How Much Do Enzymes Really Gain by Restraining Their Reacting Fragments? *J. Am. Chem. Soc.* **124**, 4097-4107 (2002).

37 Soriano, A. *et al.* Electrostatic effects in enzyme catalysis: a quantum mechanics/molecular mechanics study of the nucleophilic substitution reaction in haloalkane dehalogenase. *Theor. Chem. Accounts* **112**, 327-334 (2004).

38 Soriano, A., Silla, E., Tuñón, I. & Ruiz-Lopez, M. F. Dynamic and electrostatic effects in enzymatic processes. An analysis of the nucleophilic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.* **127**, 1946-1957 (2005).

39 Okamoto, K., Kita, T., Araki, K. & Shingu, H. Kinetic Studies of Bimolecular Nucleophilic Substitution .4. Rates of Sn2 and E2 Reactions of Beta-Substituted Ethyl Chlorides with Sodium Acetate in Aqueous Solutions. *B. Chem. Soc. Jpn.* **40**, 1913 (1967).

40 Truhlar, D. G. & Garrett, Bruce C. Variational transition-state theory. *Acc. Chem. Res.* **13**, 440-448 (1980).

41 Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **10**, 209-220 (1989).

42 Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225-11236 (1996).

43 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).

44 Gao, J. & Xia, X. A two-dimensional energy surface for a type II SN2 reaction in aqueous solution. *J. Am. Chem. Soc.* **115**, 9667-9675 (1993).

45 Verschueren, K. H., Seljée, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **363**, 693-698 (1993).

46 Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525-537 (2011).

47 Torrie, G.M. & Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187 - 199 (1977).

48 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.* **13**, 1011-1021 (1992).

49 Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Accounts* **120**, 215-241 (2008).

Acknowledgements

The authors gratefully acknowledge financial support from MEC through project CTQ2012-36253-C03. J. J. R-P. thanks a 'Juan de la Cierva' contract and R. G-M a FPU fellowship of the Ministerio de Economía y Competitividad. I. T. acknowledges helpful discussions held with Prof. D. Laage and J. T. Hynes during his sabbatical stay at the ENS. The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València in the 'Tirant' supercomputer.

Authors contributions

I.T., V. M and J. J. R-P designed the computational experiments. S. M. wrote the code and R. G-M performed the calculations. I.T., V. M and J. J. R-P wrote the first version of the paper. All the authors commented and discussed the results and the final version of the manuscript.

Additional information

The authors declare no competing financial interests. Reprints and permission information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.T. or J.J. R-P.

Figure Captions

Figure 1. Schematic representation of the S_N2 reaction catalyzed by DhlA. The QM subsystem is surrounded by a continuous line.

Figure 2: Free energy surfaces corresponding to the S_N2 reaction between an acetate anion and dichloroethane .The solute distinguished coordinate is the antisymmetric combination of the bond breaking and forming distances (ξ=d(ClC)-d(OC)); while the solvent coordinate is the antisymmetric combination of the electrostatic potential created by the environment on the leaving group and the nucleophilic oxygen, ($s=V_{Cl}-V_O$). The isoenergetic free energy lines are represented each 3 kcal·mol^{-1}. The continuous lines represent the minimum free energy path on the free energy surfaces and the dashed ones correspond to the path obtained assuming equilibrium solvation along the solute coordinate. (a) Free Energy Surface for the reaction in aqueous solution. (b) Free Energy Surface for the in the active site of DhlA. The white arrow represents the reaction path from the enzymatic Michaelis complex in a completely rigid protein environment.

Figure 3. The TS ensembles obtained from equilibrium and non-equilibrium pictures are slightly different as reflected in the respective dividing surfaces. The continuous line corresponds to the dividing surface defined according to the Free Energy Surface traced along the solute and solvent coordinates. The dashed line corresponds to the dividing surface obtained when the solvent coordinate is assumed to be at equilibrium (the equilibrium plane is then defined just as $\xi=\xi^{\ddagger}$). (a) Dividing Surfaces for the reaction in solution. The arrow represents a hypothetical trajectory recrossing the equilibrium dividing

surface. (b) Dividing surfaces for the reaction in the active site of DhlA. Note that the angle between the equilibrium and non-equilibrium plane is smaller than in solution. (c) Schematic representation of the TS wells projected on the equilibrium and non-equilibrium dividing surfaces with their characteristic frequencies (ν_{ξ} and $\nu_{s,\xi}$).

Table 1. Force constants, effective masses and characteristic frequencies associated to the solvent coordinate (s) for the S_N2 reaction in solution and in DhlA.

	Aqueous Solution	DhlA
K_s ($kcal^{-1}\cdot mol\cdot|e|^2$)	$3.3\cdot10^{-3}$	$1.4\cdot10^{-2}$
m_s ($kcal^{-1}\cdot mol\cdot|e|^2\cdot s^{-2}$)	0.40	2.30
ν_s (cm^{-1})	480	410